



PREVISION JOURNALIERE DES POLLENS SUR LE TERRITOIRE NATIONAL FRANÇAIS, AVEC UN OBJECTIF D'INFORMATION SANITAIRE DES POPULATIONS ALLERGIQUES

Etienne Cassagne

► To cite this version:

Etienne Cassagne. PREVISION JOURNALIERE DES POLLENS SUR LE TERRITOIRE NATIONAL FRANÇAIS, AVEC UN OBJECTIF D'INFORMATION SANITAIRE DES POPULATIONS ALLERGIQUES. Géographie. Université de Bourgogne, 2008. Français. NNT: . tel-00335207

HAL Id: tel-00335207

<https://theses.hal.science/tel-00335207>

Submitted on 28 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE BOURGOGNE
FACULTE DE MEDECINE

THESE

pour obtenir le grade de

DOCTEUR

Discipline : Bioclimatologie

**PREVISION JOURNALIERE DES POLLENS SUR LE TERRITOIRE
NATIONAL FRANÇAIS, AVEC UN OBJECTIF D'INFORMATION
SANITAIRE DES POPULATIONS ALLERGIQUES**

présentée et soutenue publiquement par

Etienne CASSAGNE

le 15 mai 2008

Membres du jury :

M. Philippe BONNIAUD : Professeur, Université de Bourgogne, *Directeur de thèse*.

M. Denis CAILLAUD : Professeur, Université d'Auvergne Clermont-Ferrand 1, *Rapporteur*.

Mme Martine TABEAUD : Professeur, Université de Paris 1 Panthéon-Sorbonne, *Rapporteur*.

M. Jean-Pierre BESANCENOT : Directeur de recherche CNRS, Université de Bourgogne, *Examineur*.

Mme Hélène DESQUEYROUX : Ingénieur, ADEME, *Examineur*.

M. Michel THIBAUDON : Directeur, RNSA, *Examineur*.



REMERCIEMENTS

Je tiens, tout d'abord, à manifester toute ma reconnaissance à Monsieur Jean-Pierre BESANCENOT pour m'avoir proposé cette thèse et en avoir pris la direction, ainsi que pour sa grande gentillesse et humanité. Je le remercie pour l'intérêt qu'il a su porter à cette thèse, son aide, sa disponibilité, ses conseils avisés et ses nombreuses relectures qui ont permis l'aboutissement de ce travail. Je le remercie, par ailleurs, pour sa participation au jury de thèse en tant qu'examinateur.

Une telle recherche nécessitait des moyens. Par leur soutien financier je tiens à remercier l'ADEME et le RNSA, co-financeurs de cette thèse.

Je tiens à adresser mes sincères remerciements à Monsieur Philippe BONNIAUD pour avoir bien voulu succéder à Monsieur BESANCENOT, suite à son départ en retraite en juin 2007, en tant que directeur de thèse.

Je présente mes sincères remerciements à Madame Martine TABEAUD et Monsieur Denis CAILLAUD pour m'avoir fait l'honneur de s'intéresser à ce travail et d'accepter la charge de rapporteurs.

J'associe à ces remerciements les deux derniers membres de ce jury, Madame Hélène DESQUEYROUX et Monsieur Michel THIBAUDON, qui ont, en compagnie de Monsieur CAILLAUD, pris une part active dans le suivi de ma thèse au cours, notamment, des différentes réunions programmées à ce sujet. Je les remercie pour leur soutien et les conseils qu'ils ont pu me prodiguer. Je remercie plus particulièrement Monsieur THIBAUDON pour avoir mis à ma disposition les données aérobiologiques du RNSA et pour m'avoir permis d'assurer l'analyse du contenu pollinique du site de Castres.

Je tiens, par ailleurs, à remercier Messieurs Jean-Claude COHEN et Emmanuel CLOPPET (ainsi que toute son équipe) pour l'intérêt porté aux résultats de mes travaux.

Je tiens à manifester ma reconnaissance à Monsieur Bernard FONTAINE, Directeur du Centre de Recherches de Climatologie, pour m'avoir donné la possibilité de travailler au sein du CRC et me permettre de disposer d'un compte de recherches, sans jamais avoir été officiellement rattaché à son laboratoire.

Un très grand merci à Pascal OETTLI, doctorant au CRC, pour avoir accepté de mettre au point un programme informatique qui s'est révélé essentiel dans le déroulement de ma thèse. Je le remercie par ailleurs pour sa disponibilité et ses conseils dans tous les domaines de la programmation et de l'informatique.

Je tiens également à remercier Messieurs Yves RICHARD et Pierre CAMBERLIN pour les conseils qu'ils ont pu me donner au sujet de ma thèse, et pour m'avoir donné l'opportunité d'assurer des heures d'enseignement aux étudiants de géographie. A travers eux je remercie l'ensemble des membres du CRC pour leur gentillesse avec une mention toute spéciale pour mon vieux compagnon d'armes depuis la 1^{ère} année de géographie : Benjamin POHL. Par ses conseils, par ses encouragements (un peu rudes parfois !) il s'est évertué à me motiver durant les périodes de baisse de morale. Je n'oublie pas non plus nos multiples discussions à propos de tout et de rien autour d'un café.

Je souhaite ensuite remercier tout spécialement Madame Annie PASSELEGUE pour sa disponibilité (à travers l'envoi de tous les fichiers de données aérobiologiques nécessaires à ce travail), pour avoir répondu à mes

nombreux mails et pour sa grande gentillesse. J'associe à ces remerciements Charlotte SINDT et Gilles OLIVER, pour l'aide et les précisions qu'ils ont pu m'apporter. Tous mes remerciements vont également à Mohamed LAAYDI pour sa grande gentillesse et les conseils qu'il a su me donner, notamment sur la prévision pollinique. Un grand merci à Madame Nadine DUPUY pour son enseignement précieux lors du stage de reconnaissance pollinique. Enfin, je ne peux oublier mon collègue et ami, Sylvain RIGOLLET (doctorant en géographie et analyste pollinique... entre autres !) pour son optimisme à toute épreuve, nos multiples discussions et les semaines de «galère» partagées quand les pollens avaient la mauvaise idée (sic) de s'impacter en masse sur la bande en cellophane des capteurs.

Un immense merci à ma famille. Par leur soutien moral, leur générosité et leur amour, ils m'ont soutenu durant toutes les périodes de doute et de démoralisation. Tous les week-ends à Toucy ont été pour moi des bouffées d'oxygène. A ma mère, mon père, mon frère Vincent et ma sœur Claire, je veux témoigner de toute ma tendresse et de tout mon amour.

Je remercie tous les autres membres de ma famille – et tout spécialement mes grands-parents – par la curiosité dont ils ont fait preuve à propos de mon travail.

A tous mes amis (qui se reconnaîtront) et toutes les personnes que forcément et involontairement j'oublie...
Merci !

Cette étude a été co-financée par
l'Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME)
et
le Réseau National de Surveillance Aérobiologique (RNSA)

Ce travail a été réalisé au sein du **Laboratoire Climat et Santé**, composante du
Centre Universitaire d'Epidémiologie de Population, et du **Centre de
Recherches de Climatologie**

SOMMAIRE

INTRODUCTION GENERALE.....	1
<u>PARTIE 1 : DONNEES POLLINIQUES ET TAXONS DE L'ETUDE.....</u>	5
1.1 PRESENTATION GENERALE DES DONNEES POLLINIQUES.....	7
1.2 PRESENTATION GENERALE DES TAXONS DE L'ETUDE.....	63
CONCLUSION DE LA PARTIE.....	91
<u>PARTIE 2 : CRITERES ET METHODES LIES AU DEBUT DE LA SAISON POLLINIQUE.....</u>	93
2.1 PRESENTATION ET ANALYSE DES DIFFERENTS CRITERES DE DETERMINATION DU DEBUT DE LA SAISON POLLINIQUE.....	95
2.2 PRESENTATION GENERALE DES METHODES DE PREVISION DE LA DATE D'APPARITION DU RISQUE.....	121
CONCLUSION DE LA PARTIE.....	141
<u>PARTIE 3 : PRESENTATION DES RESULTATS.....</u>	143
3.1 OPERATIONS SUR LES DONNEES POLLINIQUES ET PRESENTATION DES DONNEES METEOROLOGIQUES.....	145
3.2 PREVISION DU FRENE, DU BOULEAU, DES POACEES : LE CAS DE CINQ VILLES DE LA MOITIE NORD DE LA FRANCE.....	151
3.3 PRESENTATION DES RESULTATS POUR LES SERIES POLLINIQUES LES PLUS LONGUES.....	221
3.4 ETUDES COMPARATIVES COMPLEMENTAIRES.....	275
3.5 ETUDE SUR LES CONCENTRATIONS POLLINIQUES : LES POACEES A NANTES ET LA ROCHE-SUR-YON.....	287
CONCLUSION DE LA PARTIE.....	305
CONCLUSION GENERALE.....	307
BIBLIOGRAPHIE.....	311
TABLE DES FIGURES.....	325
TABLE DES TABLEAUX.....	331
LISTE DES SIGLES ET ABREVIATIONS.....	333
TABLE DES MATIERES.....	335

INTRODUCTION GENERALE

Le pollen, mot d'origine latine signifiant «fleur de farine», correspond, selon la définition du dictionnaire de botanique de Boullard (1999), à la production microscopique libérée par les anthères de fleurs, ou d'inflorescences tels les chatons, à la faveur de processus variés de déhiscence. En d'autres termes le pollen est contenu dans des sacs et est libéré par la fission de ces sacs qui s'ouvrent spontanément à maturité. Ce terme de pollen désigne cependant l'ensemble des grains microscopiques aussi bien que chacun d'entre eux. C'est pourquoi nous retrouverons ces deux variantes de définition dans le développement de cette étude. Le pollen constitue un mode de dispersion génétique des plantes en transmettant, lors de la fécondation, un lot de chromosomes de la plante productrice de pollens à ses descendants (Felber et Clot, 2003). De par ses caractéristiques, il est essentiel à la reproduction végétative nécessaire à l'alimentation des individus et des animaux (graines et fruits). Trois types de vecteurs de pollinisation existent (Laaidi M, 1999) : l'eau (hydrophilie), les animaux (zoïdophilie) – dont les insectes (entomophilie) sont les vecteurs les plus fréquents – et le vent (anémophilie). Ce dernier mode de dispersion, propre aux plantes anémogames, se caractérise par l'abondance du pollen présent dans l'air. Du fait d'un mode de dispersion plus aléatoire que celui des plantes zoogames, cette abondance est nécessaire dans le cadre d'une fécondation efficace. Ainsi, à maturité, les anthères de la fleur s'ouvrent et le pollen est libéré soit de façon passive (agitation par le vent des anthères qui pendent hors de la fleur), soit de façon active comme dans le cadre d'un changement du taux d'humidité où la paroi des anthères se rompt, se détend comme un ressort et entraîne l'éjection des grains de pollen (Felber et Clot, 2003). Par l'abondance de ce type de pollen, la probabilité de contact répété avec le corps humain est très importante. C'est pourquoi il est le plus souvent responsable de pollinoses, c'est-à-dire de l'ensemble des troubles liés à l'inhalation du pollen. Ainsi un peu plus de 16% des Français sont actuellement affectés par les différentes pathologies liées au pollen, proportion qui a doublé par rapport au taux couramment admis au début des années 1980 (Besancenot, 2007). Ce sont les allergènes contenus dans les pollens qui sont à l'origine des différentes pollinoses. Il faut au minimum deux saisons polliniques pour une sensibilisation qui intervient le plus souvent entre huit et vingt ans, mais continue à se manifester pendant des décennies (Michel et Bousquet, 1997). Le phénomène tend d'ailleurs à s'aggraver, les

allergologues constatant de plus en plus de symptômes allergiques, mais également une accentuation de la sévérité et de la durée des manifestations chez les patients affectés depuis plusieurs années (Besancenot, 2007). La plupart de ces pollinoses ne sont pas spécialement graves, telles les rhinites (rhumes) et les conjonctivites (inflammations de la partie postérieure des paupières et de la face antérieure du blanc de l'œil). Ce n'est pas le cas de l'asthme qui constitue un problème majeur de santé publique en causant chaque année environ 2000 décès en France (Laaidi K, 1999) et dont le coût annuel en consultations et médicaments se chiffre en plusieurs centaines de millions d'euros par an.

Face à l'ampleur de ce phénomène, l'information dont peuvent disposer, jusqu'à maintenant, les patients et leurs médecins, pour mettre en œuvre une prévention efficace, reste assez limitée et, tout du moins, inadaptée. En effet, si les données polliniques fournies par le Réseau National de Surveillance Aérobiologique (RNSA) servent de support initial à une information régulière délivrée à la partie concernée du corps médical et au grand public, elles sont avant tout des éléments de base dans des séries de mesures à visée explicative. Par la force des choses, elles se rapportent à une situation déjà révolue au moment où elles sont diffusées. En effet, il faut attendre sept jours entre le début du fonctionnement d'un tambour entouré d'une bande de cellophane, sur laquelle sont impactés les pollens présents dans l'air ambiant, et son relevé, puis encore deux jours au minimum pour qu'il soit envoyé à l'analyste, que la bande soit analysée et les résultats diffusés. Or, pour plusieurs pollens allergisants, une très forte variabilité interannuelle existe dans les dates de début de pollinisation ou, tout du moins, d'apparition de concentrations susceptibles de gêner les populations allergiques. On comprend dès lors qu'un calendrier pollinique moyen ne suffise pas à déterminer la date à laquelle doit débiter un traitement préventif, à moins de l'initier très tôt. Cela impliquerait un coût élevé tant pour le malade que pour la collectivité, des effets secondaires non négligeables et, sans doute, une moindre efficacité du fait de phénomènes d'accoutumance. Il s'avère alors nécessaire de développer des modèles qui puissent fournir des prévisions de la date d'apparition d'un risque d'exposition lié aux pollens allergisants qui soient les plus fiables possibles. De plus, idéalement, ces modèles doivent être en mesure de prévoir cette date au plus tard une semaine à dix jours avant la survenue effective de ce risque, qui correspond à une certaine concentration pollinique.

Plusieurs espèces, genres ou familles produisent des pollens allergisants, leur nombre nous impose un choix qui amène à exclure les taxons polliniques les plus précoces (noisetier, aulne,

Cupressacées-Taxacées) du fait du démarrage régulièrement trop tardif du recueil pollinique. Nous retenons pour l'étude quatre taxons polliniques qui ne sont, à priori, pas affectés par ce problème : le frêne, le bouleau, les Poacées et l'ambroisie. Ces trois derniers pollens se caractérisent, en compagnie de celui du cyprès, par un potentiel allergisant de niveau 5 (très élevé), degré le plus élevé selon la terminologie employée par le Réseau National de Surveillance Aérobiologique (RNSA) sur son site www.pollens.fr. Le pollen de frêne présente un niveau 4 de risque (élevé) à l'instar de ceux de l'aulne, du chêne, de la pariétaire et de l'armoise.

L'élaboration nécessaire de modèles de prévision capables de prévoir l'apparition d'un niveau de risque, taxon par taxon, doit ainsi s'inscrire dans une démarche sanitaire et informative afin d'aider le mieux possible les allergologues à prescrire les traitements anti-allergiques. En fin de compte, le but pratique de ce travail est de créer des modèles permettant d'initier des traitements de la manière la plus adéquate possible.

A ce titre, le laboratoire Climat et Santé, qui cessera son activité au cours de l'année universitaire 2007-2008, possède une certaine expérience en la matière. C'est en 1995 qu'il s'est résolument lancé dans l'étude des pollens et des pollinoses, essentiellement dans le cadre de la Haute-Bourgogne (Côte-d'Or, Saône-et-Loire), puis dans celui de la Bourgogne tout entière, en débordant parfois sur les régions voisines (Rhône-Alpes, par exemple, à propos de l'ambroisie). Les premiers travaux, inscrits dans le contrat Etat-Région 1995-1998, ont porté sur l'écologie des pollens et l'épidémiologie des pollinoses, et ont débouché sur deux thèses de doctorat (Laaidi K, 1999 ; Laaidi M, 1999). Le rôle des conditions météorologiques a été particulièrement approfondi. Très vite, la nécessité s'est fait sentir de développer des méthodes de prévision, orientation renforcée par une demande de la Direction Régionale des Affaires Sanitaires et Sociales (DRASS) de Bourgogne, qui a sollicité une étude relative à quatre pollens (bouleau, frêne, plantain et Poacées) (Laaidi et Laaidi, 2001). Les recherches se sont poursuivies depuis lors, débouchant sur une dizaine de publications principalement relatives à trois taxons (bouleau, Poacées et ambroisie), un peu plus marginalement à deux autres (frêne et plantain). C'est la date de début de pollinisation qui a surtout été étudiée, avec des résultats généralement très satisfaisants. Dans le cadre de cette implication dans le domaine aérobiologique, le laboratoire Climat et Santé a géré le fonctionnement d'un capteur pollinique, installé sur le toit en terrasse de la Faculté de Médecine de Dijon, depuis 1996 et qui a été intégré dans le réseau du RNSA dès 1997. Le laboratoire a été également un centre d'analyse important des pollens s'occupant, depuis

2002, des recueils polliniques de cinq sites bourguignons : Dijon, Chalon-sur-Saône, Nevers, Auxerre (en 2004 seulement) et Mâcon (durant la seule période estivale ces dernière années). L'analyse des pollens a également concerné deux sites extra-bourguignons : Annecy et Castres. Ainsi, cette thèse s'inscrit parfaitement dans la lignée des travaux effectués par le laboratoire Climat et Santé, avec un élargissement du champ d'étude au niveau national. Son contenu va s'articuler autour de trois parties.

La première partie consistera en une présentation générale des données polliniques et des taxons retenus pour l'étude où l'on s'intéressera aux modes de recueil et d'analyse des pollens, à la qualité des séries polliniques, aux paramètres techniques susceptibles d'influencer les concentrations, à la description et à la répartition géographique des pollens de frêne, de bouleau, de Poacées et d'ambroisie.

La deuxième partie, plus réduite, s'attachera à une description des différents critères de détermination du début de la saison pollinique et des différentes méthodes de prévision d'un risque d'exposition rencontrés dans la littérature. Ces deux aspects faisant partie intégrante de l'objectif final de ce travail.

La troisième partie, enfin, concernera la présentation des résultats issus des différents modèles élaborés à partir des méthodes de prévision qui seront retenues. Un passage de cette partie sera consacré à une étude des concentrations polliniques de Poacées à Nantes et La Roche-sur-Yon, passage reprenant largement l'article contenu dans le Cd-rom des Journées Interdisciplinaires de la Qualité de l'Air (JIQA), organisées par l'Association pour la Prévention de la Pollution Atmosphérique Nord-Pas-de-Calais (APPA Nord-Pas-de-Calais), qui se sont déroulées en janvier 2007 à Lille.

PARTIE 1 :
DONNEES POLLINIQUES ET TAXONS DE L'ETUDE

1.1 PRESENTATION GENERALE DES DONNEES POLLINIQUES

1.1.1 Présentation du matériel

1.1.1.1 Fonctionnement du capteur pollinique

Il s'agit d'un capteur volumétrique placé en hauteur sur le toit d'un bâtiment dont l'environnement doit être, autant que possible, dégagé. La présence de bâtiments plus élevés et/ou d'un type de végétation précis, à proximité immédiate, doit être évitée ; cela risquerait de nuire à la fonction principale de ce dispositif qui est d'être un capteur de « fond » représentatif d'une zone géographique assez étendue concernant, au moins, l'agglomération où il est implanté. Ce capteur pollinique a été mis au point par Hirst (1952). Monté sur un trépied, il se compose d'un corps central et d'une girouette (figure 1).



Figure 1 : Le capteur de pollen Hirst de Dijon.

Le corps central (figure 2) contient le mécanisme permettant la capture des pollens et autres particules contenues dans l'air ambiant.

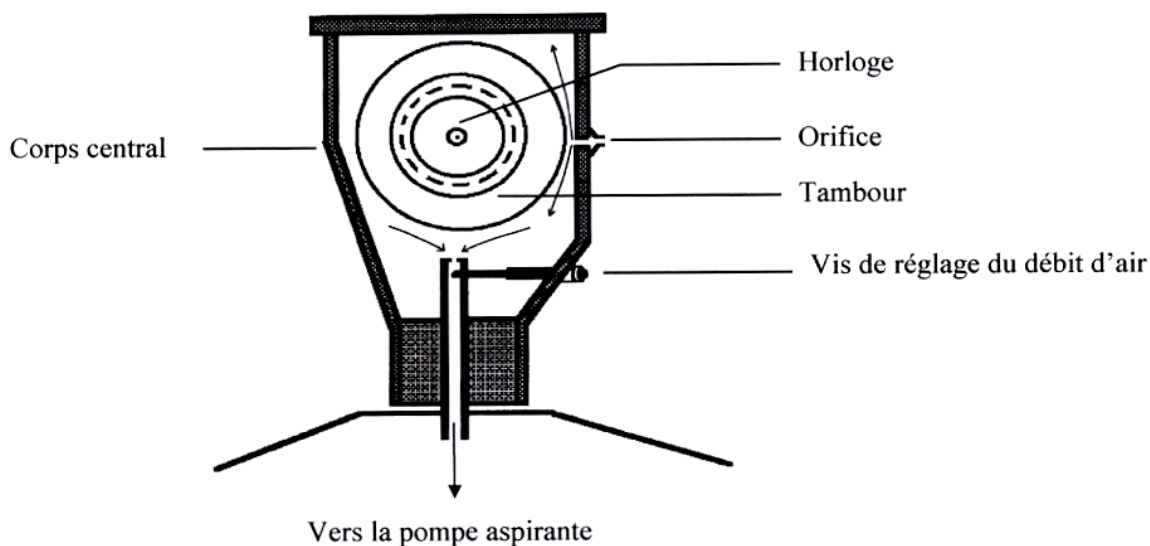


Figure 2 : Les composants du corps central du capteur pollinique.

Grâce à une pompe, l'air est aspiré à travers un orifice à un débit constant de 10 litres par minute ; le réglage du débit s'effectuant grâce à une vis placée sous l'orifice d'aspiration. Ce débit correspond à une respiration humaine moyenne et doit permettre d'évaluer la quantité de pollen que peut inhaler un individu. La girouette permet d'orienter la buse d'aspiration face au vent dominant.

Les particules présentes dans l'air aspiré sont impactées sur un tambour recouvert d'une bande de cellophane enduite d'une solution à base de silicone. Grâce à une minuterie qui se remonte mécaniquement, ce tambour tourne à la vitesse de 2 mm/heure de sorte qu'un tour complet équivaut à une semaine de données. A la fin de cette période il est relevé et remplacé par un tambour neuf.

1.1.1.2 Montage des lames et analyse des pollens

Le tambour une fois relevé, la bande de cellophane l'entourant est analysée. Elle est, tout d'abord, décollée du tambour puis montée sur une règle de découpe où elle est partitionnée en sept bandes d'égale longueur correspondant chacune à 24 heures. Les bandes sont ensuite montées entre lame et lamelle après que l'on ait ajouté une solution gélatineuse composée de fuchsine sur la bande qui, préalablement chauffée, solidifie et permet ainsi de fixer l'ensemble (figure 3). La coloration de la solution permet de teinter les pollens et autres débris végétaux de la bande d'une couleur rouge à mauve.



Figure 3 : Lames de lecture de Dijon (19-20/04/2005) et Castres (18-19/06/2005).

Une fois les lames préparées, elles sont analysées en microscopie optique. Le travail consiste alors en un décompte des différents taxons polliniques visibles sous microscope ; décompte facilité par un logiciel de reconnaissance vocale : Via Voice®. La lecture s'effectue sur deux ou

trois bandes horizontales qui suivent le sens de rotation du tambour. Il s'agit de faire défiler la bande sous le microscope et de répertorier en même temps les pollens présents qui vont s'afficher sur un document Word[®] dont le titre correspond au jour de relevé. Simultanément à la lecture, les mouvements sont enregistrés par un système de repérage orthonormé, la platine Microvision[®], qui permet de classer les pollens par tranches bihoraires.

Une fois la lecture terminée, le décompte s'effectue grâce au logiciel C-Scope[®], mis au point par le Centre Européen Médical Bioclimatique de Recherche et d'Enseignement Universitaire (CEMBREU) comme la platine Microvision[®], qui applique, aux pollens comptabilisés, un facteur de conversion dépendant à la fois du diamètre du champ du microscope et du nombre de lignes lues. Une fois cette opération établie, on peut obtenir un tableau hebdomadaire de comptes quotidiens (tableau 1) ou un tableau journalier de comptes bihoraires (tableau 2), avec l'exemple de Chalon-sur-Saône (facteur de conversion de 0,46), rapportés en grains par mètre cube d'air.

CHALON SUR SAONE du 28/03/06 au 03/04/06

FOCUS : 0.46

	28/03/06	29/03/06	30/03/06	31/03/06	01/04/06	02/04/06	03/04/06	Total/m3	E
Populus	11	21	54	156	246	235	103	826	+
Fraxinus	27	59	150	79	80	91	110	596	+
Cupressaceae-									
Taxaceae	22	8	6	23	32	11	8	110	-
Salix	9	3	4	10	6	8	8	48	-
Alnus	9	4	4	4	1	1	-	23	-
Betula	1	-	-	1	2	6	8	18	+
Ulmus	1	3	3	2	2	2	1	14	-
Corylus	4	1	2	1	1	-	-	9	-
Poaceae	1	1	1	-	-	2	-	5	-
Carpinus	-	-	-	1	1	2	-	4	+
Pinus	2	-	-	-	-	-	1	3	-
Buxus	-	-	-	-	1	-	1	2	=
Indetermines	-	-	1	-	-	-	1	2	+
Platanus	-	-	-	-	2	-	-	2	=
Cornus	-	-	1	-	-	-	-	1	=
TOTAUX	87	100	226	277	374	358	241	1663	

Tableau 1 : Tableau hebdomadaire des comptes polliniques de Chalon-sur-Saône pour la semaine du 28 mars au 3 avril 2006. E = évolution des concentrations hebdomadaires des taxons présents par rapport à la semaine précédente.

21.06.07													total	
CHALON SUR SAONE													brut	après conversion
	00-02H	02-04H	04-06H	06-08H	08-10H	10-12H	12-14H	14-16H	16-18H	18-20H	20-22H	22-24H	327	154
Castanea	2	-	-	-	4	2	2	18	12	28	23	35	126	58
Urticaceae	3	2	1	-	-	2	3	9	16	33	22	27	118	55
Poaceae	4	-	-	2	-	2	5	3	2	13	14	17	62	29
Plantaginaceae	-	-	-	-	3	2	3	3	1	-	-	-	12	6
Pinus	-	-	1	-	-	-	1	-	1	2	1	1	7	4
Indetermines	-	-	1	-	-	-	-	-	-	-	-	-	1	1
Rumex	-	-	-	-	-	-	-	-	-	1	-	-	1	1

Tableau 2 : Tableau journalier des comptes polliniques de Chalon-sur-Saône du 21 juin 2007.

Les données hebdomadaires sont ensuite envoyées au Réseau National de Surveillance Aérobiologique (RNSA) qui les diffuse dans un bulletin paraissant tous les vendredi et qui récapitule les comptes de chaque station de capture pollinique.

1.1.2 Présentation des sites du RNSA

Dans ce sous-chapitre, nous allons réaliser un historique des sites du RNSA et présenter ceux qui seront retenus pour notre étude.

1.1.2.1 Le réseau de capteurs

Les données polliniques sont issues de la base de données du RNSA qui gérait, en 2007, 63 capteurs polliniques fonctionnant à l'année. Ces sites sont présentés ci-dessous par région administrative :

- **Nord-Pas-de-Calais** : Lille, Saint-Omer
- **Picardie** : Amiens, Saint-Quentin
- **Lorraine** : Metz, Nancy
- **Alsace** : Strasbourg
- **Haute-Normandie** : Rouen
- **Basse-Normandie** : Caen, La Ferté-Macé

- **Pays de la Loire** : Le Mans, La Roche-sur-Yon, Nantes, Cholet, Angers
- **Champagne-Ardenne** : Reims, Troyes
- **Ile-de-France** : Paris
- **Centre** : Orléans, Tours
- **Bourgogne** : Nevers, Chalon-sur-Saône, Dijon
- **Franche-Comté** : Besançon
- **Bretagne** : Brest, Pontivy, Saint-Brieuc, Dinan, Rennes
- **Poitou-Charentes** : La Rochelle, Poitiers, Angoulême
- **Auvergne** : Montluçon, Clermont-Ferrand, Vichy, Aurillac
- **Aquitaine** : Périgueux, Bordeaux, Agen, Pau, Bayonne
- **Rhône-Alpes** : Lyon, Saint-Etienne, Roussillon, Annecy, Chambéry, Grenoble, Annemasse, Bourg-en-Bresse
- **Midi-Pyrénées** : Toulouse, Castres, Millau
- **Provence-Alpes-Côte-d’Azur** : Gap, Avignon, Aix-en-Provence, Briançon, Marseille, Toulon, Nice
- **Languedoc-Roussillon** : Nîmes, Perpignan, Montpellier
- **Corse** : Ajaccio

A côté de ces sites annuels, le RNSA gère un certain nombre de capteurs fonctionnant de fin juillet à début octobre. Ces sites, quasiment exclusivement situés en région Rhône-Alpes, sont mis en route afin d’enregistrer les concentrations polliniques d’ambroisie, dont la pollinisation s’effectue en août et septembre, et d’évaluer l’extension spatiale du risque d’exposition allergique à ce pollen particulièrement nuisible à la santé.

Excepté le Limousin, au moins un capteur est implanté dans chaque région. Quasiment toutes les grandes agglomérations en sont équipées d’un à l’exception de Limoges, Mulhouse et Le Havre si l’on s’intéresse aux communes de plus de 100 000 habitants. Le territoire français est, à l’heure actuelle, bien couvert grâce à un effort important dans l’optimisation du réseau, qui s’est accompagnée d’une densification de celui-ci.

Evolution globale du réseau

Depuis 1987, en effet, on constate une hausse quasi constante du nombre de capteurs polliniques comme le montre la figure 4 :

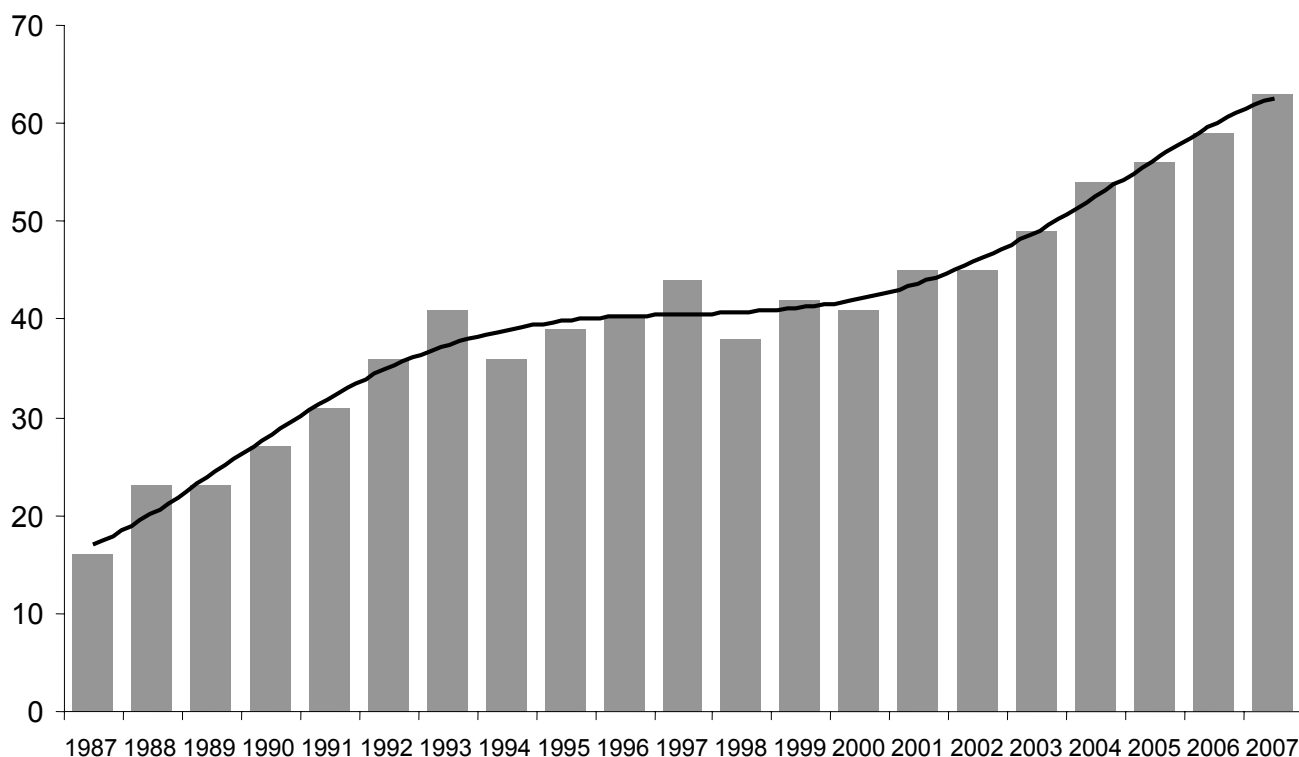


Figure 4 : Nombre annuel de capteurs en fonctionnement sur la période 1987-2007 gérés par l'Institut Pasteur puis le RNSA.

L'augmentation du parc national de capteurs s'est effectuée principalement en deux phases :

- 1987 à 1993 où l'on passe de 16 à 41 capteurs.
- 2002 à 2007 avec un réseau qui passe de 45 à 63 capteurs, augmentation qui suit une réflexion portant sur l'optimisation du réseau (Besancenot et Sulmont, 2002 ; Besancenot *et al.*, 2003) afin d'obtenir une meilleure couverture du territoire national en lien avec la démographie, tout en conservant les sites les plus anciens.

La période 1993-2002, comme le symbolise la courbe de tendance polynomiale tracée sur le graphique, se caractérise par une longue stagnation avec, parfois, des diminutions assez sensibles d'une année à l'autre du nombre de capteurs en fonctionnement (- 5 appareils entre 1993 et 1994, - 6 entre 1997 et 1998). Le changement occasionné par la création du RNSA, en 1996, à la place de l'Institut Pasteur, qui était à l'origine du réseau de capteurs, a dû s'accompagner de certaines difficultés à implanter de nouveaux appareils. Ainsi, des perspectives budgétaires pessimistes en 1998 sont très certainement à l'origine de la diminution du nombre de capteurs en fonctionnement cette année-là, qui fut « plus difficile que les précédentes » comme l'indique le bulletin des données aéropolliniques françaises, édité par le RNSA à cette époque.

En complément de cet aperçu global, nous présentons dans la figure 5 les cartes d'implantation des capteurs par département en 1996, première année d'existence du RNSA, et 2007.

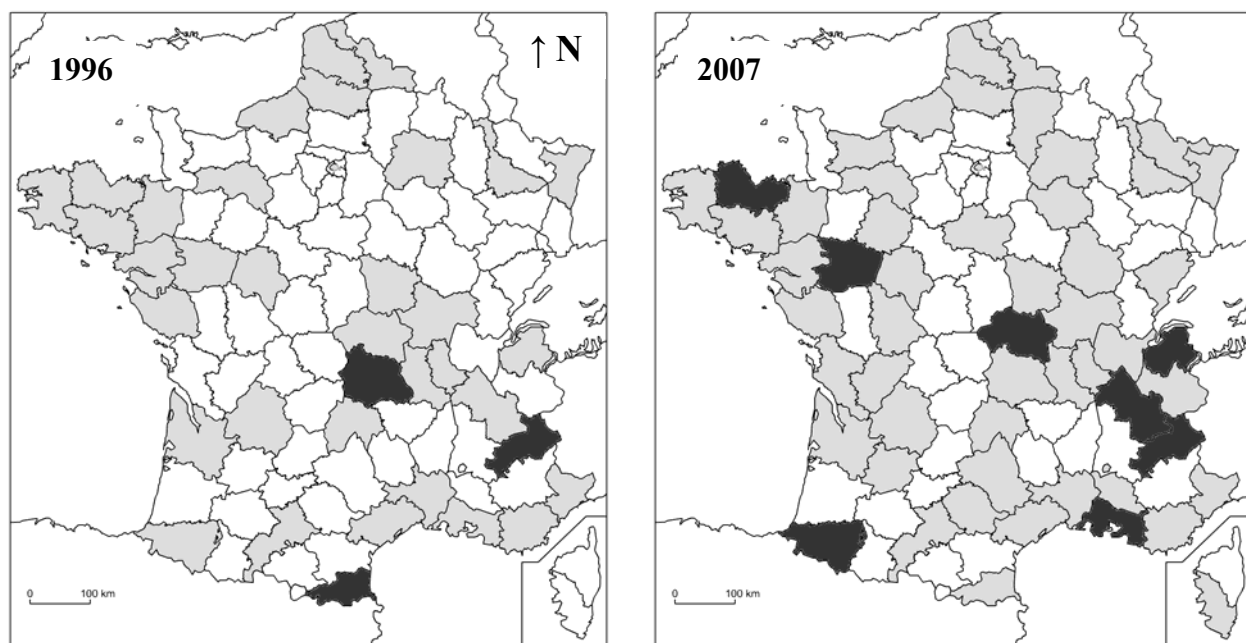


Figure 5 : Implantation des capteurs polliniques, gérés par le RNSA, par département en 1996 et 2007 (source des fonds de carte : www.hist-geo.com). Blanc = absence de capteur ; gris clair = un capteur ; gris foncé = deux capteurs.

A travers ces deux cartes nous constatons bien la diffusion progressive, sur la majeure partie du territoire national, de l'implantation des capteurs. Ainsi le nombre de départements comptant au moins un capteur est passé de 37 à 55, Paris compris. Notons, par ailleurs, que deux capteurs polliniques sont recensés, en 2007, dans huit départements contre trois en 1996. Cependant, nous

constatons que deux départements en ont perdu un entre ces deux dates : les Pyrénées-Orientales et le Puy-de-Dôme. Au vu de ce résultat, une description détaillée du réseau présente un intérêt certain.

Détail de l'évolution du réseau

La hausse générale du nombre de capteurs masque, en réalité, de nombreux cas particuliers avec, notamment, des sites abandonnés et d'autres où des capteurs ont été remis en route après une ou plusieurs années d'arrêt mais, le plus souvent, à des emplacements différents. Le tableau 3 récapitule toutes les évolutions depuis 1987 où seize capteurs, présentés ci-après en fonction de leur région d'implantation, étaient en fonctionnement.

- **Nord-Pas-de-Calais** : Lens
- **Picardie** : Amiens
- **Pays de la Loire** : Nantes
- **Aquitaine** : Bordeaux, Cambo-les-Bains
- **Midi-Pyrénées** : Toulouse
- **Languedoc-Roussillon** : Perpignan, Montpellier
- **Provence-Alpes-Côte-d'Azur** : Marseille
- **Rhône-Alpes** : Grenoble, Annecy, Lyon
- **Franche-Comté** : Lons-le-Saunier, Besançon
- **Alsace** : Strasbourg
- **Ile-de-France** : Paris

sites			
année	mise en route	remise en route	stoppés
1988	Rouen, Cholet, Nice, Saint-Etienne, Montluçon, Nevers, Reims, Nancy		Lons-le-Saunier
1989	Vernet-les-Bains, Dijon		Grenoble, Reims
1990	Périgueux, Tours	Grenoble, Reims	
1991	Briançon, Clermont-Ferrand, Nîmes, Vannes		
1992	Brest, Font-Romeu, Gap, Lille, Quimper, Rennes		Dijon
1993	Dinan, La Bourboule, La Ferté-Macé, La Roche-sur-Yon, Pontivy		
1994	Aurillac, Chalon-sur-Saône, Ploërmel		Besançon, Brest, Dinan, Pontivy, Quimper, Rennes, Vannes, Vernet-les-Bains
1995	Toulon	Brest, Dinan	
1996		Rennes	
1997	Aix-en-Provence, Ajaccio	Dijon, Vernet-les-Bains	
1998	La Rochelle		Lens, Montpellier, Nevers, Perpignan, Ploërmel, Rennes, Vernet-les-Bains
1999	Chambéry, Corte, Saint-Brieuc	Perpignan, Rennes	Ajaccio
2000		Ajaccio, Pontivy	Corte, Périgueux, Perpignan
2001	Caen, La Flèche, Saint-Quentin	Périgueux, Perpignan	Cambo-les-Bains
2002	Angers, Roussillon		Ajaccio, Saint-Brieuc
2003	Mâcon, Poitiers, Pau, Agen	Besançon, Montpellier	La Bourboule, Font-Romeu
2004	Auxerre, Troyes, Avignon, Orléans	Nevers, La Bourboule	Mâcon
2005	Metz, Castres, Millau	Font-Romeu	Auxerre, La Flèche
2006	Le Mans	Saint-Brieuc, Ajaccio	
2007	Angoulême, Annemasse, Bayonne, Bourg-en-Bresse, Saint-Omer, Vichy		La Bourboule, Font-Romeu

Tableau 3 : Récapitulatif des mises en routes, remises en route et arrêts de fonctionnement des sites de capture pollinique, de 1988 à 2007, pour un fonctionnement annuel.

Le tableau 3 montre bien que la progression générale du nombre de capteurs masque des situations assez chaotiques avec plusieurs sites qui ont été stoppés à un moment donné avant d'être remis en route après, parfois, de nombreuses années. L'exemple le plus frappant de ce cas de figure est celui de Besançon. Ce site, déjà opérationnel en 1987, a fonctionné jusqu'en 1993 puis, après une période de neuf ans, un nouveau capteur y a été réinstallé en 2003 dans une région où il n'y en avait plus aucun. Certains sites de captures ont été définitivement abandonnés après plusieurs années de fonctionnement. C'est le cas des sites de Lens et Cambo-les-Bains, opérationnels tous les deux en 1987, où les capteurs fonctionnèrent jusqu'en 1997 et 2000 respectivement. En tout et pour tout, 76 sites différents ont accueilli un capteur pollinique de 1987 à 2007.

Même si cela est anecdotique, deux fabricants se partagent le parc de capteurs : Lanzoni[®] et Burkard[®]. Le mécanisme global de fonctionnement chez les deux constructeurs est identique, excepté pour les tambours qui sont de conception légèrement différente. Ainsi un tambour de marque Lanzoni[®] sera uniquement adaptable sur un capteur de marque identique. Depuis 1997, grâce aux bulletins des données aéropolliniques françaises, il a été possible d'établir la part de chaque constructeur par rapport au nombre total de capteurs en fonctionnement. Il s'avère que la majorité du parc est composé de capteurs de marque Lanzoni[®], dont les proportions année par année sont les suivantes :

- **1997** : 56,8 % (25 sur 44)
- **1998** : 55,3 % (21 sur 38)
- **1999** : 52,4 % (22 sur 42)
- **2000** : 53,7 % (22 sur 41)
- **2001** : 57,8 % (26 sur 45)
- **2002** : 64,4 % (29 sur 45)
- **2003** : 67,3 % (33 sur 49)
- **2004** : 68,5 % (37 sur 54)
- **2005** : 71,4% (40 sur 56)
- **2006** : 72,9 % (43 sur 59)
- **2007** : 76,2% (48 sur 63)

La part des capteurs Lanzoni[®] s'est ainsi accrue au fil des années, particulièrement à partir de 2000. Depuis 2001 tous les sites où des capteurs ont été mis ou remis en route par le RNSA sont de cette marque, sauf pour le site de Besançon où le capteur remis en route en 2003 est un « Burkard[®] ». Les capteurs de cette dernière marque voient ainsi leur nombre stagner entre seize et vingt depuis 1997. En 2007, seuls quinze capteurs de marque Burkard[®] ont fonctionné à :

- Chalon-sur-Saône, Besançon, Pontivy, Dinan, Rennes, La Ferté-Macé, La Roche-sur-Yon, Nantes, Clermont-Ferrand, Aurillac, Annecy, Gap, Briançon, Toulon et Nice.

1.1.2.2 Données de l'étude

Les données les plus anciennes, disponibles dans la base de données du RNSA, remontent à 1987. Les plus longues séries temporelles atteignent donc à peine vingt ans, ce qui est relativement court. Nous avons choisi de présenter ici un certain nombre de séries démarrant au plus tard en 1997 avec, au plus, une année où le capteur n'a pas fonctionné ; la dernière année de la période d'étude étant 2005. Cependant les séries ne pourront pas toutes être utilisées pour l'élaboration de modèles de prévision, à cause d'un certain nombre de raisons qui seront évoquées ultérieurement. Les stations sont présentées en fonction de l'année de mise en route du capteur ou, tout du moins, de l'année la plus ancienne sauvegardée dans la base de données du RNSA :

- **1987** : Paris, Amiens, Lyon
- **1988** : Nancy, Rouen, Strasbourg, Marseille, Bordeaux
- **1989** : Toulouse, Nice, Montluçon, Nantes (sans l'année 1992)
- **1990** : Périgueux (sans l'année 2000)
- **1991** : Clermont-Ferrand, Saint-Etienne, Nîmes, Tours
- **1992** : Brest (sans l'année 1994), Lille, Reims, Gap
- **1993** : La Roche-sur-Yon, La Ferté-Macé
- **1994** : Cholet, Chalon-sur-Saône
- **1995** : Toulon, Aurillac
- **1996** : Grenoble

- **1997** : Aix-en-Provence, Dijon

Avant de s'intéresser aux différents taxons polliniques retenus dans l'étude, il est nécessaire de se pencher, de la manière la plus complète possible, sur la qualité des séries polliniques retenues en se focalisant, tout d'abord, sur les données manquantes affectant ces séries.

1.1.3 Etude des données manquantes

1.1.3.1 Les causes

Toutes les séries de mesures de paramètres météorologiques (température, durée d'insolation...) ou de pollution (ozone, monoxyde de carbone...) peuvent être affectées par des lacunes dues à une panne ou un dysfonctionnement du matériel de mesure ou d'enregistrement des données. Les séries polliniques ne sont évidemment pas épargnées par ce phénomène, mais il s'avère malaisé de déterminer l'origine de ces lacunes. En effet, les jours concernés par des données manquantes, c'est-à-dire par une absence de relevé, sont éliminés des fichiers de la base de données polliniques du RNSA. Mais les jours où des quantités nulles sont enregistrées du fait de conditions climatiques défavorables, comme cela peut se produire en plein hiver ou au début de l'automne, sont également supprimés de la base de données. Ainsi il est impossible de savoir précisément l'origine des lacunes, ce qui est quelque peu problématique si on ne peut disposer de l'historique de fonctionnement du site de capture. A partir des séries de Dijon et de Chalon-sur-Saône, dont les analyses sont effectuées par le laboratoire Climat et Santé depuis 2002 (la maintenance du capteur de Dijon y étant assurée depuis 1996 ; le RNSA ne disposant des données que depuis 1997), nous allons tenter de détailler les différentes causes qui sont à l'origine des lacunes enregistrées dans ces stations. Elles sont présentées ci-après :

- La cause principale provient d'une absence effective de relevé, qui peut avoir plusieurs origines :
 - **Dysfonctionnements du capteur caractérisés par une panne générale, un problème d'aspiration ou une panne de courant** : le premier cas engendre des arrêts d'environ trois semaines correspondant au temps de réparation. Les capteurs sont très majoritairement relevés une fois par semaine si bien qu'un problème intervenant pendant celle-ci (défaut d'aspiration par exemple) n'est détecté que le jour du relevé. Les pannes de courant entraînent un arrêt du capteur qui se remet en route une fois que celui est rétabli, mais, à ce moment là, la bande adhésive n'est pas «impactée» entièrement. Dans ce dernier cas, si la date et la durée de la coupure ne sont pas connues, le comptage pollinique qui suit le moment de la coupure est biaisé ; les tranches bihoraires enregistrées ne correspondent pas à la réalité.
 - **Problèmes d'envoi ou de réception des tambours** : ils concernent des sites qui gèrent l'analyse et la préparation de nouvelles bandes adhésives pour des tambours provenant de plusieurs sites. Les retards de distribution des paquets peuvent être à l'origine de lacunes de courte durée quand les tambours à changer n'arrivent pas à temps. De même les jours fériés, s'ils ont lieu le jour du relevé, sont un problème quand le service de distribution et/ou la personne relevant le capteur n'arrivent pas à s'organiser pour que l'acheminement se fasse sans accroc.
 - **Absences des personnes s'occupant du relevé des capteurs** : plusieurs sites ne sont gérés que par une seule personne dont les congés entraînent des lacunes si elle n'a pu trouver de remplaçants. Les lacunes de sept ou quatorze jours rencontrées dans les séries semblent avoir pour origine ce cas-là. Il semble donc nécessaire d'avoir deux personnes titulaires pour chaque site d'analyse afin de pallier ce problème.

Malheureusement, il s'avère impossible de discerner l'origine des lacunes parmi les trois cas décrits ci-dessus. Dans les fichiers de données du RNSA, le jour suivant une période manquante est théoriquement caractérisé, du moins pendant les premières tranches bihoraires, par le sigle EEEE s'il y a eu un problème technique se caractérisant par une absence de données (mention «jour non lu» dans les fichiers envoyés au RNSA). Or ce sigle apparaît trop peu souvent dans l'ensemble des stations étudiées, si bien qu'il n'est pas exploitable. Il semble évident, en effet, que des périodes manquantes dues à des problèmes techniques ne sont pas symbolisées par le sigle EEEE. Ainsi l'origine de trois semaines de données lacunaires à La Roche-sur-Yon en 2002, due très certainement à une panne générale du capteur qui a nécessité réparation, n'est pas clairement définie à cause de l'absence de ce sigle. L'origine des lacunes est donc impossible à déterminer de manière sûre et certaine.

1.1.3.2 Détermination des causes : le cas de Chalon-sur-Saône et de Dijon

L'origine des valeurs manquantes ne pouvant pas être définie très précisément à partir des fichiers de données, nous ne pouvons les déterminer que pour les stations dont la gestion nous est attribuée. Ainsi, le laboratoire Climat et Santé gérant l'analyse des sites de Chalon-sur-Saône et Dijon, nous allons tenter de déterminer la cause des données manquantes de ces deux séries, en fonction des différents cas de figure décrits dans la section 1.1.3.1 :

- **Chalon-sur-Saône**

- 2 au 8 juillet 1996 (sigle EEEE à la fin qui est sensé signaler un problème technique)
- 15 au 21 avril 1997
- 5 mai et 13 juillet 1999
- 21 au 26 juillet et 14 septembre 1999
- 10 au 12 juin 2003 (lacunes probablement en lien avec le lundi de Pentecôte et un problème d'envoi)
- 10 au 14 juillet 2003 (probabilité d'un problème d'envoi, le 14 juillet tombant un lundi)

- **Dijon**

- 24 septembre 1999
- 2 au 8 septembre 2002
- 28 juillet au 3 août 2003
- 15 au 21 septembre 2003

Les périodes où aucune hypothèse, inscrite entre parenthèses, n'est émise sont marquées par une des trois causes, sans qu'il soit possible d'affirmer que ce soit l'une ou l'autre. On peut supposer que les lacunes d'une semaine sont dues à une absence qui a empêché le relevé du tambour (notamment à Dijon), et que les périodes lacunaires de courte durée sont liées à des problèmes d'envoi ou de réception des tambours.

Les lacunes non présentées sont dues à trois autres cas de figure particuliers, qui sont évoqués ci-après :

- Nous avons tout d'abord constaté que des jours présentant des lacunes dues à des problèmes techniques ont été comptabilisés comme des jours où n'a été enregistré aucun pollen. La présence de concentrations non négligeables les jours entourant ces périodes lacunaires semblent indiquer à coup sûr qu'il y a eu une confusion dans la lecture de ces bandes, qui auraient dû être comptabilisées comme «jours non lu». Ce fut le cas pour les dates suivantes à Dijon en 2004 :

- 15 et 16 avril
- 12 au 16 mai, 18 mai, 20 mai, 27 mai au 6 juin

Grâce aux lames de lecture pollinique de cette époque (figure 6), nous avons pu déterminer l'origine du problème :

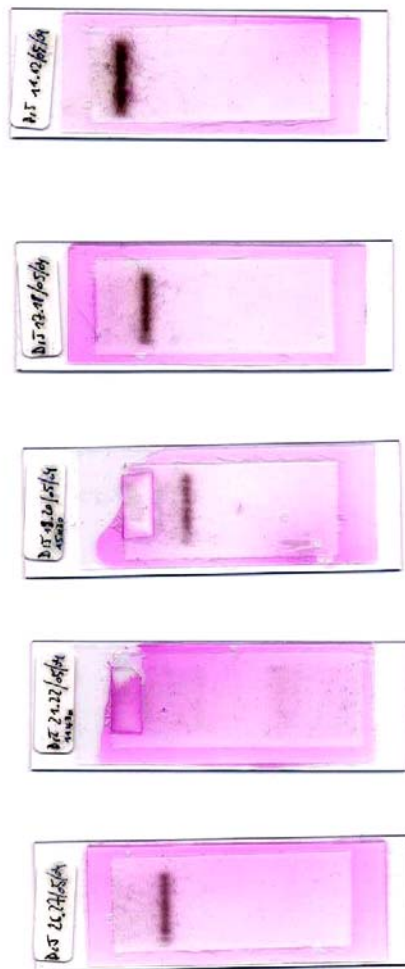


Figure 6 : Lames de lecture de Dijon (11-12 mai, 17-18 mai, 19-20 mai et 26-27 mai 2004).

Les lacunes sont ici dues à un problème d'aspiration dû à un dysfonctionnement de l'horloge mécanique qui permet de faire tourner le tambour. La bande noire, présente sur les quatre lames servant d'exemple, est causée par l'arrêt du tambour. L'aspiration se faisant toujours (donc il n'y a pas eu de problème électrique ni de panne moteur), toutes les particules aspirées se sont agglutinées sur une même zone dont la largeur correspond à celle de la buse d'aspiration. Cela corrobore l'hypothèse d'un problème d'horloge qui a dû être changée entre le 27 mai et le 6 juin. Passé cette date, il n'y eut plus de problème de ce type.

Ce dernier exemple montre que des jours présentant des données lacunaires ont été enregistrés, de manière erronée, comme des jours caractérisés par des quantités nulles (il semble que ce fut également le cas à Chalon-sur-Saône les 7 mars et 10 mai 2000). Cette erreur d'interprétation est minimisée par le fait qu'il n'y a pas de distinction entre les deux catégories dans la base de données du RNSA.

- La deuxième catégorie de données lacunaires – qui n'en sont d'ailleurs pas – rencontrée dans la base se caractérise par les jours enregistrant des concentrations nulles. Comme cela a déjà été évoqué ci-dessus, les jours présentant des conditions climatiques particulièrement défavorables, des températures particulièrement basses le plus souvent, sont comptés comme des jours où le relevé n'a pas été effectué. A Chalon-sur-Saône cela concerne les jours suivants :

- 2 et 5 février 2005

A Dijon les jours concernés sont plus nombreux :

- 23 janvier 2000
- 19 et 21 janvier 2001
- 23 janvier, 26 janvier et 8 février 2003
- 23 février 2005

- Le troisième cas de figure rencontré concerne un problème de transmission des comptages polliniques. Ainsi des données polliniques existantes n'ont pas été enregistrées dans la base de données du RNSA. Il est possible que ces données n'aient pu être analysées et envoyées à temps pour le bulletin allerge-pollinique et qu'il y ait eu un oubli dans la transmission des données de la part de l'analyste. Mais de manière systématique, à la fin de chaque année, le RNSA demande à chaque site de justifier l'existence d'éventuelles lacunes, tout du moins enregistrées comme telles, dans leur base de données. L'origine de ces problèmes semble plus vraisemblablement d'ordre informatique avec un problème de réception de la part du RNSA caractérisé, probablement, par une faille dans le traitement de certaines données par le système informatique gérant la base de données. Cette explication semble la plus plausible car d'autres types de problèmes informatiques infectèrent la base de données (absence de

29 février les années bissextiles pour certains sites, présence de 29 février pour toutes les années pour d'autres). Ce problème de réception a touché Dijon pour les jours suivants :

- 31 mars, 7 avril et 26 mai 1997
- 4 au 8 juin, 7 au 10 juillet, 28 juillet au 3 août 1998

Les concentrations des pollens allergisants de l'étude ne furent pas négligeables pendant les jours cités. Ainsi il y eut 528 grains/m³ de bouleau le 31 mars 1997 et 389 grains/m³ de Poacées le 5 juin 1998.

Disposant des données analysées d'origine pour les sites de Dijon et Chalon-sur-Saône, nous avons pu déterminer les différentes causes à l'origine de la présence de périodes ou de jours isolés manquants dans la base de données du RNSA. Malheureusement, au vu de l'organisation des fichiers, il est impossible de déterminer la cause exacte ayant conduit à des données lacunaires et plus précisément de faire la différence entre des jours «non lus» à cause de problèmes «techniques» et ceux où n'a été enregistré aucun grain de pollen. Nous pouvons, seulement, raisonnablement penser que les lacunes ayant lieu de mars à août ont des origines «techniques». Cette non-différenciation est assez problématique mais n'est pas dramatique car on peut se référer aux fichiers envoyés de manière hebdomadaire, s'ils ont été conservés, pour établir l'origine de la lacune. Plus ennuyeux est la perte des données par une transmission et une réception qui a été, à quelques occasions, défectueuse de la part du système informatique gérant la base de données du RNSA. Cela n'a concerné Dijon qu'en 1997 et 1998. Mais il faut espérer que les autres sites retenus dans ce travail n'ont pas été affectés par ce problème ou très peu.

1.1.3.3 Répartition générale des valeurs manquantes

Nous allons étudier à présent la proportion de données manquantes pour les trente villes présentées dans la section 1.1.2.2. Du fait de l'impossibilité de différencier les jours présentant des données manquantes «réelles», c'est-à-dire causées par un problème qui a empêché le capteur de fonctionner normalement, des jours enregistrant des concentrations polliniques nulles, nous avons retenu ici les mois de mars à août, car on peut raisonnablement penser que les

concentrations polliniques ne sont jamais nulles durant cette période comme cela a été évoqué à la fin de la section 1.1.3.2. Les mois de janvier et février, par des conditions de froid qui peuvent empêcher la libération des pollens, et les mois de septembre (surtout la fin) et automnaux, qui correspondent à l'entrée en dormance de la végétation, peuvent enregistrer des concentrations nulles excepté dans les stations du pourtour méditerranéen où la pollinisation se fait en quasi continu. D'ailleurs pour ces raisons, la plupart des stations, comme cela sera évoqué plus tard, sont généralement mises en route courant février et sont stoppées au tout début du mois d'octobre, à l'exception notable de quelques stations méditerranéennes dont les capteurs fonctionnent théoriquement de manière continue depuis quelques années.

Nous allons présenter la proportion de données manquantes des 30 villes de l'étude pour l'ensemble des années disponibles jusqu'en 2005. Les villes sont classées, ci-dessous, en fonction de l'année disponible la plus ancienne. Entre parenthèses sont indiquées les années sans comptes polliniques.

- **1987** : Paris, Amiens, Lyon
- **1988** : Nancy, Rouen, Strasbourg, Marseille, Bordeaux
- **1989** : Toulouse, Nice, Montluçon, Nantes (sans l'année 1992)
- **1990** : Périgueux (sans l'année 2000)
- **1991** : Clermont-Ferrand, Saint-Etienne, Nîmes, Tours
- **1992** : Brest (sans l'année 1994), Lille, Reims, Gap
- **1993** : La Roche-sur-Yon, La Ferté-Macé
- **1994** : Cholet, Chalon-sur-Saône
- **1995** : Toulon, Aurillac
- **1996** : Grenoble
- **1997** : Aix-en-Provence, Dijon

La figure 7 présente la part de données manquantes propre à chaque ville.

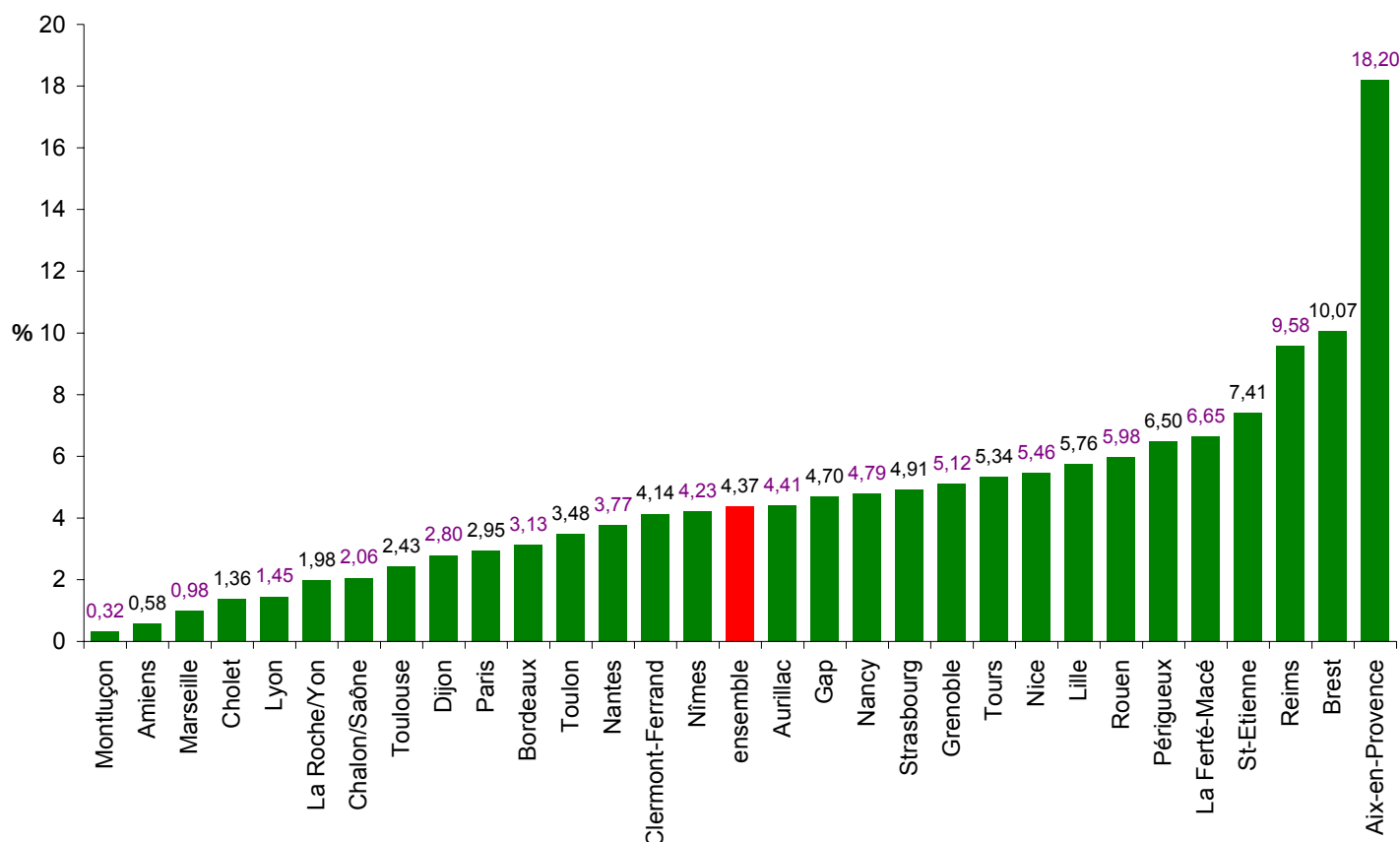


Figure 7 : Classement par ordre croissant de la proportion de données manquantes sur la totalité des années disponibles ville par ville.

Nous pouvons remarquer aisément une grande disparité au niveau de la qualité des données. Ainsi six villes présentent moins de 2% de valeurs manquantes : Montluçon, Amiens, Marseille, Cholet, Lyon et La Roche-sur-Yon. Ce chiffre est très honorable et il faut signaler tout spécialement le caractère complet des séries de Montluçon et Amiens qui n'enregistrent que 0,32 et 0,58% de valeurs manquantes et qui sont parmi les sites les plus anciens (les premières données polliniques datent de 1987 à Amiens et de 1989 à Montluçon) du réseau national de capteurs. A l'opposé trois stations se caractérisent par des lacunes importantes : Reims, Brest et surtout Aix-en-Provence (18,20%) où la qualité des fichiers est assez calamiteuse de 1997 à 2002 (25,81% de données manquantes). La proportion moyenne globale s'élève à 4,37%, cette valeur semble acceptable car sur la période mars-août complète, qui représente 184 jours, cela équivaut à huit jours affectés par une absence de données.

Afin de mieux étudier les différences en termes de données manquantes, nous avons retenu la période 1997-2005, commune à toutes les villes, excepté le site de Périgueux qui est retiré ici de l'analyse car le capteur n'a pas fonctionné en 2000. Nous allons, tout d'abord, présenter le même type de graphique que celui de la figure 7 :

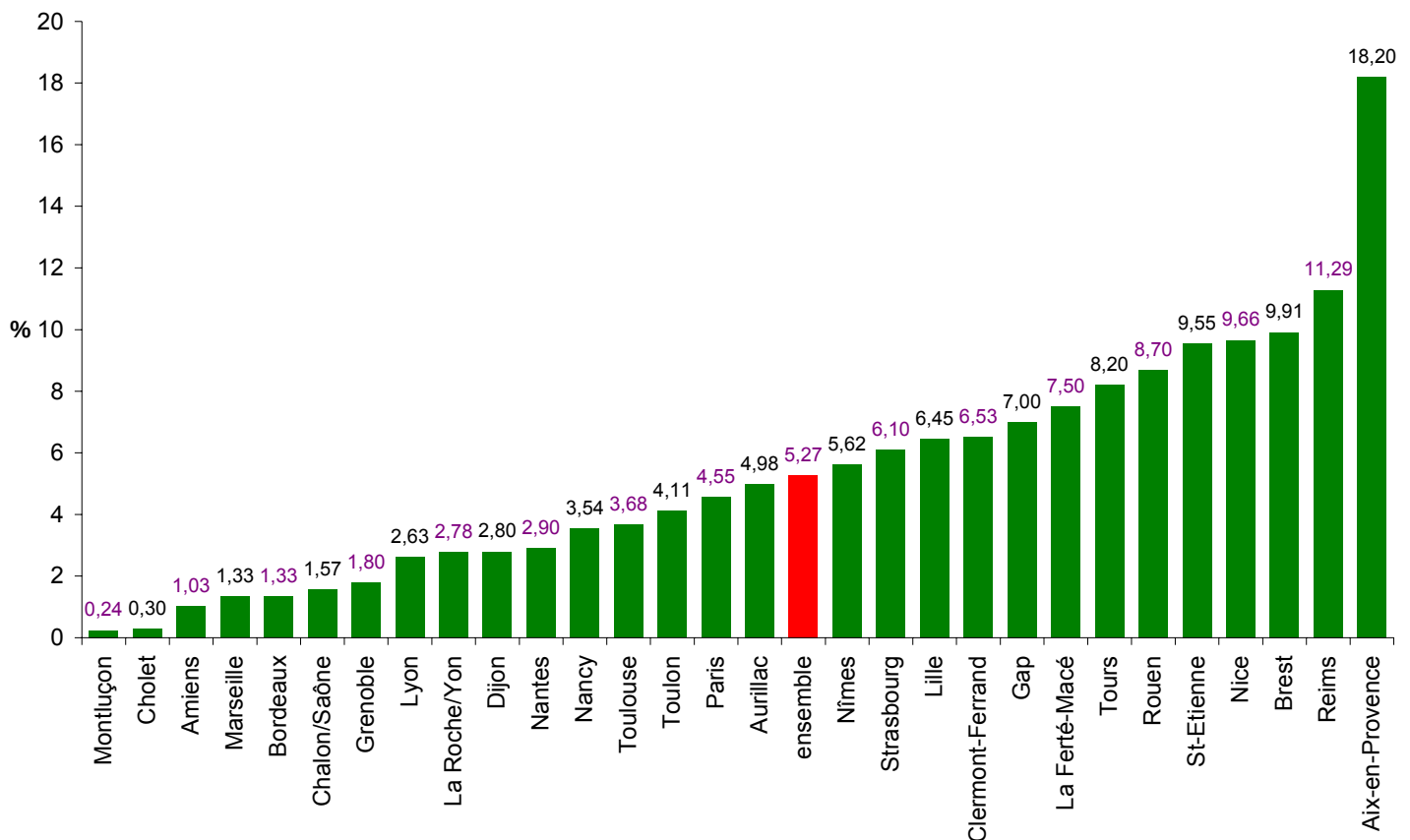


Figure 8 : Classement par ordre croissant de la proportion de données manquantes, ville par ville, sur la période 1997-2005.

On peut remarquer ici (figure 8) que la proportion moyenne globale de données manquantes sur la période 1997-2005 (5,27%) est sensiblement supérieure à celle incluant toutes les années disponibles (4,37% et 4,30% si l'on retire Périgueux). Cela semblerait indiquer une légère détérioration de la qualité des séries sur les années les plus récentes. En effet sur les 27 villes ayant des séries démarrant avant 1997, seules sept (Nancy, Brest, Bordeaux, Chalon-sur-Saône, Montluçon, Nantes et Grenoble) présentent une plus faible proportion de valeurs manquantes sur la période 1997-2005 que sur l'ensemble des années disponibles propres à chaque ville. On peut

noter de même que le classement des villes diffère quelque peu de celui présenté dans le graphique précédent.

L'évolution interannuelle de la proportion de données lacunaires va être ensuite étudiée afin de déterminer une éventuelle évolution à la hausse ou à la baisse de la proportion de valeurs manquantes (figure 9).

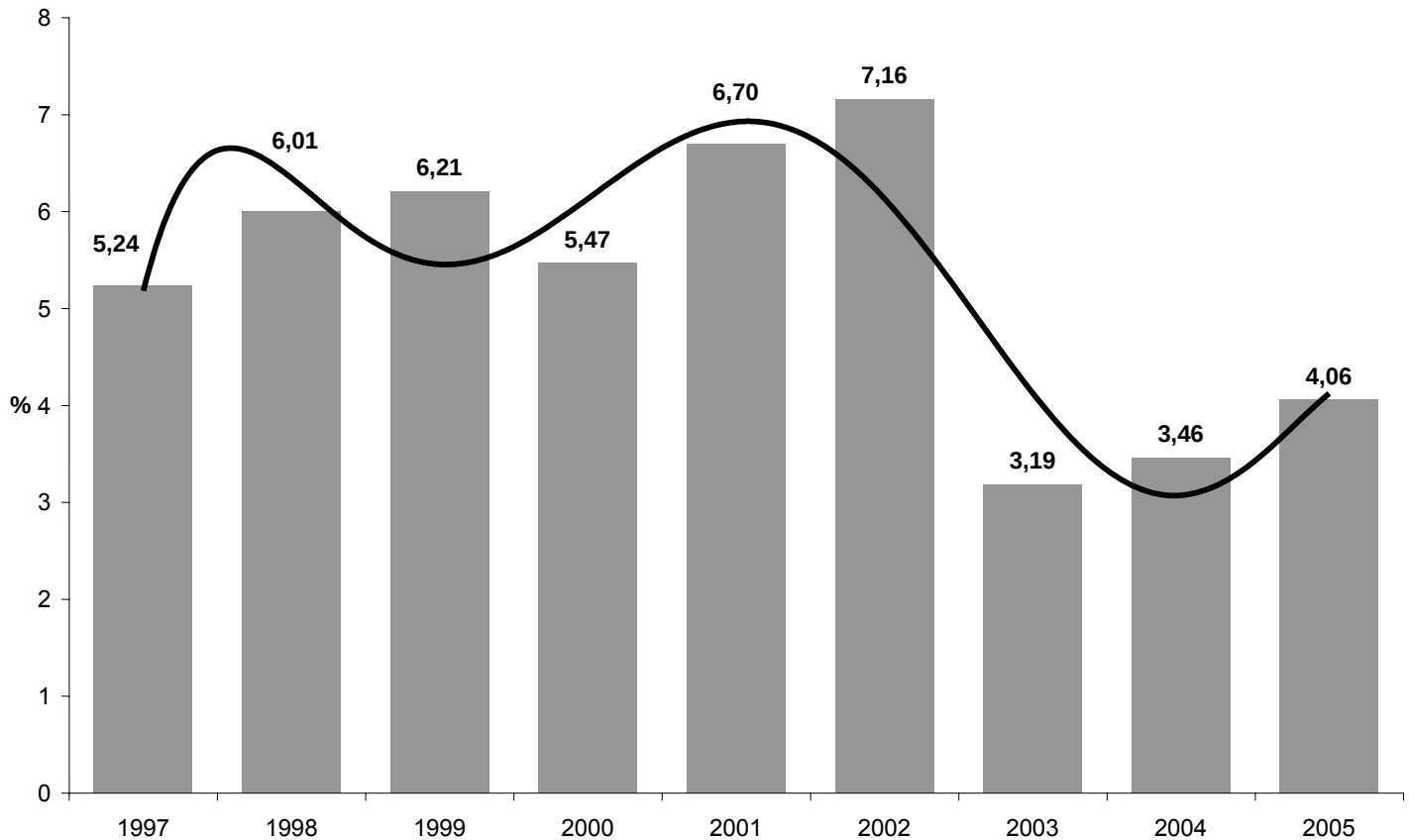


Figure 9 : Proportion annuelle de données manquantes pour les 29 villes ayant des données sur la période 1997-2005.

L'évolution de la part de données lacunaires suit une courbe polynomiale. On peut noter une proportion supérieure à 5% sur la période 1997-2005, avec un pic à 7,16% en 2002, puis une rupture dès 2003 marquée par des valeurs plus faibles. Il semble donc qu'il y ait une amélioration générale tangible du caractère complet des données polliniques dès 2003, un résultat qui restera à confirmer les années suivantes.

1.1.3.4 Etude de la variation interannuelle des valeurs manquantes par ville

Nous allons maintenant évoquer plus en détail, sur la période 1997-2005, la qualité des séries de chaque station en ne présentant pas la proportion moyenne de valeurs manquantes, mais les valeurs année par année. Les valeurs extrêmes de chaque ville sont présentées ci-après selon le classement, par ordre croissant, de la proportion de données manquantes :

- **Montluçon** : de 0% (1997 à 1999, 2002 et 2004) à 0,58% (2003)
- **Cholet** : de 0% (2001 à 2005) à 1,09% (1997)
- **Amiens** : de 0% (1997 et 2003) à 2,72% (2000)
- **Marseille** : de 0% (2000 et 2002) à 6,52% (2001)
- **Bordeaux** : de 0% (1997-1998 et 2000-2001) à 5,08% (1999)
- **Chalon-sur-Saône** : de 0% (1998, 2001, 2002 et 2005) à 4,35% (1999 et 2003)
- **Grenoble** : de 0% (1998, 2000-2001 et 2004) à 7,22% (2002)
- **Lyon** : de 0% (1998-1999 et 2002) à 9,24% (2001)
- **La Roche-sur-Yon** : de 0% (1998 et 2005) à 11,41% (2002)
- **Dijon** : de 0% (1999 à 2002, et 2005) à 10,87% (2004)
- **Nantes** : de 0% (1997-1998) à 19,28% (2001)
- **Nancy** : de 0% (1998 et 2003) à 10,99% (1997)
- **Toulouse** : de 0,54% (2000) à 7,61% (2004)
- **Toulon** : de 0% (2000 et 2001) à 10,87% (1998)
- **Paris** : de 0% (1997 et 2003) à 17,93% (2002)
- **Aurillac** : de 0,54% (2002) à 11,11% (2000)
- **Nîmes** : de 0% (1997, 2004 et 2005) à 14,67% (1998)
- **Strasbourg** : de 0% (1998) à 20,65% (1997)
- **Lille** : de 0% (1997, 2004 et 2005) à 20,61% (2001)
- **Clermont-Ferrand** : de 0% (1998) à 19,57% (2000)
- **Gap** : de 0% (1997) à 23,91% (2002)
- **La Ferté-Macé** : de 0,54% (2002) à 21,31% (2005)
- **Tours** : de 0% (1997 et 2001) à 45,51% (1998)
- **Rouen** : de 3,80% (2002) à 14,67% (1999)

- **Saint-Etienne** : de 0,54% (2003 et 2005) à 25,54% (2001)
- **Nice** : de 0% (1999 et 2005) à 53,80% (2002)
- **Brest** : de 1,20% (2003) à 30,34% (2000)
- **Reims** : de 0% (2004) à 49,07% (1999)
- **Aix-en-Provence** : de 0% (2005) à 39,47% (1997)

Tous ces résultats montrent une grande disparité d'une année à l'autre pour chaque ville. Ainsi celles présentant les séries les plus lacunaires peuvent comporter des années sans aucune valeur manquante, comme c'est le cas à Reims en 2004 et Aix-en-Provence en 2005. A partir de ce constat, nous allons recenser les années, ville par ville sur la période 1997-2005, n'ayant aucune lacune et celles dont les comptes sont affectés par, au moins, 10% de valeurs manquantes :

- **Nombre d'années sans lacunes :**
 - **5** : Montluçon, Cholet, Dijon
 - **4** : Bordeaux, Chalon-sur-Saône, Grenoble
 - **3** : Lyon, Nîmes, Lille
 - **2** : Amiens, Marseille, La Roche-sur-Yon, Nantes, Nancy, Toulon, Paris, Tours, Nice
 - **1** : Strasbourg, Clermont-Ferrand, Gap, Reims, Aix-en-Provence
 - **0** : Toulouse, Aurillac, La Ferté-Macé, Rouen, Saint-Etienne, Brest
- **Nombre d'années présentant une part de données manquantes supérieure ou égale à 10% :**
 - **0** : Montluçon, Cholet, Amiens, Marseille, Bordeaux, Chalon-sur-Saône, Grenoble, Lyon, Toulouse
 - **1** : La Roche-sur-Yon, Dijon, Nantes, Nancy, Toulon, La Ferté-Macé, Tours
 - **2** : Paris, Aurillac, Nîmes, Strasbourg, Lille, Clermont-Ferrand, Gap, Nice
 - **3** : Saint-Etienne, Brest
 - **4** : Rouen, Reims
 - **6** : Aix-en-Provence

Sur l'ensemble des 261 années de l'étude (29 stations par 9 années), 59 (22,61%) ne présentent aucune lacune alors que 43 (16,48%) sont affectées par au moins 10% de données manquantes. Les pourcentages non négligeables de ces deux extrêmes montrent bien le caractère varié des situations d'une année à l'autre. De ce fait, la proportion moyenne de valeurs manquantes, pour une ville donnée, peut être fortement influencée par une ou deux années où de fréquents problèmes de collecte des données ont été rencontrés. Pour déterminer cela, nous allons considérer que chaque année présente une saison s'étendant du 1^{er} mars au 31 août complète (c'est-à-dire 184 jours). Ainsi, pour les années où la saison mars-août est inférieure à 184 jours, nous allons calculer le nombre de jours théoriquement manquants. Pour déterminer la variabilité interannuelle des lacunes, nous allons nous baser, d'une part, sur l'écart-type qui permet d'évaluer la valeur brute, en jours, de cette variabilité et, d'autre part, sur le coefficient de variation qui correspond au rapport écart-type sur moyenne. Ces deux paramètres sont présentés dans le tableau 4 avec la moyenne annuelle du nombre de jours manquants rapportée à une saison de 184 jours ; les stations sont classées par ordre croissant de moyenne.

villes	nombre moyen de jours manquants	écart-type (en jours)	coefficient de variation (en %)
Montluçon	0,45	0,54	118,66
Cholet	0,56	0,73	130,77
Amiens	1,89	1,69	89,36
Marseille	2,44	3,71	151,85
Bordeaux	2,48	3,45	138,73
Chalon-sur-Saône	2,89	3,66	126,53
Grenoble	3,26	4,35	133,45
Lyon	4,82	5,44	113,02
La Roche-sur-Yon	5,11	7,56	147,81
Dijon	5,26	7,98	151,74
Nancy	6,49	6,96	107,30
Toulouse	6,78	4,02	59,37
Nantes	7,17	11,59	161,58
Toulon	7,56	7,18	95,01
Paris	8,26	11,42	138,28
Aurillac	9,44	6,92	73,39
Nîmes	10,33	9,82	95,07
Strasbourg	11,22	13,87	123,62
Lille	11,55	13,17	114,07
Clermont-Ferrand	12,02	11,45	95,24
Gap	12,89	13,52	104,92
La Ferté-Macé	13,81	11,10	80,41
Tours	15,53	26,22	168,88
Rouen	16,11	7,08	43,94
Brest	17,29	17,80	102,97
Nice	17,78	31,23	175,66
Saint-Etienne	17,82	16,86	94,64
Reims	22,27	28,10	126,20
Aix-en-Provence	34,65	26,85	77,50

Tableau 4 : Durée moyenne annuelle, écart-type et coefficient de variation liés aux données manquantes sur la période 1997-2005.

Le tableau 4 met en évidence la forte variabilité interannuelle de la proportion de données lacunaires, symbolisée par des écarts-types supérieurs au nombre de jours moyen enregistrant des données manquantes. La valeur des écarts-types étant étroitement liée à cette moyenne, nous allons nous intéresser plus particulièrement au coefficient de variation. Certaines stations semblent très marquées par une ou deux années particulièrement incomplètes qui gonflent la part totale de données manquantes. C'est particulièrement le cas pour les villes suivantes où il faut

noter que le nombre de jours lacunaires est celui qu'il y a eu ou aurait dû y avoir si la période mars-août était systématiquement couverte en totalité, soit 184 jours par an :

- **Dijon** avec un coefficient de variation de 151,74% à cause des années 1998 et 2004 (17,32 et 20 jours manquants sur un total de 47,32).
- **Marseille** avec un coefficient de variation de 151,85% à cause de l'année 2001 (12 jours manquants sur un total de 22).
- **Nantes** avec un coefficient de variation de 161,58% à cause de l'année 2001 (35,47 jours manquants sur un total de 64,53). En réalité, il n'y eut que 16 valeurs manquantes cette année-là mais sur une saison de seulement 83 jours. Cela a pour effet de gonfler le poids de cette année et, par conséquent, le coefficient de variation.
- **Tours** avec un coefficient de variation de 168,88% causé par l'année 1998 (83,74 jours sur un total de 139,74).
- **Nice** avec un coefficient de variation record de 175,66% causé par l'année 2002 (99 jours sur un total de 160).

Les cinq villes présentées sont celles où l'écart-type est plus d'une fois et demi supérieur à la moyenne des données manquantes. Ce sont celles qui sont les plus marquées, statistiquement parlant, par une ou deux années à problème. Le point positif pour ces stations réside dans la non-réurrence de telles lacunes qui permettent de ne pas les écarter entièrement dans le cadre d'une analyse statistique. A l'opposé un nombre minoritaire de stations présente un coefficient de variation inférieur à 100%. Les cinq villes présentant les séries avec les rapports écart-type/moyenne les plus faibles sont les suivantes :

- **La Ferté-Macé** avec un coefficient de variation de 80,41%.
- **Aix-en-Provence** avec un coefficient de variation de 77,50%.

- **Aurillac** avec un coefficient de variation de 73,39%.
- **Toulouse** avec un coefficient de variation de 59,37%.
- **Rouen** avec un coefficient de variation très faible, comparé aux autres villes, de 43,94%.

Ces valeurs plus faibles indiquent que l'on retrouve, année après année, un nombre de jours manquants relativement proche pour ces villes. Pour Toulouse ou Aurillac, dont la part de données manquantes est inférieure à 5%, cela n'est pas problématique. Par contre pour la station d'Aix-en-Provence cela démontre bien la médiocrité du caractère exhaustif des séries de cette ville. Mises à part les années 2003 à 2005, la fréquence des lacunes excédait 20% sur quatre années ce qui rend inexploitable une telle série. Le cas de Rouen est beaucoup moins dramatique (8,70% de données manquantes sur la période 1997-2005 contre 18,20% à Aix-en-Provence) mais on peut noter que depuis 1997 au moins sept jours de données manquantes par an sont enregistrés, et le plus souvent une dizaine de jours avec un pic à 27. Il semble impossible pour cette station de présenter une année sans lacune ou quasiment sans. Cela peut poser problème si ces lacunes récurrentes ont lieu à des moments critiques, tels le démarrage de la pollinisation d'une espèce allergisante.

Nous avons réalisé un aperçu assez complet sur l'importance relative des données manquantes dans les séries polliniques de chaque station ainsi que de leur évolution interannuelle. Nous allons étudier, relativement brièvement, les caractéristiques ayant trait à la durée des périodes lacunaires.

1.1.3.5 Etude de la durée moyenne des périodes lacunaires

Dans cette section nous allons étudier la durée moyenne des périodes lacunaires pour la série 1997-2005 de mars à août. Les résultats sont présentés dans la figure 10 :

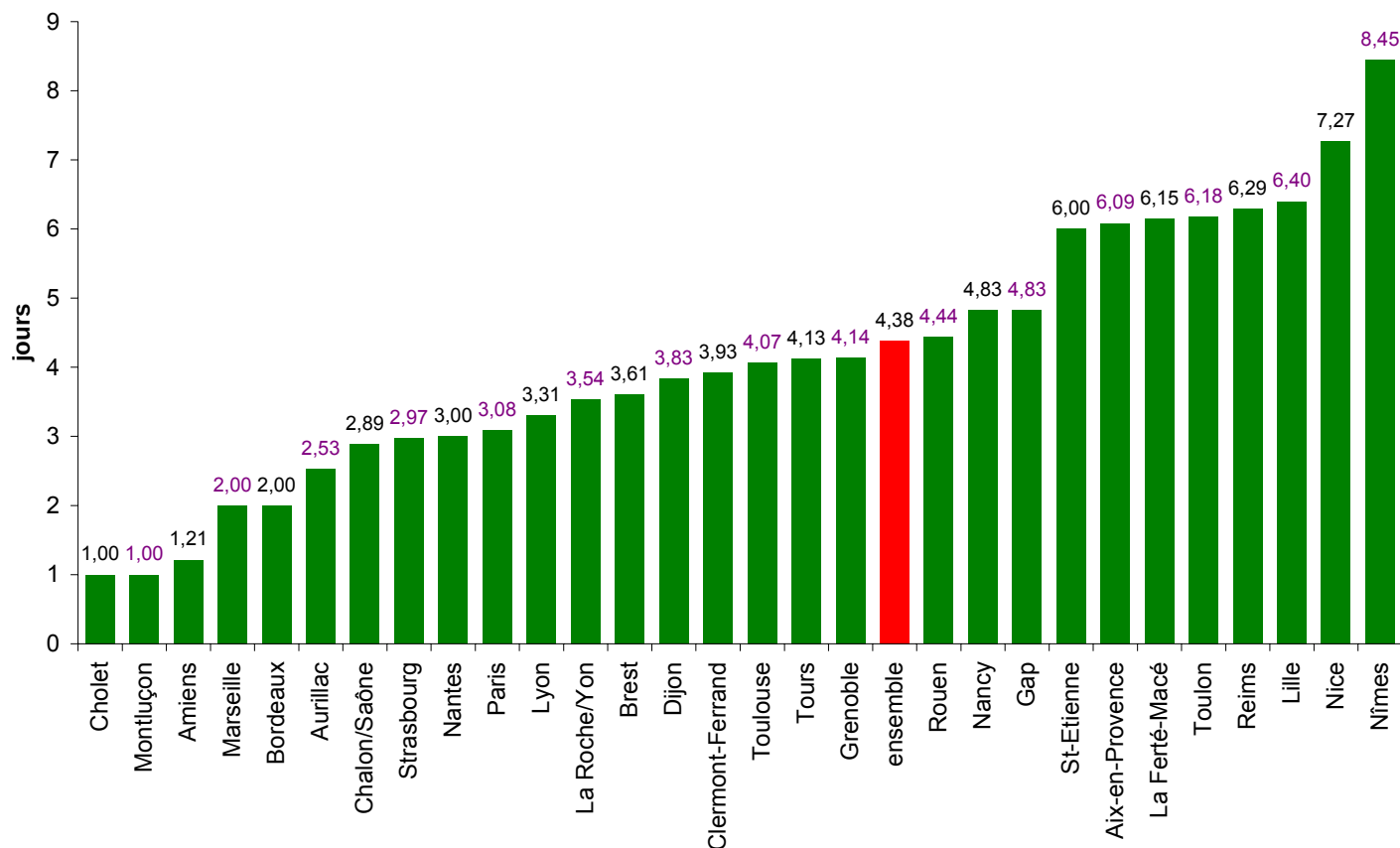


Figure 10 : Classement par ordre croissant de la durée moyenne des périodes lacunaires, ville par ville, sur la série 1997-2005.

La durée moyenne d'une période de données manquantes est de 4,38 jours de mars à août. Elle s'échelonne de un jour à Cholet et Montluçon, à 8,45 à Nîmes. La durée moyenne des périodes lacunaires est relativement liée à leur part dans les fichiers de données polliniques, à savoir que les séries les plus complètes sont caractérisées par les périodes manquantes, en moyenne, les plus brèves. Cette relation est un peu moins nette pour les séries présentant une plus grande part de valeurs lacunaires notamment avec l'exemple de Brest, troisième station la plus lacunaire de 1997 à 2005 avec 9,91% de valeurs manquantes, mais dont la durée moyenne des lacunes (3,61 jours) est inférieure à celle de l'ensemble des stations. On peut également noter, qu'à l'exception de Marseille, les stations du pourtour méditerranéen présentent des périodes lacunaires en moyenne supérieures à six jours ; Nice et Nîmes présentent d'ailleurs les valeurs les plus élevées de l'ensemble des 29 stations. La durée des périodes étant très variable, nous allons nous focaliser sur deux durées qui peuvent être explicables : un et sept jours.

Les jours isolés sont très certainement dus à une installation d'un tambour muni d'une bande neuve un jour après le relevé du précédent tambour. Ce phénomène est très certainement causé par une livraison des tambours prêts à être installés dans le capteur qui n'a pas été faite à temps. On peut aussi supposer, même si cela semble extrêmement rare de mars à août, qu'un jour manquant isolé corresponde à une absence réelle de pollens impactés. Enfin, il se peut que le tambour n'ait tourné que pendant six jours, à cause d'une panne électrique ou d'un problème d'horloge, si bien qu'il n'y aura pas de pollens impactés sur la fin de la bande. Sur les 29 stations de l'étude, pour la période 1997-2005, ces jours manquants isolés représentent 40,18% du nombre total de périodes lacunaires mais, seulement, 9,16% de la durée totale de ces périodes. Néanmoins sans ces lacunes, qui pour la plupart devraient être évitables si la distribution se fait normalement, la proportion totale de données manquantes diminuerait de quasiment 0,5 point, passant de 5,27 à 4,79%.

Les périodes manquantes de sept jours sont dues à une absence de tambour dans le capteur pendant une semaine. Le tambour relevé le jour J n'a pas été remplacé de suite mais exactement une semaine après. Cela peut être dû à un problème technique comme un décollement de la bande sur le tambour, phénomène rare survenu une fois à Dijon en 2007, mais aussi à un défaut de personnel qui a entraîné l'impossibilité de changer le tambour, voire un oubli dans l'envoi de l'analyse hebdomadaire. Dans le deuxième cas, il semble nécessaire qu'il y ait deux personnes titulaires s'occupant du capteur pour assurer un relevé en continu et que les tambours «neufs» soient prêts à l'avance... mais tout cela n'est pas évident à organiser. Ces lacunes hebdomadaires représentent 15,22% du nombre total de périodes manquantes mais 24,30% de la durée totale liée à ces périodes auxquelles il faudrait ajouter les périodes de six jours qui ont très certainement la même origine, mais où les demi-journées de comptes ont été enregistrées. Sans ces lacunes hebdomadaires, la proportion totale de données manquantes diminuerait de 1,3 point passant de 5,27 à 3,99%. En y additionnant les jours lacunaires isolés, le pourcentage s'établirait à 3,51% ce qui constituerait une assez nette amélioration du caractère complet des données nécessaire pour l'analyse statistique.

Nous avons présenté, d'une manière assez large, les problèmes que rencontraient les séries de données polliniques en termes de données lacunaires. Nous allons maintenant nous pencher sur les dates de mise en route et d'arrêt des capteurs qui peuvent aussi avoir des conséquences sur l'exploitation des données polliniques.

1.1.4 Etude sur les dates de mise en route et d'arrêt des capteurs

Ces dates jouent un rôle primordial dans le cadre d'une analyse statistique. En effet, il est nécessaire de faire démarrer le capteur assez tôt dans l'année pour que soient collectés les pollens de toutes les espèces végétales, notamment les plus allergisants et ceux qui apparaissent dès le milieu de l'hiver.

1.1.4.1 Mise en route et arrêt des capteurs : l'importance de leur choix

Le choix de la date de démarrage s'effectue, en temps normal, en fonction des conditions météorologiques dans la moitié Nord de la France. Elle sera, par exemple, avancée de une à deux semaines dans le cas d'un hiver doux (par rapport à une valeur moyenne) et si les conditions climatiques, et tout particulièrement thermiques, du mois de février sont rigoureuses, on retardera d'autant la mise en route. Il est cependant souhaitable de redémarrer, dans tous les cas, le capteur avant le 1^{er} mars. Le but principal étant de ne pas manquer le début de la pollinisation du noisetier dont le pollen, allergisant, peut apparaître dès le mois de janvier ainsi que celui de l'aulne qui survient peu après. Malheureusement, les redémarrages s'effectuent souvent trop tardivement si bien que les comptes polliniques enregistrés pour ces deux genres ne reflètent qu'une partie de leur saison propre. Le Midi méditerranéen est un cas à part, car les quantités de pollen n'y sont jamais nulles durant la fin de l'automne et l'hiver. Cela est principalement dû au pollen des familles des Cupressacées et des Taxacées qui comprend de nombreux genres comme le cyprès, le thuya, l'if et le genévrier qui induisent une pollinisation quasi continue tout au long de l'année. Cela s'explique par le fait que les différentes espèces des deux familles présentent des pollens non différenciables au microscope optique. Il est donc nécessaire que les capteurs des

viles méditerranéennes fonctionnent tout au long de l'année, sans interruption, pour avoir un aperçu complet de la pollinisation de cette famille végétale. Malheureusement cela n'est quasiment systématique que depuis quelques années, si bien qu'une grande partie des séries polliniques sont difficilement exploitables dans le cadre d'une étude sur les Cupressacées-Taxacées.

L'arrêt des capteurs, à l'exception notable de ceux du pourtour méditerranéen donc, s'effectue, de manière générale, durant la première moitié du mois d'octobre. A ce moment de l'année, les concentrations polliniques sont très faibles et la date d'arrêt se fait en fonction de la présence de pollens de cèdre. La pollinisation de ce conifère est la plus tardive de tous les taxons pouvant être «capturés». L'apparition de ce pollen, non allergisant, est le signe annonciateur de la fin de la saison et de l'entrée en repos (dormance) des arbres et de la disparition progressive des herbacées du fait du raccourcissement des jours et des conditions climatiques qui deviennent de moins en moins clémentes. Il faut noter cependant que dans la région Rhône-Alpes l'arrêt des capteurs dépend grandement de la fin de la pollinisation de l'ambroisie, herbacée très présente dans cette région et dont le pollen est hautement allergisant.

Le démarrage adéquat des capteurs est un facteur tout aussi important que celui d'avoir des séries polliniques complètes dans le cas d'une analyse statistique et, plus précisément, de la prévision de l'apparition d'un risque allergique d'exposition. Nous allons pour cela étudier les dates de démarrage et d'arrêt des capteurs, des villes présentées, hors Périgueux, dans la section 1.1.2.2 pour la période commune 1997-2005.

1.1.4.2 Comparaison des dates moyennes de démarrage

Cette section va principalement concerner les stations non méditerranéennes dont le fonctionnement du capteur n'est pas continu. Nous allons tout d'abord présenter les villes et les hivers, de 1997 à 2005, où des données polliniques sans interruption, ou avec une interruption inférieure à deux semaines, sont disponibles :

- **Marseille** : 2000-2001 et 2004-2005
- **Aix-en-Provence** : 2003-2004 et 2004-2005
- **Nice** : 2000-2001
- **Nîmes** : 1999-2000, 2001-2002, 2003-2004 et 2004-2005
- **Bordeaux** : 2000-2001 et 2001-2002

On peut noter, pour les années antérieures, des données polliniques hivernales à Toulouse en 1993-1994 et à Nîmes durant le même hiver. Le fonctionnement en continu nécessaire dans les stations méditerranéennes, pour une étude précise sur les pollens de Cupressacées-Taxacées, est loin d'être systématique. Dans ces conditions, les données polliniques de la station de Nîmes sont les plus exploitables avec cinq hivers où les relevés s'effectuèrent en continu. La comparaison des dates moyennes de démarrage va s'effectuer sur les 24 stations restantes. Nous allons tout d'abord présenter graphiquement les dates moyennes de mise en route des capteurs (figure 11).

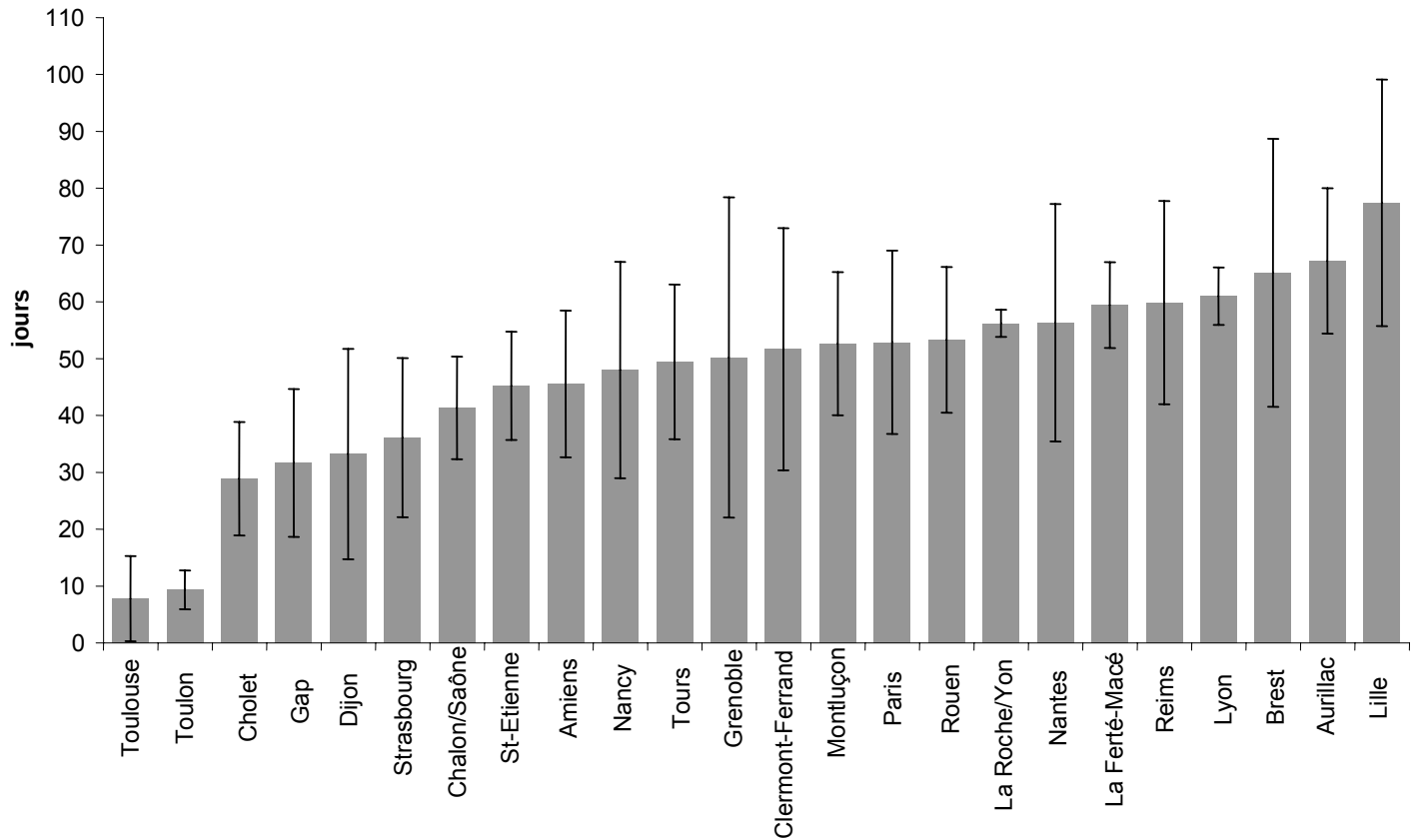


Figure 11 : Classement par ordre croissant de la date moyenne de démarrage des capteurs (avec écart-type), ville par ville et sur la période 1997-2005, en nombre de jours à partir du 1^{er} janvier.

On constate tout d'abord que les dates de démarrage de Toulouse et Toulon sont nettement plus précoces que celles de toutes les autres villes. Pour ces dernières, la date moyenne se situe entre le 10 février et le 1^{er} mars pour la plus grande part. Pour une meilleure lisibilité nous allons regrouper les 24 stations par classes de dates. Pour déterminer le nombre souhaitable de classes, nous allons utiliser la formule de Huntsberger qui est la suivante :

$$k = 1 + 3,3 \log(n)$$

k = nombre de classes ; **n** = effectif total (nombre de stations)

Pour 24 stations, la valeur de **k** est de 5,56. Nous allons donc retenir cinq classes qui vont être déterminées par une classification en nuées dynamiques. L'intérêt majeur de cette méthode réside dans l'élaboration d'un nombre **k** de classes choisies auparavant, ces dernières étant aussi

Nota : Les barres noires de la figure 11 représentent les écarts-types. Il en sera de même pour toutes les figures qui suivront où apparaîtront ces barres.

différentes entre elles que possible grâce à l'algorithme de nuées dynamiques. Ces classes sont présentées selon leur valeur moyenne, par ordre croissant :

- **Classe 1** : La date moyenne de démarrage de cette classe correspond quasiment au 9^{ème} jour de l'année (8,56 très précisément). Elle comprend Toulon et Toulouse qui présentent une variabilité interannuelle faible (écarts-types de 3,38 jours à Toulon et 7,5 jours à Toulouse). La position méridionale de ces deux sites nécessite en effet un démarrage particulièrement précoce des capteurs, qui devraient fonctionner en continu à Toulon du fait de la pollinisation des Cupressacées-Taxacées.
- **Classe 2** : La date moyenne de démarrage de cette classe correspond au 34^{ème} jour de l'année, soit le 3 février. Elle comprend les sites de Cholet, Gap, Dijon, Strasbourg et Chalon-sur-Saône. Ce sont les sites non méditerranéens où la mise en route des capteurs est, en moyenne, la plus précoce, avec une date de mise en route allant du 29 janvier (Cholet) au 10 février (Chalon-sur-Saône). On peut noter la présence des deux stations bourguignonnes dans cette classe.
- **Classe 3** : La date moyenne de mise en route de cette classe correspond au 49^{ème} jour de l'année, soit le 18 février. Elle se compose de six sites : Saint-Etienne, Amiens, Nancy, Grenoble, Clermont-Ferrand et Rouen où les dates moyennes s'étagent du 14 février (Saint-Etienne) au 22 février (Rouen). De ce groupe «moyen» se distinguent Clermont-Ferrand et Grenoble qui présentent une très forte variabilité interannuelle avec, respectivement, des écarts-types de 21,31 et 28,19 jours qui traduisent une grande amplitude entre la date de mise en route la plus précoce et la plus tardive. Ces dates sont les suivantes :
 - 17 janvier (2001) et 21 mars (2005) à Clermont-Ferrand
 - 7 janvier (1999) et 31 mars (2000) à Grenoble
- **Classe 4** : La date moyenne associée à cette classe correspond au 57^{ème} jour de l'année, c'est-à-dire le 26 février. Elle est composée de neuf sites : Tours, Montluçon, Paris, La Roche-sur-Yon, Nantes, La Ferté-Macé, Reims, Lyon et Brest. La rupture entre cette

classe et la classe 3 n'est pas très nette car Tours, Montluçon et Paris ont des dates de démarrage plus précoces que le site de Rouen qui a été classé, par l'algorithme de nuées dynamiques, dans le niveau précédent. Ce groupe peut être considéré comme «moyen» malgré le fait que les sites de Reims, Lyon et Brest présentent des dates moyennes de démarrage postérieures au 28 février (respectivement les 1^{er}, 2 et 6 mars pour une année non bissextile). Le risque de manquer le début, en fin d'hiver, de la pollinisation de plusieurs espèces précoces devient réel.

- **Classe 5** : C'est la classe «tardive» par excellence avec une date moyenne de démarrage correspondant au 72^{ème} jour de l'année soit le 13 mars pour une année non bissextile. Aurillac et Lille composent ce groupe avec des mises en route qui ont lieu, en moyenne, le 8 et le 18 mars respectivement si l'on se réfère à des années non bissextiles. Aurillac présente seulement deux années où le démarrage a eu lieu avant le 1^{er} mars alors qu'à Lille, de 2001 à 2005, la mise en route du capteur est intervenue systématiquement après la mi-mars. Il est évident que les dates de démarrage de ces deux villes, et les plus tardives de la classe 4, posent problème si l'on se penche sur l'analyse de la saison pollinique du noisetier, de l'aulne, voire même du frêne et du bouleau pour certaines années.

Cette classification a permis de mettre en avant le problème des dates de démarrage tardives. La principale cause semble être des problèmes d'ordre administratif qui peuvent empêcher la mise en route au moment adéquat, ou des problèmes techniques tels qu'un dysfonctionnement du capteur. Néanmoins, avec l'exemple de Lille, ce phénomène se répète de manière régulière, ce qui est préjudiciable dans le cadre d'une analyse statistique des données polliniques de cette ville. Ce problème semblant affecter la majorité des 29 stations retenues, nous allons présenter les années, ville par ville concernée, où la mise en route a eu lieu le 15 mars au plus tôt au cours de la période 1997-2005 :

- **Paris** : 18 mars (1998)
- **Rouen** : 25 mars (2005)
- **Clermont-Ferrand** : 21 mars (2005)

- **Brest** : 20 avril (1998), 27 mars (2001) et 18 mars (2003)
- **Aix-en-Provence** : 2 avril (1997), 15 mars (1998) et 9 avril (2002)
- **Lille** : 20 mars (2001), 15 mars (2002), 1^{er} avril (2003), 30 mars (2004) et 2 mai (2005)
- **Reims** : 24 mars (1998 et 1999)
- **Tours** : 18 mars (1998)
- **Aurillac** : 24 mars (1998), 16 mars (2004) et 21 mars (2005)
- **Nantes** : 17 avril (2001)
- **Dijon** : 15 mars (1998)
- **Grenoble** : 31 mars (2000)

23 années sur les 261 disponibles (8,81%) sont concernées, ce qui est finalement relativement peu mais, quand cela se produit, peut être particulièrement handicapant. En effet, contrairement en théorie aux valeurs manquantes, il est strictement impossible d'extrapoler les concentrations polliniques ayant lieu avant la date de démarrage. Ainsi, si à ce moment-là, la saison pollinique d'une espèce ou d'un genre est prise en cours, les données du début de la saison sont définitivement perdues. De ce fait une date de démarrage trop tardive peut être aussi problématique, sinon plus, que la présence de valeurs manquantes dans les séries.

De plus il est nécessaire de préciser qu'il pourrait y avoir, dans ces différentes séries, des dates de mise en route qui ne correspondent qu'au premier jour enregistré dans la base de données du RNSA comme cela est le cas, malheureusement, à Dijon en 1998. Ne pouvant nous fier qu'aux seules données du RNSA pour les sites non bourguignons, il nous est impossible de dissocier les dates liées à d'éventuels problèmes informatiques, des dates réelles de démarrage des capteurs. Il faut espérer que ces lacunes d'enregistrement soient très sporadiques.

1.1.4.3 Comparaison des dates moyennes d'arrêt

Nous allons nous intéresser maintenant à l'arrêt des capteurs en présentant graphiquement les dates moyennes (figure 12) puis, comme pour les dates de mise en route, nous établirons une classification en nuées dynamiques avec un découpage en cinq classes.

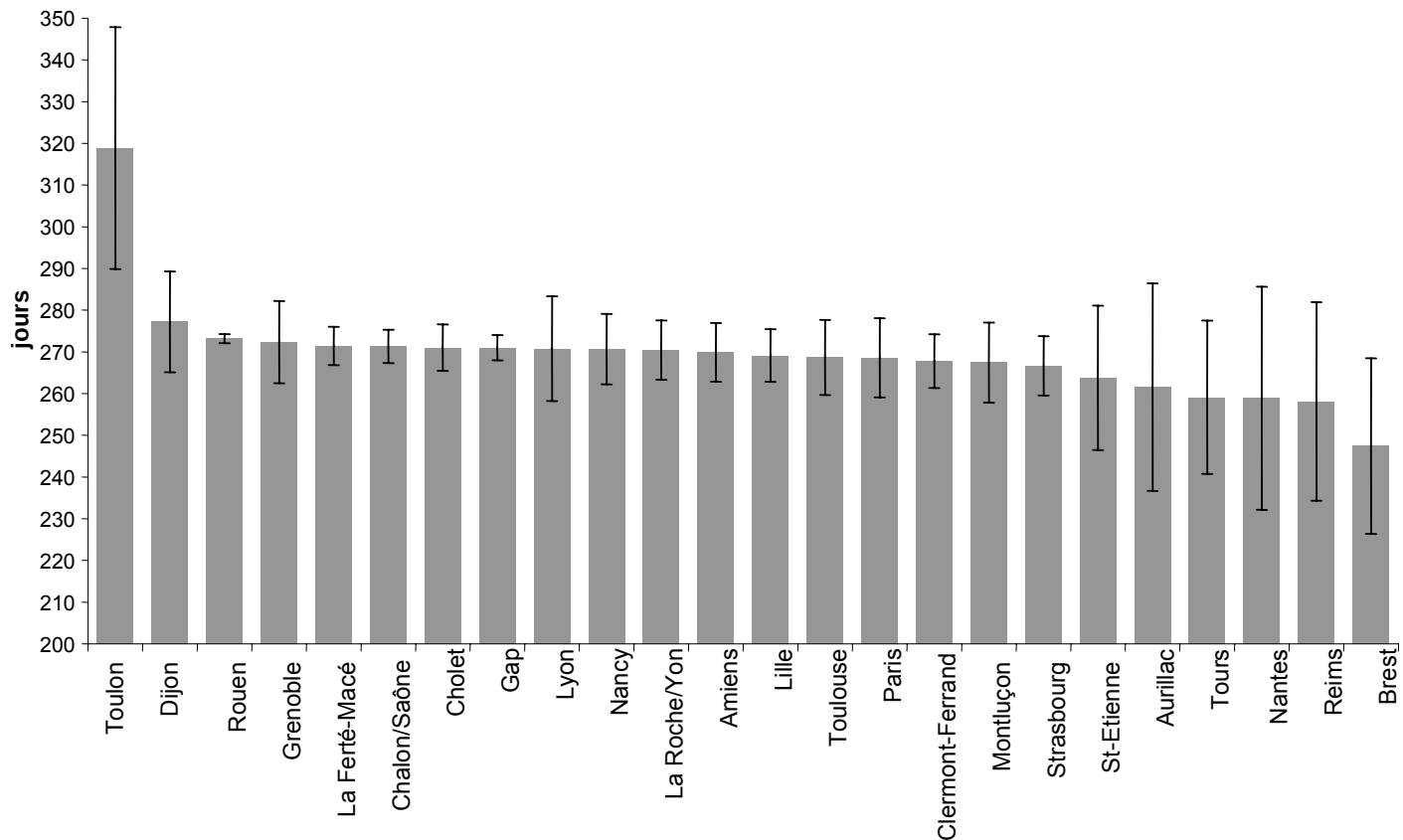


Figure 12 : Classement par ordre décroissant de la date moyenne d'arrêt des capteurs, ville par ville et sur la période 1997-2005, en nombre de jours à partir du 1^{er} janvier.

Graphiquement, on peut constater que Toulon est un cas à part avec une date moyenne d'arrêt beaucoup plus tardive que l'ensemble des 23 autres stations qui semblent former un bloc très homogène. A l'instar du graphique des dates moyennes de démarrage, le classement des villes est ici élaboré de telle sorte que celles situées sur la droite du graphique sont considérées comme ayant les séries les moins optimales, dans le cadre d'une analyse statistique. En d'autres termes on y retrouve les villes ayant les démarrages les plus tardifs et/ou les arrêts du capteur les plus précoces. La constitution des cinq classes, par l'algorithme de nuées dynamiques, est la suivante. Elles sont présentées par ordre décroissant de date :

- **Classe 1** : Elle est uniquement constituée de Toulon dont la date moyenne d'arrêt se situe le 319^{ème} jour de l'année, soit le 15 novembre pour une année non bissextile. Les conditions climatiques et la pollinisation quasiment continue des Cupressacées-

Taxacées expliquent cette date tardive même si, comme cela a déjà été évoqué, le capteur devrait fonctionner de manière continue pendant l'hiver, et ainsi éviter l'absence de données en décembre.

- **Classe 2** : La date moyenne d'arrêt de cette classe correspond au 270^{ème} jour de l'année, soit le 27 septembre pour une année non bissextile. Ce groupe peut être considéré comme médian car il comprend 17 stations dont les valeurs moyennes se suivent de Dijon (277^{ème} jour – 4 octobre) à Strasbourg (267^{ème} jour – 24 septembre). Les capteurs s'arrêtent donc de fonctionner le plus souvent avant la fin du mois de septembre. C'est notamment le cas de Lyon (28 septembre) où le capteur a été stoppé avant la dernière semaine de septembre à trois reprises depuis 1997, avec un arrêt le 31 août en 2002. Pour ces années-là, la fin de la saison de l'ambrosie est tronquée, entraînant des pertes de données dans le cas d'une étude sur les concentrations journalières et totales de ce pollen.

Les stations de ce groupe se caractérisent par les valeurs contrastées de leurs écarts-types. Dijon présente la valeur la plus élevée (12,09 jours) liée, entre autre, à une amplitude de 27 jours entre les dates les plus précoces (15 septembre en 1998 et 2002) et les plus tardives (12 octobre en 1999 et 2003). Le site de Rouen, à l'opposé, se caractérise par l'étonnante constance des dates d'arrêt avec un écart-type de 1,09 jour. L'amplitude maximale des dates pour cette ville est de trois jours, du 29 septembre (en 1999) au 2 octobre (en 2005) ; la date moyenne étant le 30 septembre pour une année non bissextile.

Malheureusement, un problème dans l'enregistrement des données est survenu, sans contestation possible, à Dijon en 2000. Le fichier RNSA a enregistré le 2 octobre comme date ultime, alors que le capteur a fonctionné jusqu'au 16 de ce mois. Le 2 octobre a été retenu pour l'analyse pour les raisons évoquées dans la section 1.1.4.2.

- **Classe 3** : La date moyenne d'arrêt de cette classe correspond au 260^{ème} jour de l'année, soit le 17 septembre pour une année non bissextile. Elle se compose de trois sites (Aurillac, Tours et Nantes) dont la caractéristique commune réside dans une variabilité interannuelle élevée de la date d'arrêt des capteurs. On peut remarquer également que

la station d'Aurillac, caractérisée par une mise en route moyenne parmi les plus tardives, se distingue par un arrêt moyen parmi les plus précoces (19 septembre).

- **Classe 4** : Elle est uniquement constituée de Reims dont la date moyenne d'arrêt se situe le 258^{ème} jour de l'année, soit le 15 septembre pour une année non bissextile. La différenciation effectuée par l'algorithme, par rapport aux composantes de la classe 3, semble purement artificielle.
- **Classe 5** : Cette classe est constituée de deux sites dont la date d'arrêt moyenne du capteur correspond au 256^{ème} jour de l'année, soit le 13 septembre pour une année non bissextile. Saint-Etienne et Brest, qui composent ce groupe, ont pourtant des dates très différentes : 21 septembre pour la première et 4 septembre pour la deuxième. La constitution de cette classe est étrange et difficilement explicable, on peut seulement noter que Brest a la date moyenne d'arrêt la plus précoce des 24 stations.

En ce qui concerne l'arrêt des capteurs, une plus grande homogénéité des dates moyennes est observée par rapport aux dates de démarrage. Cette date revêt moins d'importance pour une analyse statistique, sauf pour les stations enregistrant des concentrations importantes d'ambrosie où il serait judicieux de stopper le fonctionnement du capteur une semaine, au moins, après l'arrêt des autres sites du Nord de la France peu ou pas concernés par ce pollen.

Comme c'est le cas pour les dates de démarrage, l'étude des séries de Dijon a mis en évidence des problèmes d'enregistrement des données à la fin de la saison, pour une seule année. Il faut espérer que les dernières dates enregistrées dans les fichiers correspondent, en très grande majorité, aux dates réelles d'arrêt des capteurs.

Après avoir évoqué les lacunes, les dates de mise en route et les dates d'arrêt des capteurs, paramètres essentiels pour une exploitation optimale des données, nous allons nous intéresser aux données polliniques en elles-mêmes, à travers les différents paramètres susceptibles d'influencer les valeurs enregistrées.

1.1.5 Etude des concentrations polliniques quotidiennes : influence de différents paramètres

Les concentrations polliniques, enregistrées dans la base de données du RNSA, sont issues de l'analyse, sous microscope optique, des lames journalières préparées à partir de la bande de cellophane, sur laquelle sont impactées les particules contenues dans l'air aspiré durant une semaine complète. Les concentrations enregistrées revêtent une grande importance du point de vue allergologique. Elles permettent la mise au point de différents niveaux de risque d'exposition allergique par l'intermédiaire de valeurs-seuils ; des valeurs qui sont fonction du taxon pollinique et de la situation géographique de la station. Ces concentrations reflètent, en premier lieu, l'abondance et la répartition spatiale de telle ou telle espèce, genre ou famille. Ainsi, seuls les sites de la région lyonnaise enregistrent de fortes concentrations de pollen d'ambroisie, car cette herbacée est très présente dans cette zone. De même, le pollen de bouleau se rencontre en abondance dans un grand quart Nord-Est de la France, du fait de la bonne implantation de l'arbre dans les zones qui composent cette partie du territoire national. Ainsi, grâce aux espèces végétales présentes dans la zone de couverture du capteur, il est possible d'expliquer, pour une bonne part, la présence en plus ou moins grande abondance, année après année, d'un taxon pollinique donné. Par exemple, une étude espagnole (Trigo *et al.*, 2000) portant sur deux capteurs distants de 90 km, situés à 15 m du sol, a montré des différences significatives, pour la plupart des taxons polliniques étudiés, aux niveaux des moyennes et des distributions des concentrations, qui pourraient être expliquées par le type de végétation et sa densité. A ces causes endogènes, il faut ajouter les pollens provenant d'espèces exogènes qui, transportés par le vent sur une plus ou moins longue distance, peuvent être impactés sur les tambours des capteurs. C'est le cas de l'ambroisie à Dijon et Mâcon où les plus fortes quantités journalières sont généralement enregistrées par vent de sud (Laaidi et Laaidi, 1999), en provenance de la région lyonnaise où cette herbacée est fortement implantée.

Cependant, il apparaît évident que d'autres paramètres que la situation géographique influencent la valeur des concentrations enregistrées. Leur détermination fait l'objet de ce sous-chapitre.

1.1.5.1 Influence de la position en hauteur du capteur

Le paramètre le plus influent est, très probablement, la hauteur de la terrasse (ou du toit) où se situe le capteur, par rapport au sol. En effet, une étude sur le pollen d'ambrosie (Alcazar et Comtois, 2000) a montré que les concentrations relevées différaient significativement selon la position en hauteur du capteur, notamment entre un capteur installé au sol, qui est censé capter ce que respire un individu se déplaçant dans une rue, et un autre placé à 15 m de hauteur.

Un capteur situé au niveau du sol semblerait être la solution la plus efficace pour évaluer le plus fidèlement ce que peut respirer un individu. Mais les concentrations enregistrées seraient fortement influencées par les espèces végétales à proximité, et la girouette s'orienterait face à un flux d'air grandement affecté par la configuration du bâti alentour, ce qui n'est pas représentatif à l'échelle d'une agglomération. Enfin, la pollution automobile noircirait la bande de cellophane, la rendant quasiment illisible au microscope optique.

Selon Comtois et Gagnon (1988), la fourchette idéale de hauteur pour placer un capteur pollinique est de l'ordre de 15-20 mètres. Cet article se base sur l'hypothèse de la présence d'un nuage pollinique homogène à cette hauteur qui serait représentatif d'une zone étendue, et dont les concentrations sont significativement corrélées avec les symptômes allergique. Plusieurs études concluent que les plus fortes concentrations sont enregistrées en hauteur (Bryant *et al.*, 1989 ; Malik *et al.*, 1991), alors que d'autres démontrent que ces concentrations importantes sont enregistrées au niveau du sol (Lyon *et al.*, 1984 ; Rantio-Lethimaki *et al.*, 1991). Il semble donc difficile de déterminer une hauteur optimale au vu de ces résultats. Le positionnement au sol des capteurs présentant de nombreux inconvénients, il semble plus judicieux de se baser sur la fourchette de hauteur préconisée par l'étude de Comtois et Gagnon (1988). Cela s'avère loin d'être systématique pour les trente capteurs de notre étude, comme le montre la figure 13 :

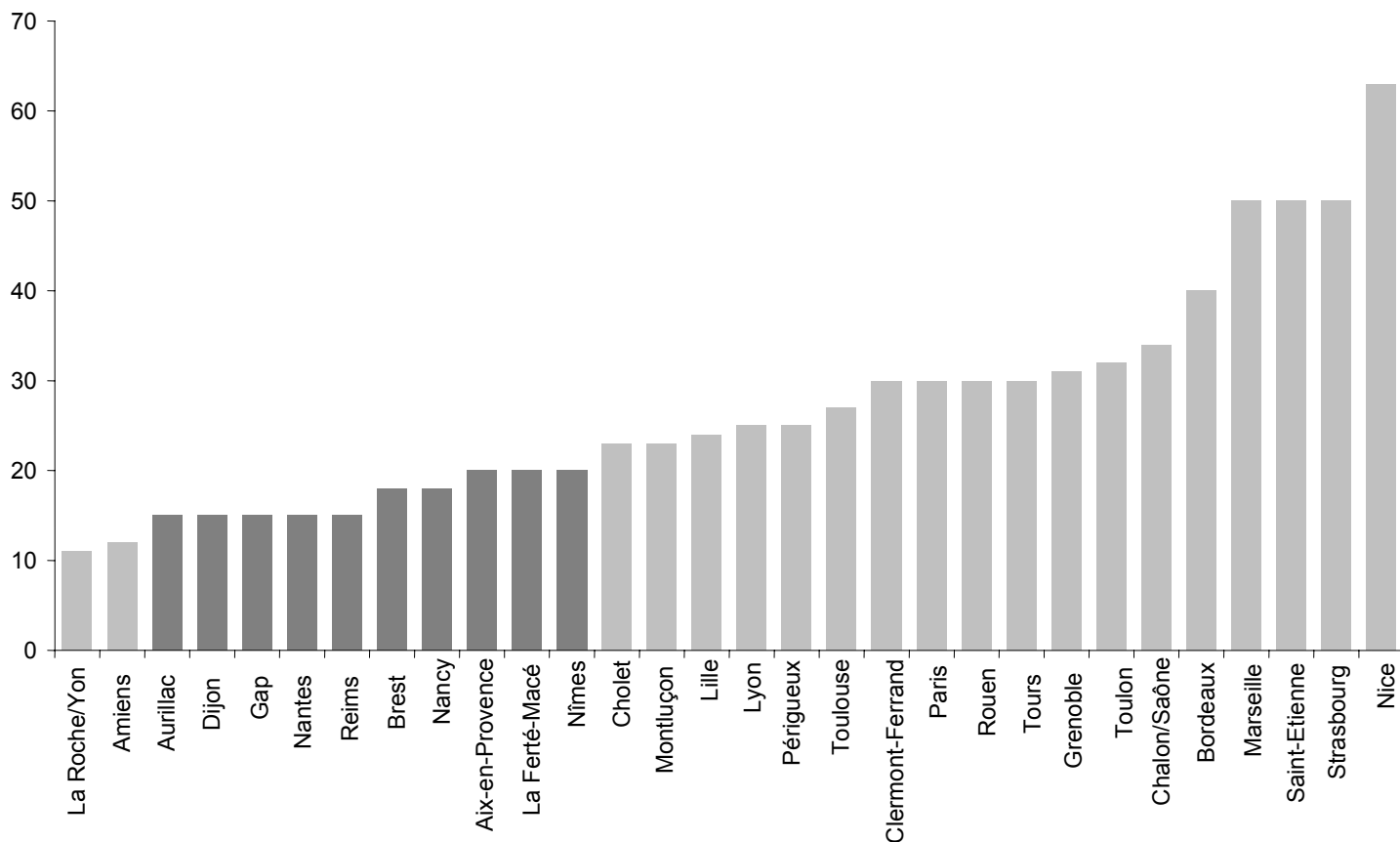


Figure 13 : Classement par ordre croissant de la hauteur (en mètres) du capteur par rapport au sol. Les capteurs situés de 15 à 20 mètres du sol sont représentés en gris foncé.

D'après la figure 13, seuls dix capteurs sur un total de trente sont à une hauteur respectant les recommandations émises par Comtois et Gagnon (1988). La hauteur moyenne pour ces trente stations est de 27 mètres avec, notamment, cinq sites se distinguant par la position élevée de leur capteur :

- Bordeaux : 40 m
- Marseille, Saint-Etienne et Strasbourg : 50 m
- Nice : 63 m

A une telle hauteur, les concentrations enregistrées risquent de n'être pas représentatives de ce que peut inhaler réellement un individu allergique aux pollens. Il est malaisé de conclure que telle ou telle hauteur est plus représentative du contenu pollinique «de fond» de l'air ambiant ; mais il

est certain que ces différences de hauteur ont une influence sur les concentrations relevées. De ce fait, pour une région présentant des conditions climatiques équivalentes, il semble, de prime abord, difficile d'estimer l'implantation d'une espèce et de comparer les quantités de pollen relevées, si la différence de hauteur entre les capteurs est trop importante.

Face à ce constat, il est nécessaire de préciser que l'implantation des capteurs n'est pas aisée. En effet, il faut trouver un bâtiment avec une terrasse dégagée, à l'abri du vandalisme et sans végétation à proximité immédiate. Il faut, de plus, qu'il soit accessible pour les services municipaux, le personnel des réseaux de surveillance atmosphérique ou le personnel universitaire qui, dans la quasi-totalité des sites, assurent la maintenance des capteurs. Les considérations administratives (et d'éventuels blocages) prennent une part aussi importante que la recherche du lieu le plus optimal, dans l'implantation des capteurs. Néanmoins, il faudrait éviter le plus possible de les placer à plus de 35 mètres du sol, de sorte que l'on puisse disposer d'un réseau relativement homogène. En s'intéressant aux capteurs installés, ou réinstallés après quelques années d'arrêt, à partir de 1998 et qui sont toujours en fonctionnement en 2006, on constate que la hauteur moyenne est de 20,4 m pour ceux dont les valeurs étaient disponibles. A part quelques exceptions (Chambéry : 40 m ; Ajaccio : 35 m), il y a eu, ces dernières années, un réel souci d'homogénéisation des hauteurs, qui sont revues à la baisse. Cette homogénéisation est absolument nécessaire, afin d'évaluer l'importance de la saison pollinique de chaque station sur des bases équivalentes. De fait, les capteurs placés à une hauteur importante font partie de ceux ayant les implantations les plus anciennes.

En étudiant les bulletins des données aéropolliniques du RNSA, on constate que la variabilité de la hauteur des capteurs n'est pas seulement inter-site, mais aussi intra-site. En effet, pour un certain nombre de raisons, des capteurs ont été déplacés, le plus souvent à des hauteurs différentes. Depuis 1997, c'est notamment le cas pour les villes suivantes:

- **Clermont-Ferrand** : en 1997 le capteur est implanté sur une terrasse de l'hôpital Sabourin, à environ 30 m du sol. Il est déplacé sur le toit d'une annexe de cet hôpital en 1998 où il se retrouve à, approximativement, 18 m de hauteur. Il est encore déplacé en 2000 sur une terrasse de l'hôpital Montpied, où il se retrouve de nouveau à 30 m du sol.

- **Grenoble** : En 1997, première année où la situation des capteurs est indiquée sur le bulletin des données aéropolliniques françaises, le capteur se situait sur le toit de l'unité de botanique de l'université à 18 m du sol. Il a été déplacé en 2000 sur le toit d'un immeuble de la ville à 31 m de hauteur.
- **Nantes** : Situé, en 1997, sur une terrasse de la Faculté de médecine à 38 m de hauteur, il a été déplacé, en 2000, sur le toit du département environnement de la mairie à 15 m du sol.
- **Reims** : Le capteur se situait sur une terrasse du centre hospitalier régional de Reims à 30 m du sol. Il est installé, en 2003, sur le toit du centre de recherches en agronomie à une hauteur de 15 m.

Ces modifications de la hauteur et du site d'implantation entraînent très certainement une rupture de stationnarité des séries concernées. C'est le cas à Reims où la concentration annuelle moyenne sur la période 1997-2002 est de 19 962 grains, et passe à 40 487 grains pour la période 2003-2005. Cela pose problème dans le cadre d'une étude se basant sur les concentrations. Néanmoins, du fait de la faible longueur des séries et de la forte variabilité interannuelle affectant les concentrations de la plupart des taxons polliniques, il est quasiment impossible de corriger les valeurs post-changement, par un coefficient simple. De plus, il n'est pas évident que les concentrations relevées à une hauteur x soient systématiquement supérieures (ou inférieures) à celles enregistrées à une hauteur y . Dans ce cas-là, associé à cette question de la hauteur, se pose le cas des concentrations enregistrées par deux capteurs séparés d'une courte distance. Une étude américaine (Frenz *et al.*, 1997) a ainsi démontré que deux capteurs placés à 5,6 km de distance capturaient, globalement, des quantités équivalentes de pollen, mais que des différences journalières importantes pouvaient apparaître, principalement quand les concentrations excédaient 100 grains/m³. Ce genre de résultats peut mettre à mal la représentativité d'un capteur unique à l'échelle d'une zone étendue ; mais si ces différences affectent les concentrations polliniques les plus élevées, il y a de fortes chances que, dans tous les cas, les seuils élevés de risque d'exposition aux pollens allergisants soient dépassés. L'interprétation des résultats, d'un point de vue allergologique, sera donc la même pour les deux capteurs. Mais cela confirme la

nécessité, quand cela est réalisable, de ne pas modifier l'implantation du capteur, ni sa hauteur afin de conserver l'homogénéité d'une série.

1.1.5.2 Influence d'autres paramètres techniques

A côté du rôle joué par la position en hauteur du capteur, deux autres critères techniques doivent être également considérés :

Le deuxième paramètre influençant les concentrations concerne le débit d'air aspiré à travers la buse du capteur. Normalement calibré à 10 litres/min, ce débit peut être sujet à variation si l'on rencontre quelques dysfonctionnements dans l'aspiration. Tout d'abord le fait de remplacer le tambour peut modifier faiblement le débit, sans que cela soit vraiment explicable. Un oubli dans le nettoyage de la buse peut entraîner, à la longue, une diminution du débit par obstruction progressive de l'orifice par les différentes particules présentes dans l'air ambiant ; ce problème a affecté la station d'Annecy pendant quelques années, nous empêchant d'utiliser les séries polliniques correspondantes. En dernier lieu, ces variations de débit peuvent être dues à des problèmes mécaniques, notamment en lien avec un dysfonctionnement de la pompe aspirante, ou à cause d'un problème d'étanchéité entre le corps central et la tête du capteur, où est fixé le tambour. Dans ces deux cas, on enregistre généralement un débit inférieur à 10 litres/min. Ainsi à Dijon, en 2001, des problèmes liés à la pompe aspirante furent notés dans les fiches de changement de tambour de l'époque. Le débit relevé au moment du changement de tambour a été, à trois reprises, de 8 litres/min, une fois à 9 litres et une dernière fois à 12 litres/min. Si ces problèmes sont connus, il est possible de corriger, en théorie, les concentrations enregistrées. Par exemple si le débit n'est que de 8 litres/min, il faudrait multiplier par 1,25 les quantités analysées, afin d'avoir une estimation pour un débit «normal». Mais cela n'est possible que si l'on peut déterminer le jour de la semaine où se produisit le dérèglement... ce qui est très rarement le cas. On peut souligner, enfin, que l'application de la solution siliconée sur la bande adhésive, par l'intermédiaire d'un pinceau, n'est pas toujours parfaitement homogène et peut ainsi influencer sur la qualité de l'impactage des pollens, notamment sur les bords extérieurs. Mais comme on passe

généralement deux couches de solution, l'influence de ce paramètre est très probablement minime par rapport aux problèmes évoqués auparavant.

Le troisième paramètre qui peut influencer les concentrations est le comptage des pollens sous microscope. Il s'effectue, à l'heure actuelle, par la lecture de deux ou trois lignes horizontales. Le facteur de conversion (cf. 1.1.1.2), différent selon le nombre de lignes, est ensuite appliqué aux pollens effectivement comptés. De ce fait, entre une lecture sur deux lignes et une autre sur trois, pour une même lame, les quantités comptabilisées sont différentes comme le montrent les tableaux 5 et 6 :

taxons	nombre de grains comptés	nombre de grains après conversion
Cupressaceae-Taxaceae	347	240
Fraxinus	181	125
Populus	81	56
Salix	56	39
Alnus	54	38
Ulmus	33	23
Corylus	23	16
Buxus	7	5
Poaceae	5	4
Chenopodiaceae	1	1
indéterminés	1	1
Ligustrum	1	1
Mercurialis	1	1
Plantaginaceae	1	1
Rosaceae	1	1
Asteraceae (Tubuliflores)	1	1
Total	794	553

Tableau 5 : Nombre de grains de pollens comptés et total après conversion (coefficient de correction de 0,69) à Chalon-sur-Saône, du 4 mars 2007 à 8 heures au 5 mars 2007 à 8 heures : résultats pour deux lignes de lecture.

taxons	nombre de grains comptés	nombre de grains après conversion
Cupressaceae-Taxaceae	622	287
Fraxinus	319	147
Populus	135	63
Alnus	93	43
Salix	89	41
Ulmus	63	29
Corylus	38	18
Buxus	15	7
Pinus	9	5
Poaceae	9	5
Ligustrum	3	2
Mercurialis	2	1
indéterminés	2	1
Rosaceae	2	1
Betula	1	1
Carpinus	1	1
Chenopodiaceae	1	1
Plantaginaceae	1	1
Asteraceae (Tubuliflores)	1	1
Urticaceae	1	1
Total	1407	656

Tableau 6 : Nombre de grains de pollens comptés et total après conversion (coefficient de correction de 0,46) à Chalon-sur-Saône, du 4 mars 2007 à 8 heures au 5 mars 2007 à 8 heures : résultats pour trois lignes de lecture.

A travers les résultats des tableaux 5 et 6, on se rend compte qu'il y a une différence non négligeable entre les concentrations enregistrées avec une lecture sur deux lignes et une autre sur trois lignes. Cette dernière semble plus pertinente car elle est plus exhaustive (20 taxa recensés au lieu de 16), sans pour autant que le total des concentrations enregistrées soit obligatoirement plus élevé qu'avec deux lignes. Pour cette bande, l'écart des concentrations entre les deux types de lecture s'explique par des différences quant au nombre de pollens recensés par ligne. En effet, sur 1407 grains comptabilisés, 441 l'ont été sur la première ligne, 353 sur la seconde et 613 sur la troisième. Cela montre que l'aspiration ne s'effectue pas vraiment de manière homogène sur la largeur de la bande. La troisième ligne, plus chargée, explique les plus fortes concentrations pour une lecture en trois lignes.

Le nombre de lignes lues et l'aspiration qui n'est pas vraiment constante sur l'ensemble de la largeur de la bande sont autant de facteurs influençant les concentrations enregistrées dans la base de données du RNSA. Depuis deux ans, le réseau impose de lire deux ou trois lignes à une distance précise par rapport au bord de la bande. Cette standardisation du positionnement de lecture devrait être étendue au nombre de lignes lues. L'idéal serait trois lignes, même si cela entraîne un surcroît de travail. Il faut signaler que le mode de lecture n'a pas toujours été le même. En effet, quand il n'y avait pas de reconnaissance vocale notamment, la lecture s'effectuait en lignes verticales. Pour établir un fichier de données bihoraires, il fallait lire douze lignes sur une bande de 24 heures ; cela nécessitait un travail fastidieux. Le passage en lecture horizontale a constitué un réel progrès dans le confort d'analyse. Mais si l'on applique la lecture horizontale sur des lames analysées verticalement, les concentrations ne seront pas identiques.

Enfin, il faut indiquer que les concentrations journalières particulièrement élevées (plus de 1 000 grains/m³ après application du coefficient de conversion) ne sont, le plus souvent, que des estimations. En effet, pour de telles concentrations, les pollens se chevauchent et « envahissent » entièrement le champ du microscope. Il est donc humainement quasi impossible de les recenser précisément, d'autant plus que par leur très grand nombre, ils ne prennent qu'une coloration très pâle au contact de la fuchsine.

1.1.5.3 Influence du nombre de lignes de lecture : le cas de quatre stations en 2007

Cette section constitue la suite du dernier point évoqué dans la section 1.1.5.2 à travers quatre exemples de stations dont l'analyse est assurée par le Laboratoire Climat et Santé. Nous allons étudier les jours présentant, après conversion pour trois lignes de lecture (cf. 1.1.1.2), des concentrations au moins égales à 300 grains à Chalon-sur-Saône, Dijon, Nevers et Castres en 2007. Malheureusement, la présentation des fichiers ne permet pas de retenir l'ensemble des jours dont la valeur atteint ou dépasse ce seuil. Nous avons pu en retenir 46 pour Chalon, 35 pour Dijon, 18 pour Nevers et 50 pour Castres.

Nous allons donc comparer ces concentrations avec celles déterminées à partir de deux lignes de lecture. Afin de choisir le test de comparaison adéquat, il est nécessaire de déterminer si la

distribution des différentes séries suit la loi normale. Pour évaluer cela, nous allons utiliser les tests W de Shapiro-Wilk et de Kolmogorov-Smirnov. Le premier consiste à mesurer la conformité de la distribution observée avec une distribution normale théorique (Legendre et Legendre, 1998). Plus la valeur W est grande, plus la distribution est proche de la normale et plus la probabilité p du test s'approche de 1. Le second repose sur le même principe mais en regroupant les données en classes, sauf pour les petits échantillons (Legendre et Legendre, 1998). Notons enfin que leur qualité propre est, selon Dutilleul et Legendre (1992), tout à fait équivalente. Par conséquent nous considérons qu'une distribution est gaussienne si au moins un des deux tests la définit comme telle. En d'autres termes, pour rejeter l'hypothèse de normalité, les résultats des deux tests doivent obligatoirement être concordants.

D'après les résultats du test de Shapiro-Wilk, confirmés par le test de Kolmogorov-Smirnov, la distribution des huit séries de concentrations issues de deux et trois lignes de lecture ne suit pas la loi normale ($p < 0,05$). Un test non-paramétrique doit, par conséquent, être employé. Nous allons donc retenir celui de Mann-Whitney, qui se base sur le calcul des différences de rangs entre chaque variable, une fois les observations classées selon une échelle ordinale. Ce test constitue l'alternative la plus puissante au test paramétrique t de Student, ce dernier permettant d'évaluer les différences entre les moyennes de deux groupes en se basant sur les valeurs brutes et la variation intra-groupe. Malgré les différences de conception, l'interprétation des résultats des deux tests est quasi identique. La valeur p , relative à la significativité du test de Mann-Whitney, de chaque ville est présentée ci-après :

- **Chalon-sur-Saône** : $p = 0,9253$
- **Dijon** : $p = 0,5183$
- **Nevers** : $p = 0,8993$
- **Castres** : $p = 0,9890$

Pour Chalon-sur-Saône, Dijon, Nevers et Castres, les fortes concentrations issues de la lecture sur trois lignes, après application du facteur de conversion, ne sont pas significativement différentes de celles issues de deux lignes de lecture. Dans le cas de Chalon, la concentration moyenne pour trois lignes est de 644,26 grains contre 643,52 grains pour celle en deux lignes. A Nevers, les

valeurs sont de 598,39 grains (trois lignes) et 599,33 grains (deux lignes). Quant à Castres, la différence enregistrée est d'à peine un grain avec 1 084,08 grains pour trois lignes de lecture et 1 083,14 grains pour deux lignes. Par contre, malgré la non significativité du test U de Mann-Whitney, nous constatons une différence plus sensible à Dijon, où les concentrations enregistrées à partir de deux lignes de lecture sont plus importantes (717,34 grains en moyenne) qu'avec trois lignes (660,83 grains). Pour cette dernière ville, il semble que la troisième ligne de lecture soit souvent moins chargée que la moyenne des deux premières, du moins pour les fortes concentrations.

Ainsi, Dijon se différencie des deux autres villes du point de vue de la différence entre les concentrations. Cependant, les valeurs moyennes de Chalon, Nevers et Castres masquent une variabilité non négligeable. Dans le cas de Chalon, en passant de deux à trois lignes de lecture, on enregistre 21 diminutions des concentrations et 25 augmentations. Les valeurs extrêmes d'évolution (à partir des concentrations sur deux lignes) sont les suivantes :

- **Le 13 avril** : – 259 grains ce qui représente une baisse de 20,97%, diminution relative la plus forte.
- **Le 8 avril** : + 192 grains correspondant à une hausse de 18,97% ; l'augmentation relative la plus élevée a lieu le 4 mars (+ 22,85%).

La valeur absolue moyenne des écarts entre les concentrations issues des deux types de lecture est de 58,09 grains, soit 9,03% de la concentration moyenne journalière pour une lecture sur deux lignes. En d'autres termes, pour des concentrations élevées, on a autant de probabilité d'avoir une hausse de 9% des concentrations qu'une baisse de même ampleur lorsque l'on passe de deux à trois lignes.

A Nevers, on enregistre le même cas de figure mais la valeur absolue moyenne des écarts est plus faible (23,94 grains) et représente 4% de la concentration moyenne journalière pour une lecture en deux lignes. En passant de deux à trois lignes, les valeurs extrêmes d'évolution sont les suivantes :

- **Le 6 avril** : – 87 grains ce qui représente une baisse de 13,64%, diminution relative la plus forte.
- **Le 10 avril** : + 54 grains correspondant à une hausse de 1,80% ; l'augmentation relative la plus élevée a lieu le 13 juillet (+ 16,49%).

Comme à Chalon, les évolutions à la hausse ou à la baisse ne sont pas propres à une période de l'année et, par conséquent, ne sont pas dépendantes des types polliniques présents dans l'atmosphère.

A Dijon, où le passage de deux à trois lignes entraîne 29 diminutions sur 35 valeurs, l'évolution des concentrations en valeur absolue représente 8,54% de la moyenne journalière pour deux lignes de lecture soit un total brut de 61,26 grains. En passant de deux à trois lignes, les valeurs extrêmes d'évolution sont les suivantes :

- **Le 14 avril** : – 296 grains ce qui représente une baisse de 14,50% ; la diminution relative la plus élevée a lieu le 9 avril (– 25,14%).
- **Le 1^{er} avril** : + 47 grains correspondant à une hausse de 18,22%, augmentation relative la plus forte.

La variabilité des concentrations, même si elle est surtout orientée à la baisse, est équivalente à celle enregistrée à Chalon-sur-Saône.

A Castres, où les concentrations moyennes sont nettement plus élevées que pour les trois villes bourguignonnes, la moyenne absolue des écarts (55,18 grains) représente 5,09% de la concentration moyenne journalière sur deux lignes. En passant de deux à trois lignes, les valeurs extrêmes d'évolution sont les suivantes :

- **Le 16 avril** : – 214 grains ce qui représente une baisse de 16,42%, diminution relative la plus forte.
- **Le 16 mars** : + 265 grains correspondant à une hausse de 5,01% ; l'augmentation relative la plus élevée a lieu le 20 juin (+ 35,32%).

Il semble que les différences de concentrations enregistrées soient liées à la vitesse du vent, problème évoqué par Hirst (Hirst, 1953), qui peut entraîner des effets de bord et une répartition inhomogène des pollens sur l'ensemble de la bande. Une étude de l' Ecole Nationale Supérieure Agronomique de Montpellier (ENSA-M), datant de 2005, avait d'ailleurs déterminé qu'une incertitude de mesure existe pour les concentrations polliniques journalières inférieures à 50 grains/m³, avec ce type de capteur, en lien avec la taille des grains et la vitesse du vent. Cette étude précisait, par ailleurs, que la méthode Hirst répondait aux critères de qualité définis dans le journal officiel des Communautés européennes de 1999 lorsque les concentrations sont supérieures à ce seuil de 50 grains.

1.1.5.4 Bilan

Les valeurs des concentrations journalières sont influencées par de nombreux paramètres, autres que la répartition spatiale des espèces végétales présentes dans la zone d'«action» du capteur. Cela a pour conséquence d'influencer les quantités enregistrées dans la base de données. De ce fait, se baser sur les valeurs brutes enregistrées pour déterminer des niveaux de risque d'exposition allergique, semble sujet à caution. La seule modification du nombre de lignes de lecture peut modifier les quantités et faire passer les concentrations d'un taxon d'un niveau de risque à un autre. Comparer les concentrations d'un taxon pollinique donné entre deux villes s'avère également délicat, si les capteurs ne sont pas situés à la même hauteur.

L'idéal serait de compter les pollens présents sur l'ensemble de la bande adhésive et que tous les capteurs soient situés à une hauteur comprise dans une fourchette de faible amplitude. Dans ces conditions, on pourrait interpréter les écarts de concentrations, d'une ville à l'autre, par les seuls critères météorologiques et de répartition géographique des espèces. Mais c'est utopique ; cela demanderait un travail d'analyse titanesque, irréalisable avec les moyens techniques actuels. Il serait alors particulièrement intéressant de placer, pour chaque station, des capteurs à différentes hauteurs, et de déterminer si les concentrations enregistrées sont significativement différentes ou non. Si c'est le deuxième cas de figure qui s'impose, la question de la représentativité des capteurs serait réglée.

Ainsi, contrairement à la mesure des polluants dans l'air ambiant, la mesure du contenu pollinique de l'air, comme celle des moisissures d'ailleurs, ne peut se faire de manière automatique. La reconnaissance des pollens à l'aide d'un logiciel informatique, déjà tentée, s'avère excessivement laborieuse du fait de la complexité de la structure des pollens, de la forte ressemblance existant entre de nombreux taxons, et de la position des pollens sous la lame qui les rend déjà, assez souvent, difficilement identifiables en reconnaissance visuelle.

1.1.6 Conclusion

Cette présentation des données polliniques a permis de détailler tout le processus de capture, de relevé et d'analyse qui nécessite une bonne gestion des différentes actions nécessaires au bon fonctionnement des capteurs avec, par exemple, une bonne coordination dans l'envoi et la réception des tambours des sites extra locaux. Etant donné la nature du contenu à analyser (les pollens), un décompte automatique est impossible. Ainsi, ne pouvant étudier une bande complète, la lecture ne s'effectue que sur deux ou trois lignes, ce qui aboutit à des concentrations qui ne seraient que des approximations. De plus, la position des capteurs en hauteur par rapport au sol, différente d'un site à l'autre, influence très probablement les concentrations relevées et rend la comparaison inter-site délicate. Ainsi, ces concentrations ne reflètent pas exclusivement la géographie de la végétation de la zone d'action des capteurs et les apports exogènes. Cela peut poser problème car il n'est pas certain que les concentrations enregistrées représentent bien ce que peut respirer n'importe quel individu. La standardisation de la hauteur des capteurs n'est cependant pas évidente. En effet, il n'est pas facile de trouver un bâtiment ayant une terrasse sommitale bien dégagée, à la fourchette de hauteur préconisée de 15-20 mètres.

Mis à part ce point-là, les séries polliniques présentent, en moyenne, une proportion raisonnable de lacunes (moins de 5% sur les sites étudiés), même si cela cache de grandes disparités. Ces lacunes sont, notamment, dues à des problèmes mécaniques que l'on ne peut pas toujours éviter. Enfin des problèmes de réception des données par le système informatique du RNSA ont été mis à jour, entraînant des lacunes dans la base de données, alors que les comptes polliniques existent bien. Le dernier point à améliorer serait la date de démarrage des capteurs qui n'est pas toujours

appropriée, et qui tronque la saison pollinique de plusieurs genres printaniers tels le noisetier, l'aulne, voire même le frêne.

Ces données polliniques, comme d'autres types de données d'ailleurs, ne sont donc pas parfaites. Néanmoins certains problèmes rencontrés pourraient être corrigés comme, par exemple, s'assurer d'une mise en route adéquate des capteurs en début de saison.

Pour notre étude, nous ne pouvons nous baser que sur ces données qui ont le mérite d'exister, car la gestion d'un tel réseau par le RNSA, qui est une association, représente un vrai challenge. Tout en restant prudent sur la représentativité des concentrations relevées, l'analyse statistique, dans un but préventif pour les personnes allergiques, va s'avérer réalisable excepté pour les stations et/ou années présentant de trop nombreuses lacunes.

1.2 PRESENTATION GENERALE DES TAXONS DE L'ETUDE

1.2.1 Les types de taxons retenus

Les taxons originellement retenus pour ce travail sont les suivants : les Cupressacées-Taxacées, le bouleau, le frêne, les Poacées et l'ambroisie. Ils s'inscrivent parmi les plus allergisants en France et en Europe et font l'objet de nombreux articles. Malheureusement, nous avons dû retirer les Cupressacées-Taxacées de l'étude, car la prévision de la date d'apparition du risque s'avère quasiment impossible. Pour ce taxon, nous nous heurtons en effet à deux problèmes :

- La pollinisation de ces deux familles, comprenant de nombreux genres (cyprès, if, thuya et genévrier entre autres), s'effectue de manière quasi continue tout au long de l'année pour les stations du pourtour méditerranéen avec, cependant, un optimum en février et mars. Cela rend difficile la prévision de l'apparition du risque, d'autant plus que le pollen de cyprès, seul taxon pollinique réellement allergisant de cette famille, ne peut être distingué, rappelons-le, de celui des autres genres de Cupressacées-Taxacées, en microscopie optique. Un article traitant de la prévision de la date de début de pollinisation de ce taxon (Fuertes-Rodríguez *et al.*, 2007) a débouché sur des prévisions très mauvaises pour une année (43 jours d'écart entre date prévue et observée).
- En lien avec cette pollinisation quasi continue, se pose le problème, déjà évoqué (cf.1.1.4.1), du fonctionnement des capteurs des sites du pourtour méditerranéen. En effet, ces derniers n'ont pas toujours fonctionné de manière continue, si bien que les concentrations de Cupressacées-Taxacées, susceptibles d'avoir lieu en décembre ou janvier, n'étaient pas enregistrées.

Nous allons donc nous focaliser sur les quatre autres taxons en présentant rapidement le genre ou la famille correspondante, le pollen et le risque allergénique qui lui est associé.

1.2.1.1 Les Poacées

Les Poacées forment une importante famille botanique. Elle comporte près de 12 000 espèces regroupées dans environ 700 genres. On y retrouve la plupart des espèces de plantes qu'on appelle communément «herbes» dont la majorité sont sauvages, comme le dactyle pelotonné (figure 14). C'est une famille cosmopolite qui présente un intérêt économique majeur, car un grand nombre de ces espèces joue un rôle fondamental dans l'alimentation de l'homme et des animaux (blé, avoine, maïs, canne à sucre). Ce sont des plantes essentiellement anémogames, bien que certaines espèces soient autogames. Elles se rencontrent sur tous les continents et sont bien représentées dans toutes les zones climatiques (sauf l'Antarctique). Cette famille représente environ 20% de la couverture générale du globe et elle est l'élément dominant de plusieurs formations végétales très étendues comme la steppe, la savane, la pampa, les Grandes Plaines américaines et la pelouse alpine.



Figure 14 : Une espèce de la famille des Poacées : le dactyle pelotonné (*Dactylis glomerata* L.) (Guérin, 1993).

Cette couverture végétale importante, associée au caractère essentiellement anémogame de la famille, explique que les pollens de Poacées arrivent en tête des émissions polliniques dans l'atmosphère européenne (Laaidi et Laaidi, 2001), essentiellement par l'entremise de quelques espèces particulièrement productives selon une étude réalisée en Espagne (Prieto-Baenapablo *et al*, 2003). La période de pollinisation des Poacées sauvages est très longue et s'étend souvent sur 7 à 8 mois du fait de la floraison successive des espèces de cette famille. Mais c'est principalement au mois de mai que les concentrations sont les plus élevées dans les régions méditerranéennes, en juin dans le nord-ouest et le centre de l'Europe et en juillet dans la partie septentrionale du continent (Spieksma *et al.*, 1993). En France, le début de la pollinisation a lieu peu avant la mi-mai dans le quart nord-est du pays et dès la fin avril dans le sud, le sud-ouest et les régions proches de l'Atlantique ; les plus fortes concentrations sont relevées pendant la première quinzaine de juin. A Amiens, site de référence de par le caractère complet et la relative longueur de ses séries polliniques, la date de début de pollinisation moyenne (sur la période 1989-2005), selon le critère de fréquence cumulée de 5% (Nilsson et Persson, 1981), intervient le 20 mai avec un écart-type de 4,5 jours. La variation interannuelle maximale de la date de début de pollinisation dépasse les deux semaines, allant du 13 mai (1990) au 30 mai (1996). La concentration moyenne annuelle est de 4 032 grains/m³ avec un coefficient de variation des concentrations de 34,5%, sur la période 1989-2005. Ce qui montre une relative homogénéité, d'une année à l'autre, du contenu en pollens de Poacées de l'air d'Amiens.



Figure 15 : Photographies du pollen de Poaceae (ϕ de 10 à 80 μ m) en microscopie optique (à gauche) et microscopie électronique (à droite) (Laaidi et Laaidi, 2001).

Le grain de pollen de Poacées est monoporé (présence d'un pore sur sa paroi), de forme sphérique, elliptique ou ovoïde (figure 15). Le pollen des Poacées sauvages est plus petit (10 à 40 μm) que celui des céréales (40 à 80 μm). La plus grande allergénicité des premiers peut s'expliquer par leur petite taille (Guérin, 1993), effet amplifié par la plus grande abondance de ces pollens «sauvages» dans l'atmosphère par rapport à ceux des céréales. Cela se confirme par les concentrations et, surtout, par le fait que ce pollen est la principale cause de pollinoses en Europe (D'Amato *et al.*, 1998). Des études réalisées sur le Centre-Est de la France (Laaïdi K, 1997 ; Laaïdi *et al.*, 1998 ; Laaïdi *et al.*, 2003c) confirment la première place du pollen de Poacées dans le déclenchement de pollinoses. D'ailleurs en France, et également aux Pays-Bas, la sensibilisation allergique à ce pollen atteint des pointes à 80% (Spieksma *et al.*, 1993).

1.2.1.2 Le bouleau

Faisant partie des Bétulacées-Corylacées, au même titre que le noisetier, l'aulne et le charme, il est l'un des arbres à feuilles caduques les plus caractéristiques des régions du nord, du nord-ouest et du centre de l'Europe. Il peut atteindre 20 à 30 mètres de hauteur (figure 16). Les fleurs mâles et femelles sont présentes sur le même individu (arbres monoïques), et groupées en chatons (Guérin, 1993). On rencontre essentiellement deux espèces:

- Le bouleau blanc (*Betula verrucosa* E., *Betula alba* C. ou *Betula pendula* R.) est un arbre de peu de longévité pouvant atteindre 30 mètres. Il se rencontre dans la majeure partie de l'Europe, sur sols pauvres et sableux.
- Le bouleau pubescent (*Betula pubescens* E.) qui est une espèce plutôt nordique. Peu abondant en France, on le rencontre dans les tourbières, en montagne, jusqu'à 2 000 mètres d'altitude. Il ne dépasse guère 20 mètres de hauteur et apprécie les terrains détrempés (Guérin, 1993).



Figure 16 : Un spécimen de bouleau : le bouleau pleureur (source : www.web-provence.com).

Du fait de sa répartition spatiale étendue, associé au caractère anémogame de la famille, les pollens de bouleau arrivent en 4^{ème} position des émissions polliniques (Guérin, 1993) et, régulièrement, dans les trois premiers taxons dans la moitié Nord de la France. La période de pollinisation du bouleau s'étend de la mi-mars au mois de mai, et coïncide avec la pousse des feuilles. En France, la date de début de pollinisation connaît une variation maximale de l'ordre d'un mois, d'une année à l'autre. Par exemple à Amiens, selon le seuil de fréquence cumulée de 5% (Nilsson et Persson, 1981), la date de démarrage oscille du 18 mars (1990) au 18 avril (1996) sur la période 1989-2005 avec une moyenne au 1^{er} avril et un écart-type de 6,9 jours. La concentration moyenne annuelle (1989-2005) est de 3 699 grains/m³ avec un coefficient de variation de 72,4%. Cette variabilité interannuelle élevée confirme ce qui est observé au niveau européen (Spieksma *et al.*, 1993).

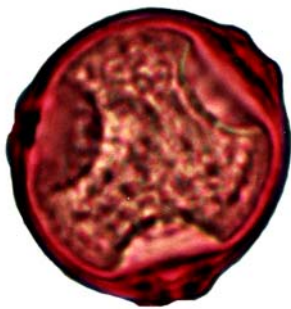


Figure 17 : Photographies du pollen de bouleau (σ de 18 à 24 μm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Laaïdi et Laaïdi, 2001).

Le grain de pollen de bouleau est triporé (présence de trois pores disposés sur la partie équatoriale du pollen), subsphérique à pores protubérants (figure 17). De petite taille (18 à 24 μm), il est l'un des principaux taxons allergisants en Europe septentrionale et centrale (Wallin *et al.*, 1991 ; D'Amato et Spieksma, 1992 ; Wihl *et al.*, 1998) à l'origine des pollinoses. Ces dernières ont une telle importance en Europe du Nord qu'un modèle complexe de prévision de la dispersion du pollen de bouleau a été initié par l'Institut Météorologique Finlandais (Sofiev *et al.*, 2006). Dans le Centre-Est de la France (Laaïdi K, 1997 ; Laaïdi *et al.*, 1998 ; Laaïdi *et al.*, 2003c), le bouleau apparaît comme le deuxième taxon le plus allergisant après les Poacées. La sensibilité allergique à ce pollen est bien connue avec, pour preuve, un nombre important de publications et d'articles à son sujet. Elle est d'autant mieux connue qu'elle se croise avec une sensibilité aux légumes et/ ou aux fruits, telles les pommes (Lahti *et al.*, 1980 ; Eriksson *et al.*, 1982 ; Dreborg et Foucard, 1983 ; Spieksma *et al.*, 1993 ; Malandain, 2003) ce qui a conduit à rechercher les variétés de pommes faiblement allergéniques (Asero *et al.*, 2006).

1.2.1.3 Le frêne

Cet arbre fait partie de la famille des Oléacées au même titre que le troène et, surtout, l'olivier. Le frêne commun (*Fraxinus excelsior* L.), le plus répandu, est l'un des grands arbres des forêts françaises et, plus largement, des régions tempérées d'Europe (Spieksma *et al.*, 1993). Pouvant atteindre une hauteur de 40 mètres et un diamètre de 2 m, on le retrouve jusqu'à 1 500 mètres

d'altitude (Guérin, 1993). L'autre type le plus souvent rencontré est le frêne à fleurs (*Fraxinus ornus* L.). Le frêne pousse spontanément dans les forêts, mais il est également cultivé comme arbre ornemental (figure 18) ou afin de servir de bois de construction et d'ameublement (en ébénisterie notamment). Les fleurs sont unisexuées ou bisexuées et poussent en groupes ou en grappes de couleur violacée. En hiver les bourgeons, noirs, sont particulièrement caractéristiques de cet arbre.



Figure 18 : Un spécimen de frêne : le frêne commun (source : www.web-provence.com).

Au niveau des émissions polliniques en Europe, le pollen de frêne n'apparaît qu'en 10^{ème} position (Guérin, 1993). C'est le taxon pollinique de la famille des Oléacées le plus fréquemment rencontré dans l'atmosphère. En France, il se place généralement derrière le bouleau en termes d'émission pollinique, excepté dans la moitié sud du pays ; sa répartition est ainsi plus homogène sur l'ensemble du territoire français. Le frêne fleurit au printemps, de février à mai, en fonction du climat et des conditions météorologiques et les fleurs apparaissent un peu avant les feuilles.

En France, le début de la pollinisation du frêne s'effectue, en moyenne, 7 à 14 jours avant celle du bouleau dans la moitié Nord du pays, là où les émissions polliniques des deux genres sont importantes. Ainsi à Amiens, la pollinisation du frêne débute en moyenne 12 jours avant celle du bouleau (20 mars contre 1^{er} avril, respectivement, pour une année non bissextile) sur la période 1989-2005. Cependant les écarts entre les deux taxons sont très variables d'une année à l'autre, ils vont ainsi de 2 jours en 2005 (30 mars pour le frêne, 1^{er} avril pour le bouleau) à 27 jours en 1995 (11 mars pour le frêne, 7 avril pour le bouleau). La variation interannuelle maximale de la date de début de pollinisation dépasse le mois, allant du 11 mars (1995, 1997 et 2001) au 15 avril (1996) ; cette forte variabilité se traduit par un écart-type plus élevé que celui du bouleau (8,6 au lieu de 6,9 jours). La variabilité interannuelle des concentrations est aussi élevée que celle du bouleau (coefficient de variation de 75,9%) avec une moyenne de 2 159 grains/m³/an.

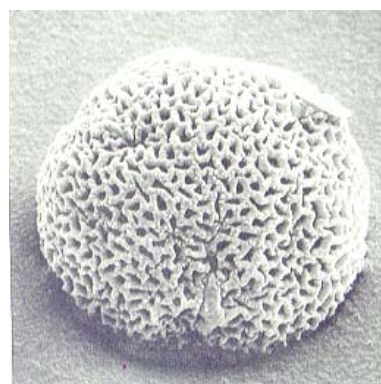
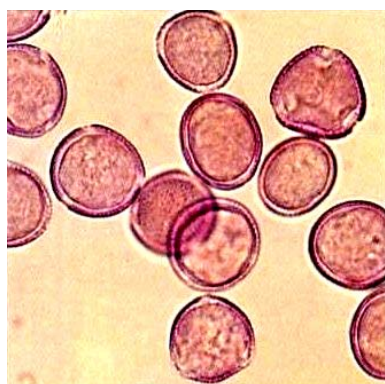


Figure 19 : Photographies du pollen de frêne (ø de 21 à 26 µm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Spieksma *et al.*, 1993).

Le grain de pollen de frêne est tricolpé (présence de trois sillons sur la partie équatoriale du pollen) et de forme bréviaxe (figure 19), c'est-à-dire que l'axe équatorial (transversal) du pollen est plus long que l'axe polaire (vertical). Ce pollen se caractérise par un diamètre légèrement supérieur au pollen de bouleau, compris entre 21 et 26 µm. Sa morphologie maillée est très proche de celle des autres genres et espèces de la famille des Oléacées, dont l'olivier. Le pollen de frêne, contrairement à son «cousin» méridional, a été souvent considéré comme peu susceptible de déclencher des allergies (Spieksma *et al.*, 1993). Mais, en Suisse dès 1991, il est considéré comme faisant partie des pollens les plus allergisants au même titre que le bouleau ou les Poacées (Frei *et al.*, 1995 ; Wüthrich, 1995 ; Wüthrich *et al.*, 1995). Dans le Centre-Est de la France

(Laaidi K, 1997 ; Laaidi *et al.*, 1998 ; Laaidi *et al.*, 2003c), le frêne apparaît comme le sixième taxon le plus allergisant après les Poacées, et les quatre taxons des Bétulacées-Corylacées que sont le bouleau (2^{ème}), le noisetier (3^{ème}), l'aulne et le charme (4^{ème} ex aequo). La monosensibilisation aux allergènes de frêne est rare, et les polysensibilisations avec d'autres taxons sont, par contre, fréquentes. Mais on a pu relever uniquement des sensibilisations à l'allergène majeur des Oléacées (Pauli *et al.*, 2003), semblant ainsi montrer que Spieksma *et al.* (1993) sous-estimait le potentiel allergisant de ce pollen. Il semble, malgré tout, rester en retrait par rapport au bouleau ou aux Poacées.

1.2.1.4 L'ambroisie

Cette plante fait partie de la famille des Astéracées qui comprend, entre autres genres connus, l'armoise, la marguerite, le pissenlit et le tournesol. Deux principales espèces d'ambroisie se distinguent :

- *Ambrosia maritima* L., originaire des régions méditerranéennes (Laaidi *et al.*, 2003b).
- *Ambrosia artemisiifolia* L., originaire d'Amérique du Nord.

Signalée, pour la première fois, dans un milieu naturel (un champ de l'Allier) en 1865 (Touraine *et al.*, 1966), cette dernière espèce a été introduite principalement en France avec des plants de pommes de terre dans les années 1930, puis avec les réserves de guerre américaines en 1944 et, enfin, par l'entremise de sacs de céréales dans les années 1960 (Touraine, 1984 ; Déchamp, 1995). Cette espèce a pour caractéristique majeure de coloniser une grande variété de zones grâce à la probable existence de multiples moyens de dispersion (Dessaint *et al.*, 2005). Pouvant atteindre 1,5 mètre de hauteur et se caractérisant par ses feuilles lancéolées (en forme de lance) et épineuses (figure 20), elle colonise les parcelles agricoles et, notamment, celles de tournesol, qui fait partie de la même famille. D'autres mécanismes de propagation existent (Dessaint *et al.*, 2005) tels que l'utilisation de graviers pour la construction, de marne pour le chaulage (technique servant à améliorer le rendement d'un sol par l'ajout de chaux), de compost sur les terres cultivées ou par le déplacement de terre végétale. En France, cette espèce concerne essentiellement la

région lyonnaise et la vallée du Rhône avec une évolution croissante des concentrations depuis 1987 jusqu'à un certain palier au début des années 2000 (Thibaudon *et al.*, 2003 ; Thibaudon et Lachasse, 2004). En Europe, on retrouve plusieurs foyers de colonisation (Spieksma *et al.*, 1993) tels la plaine du Pô (Italie), l'Europe Centrale de manière générale avec l'Autriche, la République Tchèque et la Hongrie plus particulièrement.



Figure 20 : Aperçu de quelques branches et feuilles d'un plant d'ambrosie (source : www.univ-ubs.fr).

Au niveau des émissions polliniques en Europe, le pollen d'ambrosie n'apparaît qu'à une modeste 41^{ème} place (Guérin, 1993). Il n'apparaît en quantité importante que sur le site de Lyon où il avoisine les 4% du total global en pollens sur la période 1997-2005. La période de floraison et de pollinisation s'étale d'août à septembre. A Lyon, la date de début de pollinisation se situe, généralement, aux alentours de la mi-août.

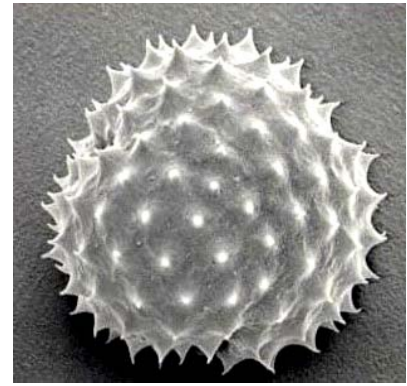
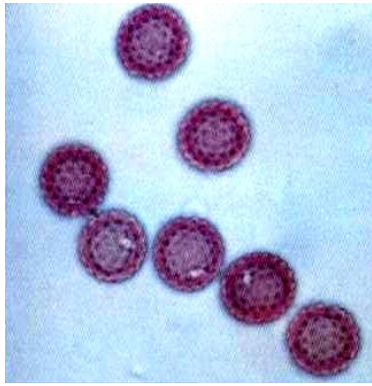


Figure 21 : Photographies du pollen d'ambrosie (ϕ de 19 μm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Spieksma *et al.*, 1993).

Le grain de pollen d'ambrosie est tricolporé (présence simultanée de trois sillons et trois pores sur la partie équatoriale du pollen) et de forme quasi sphérique. Il se caractérise par des épines triangulaires, régulièrement réparties à la surface, et qui semblent plus ou moins émoussées (figure 21). Cet aspect épineux est la principale caractéristique des pollens des espèces appartenant à la famille des Astéracées. C'est un pollen particulièrement allergisant ne contenant pas moins de six allergènes majeurs à l'origine du déclenchement de pollinoses (Bagarozzi et Travies, 1998). Dans le Sud de la Hongrie (Kadocsá et Juhász, 2002), il apparaît en tête des pollens à l'origine de rhinites allergiques (85% de tests positifs), loin devant les Poacées. Une étude européenne sur l'allergie à l'ambrosie, établie à partir d'un panel de patients dans les années 1990 (Rybníček et Jaeger, 2001), a donné des résultats positifs de plus de 80% en Hongrie, environ 70% en Italie du Nord, 35% à Prague et de 25 à 30% à Vienne (Autriche). Les résultats de la région lyonnaise ont été de l'ordre de 30 à 40%, justifiant la parution d'articles récents à propos de la progression de la plante, de l'allergénicité de son pollen et des différents moyens de limiter ses effets néfastes sur la santé (Laaidi *et al.*, 2003a ; Laaidi *et al.*, 2003b ; Laaidi *et al.*, 2004 ; Thibaudon *et al.*, 2004). De plus, il s'est avéré que la consommation d'antihistaminiques, et autres traitements contre les symptômes allergiques, dans la région Rhône-Alpes (2000 à 2003) a été multipliée par un ratio de 2 à 4 durant la pollinisation de l'ambrosie ; les remboursements de ces médicaments sont donc un bon indicateur de l'impact épidémiologique de ce pollen (Christophe *et al.*, 2004).

1.2.2 Concentration et répartition des quatre taxons de l'étude sur l'ensemble du territoire français

A partir des 29 villes de l'étude disposant de séries polliniques sur la période 1997-2005, nous allons étudier la part des concentrations des quatre taxons par rapport au total global de l'ensemble des pollens recensés. Les résultats que nous allons présenter doivent être considérés avec certaines précautions. En effet, les valeurs manquantes, et surtout les mises en route trop tardives des capteurs certaines années, ont une influence certaine sur la proportion d'un taxon pollinique par rapport au total tous taxons confondus, nommé index pollinique par le RNSA. Ainsi, la part occupée par des taxons précoces tels que le noisetier, l'aulne et les Cupressacées-Taxacées est notamment affectée par le deuxième cas de figure, précédemment cité, avec une sous-estimation de la proportion théoriquement occupée. De même les taxons plus tardifs, dans le cadre d'un démarrage des capteurs empiétant largement sur les saisons polliniques du noisetier, de l'aulne ou des Cupressacées (et Taxacées), voient la proportion qu'ils occupent gonflée artificiellement.

Si l'on souhaite parler de la «signature pollinique» d'un site, il faudrait que le capteur fonctionne de manière continue sans qu'il soit déplacé durant la période d'étude et, si possible, sans lacunes de plusieurs jours. Les résultats taxon par taxon présentés ici ne devraient être, au sens le plus strict, interprétés que comme constituant une part de tous les pollens comptabilisés.

Cependant, malgré les différents problèmes qui peuvent affecter les séries, les résultats qui seront présentés s'avéreront exploitables, notamment pour la mise en avant de particularismes zonaux ou régionaux, et globalement conformes à la répartition spatiale des différents genres et familles sur l'ensemble du territoire français, telle qu'elle a pu être esquissée dans le sous-chapitre 1.2.1.

1.2.2.1 Présentation générale des résultats

Durant les neuf années de la période d'étude, 9 341 731 grains de pollen ont été enregistrés – après application des facteurs de conversion – sur l'ensemble des 29 villes ; cela représente une moyenne de 35 792 grains par an et par ville. Les trois sites où ont été enregistrés le plus grand nombre de grains sont, dans l'ordre décroissant, Toulon (58 771 grains/m³/an), Nîmes (56 173

grains/m³/an) et Marseille (54 613 grains/m³/an). Ces trois sites méditerranéens se distinguent par leurs concentrations élevées essentiellement dues aux pollens de Cupressacées et de Taxacées, beaucoup plus abondants dans cette région que sur le reste du territoire, mais aussi par la durée de fonctionnement des capteurs plus longue que dans les zones non méditerranéennes. Brest (9 677 grains/m³/an), Nantes (19 966 grains/m³/an) et Lille (20 787 grains/m³/an) forment le trio enregistrant les comptes les plus faibles.

A partir des comptes globaux, nous avons élaboré un classement des taxons en fonction de la part qu'ils occupent sur l'ensemble des 29 villes. Les dix premiers taxons sont présentés dans le tableau 7 :

Cupressacées-Taxacées	15,63%
Urticacées	14,67%
Poacées	13,99%
chêne	10,29%
pin	6,96%
bouleau	6,68%
platane	5,33%
frêne	4,96%
châtaignier	3,70%
aulne	2,54%

Tableau 7 : Classement des dix premiers taxons par ordre décroissant sur l'ensemble des 29 villes de l'étude.
En gras : les taxons retenus pour notre travail.

Les Poacées représentent le troisième taxon par ordre d'importance avec 13,99% du total, tandis que le bouleau apparaît en sixième position et le frêne en huitième. Le pollen d'ambrosie, quant à lui, ne représente qu'une part infime du total global (0,39%).

Nous allons maintenant présenter les tableaux relatifs aux villes situées au sud et au nord du 45^{ème} parallèle. Neuf des 29 villes se situent en deçà de cette latitude : Bordeaux, Aurillac, Toulouse, Aix-en-Provence, Marseille, Nîmes, Toulon, Nice et Gap. Les dix premiers taxons par zone sont présentés dans le tableau 8 :

partie Sud		partie Nord	
Cupressacées-Taxacées	28,45%	Urticacées	18,16%
chêne	14,35%	Poacées	18,09%
pin	9,44%	bouleau	9,38%
Urticacées	9,14%	chêne	7,72%
Poacées	7,49%	Cupressacées-Taxacées	7,55%
platane	5,85%	frêne	5,81%
frêne	3,60%	pin	5,39%
châtaignier	2,73%	platane	5,00%
bouleau	2,40%	châtaignier	4,31%
peuplier	2,05%	aulne	2,91%

Tableau 8 : Classement des dix premiers taxons par ordre décroissant sur les villes situées au sud et au nord du 45^{ème} parallèle. En gras : les taxons retenus pour notre travail.

Dans la partie Sud, nous constatons le poids considérable des Cupressacées-Taxacées et le positionnement plus en retrait dans le classement des Poacées et du bouleau, dont le pourcentage est gonflé par les valeurs d'Aurillac, où il représente 11,56% des pollens comptés, et dans une moindre mesure par celles de Gap (3,89%). Le frêne, quant à lui, se place en 7^{ème} position avec un pourcentage de 3,60%. Le classement est tout autre dans la partie Nord où les Poacées arrivent en 2^{ème} position juste derrière les Urticacées, ces deux taxons se distinguant nettement des autres. Le bouleau, quant à lui, se révèle être le premier taxon arboréen de la zone, devant le chêne et les Cupressacées-Taxacées, avec une proportion quasiment quatre fois supérieure à celle enregistrée dans la partie Sud. Le frêne apparaît quant à lui en 6^{ème} position et l'écart entre les deux pourcentages est nettement plus réduit que pour les Poacées et le bouleau, semblant montrer une répartition plus homogène de ce taxon sur le territoire. Quant à l'ambroisie, elle ne représente que 0,22% du total de la partie Sud et 0,49% de la partie Nord.

Nous allons maintenant présenter les résultats, taxon par taxon, par ordre décroissant d'importance.

1.2.2.2 Les Poacées

Représentant 13,99% du total global de pollens recensés, les Poacées se classent en 3^{ème} position par ordre d'importance. Le total annuel moyen par ville s'élève ainsi à 5 007 grains. Les trois villes présentant les plus fortes quantités sont les suivantes :

- **Montluçon** : 11 870 grains/m³/an
- **Aurillac** : 9 672 grains/m³/an
- **Cholet** : 9 617 grains/m³/an

Celles où sont recensées les plus faibles quantités sont les suivantes :

- **Nice** : 1 014 grains/m³/an
- **Aix-en-Provence** : 1 546 grains/m³/an
- **Toulon** : 1 578 grains/m³/an

Nous constatons ainsi que trois villes méditerranéennes ferment la marche. Nîmes avec 1 977 grains et Marseille avec 2 040 grains sont classées juste avant. Cela semble montrer la faible abondance de ce taxon dans cette région par rapport aux autres sites.

Après cette présentation rapide des quantités brutes, nous allons nous attarder plus longuement sur les proportions de pollens de Poacées par rapport aux totaux de l'ensemble des taxons recueillis site par site.

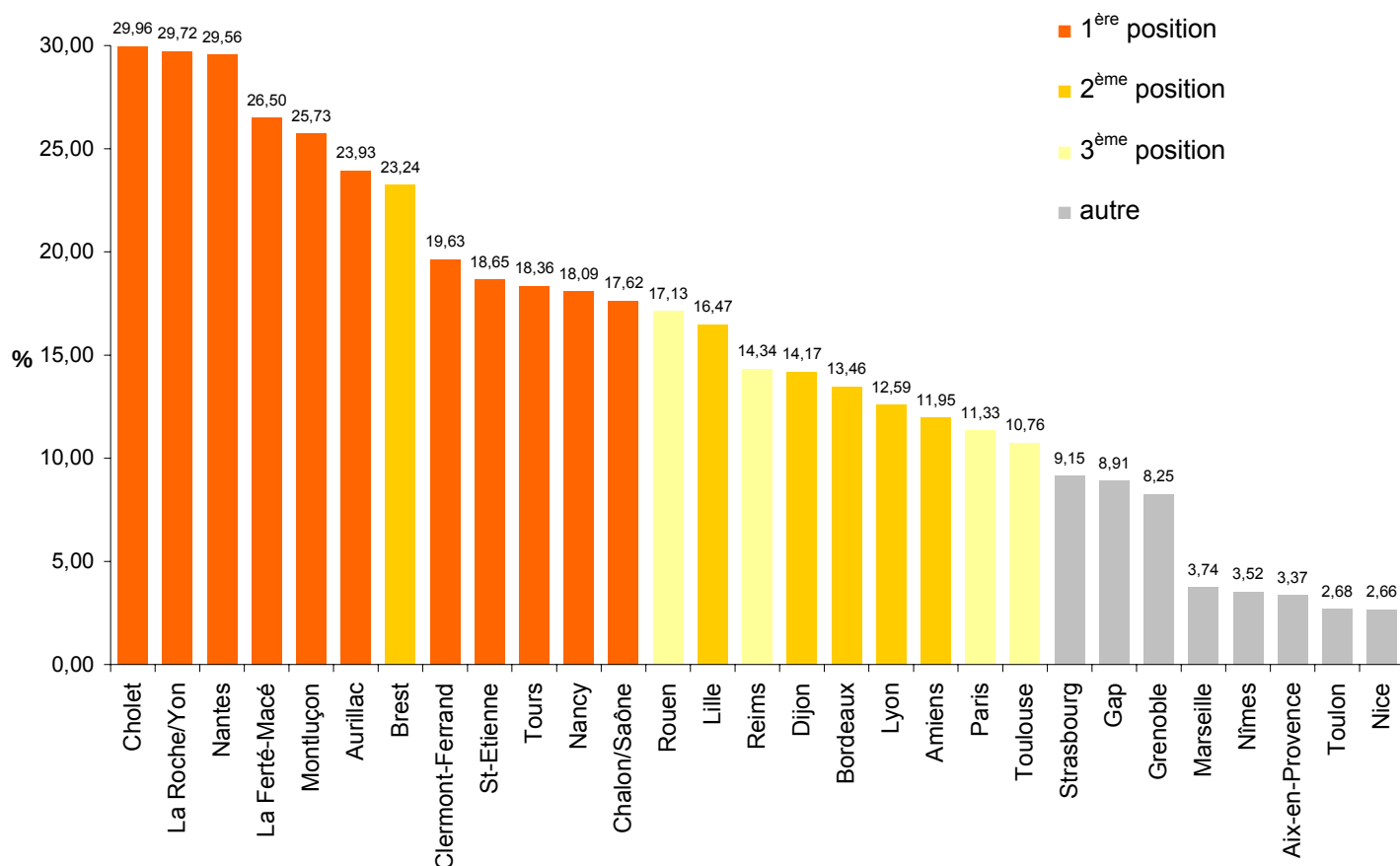


Figure 22 : Classement par ordre décroissant de la proportion de pollens de Poacées, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.

A travers la figure 22 nous pouvons distinguer trois zones géographiques homogènes présentant des pourcentages similaires. Nous constatons ainsi que les trois premières places sont occupées par les stations des Pays de la Loire (Cholet, La Roche-sur-Yon et Nantes) avec des valeurs comprises entre 29 et 30%. Nous notons également que les stations de la Ferté-Macé (Orne) et Brest présentent des pourcentages supérieurs à 20% et, associées aux trois stations précédentes, elles permettent de délimiter une zone «Atlantique» où les pollens de Poacées sont fortement représentés par rapport à l'index pollinique annuel moyen.

Le Massif Central, au sens large, se distingue également par des proportions élevées de Poacées. Cette zone contenant les sites de Montluçon, Clermont-Ferrand, Aurillac et Saint-Etienne se caractérise par une proportion de 22,46%, avec des valeurs s'échelonnant de 18,65% à Saint-Etienne à 25,73% à Montluçon.

A l'opposé, les cinq sites du pourtour méditerranéen marquent une rupture avec les autres en présentant des valeurs particulièrement faibles, s'échelonnant de 2,66% à Nice à 3,74% à Marseille. Les faibles quantités enregistrées associées au fait que ces villes se situent parmi celles où l'on comptabilise le plus de pollens, expliquent cette différenciation nette observée sur le graphique.

Nous constatons que pour 11 des 29 villes, la famille des Poacées constitue le taxon le plus fréquemment impacté par le capteur. Il est par ailleurs classé dans les trois premiers taxons dans 21 villes. Il faut noter que sur les six cas où ce taxon se classe en deuxième position, il est devancé à quatre reprises par les Urticacées, une fois par la platane (Lyon) et une fois par le chêne (Bordeaux). A l'exception de Toulouse, où il est précédé par les Cupressacées-Taxacées et le chêne, il est devancé par les Urticacées et le bouleau dans les trois autres villes où il apparaît en troisième position.

1.2.2.3 Le bouleau

Avec 6,68% du total global de pollens recensés, le bouleau se classe en 6^{ème} position par ordre d'importance. Le total annuel moyen par ville s'élève ainsi à 2 391 grains. Les trois villes présentant les plus fortes quantités sont les suivantes :

- **Rouen** : 5 867 grains/m³/an
- **Aurillac** : 4 673 grains/m³/an
- **Dijon** : 4 291 grains/m³/an

Celles où sont recensées les plus faibles quantités sont les suivantes :

- **Aix-en-Provence** : 244 grains/m³/an
- **Nîmes** : 309 grains/m³/an
- **Toulon** : 340 grains/m³/an

Nous constatons de nouveau que trois villes méditerranéennes ferment la marche, et qu'elles sont immédiatement précédées de Marseille (568 grains) et Nice (566 grains). Cela confirme bien le fait que le bouleau n'est pas un arbre méridional.

Après ce bref aperçu des concentrations, nous allons nous intéresser à la part occupée par ce taxon par rapport à l'index pollinique de chaque site.

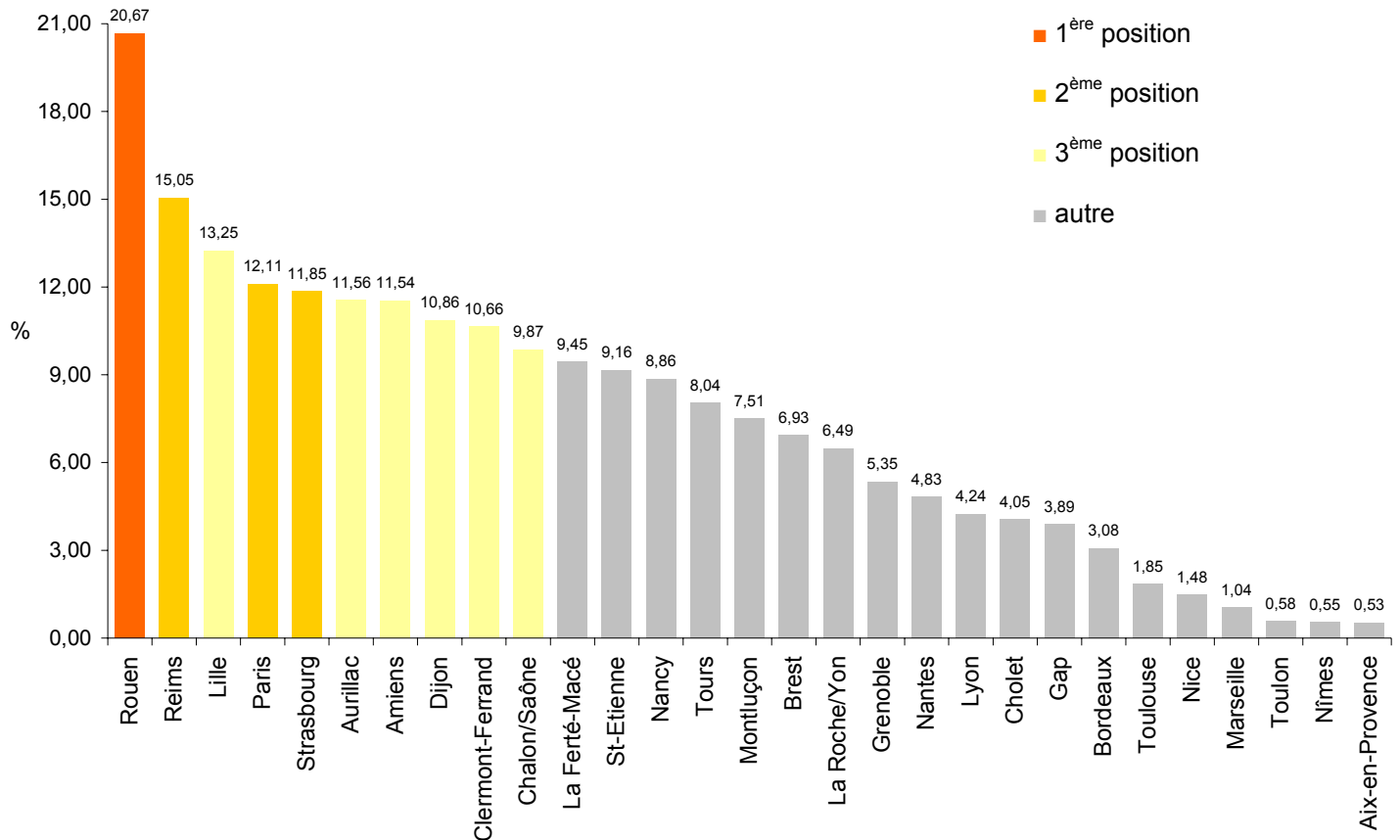


Figure 23 : Classement par ordre décroissant de la proportion de pollens de bouleau, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.

Nous pouvons remarquer sur la figure 23 que Rouen se distingue particulièrement des autres villes par la proportion de pollens de bouleau, qui dépasse 20%, confirmant les concentrations élevées enregistrées. Nous constatons que les plus fortes proportions sont enregistrées dans les stations du grand quart Nord-Est de la France et du Massif Central. Les plus faibles sont fournies par le pourtour méditerranéen avec des valeurs s'échelonnant de 0,53% (Aix-en-Provence) à 1,48% (Nice), ainsi que par le Sud-Ouest (3,08% à Bordeaux et 1,85% à Toulouse).

Le bouleau n'arrive en première position qu'à Rouen et se classe parmi les trois premiers taxons dans dix villes, parmi lesquelles toutes les villes du grand quart Nord-Est de la France à l'exception de Nancy où il ne se positionne qu'en quatrième position.

1.2.2.4 Le frêne

Avec 4,96% du total global de pollens recensés, le frêne se classe en 8^{ème} position par ordre d'importance. Le total annuel moyen par ville s'élève ainsi à 1 774 grains. Les trois villes présentant les plus fortes quantités sont les suivantes :

- **Aurillac** : 4 880 grains/m³/an
- **Grenoble** : 4 427 grains/m³/an
- **Clermont-Ferrand** : 3 036 grains/m³/an

Celles où sont recensées les plus faibles quantités sont les suivantes :

- **Brest** : 217 grains/m³/an
- **Nantes** : 532 grains/m³/an
- **Aix-en-Provence** : 600 grains/m³/an

Contrairement à ce qui a été constaté pour les deux précédents taxons, les sites méditerranéens n'occupent pas systématiquement les dernières places du classement par concentrations. Nous enregistrons entre 1 000 et 1 200 grains par an à Marseille, Nîmes et Toulon ; des valeurs supérieures à celles de plusieurs villes de la moitié Nord de la France telles Tours, Cholet et Lille. Suite à ce bref aperçu des concentrations en pollen de frêne, nous allons nous intéresser à la part occupée par ce taxon dans le total, tous taxons confondus, de chaque site.

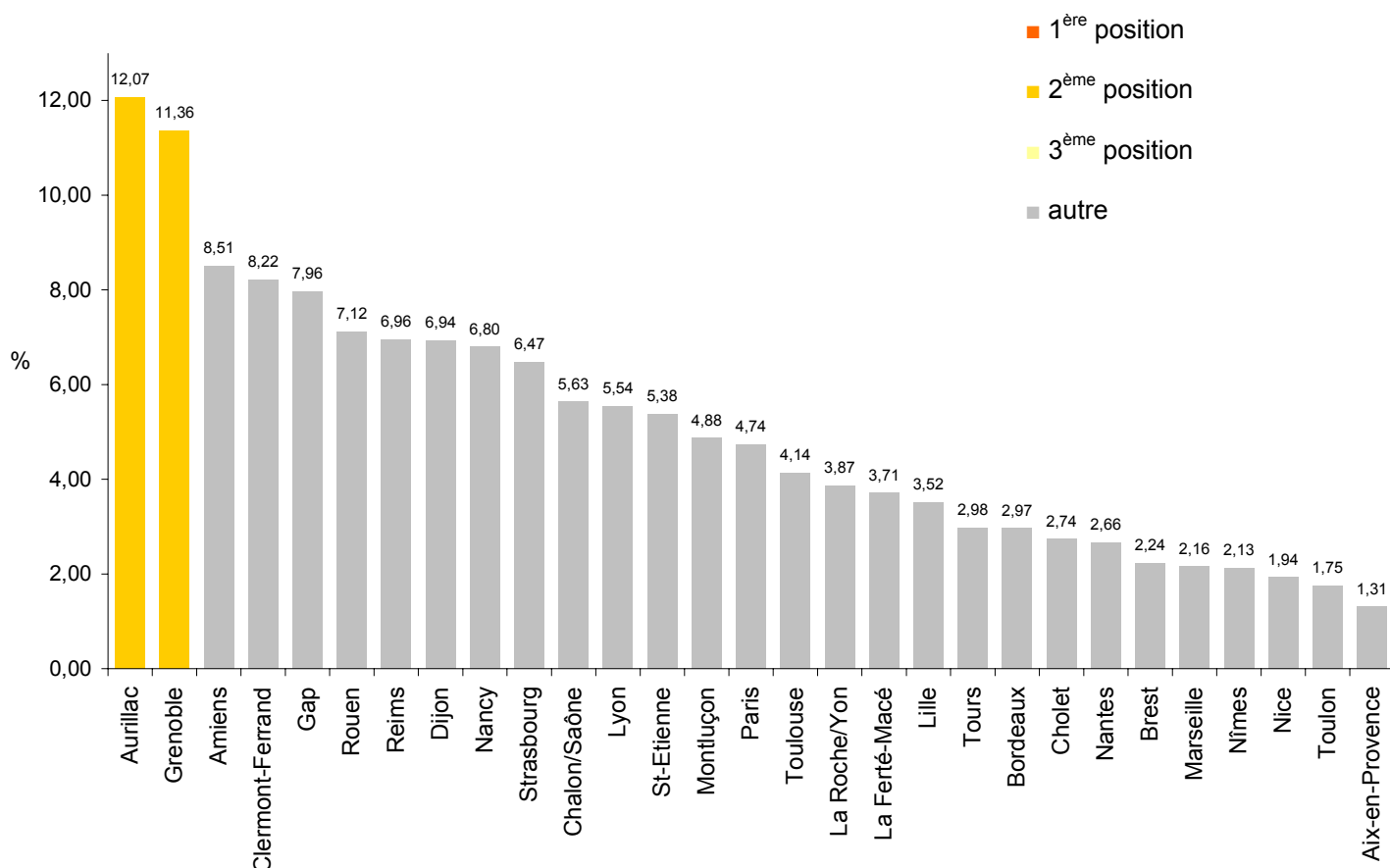


Figure 24 : Classement par ordre décroissant de la proportion de pollens de frêne, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.

Le frêne se caractérise par une présence globalement plus importante dans la moitié Nord du territoire et, pour 19 des 29 villes, il se positionne après le bouleau (figure 24). Ne font exception que les cinq villes méditerranéennes, Toulouse, Gap, Grenoble, Lyon et Aurillac où il est globalement plus abondant que le bouleau. Avec des valeurs s'échelonnant de 1,31% (Aix-en-Provence) à 2,16% (Marseille), les cinq villes méditerranéennes occupent les dernières places mais, contrairement au bouleau et surtout aux Poacées, il n'y a pas de rupture avec les sites non méditerranéens les précédant. Une plus grande homogénéité de la présence du pollen de frêne sur le territoire français est mise en avant en comparaison de la répartition du pollen de bouleau.

Une particularité géographique affectant la répartition de ce taxon se dessine à travers ce graphique. Aurillac, Grenoble, Clermont-Ferrand et Gap se situent parmi les cinq villes caractérisées par les plus forts pourcentages. Le frêne est d'ailleurs le deuxième taxon derrière les

Poacées à Aurillac et les Cupressacées-Taxacées à Grenoble. Or ces villes se situent toutes au pied de massifs de moyenne montagne (les Monts du Cantal à Aurillac et la chaîne des Puys à Clermont-Ferrand) ou de haute montagne (le Dévoluy et les Ecrins à Gap, les massifs de Belledonne, de la Chartreuse et du Vercors à Grenoble). Cette particularité ne semble pas fortuite, malgré les nécessaires précautions d'interprétation des résultats liées aux problèmes évoqués en début de chapitre. Le frêne semble ainsi être un genre courant de l'étage collinéen de ces zones (limite supérieure de 800 à 1 000 mètres d'altitude). Pouvant être présent jusqu'à 1 500 mètres d'altitude (Guérin, 1993), il doit aussi occuper l'étage montagnard (limite supérieure 1 500-1 700 mètres d'altitude) des différents massifs environnants.

1.2.2.5 L'ambroisie

Ne représentant que 0,39% de l'index pollinique, les concentrations polliniques d'ambroisie apparaissent relativement confidentielles à l'échelle du territoire français. Au niveau des herbacées, elles se situent derrière plusieurs genres ou familles telles les Plantaginacées (1,08%), l'oseille (0,98%), la mercuriale (0,54%), les Chénopodiacées (0,45%) et l'armoïse (0,42%). Le total annuel moyen par ville s'élève ainsi à 139 grains. Les trois villes présentant les plus fortes quantités sont les suivantes :

- **Lyon** : 1 819 grains/m³/an
- **Grenoble** : 283 grains/m³/an
- **Marseille** : 223 grains/m³/an

Celles où sont recensées les plus faibles quantités sont les suivantes :

- **Brest** : 1 grain/m³/an
- **Rouen** : 3 grains/m³/an
- **Tours** : 4 grains/m³/an

Nous constatons que les trois dernières villes présentent des concentrations anecdotiques et que 11 villes présentent des comptes annuels moyens inférieurs à 20 grains. Seule Lyon présente des quantités importantes symbolisant bien le caractère régional de l'implantation de cette herbacée. Suite à cet aperçu des concentrations polliniques d'ambroisie, nous allons nous intéresser à la part occupée par ce taxon dans le total, tous taxons confondus, de chaque site.

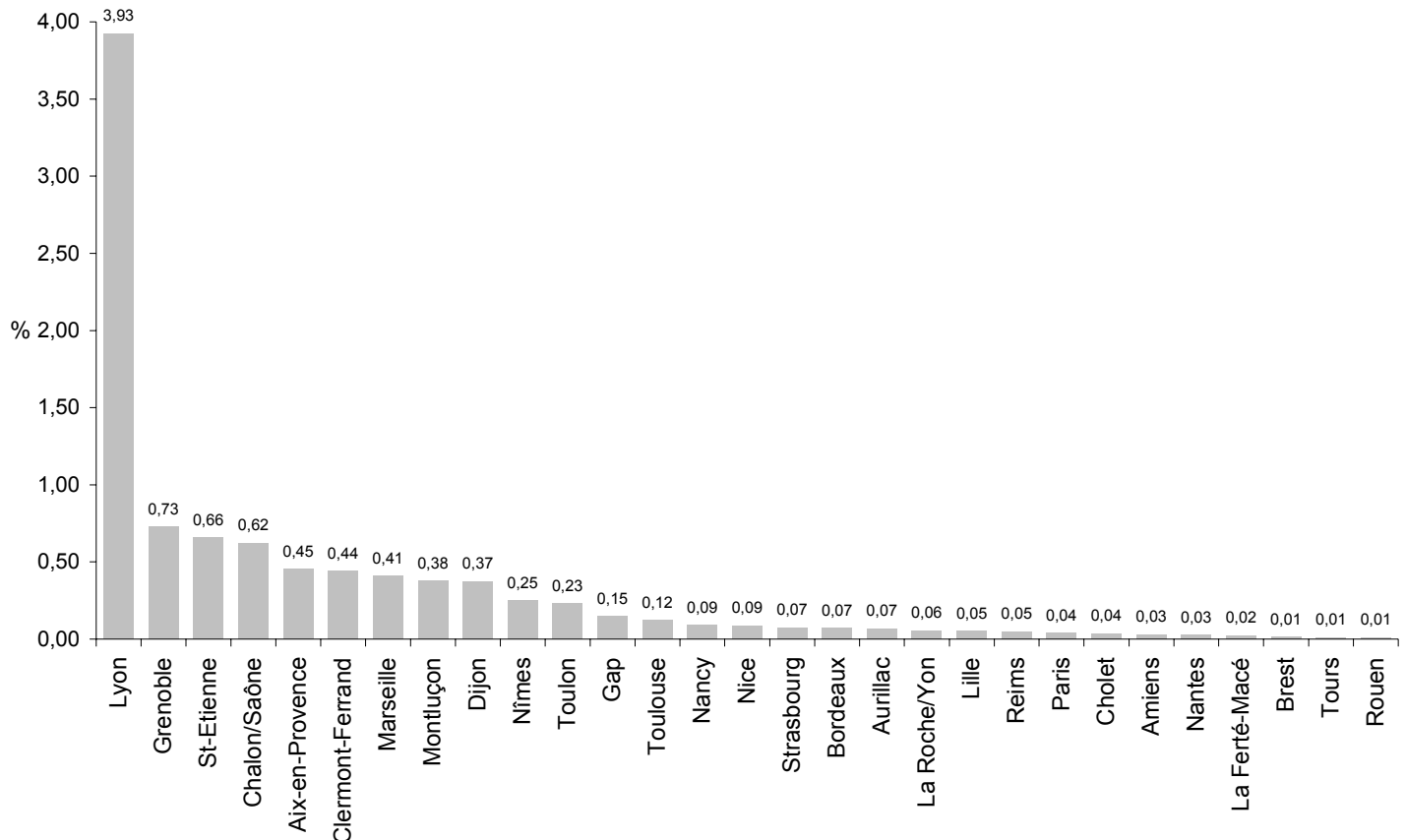


Figure 25 : Classement par ordre décroissant de la proportion de pollens d'ambroisie, ville par ville, sur la période 1997-2005.

Lyon se distingue de nouveau avec une proportion d'ambroisie proche de 4% ce qui en fait le 11^{ème} taxon dans cette ville (figure 25). Toutes les autres villes présentent des scores inférieurs à 1% où se positionnent, en tête, les sites des régions Rhône-Alpes (Grenoble, Saint-Etienne) et Auvergne (Clermont-Ferrand, Montluçon) ainsi que ceux se situant sur l'axe Rhône-Saône (Dijon et Chalon-sur-Saône au nord, Aix-en-Provence, Marseille et Nîmes au sud). Cela confirme l'implantation très spécifique de cette espèce dans la région lyonnaise et l'influence du vent à

l'origine des concentrations enregistrées dans les sites cités ci-dessus avec, par exemple, le rôle très probable du mistral pour les trois sites méditerranéens.

1.2.2.6 Une autre approche de la présentation générale des résultats : application d'une Analyse en Composantes Principales (ACP)

L'objectif est d'effectuer une présentation globale de la répartition des trois taxons présentant des pourcentages non négligeables d'une ville à l'autre : les Poacées, le bouleau et le frêne. L'ambrosie n'est pas retenue du fait des très faibles concentrations enregistrées dans tous les sites, à l'exception de Lyon. Afin de présenter des résultats prenant en compte ces trois taxons, il est nécessaire de mettre en œuvre une Analyse en Composantes Principales (ACP). Cette méthode a pour but de détecter la structure des relations pouvant exister entre les variables et de les synthétiser en un nombre de composantes principales plus faible que le nombre initial de variables. Chaque composante va correspondre à un type de relation donnant une information qui représente un certain pourcentage de la variance totale issue de l'ensemble des variables. En d'autres termes, le but de cette méthode est de faire ressortir les relations essentielles existant entre l'ensemble des variables là où une étude variable par variable serait fastidieuse ou improductive.

Dans le cas présent nous ne prenons en compte que trois variables, nombre qui ne nécessiterait pas, à priori, l'utilisation de cette méthode. Elle n'est établie ici que dans un but de synthèse des résultats.

Un maximum de trois composantes principales peut être extrait. Les deux premières composantes expliquent 86,67% de la variance totale et sont donc retenues. Nous allons ainsi présenter graphiquement les poids factoriels, qui peuvent être considérés comme les corrélations entre les variables et les composantes respectives, et ainsi permettre l'interprétation de ces dernières (figure 26).

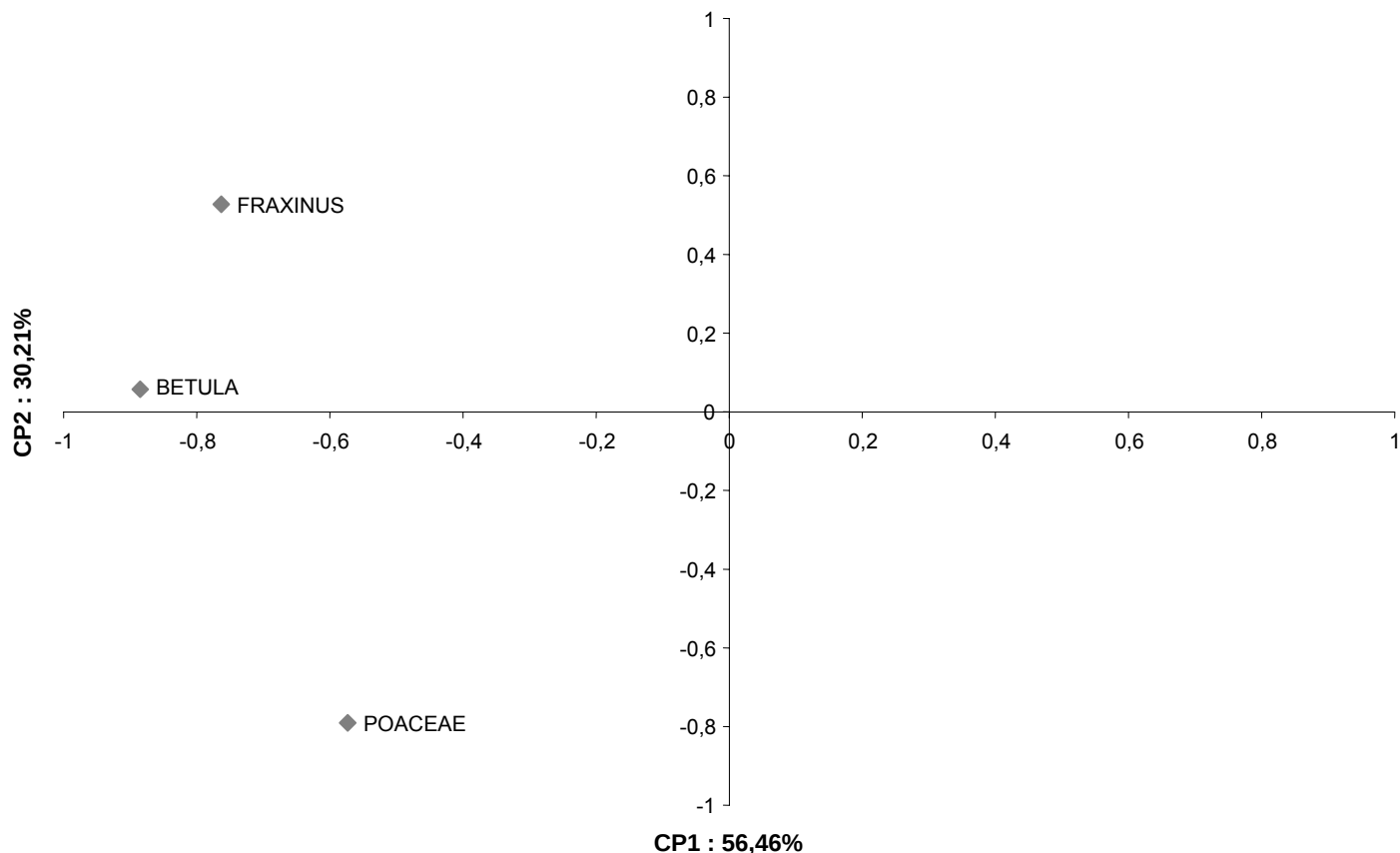


Figure 26 : Poids factoriels des trois variables sur les deux premières composantes principales.

56,46% de la variance totale est expliquée par la première composante principale. Les poids factoriels des trois variables associés à celle-ci sont négatifs et de valeurs relativement proches. Ces résultats traduisent l'existence d'une opposition entre, d'une part, des stations où ces trois taxons sont bien représentés, et constituent une part relativement élevée par rapport au total global des pollens comptabilisés, et les stations où ils n'y représentent qu'une faible part. Cette opposition apparaît comme la principale relation liant ces variables, c'est l'information statistiquement la plus importante extraite par l'intermédiaire de l'ACP.

La part de variance expliquée par la deuxième composante (30,21%) n'est pas négligeable, elle semble montrer que la force de la représentation de chaque taxon n'est pas homogène. Cette composante principale semble mettre en avant le poids important constitué par les pollens de Poacées sur le total global de certaines stations.

Afin d'affiner l'interprétation des résultats, nous allons présenter les résultats factoriels qui correspondent aux valeurs associées à chaque observation (ici nos 29 villes). Le signe de ces résultats factoriels, pour une composante principale donnée, est à mettre en relation avec le signe des poids factoriels et permet de classer les différentes stations en fonction des informations déterminées à partir des composantes principales. Nous allons les présenter à l'aide de deux cartes de France (figures 26 et 27). Pour éviter une surcharge de ces dernières, nous avons attribué à chaque ville un numéro dont voici la liste :

1 – Brest	16 – Montluçon
2 – La Ferté-Macé	17 – Clermont-Ferrand
3 – Rouen	18 – Saint-Etienne
4 – Amiens	19 – Lyon
5 – Lille	20 – Grenoble
6 – Paris	21 – Bordeaux
7 – Reims	22 – Toulouse
8 – Nancy	23 – Aurillac
9 – Strasbourg	24 – Nîmes
10 – Nantes	25 – Marseille
11 – Cholet	26 – Aix-en-Provence
12 – La Roche-sur-Yon	27 – Toulon
13 – Tours	28 – Nice
14 – Dijon	29 – Gap
15 – Chalon-sur-Saône	

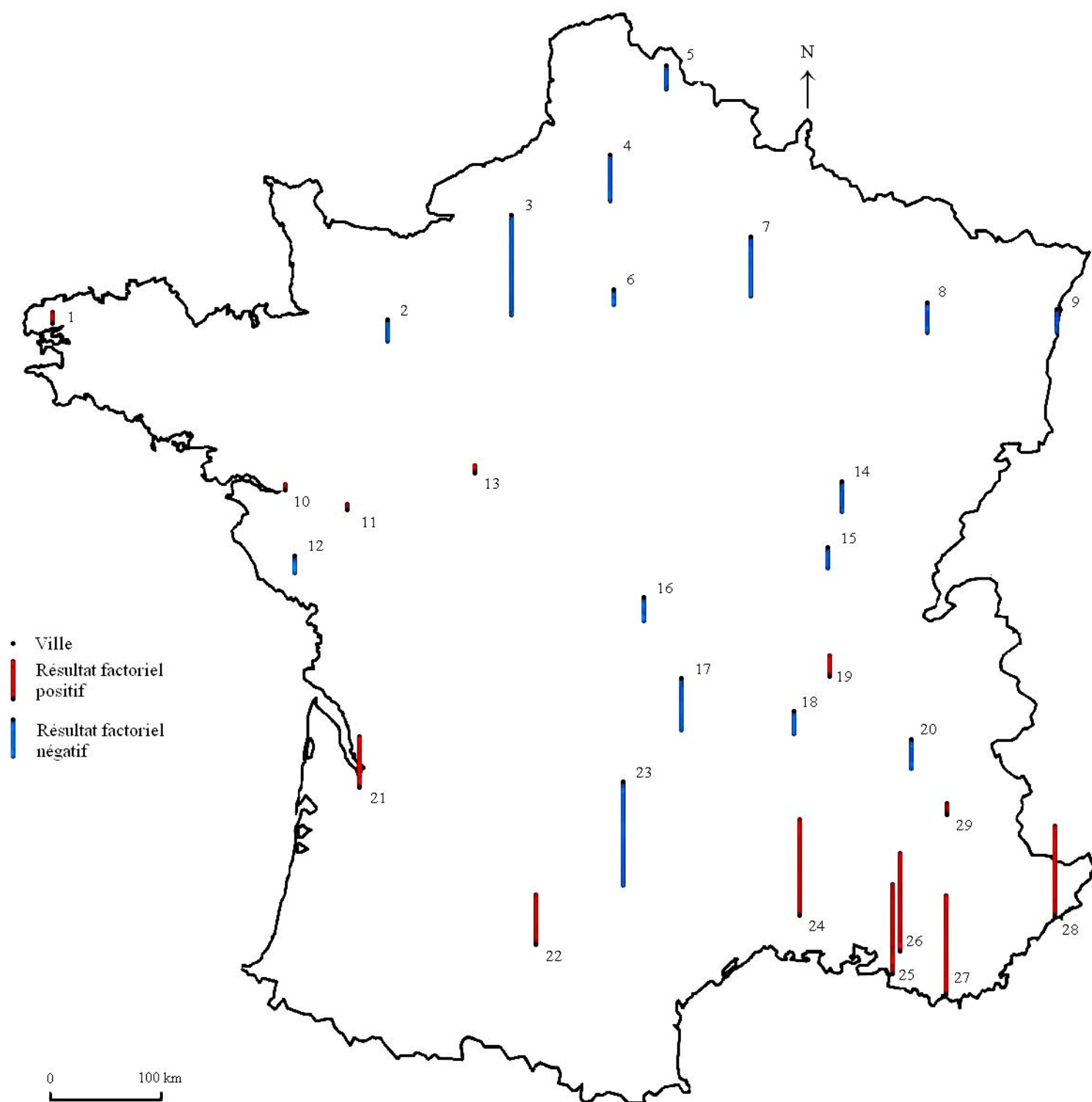


Figure 27 : Résultats factoriels de la première composante principale où 1 cm = 1 unité (source du fond de carte : www.hist-geo.com).

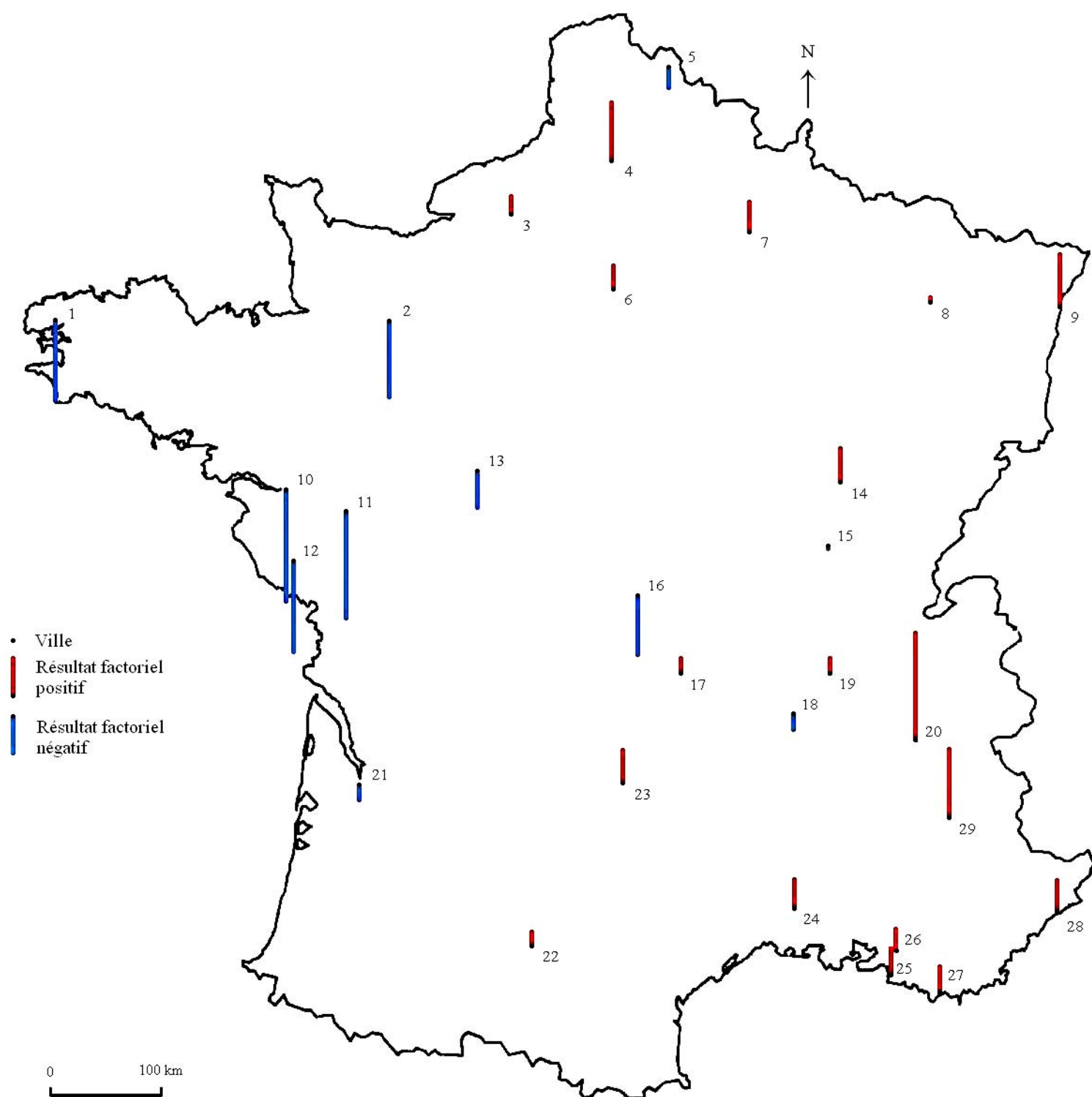


Figure 28 : Résultats factoriels de la deuxième composante principale où 1 cm = 1 unité (source du fond de carte : www.hist-geo.com).

La première composante a mis en avant une opposition entre des stations où les pollens de frêne, de bouleau et de Poacées sont bien représentés et d'autres stations où ces taxons ne le sont pas. Cela se traduit géographiquement par une opposition entre les sites méditerranéens et les autres (figure 27). Cette composante met en avant cinq villes où les trois taxons sont faiblement représentés, cas de figure que l'on retrouve dans le Sud-Ouest (Bordeaux et Toulouse) mais dans une moindre mesure. Cette opposition souligne le rôle de la latitude sur la part de ces trois taxons sur laquelle se superpose une influence régionale liée à un climat bien spécifique, en l'occurrence le climat méditerranéen.

La deuxième composante semble mettre en avant une opposition ouest-est où ressortent les résultats factoriels négatifs des stations de l'Ouest de la France. Ces stations se caractérisent par de fortes proportions de Poacées en termes de valeur, mais qui sont par la même occasion nettement supérieures à celles enregistrées pour le frêne et le bouleau (figure 28). Cette prédominance des Poacées est globalement moins sensible sur la partie Est de la France expliquant les résultats factoriels positifs. C'est très nettement le cas à Strasbourg, Gap et Grenoble où les proportions de pollens de frêne, ou de bouleau, sont proches de celles des Poacées. Cette composante met en exergue le caractère très spécifique des stations «atlantiques», et de Montluçon, où les Poacées sont très prédominantes par rapport aux deux autres taxons.

1.2.3 Conclusion

Nous avons présenté, dans le chapitre 1.2, les caractéristiques des taxons polliniques des genres et familles retenues pour notre étude. Malgré le fait que l'index pollinique ne traduise pas de manière précise le contenu pollinique annuel de l'air d'une année à l'autre, il permet néanmoins de faire ressortir l'implantation géographique des plantes concernées, et de confirmer globalement la validité des séries polliniques d'un point de vue aérobiologique. Ainsi, nous avons pu voir, entre autres, l'abondance des pollens de bouleau et de Poacées dans la moitié Nord de la France – ces derniers étant particulièrement représentés dans le Centre et l'Ouest du territoire – caractéristique qui les rend d'autant plus néfastes pour les personnes allergiques.

CONCLUSION DE LA PARTIE

Nous avons présenté dans cette première partie les procédés de capture et d'analyse des pollens, le réseau de capteurs et différents points liés à la qualité des données polliniques. Ces dernières sont loin d'être parfaites et présentent des faiblesses sur divers points, notamment par rapport à l'homogénéisation des hauteurs des capteurs dont l'influence, très probable, sur les concentrations n'est pas quantifiable. La non-disponibilité de certaines données peut, quant à elle, être handicapante lorsque les lacunes se situent au début de la pollinisation d'une espèce, d'un genre ou d'une famille donnée. Mais au vu de la complexité de la gestion de ces capteurs, tous ces cas de figure s'avèrent être un moindre mal. L'analyse statistique de ces données est, au bout du compte, réalisable car, au-delà du fait que ces concentrations polliniques journalières sont indispensables à notre étude, l'analyse des concentrations des quatre taxons polliniques a permis de mettre en avant des particularités régionales. La part occupée par chacun des quatre taxons par rapport au cumul annuel total tous taxons confondus, sujet du sous-chapitre 1.2.2, permet, au bout du compte, de confirmer la capacité des capteurs à retranscrire l'implantation des genres ou familles correspondantes. Cela renforce donc la pertinence de l'utilisation de ces concentrations journalières en pollen pour les familles ou genres anémophiles.

La présentation générale des données polliniques étant accomplie, nous allons passer à la deuxième partie de ce travail qui consiste en la description des différents critères de détermination du début de la saison pollinique et des méthodes de prévision de la date d'apparition d'un risque d'exposition allergique aux pollens.

PARTIE 2 : CRITERES ET METHODES LIES AU **DEBUT DE LA SAISON POLLINIQUE**

2.1 PRESENTATION ET ANALYSE DES DIFFERENTS CRITERES DE DETERMINATION DU DEBUT DE LA SAISON POLLINIQUE

La détermination de la date de début de pollinisation ou, plus exactement, de l'apparition d'un risque allergique d'exposition aux pollens, est primordiale dans le cadre d'un dispositif préventif. En effet, afin d'initier un traitement anti-allergique qui soit efficace, il est nécessaire de prévoir le plus exactement possible la date d'apparition de ce risque. Il est donc pertinent de s'attarder sur le sujet avant d'évoquer les méthodes de prévision à proprement parler et les résultats qui en découlent.

De nombreux critères de détermination sont recensés dans les articles traitant de la prévision des pollens. Ils se basent soit sur une valeur brute, soit sur une concentration correspondant à un certain pourcentage du total saisonnier d'un taxon donné. Nous allons d'abord présenter les critères s'appuyant sur des seuils bruts, puis ceux s'appuyant sur des seuils relatifs.

2.1.1 Revue des différents critères s'appuyant sur des valeurs brutes

- Le principal seuil brut rencontré dans la littérature est de 30 grains/m³ par jour. Le début de la pollinisation, ou plutôt de l'apparition du risque lié aux pollens, a lieu quand cette concentration est atteinte. La genèse de ce seuil se base sur des études cliniques portant sur un certain nombre de patients. Plusieurs d'entre elles (Davies et Smith, 1973 ; Viander et Koivikko, 1978 ; Corsico, 1993 ; Antépara *et al.*, 1995) concluant, semble-t-il, au même seuil ont servi de base d'étude à des articles postérieurs traitant du bouleau (Clot, 2001 ; Cotos-Yanez *et al.*, 2004), du frêne (Peeters, 2000) et des Poacées (Sanchez-Mesa *et al.*, 2003 ; Smith et Emberlin, 2005). Un seuil de 50 grains est utilisé pour une étude australienne sur les Poacées (Ong *et al.*, 1997), se basant sur un seuil qui a été établi par Davies et Smith (1973). Cela montre que les auteurs ont réutilisé ce seuil pour différents taxons. La valeur utilisée semble, en effet, relativement médiane, et pourrait constituer un bon critère de détermination pour des espèces dont la pollinisation est relativement abondante.

- Le deuxième seuil brut principalement rencontré dans la littérature consiste en une valeur cumulée. Le début de la pollinisation, ou l'apparition d'un niveau de risque allergique, correspond au jour où l'on atteint, au moins, 75 grains cumulés, depuis l'apparition du premier grain. Ce critère permet, en partie, de ne pas tenir compte des concentrations extrêmement faibles rencontrées à l'amorce de la saison pollinique d'un taxon donné. Avec un tel critère de détermination, il suffit d'une concentration journalière d'un grain pour permettre d'atteindre le seuil requis, en passant de 74 à 75 grains cumulés. L'article semblant être à l'origine de ce critère traite des Poacées (Driessen *et al.*, 1990) et a été repris par la suite dans une thèse (Jones, 1995). Il a été surtout utilisé, ultérieurement, dans des articles traitant de la pollinisation du bouleau (Emberlin *et al.*, 1997 ; Corden *et al.*, 2000 ; Adams-Groom *et al.*, 2002 ; Corden *et al.*, 2002 ; Radisic et Sikoparija, 2005).
- Le dernier critère utilisant des seuils bruts se base sur des concentrations journalières très faibles, mais qui ne doivent pas être isolées. Ainsi, le début de la pollinisation d'une espèce donnée correspond au premier jour atteignant un grain, à condition qu'il soit suivi d'au moins cinq jours consécutifs présentant, au moins, la même concentration. Il apparaît d'ailleurs que la deuxième condition n'est pas toujours aussi clairement présentée dans certains articles utilisant ce critère. Celui-ci semble spécialement adapté à l'étude du pollen d'olivier qui pose un problème sanitaire majeur dans les pays méditerranéens, et plus particulièrement l'Espagne. La plupart des articles, utilisant ce critère, traitent de ce pollen (Galán *et al.*, 2001a, 2001b, 2005 ; Vazquez *et al.*, 2003). Il est également choisi parfois dans l'étude d'autres taxons tels que le bouleau (Radisic et Sikoparija, 2005), les Poacées (Sanchez-Mesa *et al.*, 2005) et le chêne (García-Mozo *et al.*, 1999, 2000, 2002). Ce critère, quasi exclusivement retenu par les auteurs espagnols, semble difficile à utiliser pour des plantes dont la pollinisation est très progressive, durant la période précédant le pic maximal. En effet, dans le cas des Poacées par exemple, de très faibles concentrations peuvent apparaître dès le mois de mars dans la moitié Sud du territoire, sans pour autant avoir d'incidence sur les allergiques, et sans correspondre à la pleine pollinisation de ce taxon qui a lieu en mai-juin.

2.1.2 Revue des différents critères s'appuyant sur des valeurs relatives

La grande majorité des articles traitant de la prévision des pollens utilisent des critères de détermination s'appuyant sur des pourcentages :

- Le critère le plus souvent rencontré est celui de Nilsson et Persson (1981). Il considère que la date de début de pollinisation est le jour où la production cumulée, depuis l'apparition du premier grain, atteint 5% de la production totale de la saison. Comme pour le critère des 75 grains cumulés, il permet de ne pas tenir compte des faibles concentrations enregistrées en début de saison, qui ne sont probablement pas en accord avec la phénologie locale des espèces étudiées. Ce seuil peut être atteint pour un jour où les concentrations sont extrêmement faibles par rapport au total saisonnier. Ainsi, si l'on enregistre 5 000 grains d'un taxon donné, on peut passer de 4,98% à 5% de ce total avec une concentration journalière d'un seul grain, correspondant à la différence de 0,02%. Ce critère est utilisé pour étudier la pollinisation de plusieurs espèces simultanément (Larsson, 1993 ; Gottardini et Cristofolini, 1997 ; Valencia-Barrera *et al.*, 2001 ; Clot, 2003), mais aussi pour des taxons aussi variés que le bouleau (Dahl et Strandhede, 1996 ; Latalowa *et al.*, 2002), les Urticacées (Emberlin et Norris-Hill, 1991), l'armoise (Munuera-Giner *et al.*, 1999), le pin (Jato *et al.*, 2000b ; Frenguelli *et al.*, 2002), l'ambroisie (Makra *et al.*, 2004), les Cupressacées-Taxacées (Galán *et al.*, 1998) et le chêne (Jato *et al.*, 2002).
- Le deuxième critère le plus usité, après celui de Nilsson et Persson, est celui d'Andersen (1991). Il se calcule de manière identique, mais le seuil retenu est plus bas (2,5%). Comme pour les critères présentés précédemment, ce seuil est utilisé pour différents taxons tels les Poacées (Norris-Hill, 1997), le bouleau (Corden *et al.*, 2000, 2002 ; Rodríguez-Rajo *et al.*, 2003), l'aulne (Jato *et al.*, 2000a ; González-Parrado *et al.*, 2006), le chêne (Jato *et al.*, 2002) et les Cupressacées-Taxacées (Fuertes-Rodríguez *et al.*, 2007).
- Quelques articles vont présenter des seuils relatifs journaliers où la date de début de pollinisation correspond au premier jour où l'on atteint 1% du total annuel pour un taxon donné (Emberlin *et al.*, 2002) ou 2,5% (Emberlin *et al.*, 1997, 2002 ; Radisic

et Sikoparija, 2005). Cela peut correspondre à des concentrations journalières non négligeables pour un taxon donné, dans le cas d'une pollinisation très abondante. Par exemple, si l'on relève 10 000 grains/m³ pour un taxon donné, cela représente une concentration de 250 grains avec le seuil de 2,5%, bien au-delà du seuil de 30 grains.

- Le quatrième critère, développé par Lejoly-Gabriel (1978), s'avère être une combinaison des deux types précédemment exposés. Cette technique consiste à prendre pour date de début de pollinisation d'un taxon donné le premier jour où sont remplies les deux conditions suivantes :
 - La production pollinique du jour atteint, au moins, 1% de la production totale de la saison.
 - La production cumulée jusqu'à ce jour est égale à 5% de la production totale de la saison.

Cette méthode est plébiscitée dans les travaux français traitant de plusieurs taxons (Laaidi M, 1999 ; Laaidi et Laaidi, 2001), plus particulièrement du bouleau (Sutra *et al.*, 1992 ; Laaidi M, 2001a) et des Poacées (Laaidi M, 2001b). Ce critère a pour double avantage de distinguer un jour qui soit, à la fois, bien corrélé à la phénologie locale en écartant les cinq premiers %, et où l'on enregistre une concentration non négligeable par rapport au cumul total. Par rapport au critère de Nilsson et Persson, cela permet d'éviter de faire démarrer la saison par un jour où la concentration du taxon étudié est infime, et qui ne présente, par conséquent, aucun risque allergique d'exposition.

- Dans ce dernier point, nous regroupons quatre autres critères spécifiques, mais peu employés, qui sont les suivants :
 - Le premier considère que la saison débute quand la concentration journalière moyenne pentadaire d'un taxon donné atteint 1% (Mullenders *et al.*, 1972). On peut légitimement supposer que le début de la pollinisation correspond au premier jour de la pentade en question, à condition que l'on enregistre au

moins un grain de pollen. Ce critère est repris pour l'aulne (Jato *et al.*, 2000a) et le bouleau (Jato *et al.*, 1999).

- Le deuxième critère se base sur le seuil journalier de 1%, mais le jour atteignant cette valeur doit être suivi d'un maximum six jours avec des concentrations nulles. Ce critère, utilisé pour des études à l'échelle européenne (Jaeger *et al.*, 1996 ; Emberlin *et al.*, 2000), n'apporte, à priori, rien de plus que l'utilisation seule du critère de 1% car, avec un tel pourcentage, on enregistre des concentrations de l'ordre de 20 à 50 grains dans la plupart des cas. Il est donc rarissime que de telles concentrations journalières soient suivies de plusieurs jours enregistrant une absence de pollens.
- Le troisième critère considère que la date de début de pollinisation correspond au premier jour où la production cumulée, depuis l'apparition du premier grain, atteint 1% de la moyenne logarithmique du total annuel d'un taxon donné (Spieksma *et al.*, 1995). Ce critère, quelque peu complexe, est seulement évoqué dans un article danois traitant du bouleau (Rasmussen, 2002).
- Le quatrième critère consiste en une détermination graphique de la date de début de pollinisation (Pathirane, 1975). Il s'agit de représenter la fréquence cumulée journalière d'un taxon donné, qui va déboucher sur une courbe d'allure sinusoïdale. La tangente à la première inflexion de la courbe correspond à la date de début de pollinisation. Notamment reprise pour la prévision de l'ambrosie à Lyon (Laaidi *et al.*, 2003a, 2004), elle correspond, dans ce cas précis, à une fréquence cumulée de 2%. Cette méthode originale présente, comme principal inconvénient, le risque que le moment de l'inflexion soit lié à une fréquence différente d'une année sur l'autre. Ce qui ne permet aucune comparaison entre deux séries polliniques.

2.1.3 Avantages et inconvénients des deux types de critères

Il est difficile d'affirmer que tel critère est plus pertinent qu'un autre, et que l'utilisation d'un seuil relatif s'avère plus judicieuse que celle d'un seuil brut. Cela dépend, pour une part importante, du but de l'étude. Nous allons présenter les points forts et les points faibles des deux types de critères en nous positionnant par rapport aux critères de détermination du début de pollinisation employant des seuils relatifs.

Le principal avantage dans l'utilisation d'un seuil relatif, c'est la possibilité d'effectuer une comparaison inter-sites. Les concentrations journalières enregistrées dépendant de nombreux paramètres, il est malaisé de déterminer le début de la pollinisation en se basant sur les valeurs brutes. En supposant que la cinétique de pollinisation d'une espèce est équivalente, on peut établir des comparaisons d'une ville à l'autre, en limitant l'assujettissement aux concentrations. Il semblerait que le critère relatif se rapproche plus de la phénologie de la plante, ce qui le rend plus judicieux dans le cadre d'une étude agronomique. A cette meilleure adéquation à la phénologie locale, on peut émettre une réserve quant à la prise en compte, par ce type de critère, des concentrations de tout début et de toute fin de saison. Ces petites concentrations, décorréliées de la période de pollinisation de la plante étudiée, peuvent influencer l'élaboration des pourcentages et introduire un biais d'interprétation, surtout si l'on n'utilise que des critères de fréquence cumulée. Le critère de Lejoly-Gabriel (1978), prenant en compte à la fois une fréquence cumulée et une fréquence journalière, permet, en partie, d'éviter ce problème et peut être considéré comme le plus pertinent des critères employant un seuil relatif.

Outre la prise en compte des très faibles concentrations enregistrées hors de la période «phénologique» de pollinisation, les seuils relatifs présentent d'autres inconvénients, notamment dans le cadre d'une étude aérobiologique à visée sanitaire et préventive :

- Le principal inconvénient, d'ordre général, est en lien direct avec les données lacunaires rencontrées dans les séries polliniques. Leur présence et la reconstitution nécessairement sommaire, du fait du mode de lecture, qui en découle rend impossible la détermination du cumul total exact, pour un taxon donné. En effet, si les lacunes interviennent pendant la période de présence dans l'air ambiant des pollens d'une espèce ou d'un genre donnés et qu'aucune reconstitution n'est effectuée, le cumul total est biaisé, et la fréquence cumulée, qui en découle,

également, surtout si les lacunes interviennent pendant la pleine saison pollinique de l'espèce ou du genre en question. L'emploi de seuils bruts est moins problématique de ce point de vue, sauf si les lacunes sont présentes en début de saison, et ont de fortes chances d'englober la date de démarrage.

- Le deuxième inconvénient des seuils relatifs par rapport aux seuils bruts est d'ordre pratique. En effet, avec un seuil relatif, il faut attendre la fin de la saison pollinique, et l'apparition des derniers grains d'un taxon donné, pour déterminer la date de début de pollinisation. Cet aspect n'est pas commode dans le cadre d'une prévision de l'apparition d'un risque d'exposition allergique, mais n'entrave en aucune façon l'élaboration d'un modèle de prévision se basant sur ces seuils.
- Le troisième inconvénient des seuils relatifs joue un rôle majeur, si on se place dans le cadre de la détermination de niveaux de risque d'exposition allergique aux pollens. En effet, la concentration en pollen, en lien avec un seuil relatif, fluctue d'une année à l'autre. Or la population allergique à un pollen est sensible, à priori, aux mêmes concentrations année après année. Pour un taxon allergisant donné (comme le bouleau), l'impact d'un certain niveau de concentration journalière reste le même, que la saison pollinique soit abondante ou non. De plus, dans le cadre d'une forte pollinisation saisonnière, les concentrations liées aux seuils relatifs risquent, pour la plupart des taxons allergisants, d'être supérieures aux concentrations susceptibles de déclencher des symptômes allergiques. Dans ce cas, la prévision de l'apparition du risque serait trop tardive et, par conséquent, totalement inefficace dans un but préventif.

Le choix d'un type de critère par rapport à l'autre dépend principalement du but de l'étude. Dans le cadre d'une élaboration de modèles de prévision à destination des personnes allergiques au pollen, l'utilisation de seuils bruts semble, de loin, la plus adéquate. Dans ce cas-ci, la date fixée par de tels seuils doit être définie comme celle où apparaît un certain niveau de risque d'exposition allergique aux pollens. Utiliser le terme de date de début de pollinisation dans ce cas serait impropre, car les concentrations journalières enregistrées sont influencées par divers paramètres, et un même seuil ne correspond pas tout à fait, d'une année sur l'autre, au même stade phénologique de la plante.

2.1.4 Application de différents seuils pour le bouleau, le frêne et les Poacées : exemples d'Amiens et de Montluçon

Nous allons, dans ce sous-chapitre, comparer les dates d'apparition du risque allergique lié aux pollens, déterminées à partir de la plupart des seuils de détermination du début de la saison pollinique présentés ci-dessus (cf. 2.1.1 et 2.1.2). Nous avons retenu deux sites, Montluçon et Amiens, car ils présentent les séries polliniques les plus complètes, avec des données lacunaires seulement isolées et éparses. L'étude porte sur les trois taxons allergisants, retenus pour ce travail, présents en plus ou moins grande abondance dans ces deux sites. Les huit seuils les plus marquants sont retenus ; à chacun d'eux est associé une abréviation :

- Fréquence cumulée de 5% (**5%**)
- Fréquence cumulée de 2,5% (**2,5%**)
- Somme cumulée de 75 grains (**sum75**)
- Fréquence journalière de 1% (**1%/j**)
- Fréquence journalière de 2,5% (**2,5%/j**)
- Concentration journalière de 30 grains (**30 g/j**)
- Fréquence journalière moyenne de 1% sur une pentade (**1%/pen**)
- Fréquence cumulée de 5% et fréquence journalière de 1% (**5%_1%/j**)

Le but est ici de déterminer s'il existe des différences significatives entre les dates moyennes de démarrage selon le critère employé. La normalité de la distribution des séries est déterminée à l'aide des tests de Shapiro-Wilk et de Kolmogorov-Smirnov, pour un seuil de significativité de 95% selon la modalité définie dans la section 1.1.5.3. Cette dernière, rappelons-le, consiste en une acceptation de l'hypothèse de normalité de la distribution d'une série par, au moins, un des deux tests. Nous tiendrons compte de cette modalité pour l'ensemble des sections de ce sous-chapitre.

2.1.4.1 Le frêne à Amiens (1989-2005)

Les données polliniques sont disponibles depuis 1987, mais les quantités de pollen de frêne enregistrées en 1987 et 1988 ont été très faibles. Le seuil de 30 grains journalier n'a jamais été

atteint la première année, et seulement une fois en 1988 ; ces années ont donc été retirées de l'analyse.

Dans le cas présent trois des huit séries ne présentent pas une distribution normale, selon les résultats des tests de Shapiro-Wilk et de Kolmogorov-Smirnov : **5%**, **1%/j** et **5%_1%/j**. La raison de la non-normalité de ces séries, comparées aux cinq autres, est difficilement explicable car on retrouve aussi bien des seuils relatifs que bruts, ainsi que des fréquences journalières et cumulées dans les deux «camps». Le faible nombre d'observations (17) en est peut-être la cause.

A travers ce constat, nous utilisons un test t de Student sauf si le s'on s'intéresse à une des trois séries évoquées ci-dessus. Dans ce cas-là le test U de Mann-Whitney est employé.

Avant d'appliquer ce test nous allons présenter graphiquement les dates moyennes d'apparition du risque allergique, ou de la date de début de pollinisation selon le type de critère employé (figure 29).

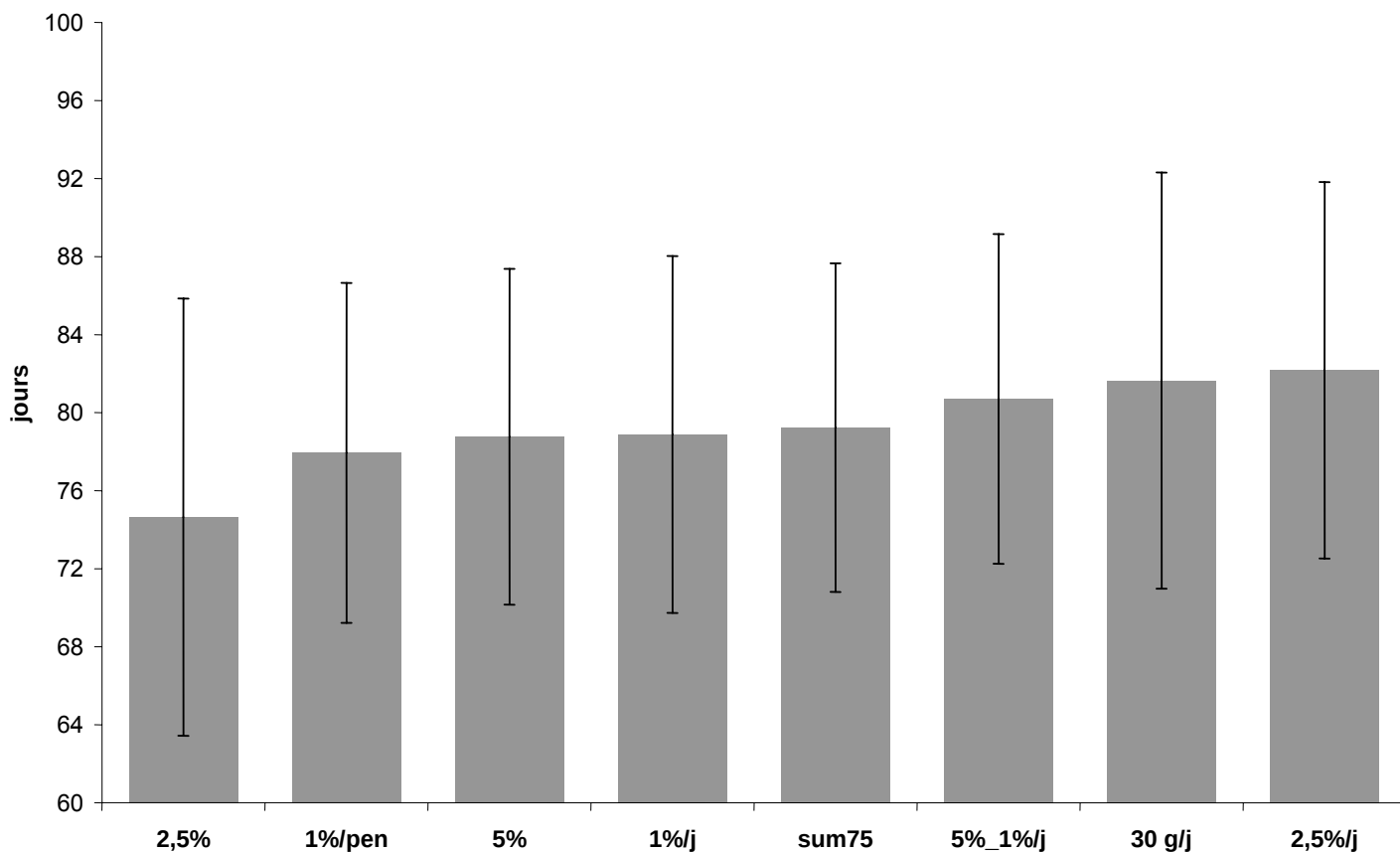


Figure 29 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au frêne à Amiens (1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.

Les dates moyennes s'étagent du 16 mars (en arrondissant), pour une année non bissextile, avec le critère de 2,5% du total cumulé, au 23 mars avec celui de 2,5% journalier. Les écarts-types sont, quant à eux, de l'ordre de 8,5 à 11 jours.

Le test de Student, ou celui de Mann-Whitney dans le cadre des comparaisons faisant intervenir les séries **5%**, **1%/j** ou **5%_1%/j**, est appliqué sur des couples de seuils présentant la même définition «temporelle», c'est-à-dire définis à partir d'une valeur journalière ou d'une valeur cumulée. Il en sera de même pour les sections 2.1.4.2 à 2.1.4.6. La valeur p des tests est inscrite entre parenthèses et l'on considère qu'il y a une différence significative pour p inférieur à 0,05 :

- **2,5% et 5%** ($p = 0,1222$)
- **5% et 5%_1%/j** ($p = 0,3057$)
- **1%/pen et 1%/j** ($p = 0,6832$)

- **1%/j et 2,5%/j** ($p = 0,1401$)
- **30 g/j et 1%/j** ($p = 0,3940$)
- **30 g/j et 2,5%/j** ($p = 0,8803$)
- **sum75 et 2,5%** ($p = 0,1867$)
- **sum75 et 5%** ($p = 0,8384$)

Aucune différence significative n'est enregistrée à cause, en partie, de la variabilité interannuelle des dates de démarrage qui diffère d'un critère à l'autre, symbolisée en cela par des écarts-types assez dissemblables. Il convient cependant de noter que les dates moyennes des critères **2,5%** et **5%** sont distantes de 4,2 jours et que celles des critères **1%/j** et **2,5%/j** de 3,3 jours, écarts relativement faibles mais bien réels. Enfin précisons que la seule différence significative est enregistrée entre les dates moyennes issues des critères **2,5%** et **2,5%/j**.

2.1.4.2 Le frêne à Montluçon (1990-2002 et 2004)

Les années 1989 et 2005, disponibles, sont retirées car les faibles concentrations enregistrées de frêne ne permettent pas l'utilisation de tous les critères présentés. La mise en route tardive du capteur en 2003 a tronqué le début de la saison pollinique de ce taxon ; aucune détermination de la date d'apparition d'un risque d'exposition allergique n'est alors possible.

L'application des tests de Shapiro-Wilk et Kolmogorov-Smirnov aux huit séries a permis de confirmer l'hypothèse de normalité pour chacune d'entre elles, au contraire d'Amiens. Nous allons donc appliquer le test de Student dont les résultats sont précédés de la présentation (figure 30) des dates moyennes d'apparition du risque allergique ou du début de pollinisation, selon que l'on emploie un seuil brut ou relatif.

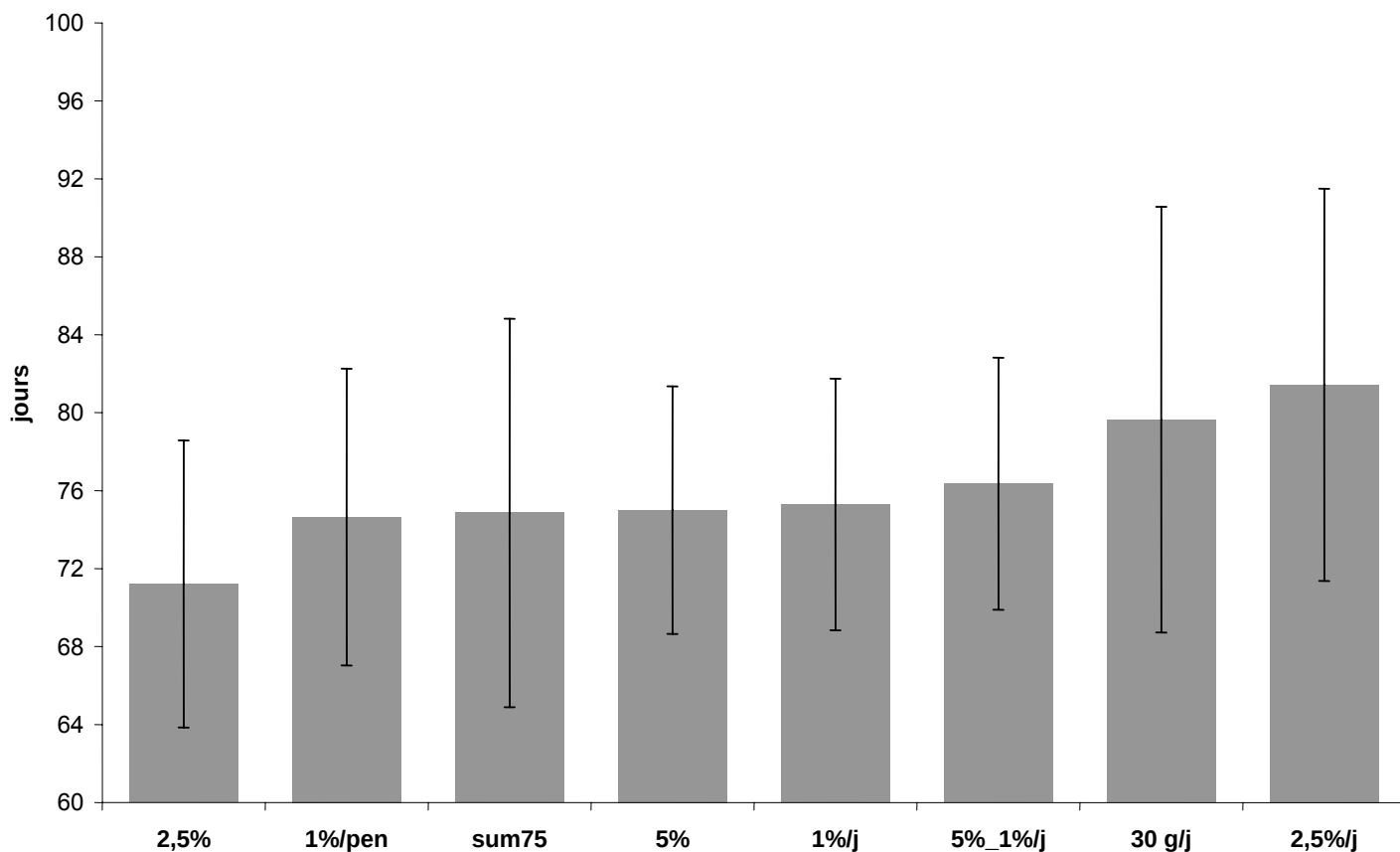


Figure 30 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au frêne à Montluçon (1990-2002 et 2004), critère par critère, en nombre de jours à partir du 1^{er} janvier.

Les dates moyennes s'échelonnent du 12 mars (en arrondissant), pour une année non bissextile, avec le critère de 2,5% du total cumulé, au 22 mars avec celui de 2,5% journalier. Les écarts-types s'étendent de 6,4 jours à plus de 10 jours pour **30 g/j** et **2,5%/j**. Il est intéressant de constater que le classement des critères, selon les dates moyennes qui en découlent, est très semblable entre Amiens et Montluçon.

Le test de Student est appliqué sur les duos de critères présentés pour Amiens, la valeur p est inscrite entre parenthèses :

- **2,5% et 5%** ($p = 0,1570$)
- **5% et 5%_1%/j** ($p = 0,5800$)
- **1%/pen et 1%/j** ($p = 0,8114$)
- **1%/j et 2,5%/j** ($p = 0,0654$)

- **30 g/j** et **1%/j** ($p = 0,2101$)
- **30 g/j** et **2,5%/j** ($p = 0,6565$)
- **sum75** et **2,5%** ($p = 0,2815$)
- **sum75** et **5%** ($p = 0,9643$)

Le test de Student n'est significatif pour aucune des combinaisons testées, même s'il est proche du seuil dans le cas de la comparaison entre les dates moyennes issues de **1%/j** et **2,5%/j**, séparées de 6,1 jours. Les valeurs des écarts-types variables d'un critère à l'autre semblent expliquer pour une bonne part l'absence de significativité enregistrée.

L'écart extrêmement faible enregistré entre les dates moyennes de **sum75** et **5%**, 0,1 jour d'écart, mérite que l'on s'y attarde. Le total annuel moyen de pollens de frêne, sur la période étudiée, est de 1 761 grains avec un coefficient de variation de 56,90%. Comme 75 grains cumulés équivalent à une fréquence cumulée de 5%, pour un total de 1 500 grains, cela explique la quasi-égalité des deux dates moyennes.

2.1.4.3 Le bouleau à Amiens (1987 et 1989-2005)

L'année 1988 est retirée de l'analyse à cause de concentrations en pollen de bouleau extrêmement faibles, où le seuil de 30 grains journalier n'est jamais atteint. L'emploi des deux tests de normalité a permis de déterminer le caractère gaussien de la distribution des huit séries.

Nous allons appliquer le test t de Student et, tout d'abord présenter les dates moyennes d'apparition du risque allergique, liées à chaque critère (figure 31).

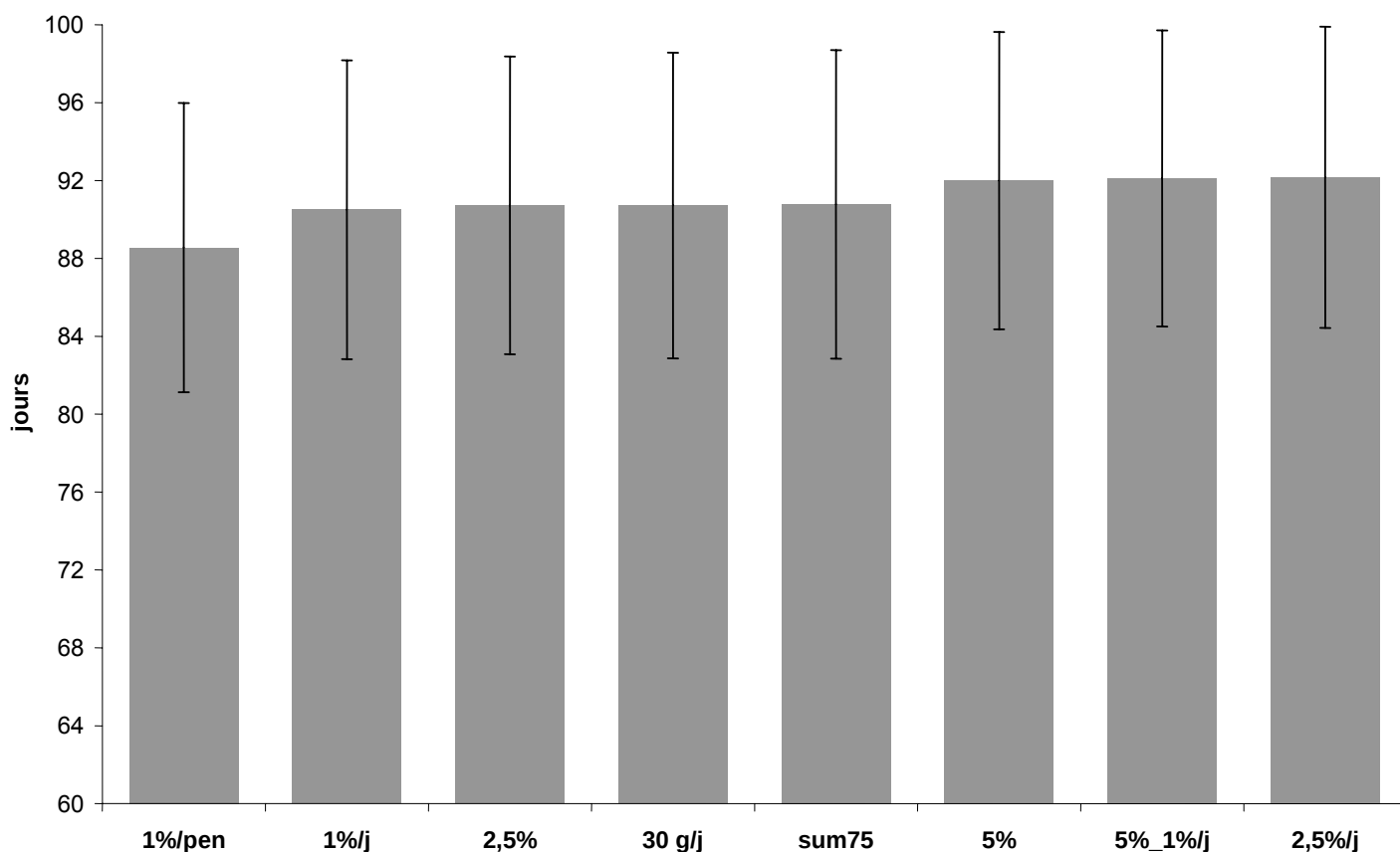


Figure 31 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au bouleau à Amiens (1987 et 1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.

Graphiquement, on se rend compte que la différence est minimale entre toutes les dates. En effet elles s'étendent du 30 mars, pour une année non bissextile, avec le critère **1%/pen** au 2 avril avec **5%/j**, **5%_1%/j** et **2,5%/j**. Les écart-types sont similaires et compris entre 7,6 et 8 jours.

Le test de Student est appliqué sur les couples de critères suivants, la valeur p est inscrite entre parenthèses :

- **2,5%** et **5%** ($p = 0,6193$)
- **5%** et **5%_1%/j** ($p = 0,9654$)
- **1%/pen** et **1%/j** ($p = 0,4452$)
- **1%/j** et **2,5%/j** ($p = 0,9311$)

- **30 g/j** et **1%/j** ($p = 0,9321$)

- **30 g/j** et **2,5%/j** ($p = 0,5817$)
- **sum75** et **2,5%** ($p = 0,9830$)
- **sum75** et **5%** ($p = 0,6405$)

Le test de Student n'enregistre aucune différence significative. Cela est tout fait logique car les dates comparées sont très similaires avec, au maximum, 1,9 jour d'écart entre les dates moyennes issues de **1%/pen** et **1%/j**. Le démarrage de la saison du bouleau s'avère ainsi plus brutal que celui du frêne, avec des concentrations importantes atteintes rapidement. De ce fait le choix du critère de détermination importe assez peu.

2.1.4.4 Le bouleau à Montluçon (1989-2005)

Selon les tests de Shapiro-Wilk et Kolmogorov-Smirnov, les huit séries présentent une distribution qui suit la loi normale pour, rappelons-le, un seuil de significativité de 95%. Nous allons appliquer le test de Student et, tout d'abord présenter les dates moyennes d'apparition du risque allergique, liées à chaque critère (figure 32).

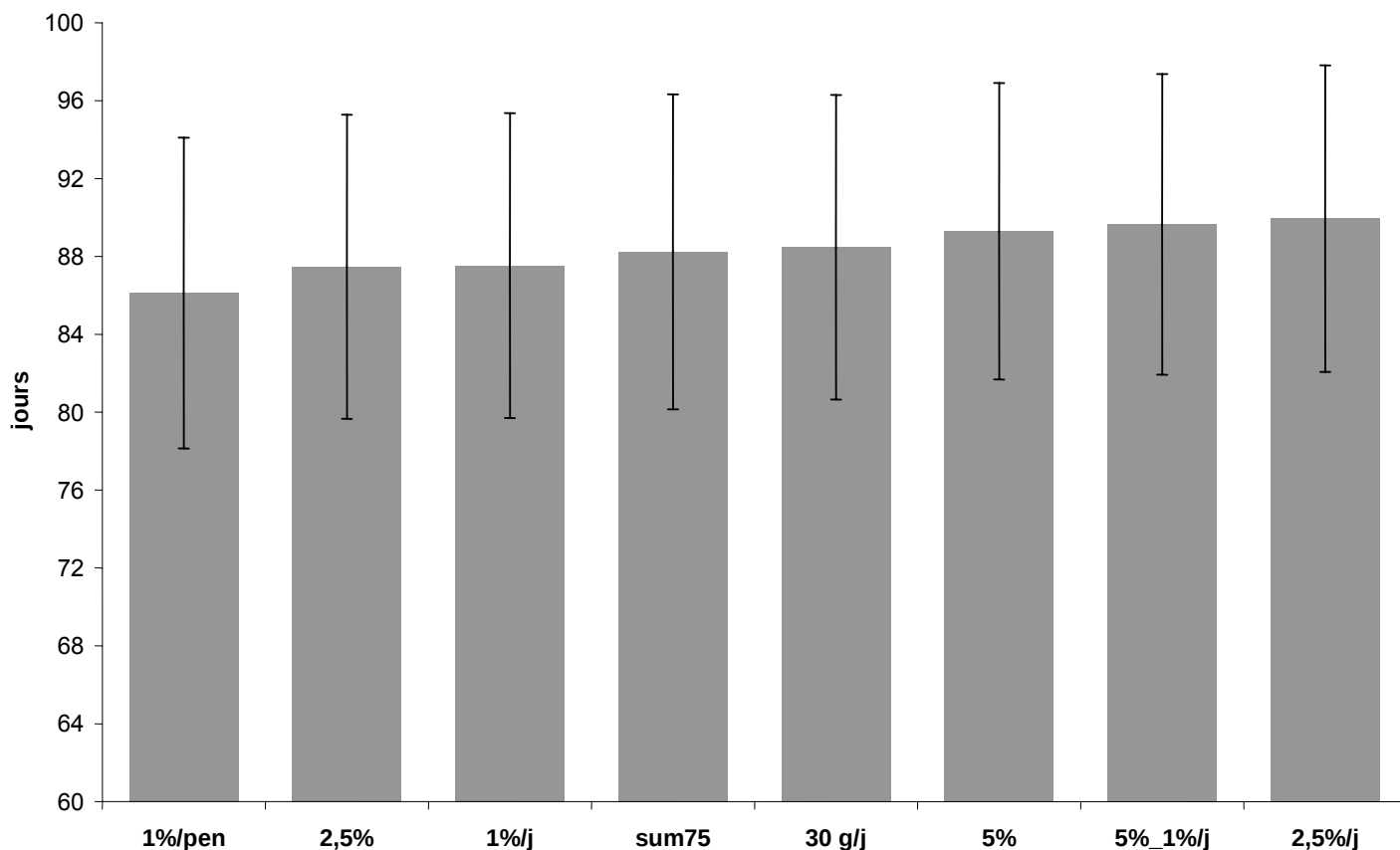


Figure 32 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au bouleau à Montluçon (1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.

Comme à Amiens, on constate que la différence est minime entre toutes les dates. En effet elles s'étendent du 27 mars, pour une année non bissextile, avec le critère **1%/pen** au 31 mars avec **5%_1%/j** et **2,5%/j** ; toutes ces valeurs étant arrondies. Les écart-types sont similaires et compris entre 7,6 et 8,1 jours.

Suite à ce constat nous appliquons le test de Student, la valeur p est inscrite entre parenthèses :

- **2,5%** et **5%** ($p = 0,4954$)
- **5%** et **5%_1%/j** ($p = 0,8941$)
- **1%/pen** et **1%/j** ($p = 0,6062$)
- **1%/j** et **2,5%/j** ($p = 0,3770$)
- **30 g/j** et **1%/j** ($p = 0,7281$)

- **30 g/j** et **2,5%/j** ($p = 0,5885$)
- **sum75** et **2,5%** ($p = 0,7809$)
- **sum75** et **5%** ($p = 0,6969$)

Comme à Amiens, aucune différence significative n'est enregistrée entre les différents critères comparés. Le plus grand écart (2,4 jours) est constaté entre les dates moyennes issues des critères **1%/j** et **2,5%/j**. Tous les critères de détermination donnent ainsi des dates de démarrage similaires et la cinétique de pollinisation du bouleau en est, à nouveau, la cause majeure.

2.1.4.5 Les Poacées à Amiens (1987-2005)

D'après les tests de Shapiro-Wilk et Kolmogorov-Smirnov, l'hypothèse de normalité doit être rejetée pour les séries de dates définies par les critères **2,5%** et **sum75**. La raison de la non-normalité de la distribution de ces séries n'est pas clairement explicable, même si, comme ce sera montré ultérieurement, ce sont les deux critères débouchant sur les dates moyennes les plus précoces. Le test de Student est ainsi appliqué sur les diverses combinaisons à l'exception de celles employant, au moins, l'un des deux critères non gaussien. Il a été jugé préférable, dans ce dernier cas de figure, d'utiliser le test de Mann-Whitney.

Avant d'appliquer ce test nous allons présenter graphiquement les dates moyennes d'apparition du risque allergique, critère par critère (figure 33).

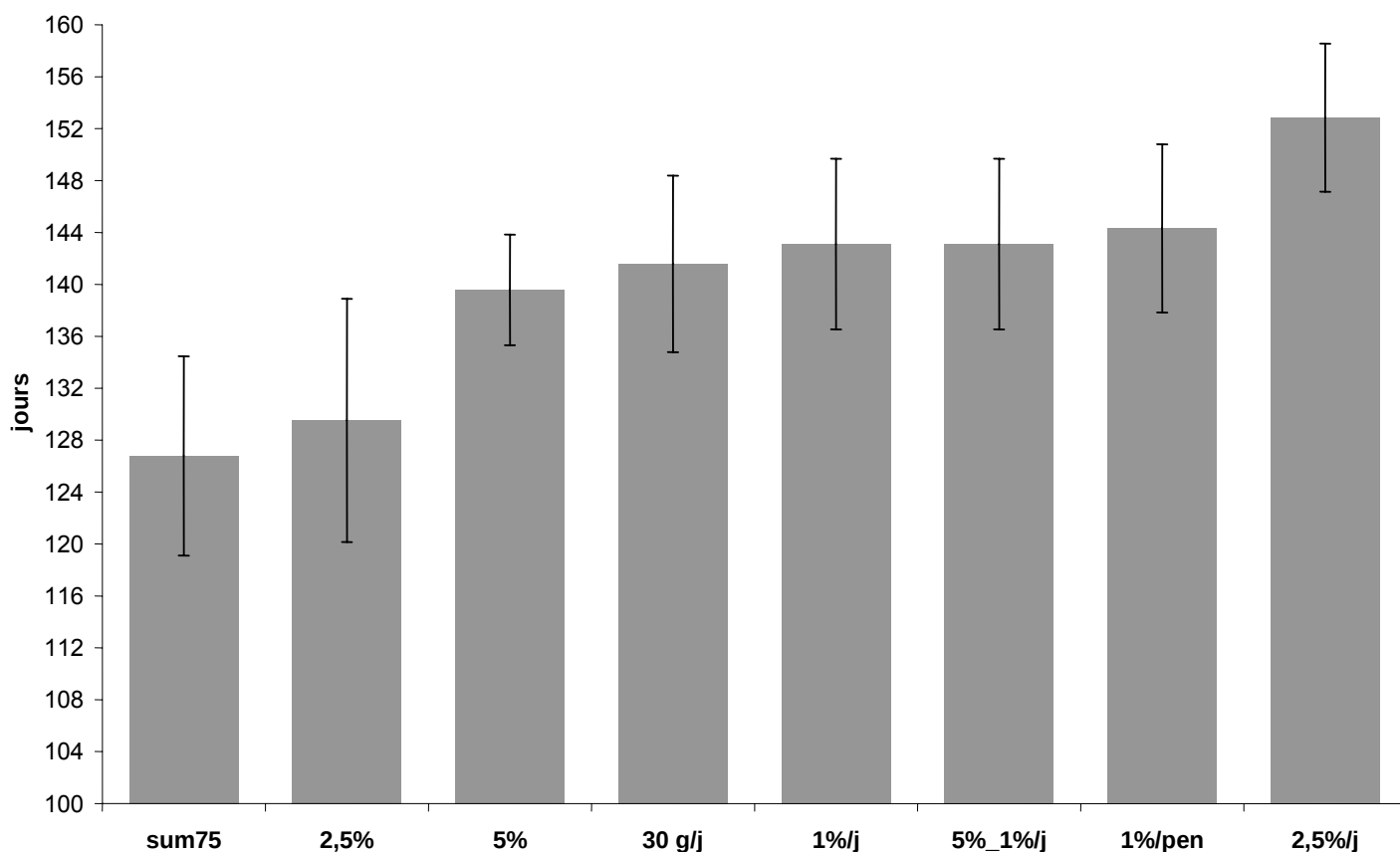


Figure 33 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé aux Poacées à Amiens (1987-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.

L'amplitude maximale des dates moyennes s'étend sur un pas de temps relativement long : 25 jours. La date moyenne la plus précoce est établie à partir du critère **sum75** (7 mai pour une année non bissextile) et la plus tardive à partir du critère **2,5%/j** (2 juin pour une année non bissextile). Les écarts-types les plus importants apparaissent pour les séries définies à partir de **sum75** et **2,5%** (7,7 et 9,4 jours), qui donnent les dates moyennes les plus précoces. La valeur des écarts-types n'excède pas 6,8 jours pour les autres critères, et n'atteint que 4,3 jours pour le critère **5%**.

Le test de Student, ou celui de Mann-Whitney dans le cadre des comparaisons faisant intervenir les séries **2,5%** ou **sum75**, est appliqué sur les mêmes duos de critères que ceux évoqués pour le frêne et le bouleau. La valeur p est inscrite entre parenthèses :

- **2,5%** et **5%** ($p < 0,0001$)
- **5%** et **5%_1%/j** ($p = 0,0577$)

- **1%/pen** et **1%/j** ($p = 0,5714$)
- **1%/j** et **2,5%/j** ($p < 0,0001$)
- **30 g/j** et **1%/j** ($p = 0,4863$)
- **30 g/j** et **2,5%/j** ($p < 0,0001$)
- **sum75** et **2,5%** ($p = 0,1227$)
- **sum75** et **5%** ($p < 0,0001$)

Contrairement à ce qui a été constaté pour les deux taxons précédents, nous enregistrons quatre différences significatives se traduisant par des écarts allant de 9,7 à 12,8 jours. Ainsi le passage de **2,5%** à **5%** et de **1%/j** à **2,5%/j**, critères cousins, occasionne une date d'apparition moyenne de 10,1 et 9,7 jours plus tardive. En effet, à l'opposé du bouleau, la pollinisation des Poacées est très progressive. Cela se caractérise par d'assez longues périodes de faibles et moyennes quantités précédant la pleine saison pollinique, qui a lieu généralement à partir de fin mai. Le choix du critère de détermination, contrairement au bouleau, est loin d'être anodin.

Nous pouvons noter ici que les dates moyennes issues du seuil **30 g/j**, et du critère **sum75**, sont significativement équivalentes à celles déterminées par les critères **1%/j** et **2,5%** respectivement. En théorie, les dates déterminées par ces quatre critères sont identiques si l'on enregistre exactement 3 000 grains de pollen annuellement. A Amiens, le cumul total moyen annuel est de 3 882 grains sur la période 1987-2005, ce qui est du même ordre de grandeur que la valeur exposée précédemment et explique l'absence de différence significative.

2.1.4.6 Les Poacées à Montluçon (1989-2005)

Selon les deux tests de normalité, l'ensemble des séries présente une distribution gaussienne, malgré les importantes disparités rencontrées au niveau des dates. Le test de Student est par conséquent utilisé. La présentation graphique des dates moyennes d'apparition du risque (figure 34) précède l'application du test.

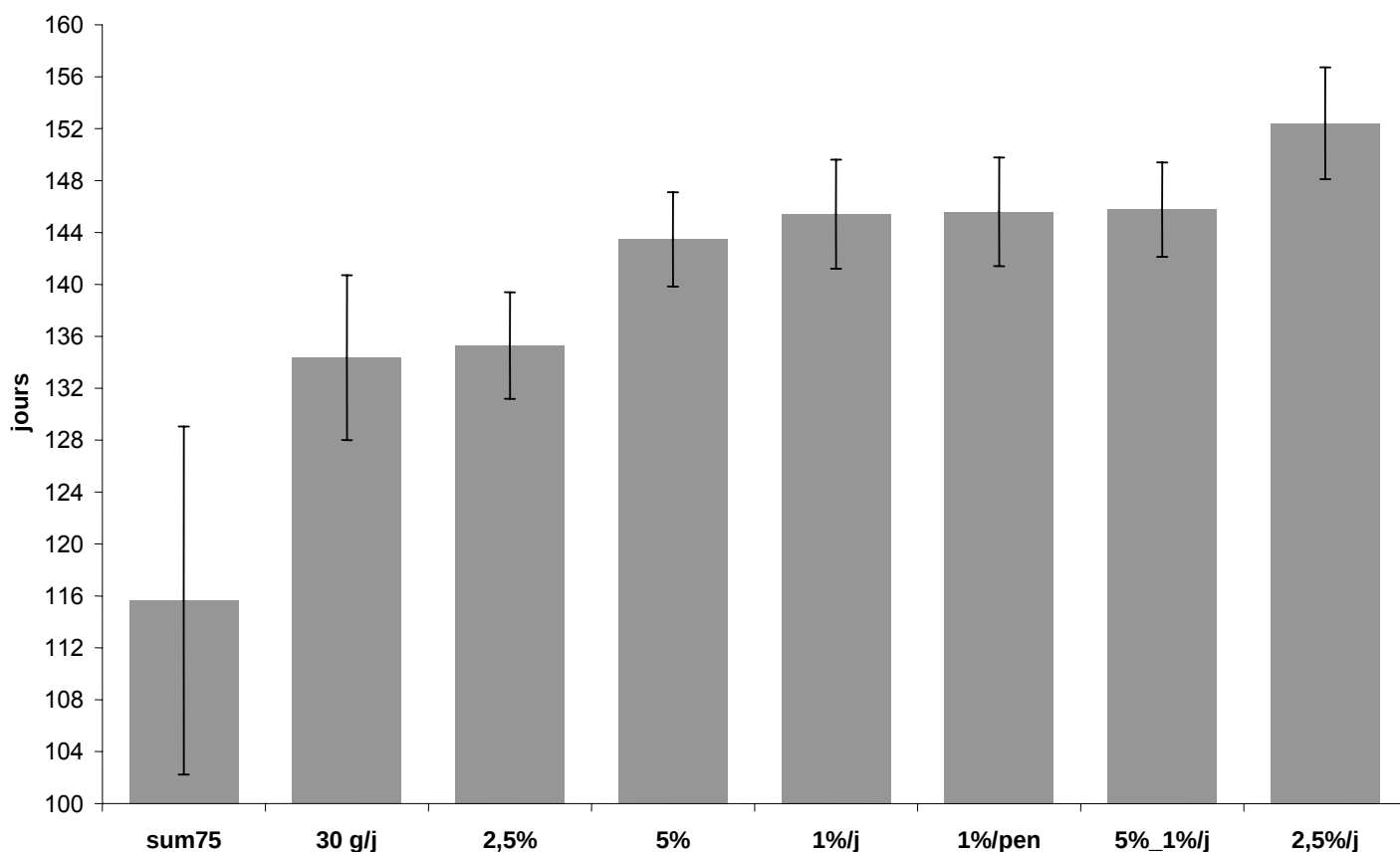


Figure 34 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé aux Poacées à Montluçon (1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.

Comme à Amiens, l'amplitude maximale des dates moyennes s'étend sur un pas de temps long (36 jours). La date moyenne la plus précoce est établie à partir du critère **sum75** (26 avril pour une année non bissextile) et la plus tardive à partir du critère **2,5%/j** (1^{er} juin pour une année non bissextile). Les écarts-types, compris entre 3,6 et 4,3 jours, sont faibles à l'exception des écarts-types des séries issues du critère **30 g/j** (6,4 jours) et, surtout, de celles déterminées à partir du critère **sum75** (13,4 jours). Ce sont les deux critères débouchant sur les dates moyennes les plus précoces.

Le test de Student est appliqué sur les duos de critères définis dans la section 2.1.4.1. La valeur p est inscrite entre parenthèses :

- **2,5% et 5%** ($p < 0,0001$)
- **5% et 5%_1%/j** ($p = 0,0752$)
- **1%/pen et 1%/j** ($p = 0,9031$)

- **1%/j et 2,5%/j** ($p < 0,0001$)
- **30 g/j et 1%/j** ($p < 0,0001$)
- **30 g/j et 2,5%/j** ($p < 0,0001$)
- **sum75 et 2,5%** ($p < 0,0001$)
- **sum75 et 5%** ($p < 0,0001$)

Ce test est significatif, pour un risque d'erreur de 5% pour six des huit combinaisons. On peut noter, entre autres, un écart de 8,2 jours entre les dates moyennes déterminées par les critères **2,5%** et **5%**, et une différence de sept jours exactement entre celles issues des critères **1%/j** et **2,5%/j**. Le caractère progressif de la pollinisation des Poacées est confirmé, ce qui ne rend pas anodin le choix du critère de détermination.

La particularité de Montluçon réside dans l'opposition entre les dates déterminées par les seuils bruts, et celles déterminées par des seuils relatifs. Les premières sont, en moyenne, significativement plus précoces et cela de manière très nette (jusqu'à 27,9 jours entre **sum75** et **5%**). Cela s'explique, en grande partie, par la pollinisation très abondante des Poacées, caractérisée par des cumuls annuels régulièrement supérieurs à 10 000 grains. La moyenne annuelle des pollens impactés est ainsi de 10 970 grains/m³ avec un pic, en 2001, à 18 435 grains. Pour une fréquence journalière de 1%, cela représente une moyenne de 110 grains et pour une fréquence cumulée de 5%, quasiment 550 grains. Des valeurs très largement supérieures aux deux seuils bruts utilisés.

2.1.5 Quel critère choisir ?

A travers les exemples d'Amiens et Montluçon, il est impossible d'affirmer que tel ou tel critère est meilleur qu'un autre. Le choix doit s'effectuer, avant tout, selon le but recherché. Selon le taxon, ce choix aura une plus ou moins grande incidence sur les dates déterminées. La pollinisation courte et brutale du bouleau n'entraîne que de très faibles différences, entre les critères, au niveau des dates de démarrage. Le critère retenu n'apparaît pas primordial contrairement aux Poacées, dont la pollinisation progressive entraîne des différences importantes entre les dates découlant des différents critères. Les seuils bruts, dans le cas d'une pollinisation abondante de ces Poacées, donnent des dates plus précoces que les seuils relatifs,

du fait des longues périodes de faibles quantités, pouvant faire leur apparition dès le début du mois d'avril.

L'étude portant sur les sites d'Amiens et de Montluçon a permis de montrer que chaque taxon a sa propre cinétique de pollinisation : rapide pour le bouleau, «intermédiaire» pour le frêne, progressive pour les Poacées. Afin de prévenir la survenue des pollinoses, l'utilisation de seuils bruts est plus pertinente car les symptômes allergiques se déclenchent, pour les personnes qui en sont affectées, à partir d'une certaine quantité qui n'est pas censée évoluer d'une année à l'autre. Etant donné que notre étude porte sur des stations françaises, nous allons utiliser les seuils liés aux niveaux de risque déterminés par le RNSA (Ickovic *et al.*, 1988 ; Thibaudon, 2003). Il y a six niveaux de risque qui sont les suivants :

- 0 = risque nul
- 1 = risque très faible
- 2 = risque faible
- 3 = risque moyen
- 4 = risque élevé
- 5 = risque très élevé

Nous allons, tout d'abord, récapituler les seuils associés aux niveaux de risque moyen à très élevé pour les taxons de notre étude :

- **Poacées :**
 - Niveau 3 $\geq$ 10 grains
 - Niveau 4 $\geq$ 35 grains
 - Niveau 5 $\geq$ 100 grains
- **Bouleau :**
 - Niveau 3 $\geq$ 50 grains
 - Niveau 4 $\geq$ 100 grains
 - Niveau 5 $\geq$ 171 grains
- **Frêne :**
 - Niveau 3 $\geq$ 71 grains

- Niveau 4 ≥ 228 grains
 - Niveau 5 ≥ 457 grains
- **Ambroisie :**
 - Niveau 3 ≥ 6 grains
 - Niveau 4 ≥ 10 grains
 - Niveau 5 ≥ 35 grains

Les concentrations journalières, correspondant aux divers niveaux, diffèrent selon le taxon et/ou la zone géographique. Les critères retenus, pour les trois taxons étudiés ci-dessus et l'ambroisie, dépendent à la fois de la cinétique de pollinisation, et du caractère allergisant du pollen :

- L'ambroisie étant un taxon au très fort potentiel allergisant, le RNSA considère qu'un risque moyen (niveau 3) d'exposition à ce taxon a lieu pour une quantité journalière supérieure ou égale à 6 grains/m³. Nous avons retenu, comme date d'apparition d'un risque d'exposition allergique, le premier jour (J) atteignant ce seuil avec, au moins, un total cumulé de 18 grains de J à J+2. Cette légère restriction permet d'éviter de faire démarrer la saison par un jour qui soit trop isolé, même si ce taxon commence sa pollinisation assez brusquement.
- En ce qui concerne les Poacées, le risque 3 survient pour une concentration journalière supérieure ou égale à 10 grains/m³. Ce seuil est volontairement bas pour initier des traitements anti-allergiques préventifs qui soient assez précoces, car ce taxon est la première cause de pollinoses en France. C'est pour cela que nous avons retenu comme date d'apparition du risque, le premier jour (J) atteignant 10 grains/m³ avec, au moins, un total cumulé de 50 grains de J à J+4 ; à condition que trois des cinq jours enregistrent des concentrations supérieures ou égales à 10 grains.
- Le risque 3 pour le bouleau correspond à un seuil de 50 grains/m³/jour. Du fait de la pollinisation courte et brutale de ce taxon, les écarts entre les dates moyennes issues des différents critères de détermination (bruts et relatifs) sont très faibles. C'est pour cela que nous allons retenir, comme date d'apparition du risque, le

premier jour (J) atteignant 20 grains/m³ (soit un risque 2), avec un total cumulé d'au moins 60 grains de J à J+2 .

- Les seuils de risque associés au frêne sont, quant à eux, difficilement exploitables car beaucoup trop élevés. C'est le cas de la concentration correspondant au risque 3, 71 grains/m³, qui peut ne jamais être atteinte pour des années à faible pollinisation. C'est pour cela que nous avons pris le même critère de détermination que celui du bouleau, étant donné que ces deux genres commencent leur pollinisation à une dizaine de jours d'intervalle, et qu'il y a chevauchement entre les deux saisons polliniques.

Nous avons retenu des critères de détermination inspirés par les seuils de risque, mis au point par le RNSA, pour leur aspect pratique. Comme cela a déjà été évoqué, ces critères, difficiles à déterminer, ne sont ni meilleurs, ni moins bons que ceux rencontrés dans la littérature mais ils ont l'indéniable avantage d'avoir une application pratique pour notre sujet. De plus, les concentrations retenues étant plutôt faibles, on minimise le risque de rater le démarrage de la pleine saison pollinique de ces différents taxons et des symptômes allergiques qui leur sont associés.

Il faut, cependant, préciser que les seuils de risque ont tendance à évoluer selon l'état d'avancement de la saison pollinique (Laaidi K, 2001 ; Thibaudon, 2003) avec une sensibilité plus grande des patients en cours de saison pollinique (Taudorf et Moseholm, 1988 ; Obtulowicz *et al.*, 1991 ; Laaidi K, 2001). De plus, la dernière étude citée en référence montre la difficulté d'établir des seuils cliniques en fonction des divers symptômes allergiques que sont la rhinite, la conjonctivite, la toux et l'asthme. Ainsi, pour le frêne, les seuils s'étendent de 10 à 247 grains pour l'ensemble des quatre symptômes, 35 à 252 grains pour le bouleau, et 3 à 81 pour les Poacées. Malgré les fluctuations enregistrées et la difficulté de détermination de seuils fixes inhérente à ces variations, ces valeurs brutes de concentration apparaissent, de nouveau, essentielles pour apprécier la validité de modèles de prévision à visée préventive.

2.1.6 Conclusion

Il semble évident que le choix du critère de détermination de la date de démarrage d'une saison pollinique dépend en grande partie du but de l'étude. En agronomie, retenir un seuil relatif semble beaucoup plus pertinent, surtout si l'on veut évoquer une date de début de pollinisation. En allergologie, il est beaucoup plus utile de retenir un seuil de détermination correspondant à une concentration brute et, si possible, que ce critère repose sur des niveaux de risque élaborés pour la zone d'étude. Ainsi, pour répondre à l'objectif préventif de notre travail, il est pertinent de se baser sur les critères du RNSA élaborés pour chaque taxon allergisant.

Suite à cette présentation des critères de détermination de la date d'apparition d'un risque d'exposition allergique aux pollens, nous allons évoquer les différentes méthodes issues de la littérature permettant de prévoir ce moment crucial pour les populations allergiques.

2.2 PRESENTATION GENERALE DES METHODES DE PREVISION DE LA DATE D'APPARITION DU RISQUE

La prévision de la date d'apparition d'un certain niveau de risque d'exposition aux pollens allergisants, ou de la date de début de pollinisation si l'on emploie un seuil relatif, constitue le point le plus important dans la prévention des symptômes allergiques liés aux pollens. En effet, il est nécessaire de donner une prévision de la date d'apparition du risque qui soit la plus précise possible, afin d'initier un traitement qui soit pris de la manière la plus efficace. En effet, si le début du traitement a lieu trop en avance par rapport à l'apparition du risque, cela entraîne un coût important pour le malade et la collectivité, des effets secondaires qui peuvent être non négligeables si le traitement se prolonge et, enfin, le risque d'apparition d'un phénomène d'accoutumance. Ce dernier inconvénient risque de rendre inefficace le traitement dès la survenue des premières concentrations susceptibles de déclencher des symptômes allergiques. De même, il n'a que peu d'utilité s'il est pris trop tardivement ; les patients auront eu le temps de souffrir de leur allergie, ce qui n'est évidemment pas le but recherché. Si le traitement est initié à temps, l'impact des concentrations journalières, relevées pendant la pollinisation de l'espèce allergisante étudiée, revêt une moins grande importance tant que le patient suit de manière assidue les prescriptions de l'allergologue.

Afin de déterminer cette date de début de pollinisation, ou d'apparition d'un niveau de risque dans notre cas, plusieurs méthodes sont utilisées dans la littérature. La plupart des articles, traitant de ce sujet, se basent uniquement sur la température de l'air, qui s'avère être le paramètre météorologique majeur dans l'établissement de ce genre de prévision. Dans ces articles, assez fréquemment, la période précédant le début de la pollinisation va être divisée en deux parties où les températures ne jouent pas le même rôle, et sont, par conséquent, exploitées de manière différente : la dormance (période de repos) suivie d'une période de «forçage» consistant en une accumulation d'unités de «chaleur». D'autres articles, enfin, vont employer des méthodes permettant de prendre en compte d'autres paramètres, telles les précipitations et l'humidité relative, qui peuvent avoir une influence sur le démarrage d'une saison pollinique.

Nous allons présenter, dans l'ordre, les méthodes de calcul de la dormance, de l'accumulation de chaleur une fois la dormance levée et, enfin, des méthodes permettant de tenir compte de plusieurs paramètres météorologiques.

2.2.1 Détermination de la dormance

2.2.1.1 Intérêt de la détermination

La dormance est un phénomène affectant les plantes pendant l'automne et une partie de l'hiver. Elle se caractérise par une période de repos déclenchée par la baisse de la température et par la diminution de la photopériode, qui correspond à la durée d'exposition de la plante à la lumière du jour, en lien avec des facteurs internes tels des régulateurs hormonaux. Cette dormance se manifeste par un quasi-arrêt ou un ralentissement du fonctionnement interne de la plante, et s'accompagne d'une plus grande résistance aux conditions ambiantes, même si ces dernières sont favorables à son développement (Sutra *et al.*, 1992). L'interruption ou levée de dormance s'effectue une fois que la plante a été exposée à des températures basses pendant un certain temps, par le truchement d'une élévation de la température et d'un allongement de la durée du jour. Le principe général d'évaluation de la dormance d'une plante consiste à calculer le nombre d'unités de «froid» (*chilling units* ou *chilling hours* selon les méthodes) nécessaire pour son complet accomplissement, en fonction de seuils de température adéquats.

2.2.1.2 Présentation de différentes méthodes

Plusieurs modes de calcul de la dormance sont rencontrés dans la littérature :

- **La méthode d'Aron (1983)** : elle constitue la méthode la plus couramment utilisée et s'avère particulièrement prisée dans les articles espagnols (García-Mozo *et al.*, 2000 ; Jato *et al.*, 2000a ; Galán *et al.*, 2001a, 2001b ; Jato *et al.*, 2002 ; Rodríguez-Rajo *et al.*, 2003, 2004 ; González-Parrado *et al.*, 2006 ; Fuertes-Rodríguez *et al.*, 2007). Cette méthode, employée à la base en agronomie et déterminée de manière empirique, détermine le nombre d'heures de «chilling» cumulées (comprises entre 0 et 7,2°C) pendant la période de dormance à partir d'équations de régression. Version un peu simplifiée de la formule exposée dans un article précédent du même auteur (Aron, 1975), cette méthode prend en compte la moyenne des températures minimales et maximales de la période et un critère se basant sur la longueur de cette dernière. Ce critère est calculé en fonction du seuil supérieur de température testé, de la durée de la période de dormance et de la moyenne des températures

minimales et maximales correspondante. Différents seuils de température (inférieurs et/ou supérieurs à 7,2°C) sont généralement testés, notamment par Aron et Gat (1991) qui ont employé les seuils de 7,2, 8,4 et 10,4°C, sans utiliser de seuil inférieur cette fois, pour déterminer le besoin en «froid» des plantes. A côté de ces critères d'estimation des heures de «chilling», la détermination de la durée totale de la période, telle qu'elle est présentée dans les articles reprenant le critère d'Aron (1983), s'effectue selon des critères qui peuvent conduire à l'imprécision. Selon les articles, le calcul commence :

- En octobre ou novembre, quand les températures moyennes descendent en dessous de 12,5°C et se termine quand un changement de tendance dans les températures est enregistré (Jato *et al.*, 2000a ; González-Parrado *et al.*, 2006 ; Fuertes-Rodríguez *et al.*, 2007).
- Dès que la température minimale descend en dessous du seuil testé. Il se termine quand une valeur minimale (durant la première «quinzaine» de janvier) est atteinte et qu'un changement de tendance est détecté (Jato *et al.*, 2002).
- Dès que les températures moyenne et minimale descendent sous la valeur du seuil testé (Rodríguez-Rajo *et al.*, 2003).
- A partir d'une date fixe comme le 1^{er} décembre dans García-Mozo *et al.* (2000).
- En sélectionnant les jours où les conditions météorologiques sont censées conduire à une absence d'activité physiologique selon la définition de Montero et González (1983), c'est-à-dire quand la température maximale est inférieure ou égale au seuil testé.

Les critères de détermination semblent ainsi sujets à interprétations. C'est surtout le cas de la détermination de la fin de la dormance. En effet, on peut rencontrer plusieurs périodes, pendant celle-ci, où l'on a des évolutions à la hausse des températures qui peuvent être suivies d'une baisse. A ce moment-là comment

savoir si la date d'arrêt retenue est juste ou non ? Il faudrait fixer de manière empirique un moment de l'année où la période de dormance est censée se finir ; cela permettrait de cadrer quelque peu la date, comme cela est préconisé dans Jato *et al.* (2002). La détermination au préalable de la durée de la dormance à partir de critères (surtout celui de fin de période) aussi sujets à interprétations risque de conduire à des cumuls de «chilling» assez peu fiables, et influencer la qualité des prévisions du début de pollinisation.

- **La méthode d'Andersen** : Combinant calculs de la dormance et d'unités de chaleur, elle se base sur la notion d'unité de «froid» (*chill unit*) qui s'avère être maximale pour une valeur de température donnée. Elle se réfère en fait au seuil défini par Richardson *et al.* (1974) où une unité de «chilling» correspond à une heure d'exposition à 6°C ; au-dessus et en dessous de ce seuil, la valeur est inférieure à une unité. Une fonction sinusoïdale permet alors de déterminer graphiquement et facilement la valeur du «chilling» par rapport à la température (Linville, 1990). Ce paramètre, élaboré à la base pour le pêcher, a été repris tel quel pour d'autres arbres tels que le bouleau (Sutra *et al.*, 1992) et le noisetier (Frenguelli *et al.*, 1992). Selon Sutra *et al.* (1992), la somme de *chill units* nécessaire est définie à partir de la méthode statistique d'Ashcroft *et al.* (1977). Elle consiste à prendre une gamme raisonnable de cumuls de *chill units*, à partir du mois de septembre, puis de calculer, pour chaque cumul, la date de fin de repos et la somme des unités de «chaleur» qui suivent (Sutra *et al.*, 1992). On cherche alors, pour chaque cumul de *chill units* (par exemple 250 CH, 275 CH, etc....), le plus faible écart-type des sommes d'unités de chaleur, et on retient le cumul de «froid» correspondant ainsi que la somme moyenne d'unités de «chaleur» associée à ce plus faible écart-type (Sutra *et al.*, 1992). Le choix final dépend donc du cumul d'unités de «chaleur» qui suit.
- **La méthode dite de Cesaraccio (2004)** : Développée pour des espèces forestières et culturales, elle s'appuie sur la notion de jours de «chilling» et d'«anti-chilling» (Cesaraccio *et al.* 2004), ces derniers correspondant en fait à l'accumulation d'unités de «chaleur». Les unités de «froid» correspondent à des valeurs négatives qui sont cumulées jusqu'à une valeur présélectionnée, identifiée comme la quantité de «chilling» à atteindre. A partir de là, on cumule des unités de «chaleur»

(positives) jusqu'à ce que le cumul des deux types d'unités atteigne 0, ce qui correspond à la date prévue de début de floraison. Les unités de froid sont calculées en fonction de la température minimale et maximale, par l'intermédiaire de différents seuils de température testés. D'une manière assez semblable à la méthode d'Ashcroft *et al.* (1977), les unités de «froid» sont déterminées après avoir au préalable calculé des degrés-jours au-dessus d'un seuil de 0°C, desquels on soustrait le total des unités d'«anti-chilling». Le signe de la différence est modifié afin d'avoir une valeur négative qui correspond alors au cumul d'unités de «froid». Selon ces auteurs cette méthode, approfondie par la suite (Cesaraccio *et al.*, 2006), est plus performante que les méthodes classiques évoquées précédemment. Une des principales causes de cette meilleure performance proviendrait du critère de démarrage de l'accumulation des unités de «froid». Dans les méthodes «classiques», les cumuls sont entamés en dessous d'un certain seuil de température, alors que les unités de «froid» du modèle de Cesaraccio *et al.* (2004) sont cumulées à partir d'un stade phénologique précis, comme la chute des feuilles. Ce critère semble en effet pertinent, et moins sujet à caution, à condition que l'on définisse un critère précis comme, par exemple, un certain pourcentage de feuilles jaunissantes ou qui sont déjà tombées. Malheureusement, il ne peut être utilisé que si une base de données phénologiques existe, ce qui semble difficilement applicable en-dehors d'un espace réduit (tel un parc) où l'on peut contrôler aisément les différents stades. Cependant si un réseau dense existe à l'échelle nationale, comme c'est le cas en Allemagne avec environ 1 600 sites d'observation (Estrella *et al.*, 2006), cette opération semble tout fait réalisable.

- D'autres méthodes, reprises dans des articles traitant de la prévision de l'ouverture des bourgeons (Chuine *et al.*, 1998, 1999), s'intéressent également à la détermination du besoin en «froid». La méthode dite «alternative» (Cannell et Smith, 1983 ; Murray *et al.*, 1989 ; Kramer, 1994) considère ainsi qu'en fonction d'un seuil de température donné, on cumule soit des unités de «froid», soit des unités de «chaleur». Les méthodes dites «séquentielle» (Sarvas, 1974 ; Hänninen, 1987, 1990 ; Kramer, 1994) et «parallèle» (Landsberg, 1974 ; Hänninen, 1987, 1990 ; Kramer, 1994) s'appuient sur le critère de Sarvas (1974) qui considère que les unités de «froid» sont calculées entre les seuils - 3,4°C et 10,4°C. La différence majeure entre les deux méthodes réside dans l'utilisation de ces unités de «froid».

Dans le cas de la méthode «séquentielle», les unités de «chaleur» ne peuvent être cumulées qu'une fois la dormance accomplie, alors que pour la méthode «parallèle» on considère que les unités de «chaleur» peuvent être actives de manière concomitante à des conditions nécessitant également des cumuls d'unités de «froid». Toutefois ces unités de «chaleur» ne sont pas pleinement efficaces tant que la période de dormance n'est pas complètement accomplie. Plusieurs hypothèses d'utilisation du «chilling» sont ainsi mises à jour. Dans le cadre de l'étude de Chuine *et al.* (1999), leur utilisation couplée à des méthodes de cumul d'unités de «chaleur» prévoit de manière plus précise la floraison des espèces étudiées qu'avec les seules méthodes de cumul d'unités de «chaleur».

2.2.1.3 La détermination des unités de «froid» est-elle essentielle ?

La dormance est un stade essentiel de repos de la plante, nécessaire à son bon développement. A première vue, il semble intéressant d'utiliser les unités de «froid» afin de prévoir la date de début de pollinisation. Malheureusement, les différentes méthodes de cumul présentent un problème principal qui réside dans l'imprécision de certains des critères de détermination de cette période, et dans le fait qu'il faut, pour certaines, se servir des unités de «chaleur» au préalable pour déterminer les unités de «froid», du moins pour fixer la durée de cette période. Dans ce cas-là, on peut discuter de l'intérêt d'une complication des modèles de cumul par une estimation du besoin en «froid».

De plus on peut considérer que dans un climat comme celui de la France, la dormance, qui dépend à la fois de la photopériode et de la température, est accomplie à une date qui varie assez peu d'une année sur l'autre, sauf pour une année avec un hiver exceptionnellement doux si on cumule les «chilling» sous un seuil donné de température. Mais si l'on considère, comme certaines des méthodes ci-dessus, que l'accumulation optimale de «froid» s'effectue entre deux seuils de température, la variabilité interannuelle de la fin de dormance doit être, théoriquement, plus faible qu'avec le critère précédent. On peut alors se focaliser sur l'accumulation d'unités de «chaleur», notamment dans le cadre de comparaisons entre plusieurs sites comme c'est le cas d'articles (García-Mozo *et al.*, 2002 ; Orlandi *et al.*, 2005), où les auteurs qui utilisent habituellement et favorablement la méthode d'Aron (1983), ne l'ont pas retenu.

L'amélioration que pourrait apporter l'utilisation d'une méthode de cumul d'unités de «froid» serait une date de fin de dormance et, donc, de début de cumul d'unités de «chaleur» qui ne soit pas fixe d'une année à l'autre. Pour répondre à cet objectif la détermination de la dormance, telle qu'elle est exposée par Sutra *et al.* (1992), semble la plus pertinente par le fait de tester différents cumuls fixes de «chilling» – en fonction desquels sont déterminés la fin de la dormance –, à condition de choisir la date de début de dormance à partir d'un seuil de température pris au-delà d'une certaine date (l'équinoxe d'automne par exemple), solution la moins mauvaise si nous ne disposons pas de données phénologiques. Cependant, si l'on veut être le plus complet possible, il faudrait tester différents seuils de température débouchant sur l'entrée en dormance. Ensuite, pour chaque critère de début de dormance, il faudrait tester différents seuils d'accumulation d'unités de «froid», partant du principe que ceux accompagnant les modèles ne sont pas adaptés à toutes les genres ou espèces. Enfin, pour chaque seuil d'accumulation d'unités de «froid», il faudrait tester différents cumuls afin de déterminer la fin de la dormance. Avec l'ajout de la détermination de cumuls d'unités de «chaleur», selon différents seuils de température, cela constituerait un travail extrêmement lourd et fastidieux, sans assurance d'amélioration des prévisions par rapport à la simple utilisation de degrés-jours. Ainsi dans une optique préventive, il ne semble pas nécessaire de calculer la dormance. De plus, dans le cadre d'une comparaison inter-sites, l'emploi des unités de «froid» serait particulièrement malaisé. Il le serait également dans la perspective d'une généralisation des modèles sur une large échelle.

2.2.2 La période de forçage des températures : présentation de plusieurs méthodes

2.2.2.1 Intérêt de la détermination de cette période

Cette période, qui s'étend de la fin de la dormance jusqu'à la date de début de pollinisation d'une espèce, est bien évidemment primordiale pour la détermination de cette dernière. Se basant uniquement sur la température de l'air, les différentes méthodes qui vont être présentées mettent en avant le rôle bénéfique des conditions thermiques sur le développement de la plante. En d'autres termes, plus la température est élevée, plus la somme de «chaleur» nécessaire à l'initiation de la pollinisation d'une espèce (ou de tout autre stade phénologique) est atteinte rapidement, débouchant alors sur une date précoce. Cette période de forçage est appréhendée de manière différente selon les méthodes. La relation entre la température et le

développement de la plante n'est pas toujours considérée comme linéaire ; certaines méthodes, présentées par la suite, s'appuient en effet sur une relation température/développement considérée comme exponentielle ou sinusoïdale. Six méthodes vont être préalablement retenues pour ce travail.

2.2.2.2 Les six méthodes présélectionnées dans ce travail

Leur choix s'est effectué selon plusieurs critères. Les méthodes les plus couramment utilisées ont été retenues, à condition que leur application ne soit pas problématique. Cela suggère que les équations soient correctement présentées et expliquées, et qu'elles ne nécessitent que des paramètres météorologiques. De même, dans ce choix, nous avons privilégié une certaine diversité quant au type de relation unissant le développement de la plante à la température. Les six méthodes retenues sont exposées ci-après.

La méthode des GDD (Growing Degree Days) par triangulation

Cette méthode linéaire a été développée, à l'origine, pour calculer les unités de chaleur permettant de gérer l'influence des insectes nuisibles dans les cultures (Zalom *et al.*, 1983), en partant du principe que la plupart des organismes se développent et grandissent en fonction de la température. Le but de cette méthode est d'estimer les degrés-jours d'une journée entière en fonction de la température minimale et maximale. Les auteurs considèrent que l'évolution diurne des températures forme un triangle équilatéral, dont le sommet correspond à la température maximale. A partir de là, ils vont estimer les degrés-jours en fonction de deux seuils de température : un seuil inférieur et un seuil supérieur. Ces degrés-jours correspondent à la surface du triangle comprise entre les deux seuils. Selon la position de la température minimale et de la température maximale par rapport à ces seuils, différentes formules sont élaborées pour calculer ces unités de «chaleur». Cela impose donc de tester deux seuils de température simultanément afin de déterminer la quantité d'unités de «chaleur» cumulées qui fournisse la meilleure prévision. Cela est donc particulièrement fastidieux et demande un temps de calcul considérable. Mais les différents articles, s'appuyant sur cette méthode, n'utilisent en pratique que le seuil inférieur de température en reprenant les formules développées par Zalom *et al.* (1983) quand la température maximale est inférieure au seuil

supérieur. Ce choix est, en partie, justifié grâce à une étude menée sur des observations au champ (Snyder *et al.*, 1999) qui considère que le développement de la plante est négligeable quand la température est inférieure au seuil minimal, et qu'il n'y a pas d'augmentation du taux de croissance quand la température est supérieure au seuil maximal. De manière sous-entendue, cet article considère que l'utilisation d'un seuil unique de température, en fonction duquel sont calculés les degrés-jours, est amplement suffisante. Il n'y a plus alors qu'à tester différentes valeurs de température pour ce seul seuil, afin de déterminer la somme d'unités donnant les prévisions les plus précises. Les *Growing Degree Days* se calculent dès lors de la manière suivante :

$$\text{GDD} = 0 \text{ quand } T_x \leq x$$

$$\text{GDD} = [(T_x - x)/2][(T_x - x)/(T_x - T_n)] \text{ quand } T_n < x < T_x$$

$$\text{GDD} = T_m - x \text{ quand } T_n \geq x$$

T_n , T_x et T_m correspondent, respectivement, à la température minimale, maximale et moyenne ; x est le seuil de température testé.

La méthode des GDD (Growing Degree Days) par une fonction sinusoïdale

Elle est le pendant de la précédente car elle a également été développée par Zalom *et al.* (1983). Elle s'appuie aussi sur les températures minimale et maximale mais, contrairement à la méthode se basant sur la triangulation, elle considère que l'évolution diurne ne forme pas un triangle, mais suit une courbe sinusoïdale. Les degrés-jours correspondent à la surface délimitée par la courbe et les seuils inférieur et supérieur de température. Cette évolution estimée apparaît plus proche de la réalité, mais en contrepartie elle débouche sur des formules plus complexes. Comme pour la méthode précédente, une simplification a été réalisée selon les critères évoqués par Snyder *et al.* (1999). Les formules sont identiques à celles du modèle précédent, excepté quand le seuil de température testé est compris entre la température minimale et maximale. Dans ce dernier cas les degrés-jours se calculent de la manière suivante :

$$GDD = \frac{1}{\pi}[(T_m - x)(\frac{\pi}{2} - \theta) + \alpha \cos(\theta)]$$

$$\alpha = (T_x - T_n)/2$$

$$\theta = \sin^{-1}[(x - T_m)/\alpha]$$

Ces deux méthodes sont utilisées dans plusieurs articles traitant, notamment, de la prévision de la date de début de pollinisation de l'olivier (Galán *et al.*, 2001a, 2005 ; Orlandi *et al.*, 2005) et du chêne (García-Mozo *et al.*, 2000, 2002 ; Jato *et al.*, 2002) avec une plus grande prédilection pour la méthode sinusoïdale. A ce propos, notons que plusieurs de ces auteurs citent, comme référence pour cette méthode, Snyder (1985) au lieu de Zalom *et al.* (1983).

La méthode de Lejoly-Gabriel

Elle est issue des travaux de Lejoly-Gabriel (1978) réalisés sur le site de Louvain en Belgique. Il s'agit d'une version, également linéaire mais beaucoup plus simple, des méthodes de GDD. Le principe est de cumuler les températures moyennes journalières au-dessus d'un certain seuil de température ; plusieurs seuils sont évidemment testés. A côté de la température moyenne, nous avons testé les cumuls de température minimale et maximale. Ce dernier paramètre s'est avéré être le plus efficace et a été retenu préférentiellement à la température moyenne. La formule est donc la suivante :

$$\text{Lejoly} = T_x \text{ quand } T_x > x$$

Comme précédemment, T_x correspond à la température maximale et x est le seuil de température testé. Il faut souligner que si T_x est inférieure ou égale au seuil testé, la valeur du critère est égale à 0. Cette méthode est très majoritairement utilisée dans des articles d'auteurs francophones traitant aussi bien du bouleau (Sutra *et al.*, 1992 ; Laaidi M, 2001a), que des Poacées (Laaidi M, 2001b), de l'ambrosie (Laaidi *et al.*, 2003a ; Thibaudon *et al.*, 2004) ou de plusieurs taxons (Laaidi M, 1999 ; Laaidi et Laaidi, 2001) à partir de la température moyenne. Elle a très probablement inspiré d'autres auteurs, qui ont également utilisé des cumuls de température maximale (García-Mozo *et al.*, 2000 ; Galán *et al.*, 2001a).

La méthode FT (Fraction-Time)

Cette méthode, comme les deux s'appuyant sur les *Growing Degree Days*, se base sur les températures minimale et maximale. Développée en lien avec une méthode de cumul d'unités de «froid» par Cesaraccio *et al.* (2006), elle a été élaborée, dans un cadre agronomique, pour la prévision de la floraison de plusieurs arbres fruitiers. Les formules sont les suivantes :

$$FT = 0 \text{ quand } T_x \leq x$$

$$FT = (T_x - x)/(T_x - T_n) \text{ quand } T_n < x < T_x$$

$$FT = 1 \text{ quand } T_n \geq x$$

Comme précédemment, T_n et T_x correspondent respectivement à la température minimale et maximale, et x au seuil de température testé. Les valeurs d'unités de chaleur de cette méthode sont comprises entre 0 et 1. Ainsi, contrairement aux méthodes GDD, l'unité de chaleur ne varie pas quand la température minimale est supérieure ou égale au seuil testé. En d'autres termes on considère, à travers cette méthode, qu'au-delà d'une certaine valeur l'influence de la température reste la même, et que la relation linéaire entre cette dernière et le développement de la plante ne s'effectue qu'en présence d'un seuil de température compris entre T_n et T_x .

La méthode du Q_{10}

Cette méthode a été mise au point à la fin du XIX^{ème} siècle et s'inspirait des récentes théories de la chimie (Laaidi et Laaidi, 2001). Arrhenius supposait que la vitesse d'une réaction chimique est multipliée par un facteur constant (le Q_{10}) chaque fois que la température croît de 10°C (Durand, 1967, 1969). En lien avec la loi de J.H. Van't Hoff en cinétique chimique, $Q_{10}(t)$ s'identifie ainsi au rapport existant entre la vitesse de croissance V à une température T et celle qui serait observée à la température $T+10^\circ\text{C}$ (Sutra *et al.*, 1992), selon la formule suivante où V représente la vitesse de développement de la plante et T la température :

$$Q_{10}(t) = V(T+10)/V(T)$$

Nous avons choisi de retenir ici le critère de Bidabe (1967) d'«action chaude» qui se base sur la notion de Q_{10} , et qui utilise les températures minimale et maximale pour l'élaboration de la formule :

$$AC = Q_{10}^{T_n/10} + Q_{10}^{T_x/10}$$

Par parenthèse, Il faut noter que Bidabe (1967) a également mis au point une formule d'«action froide» qui correspond à une accumulation d'unités de «froid», et dont la formule est identique à AC sauf que les températures en exposant sont précédées d'un signe négatif. Pour en revenir au calcul de AC, contrairement aux méthodes précédentes, on ne va pas tester différents seuils de température, mais différents coefficients du Q_{10} . En outre, et là encore contrairement aux modèles précédents qui se basaient sur l'hypothèse que la température ne joue aucun rôle sur le développement de la plante sous un certain seuil, cette méthode d'action exponentielle considère que toutes les températures, mêmes négatives à priori, ont une influence sur le développement de la plante, aussi minime soit-elle. Cette méthode, dont la philosophie diffère des modèles plus «classiques», est reprise dans des articles d'auteurs francophones (Laaidi et Laaidi, 2001 ; Laaidi M, 2001b ; Chappard *et al.*, 2004).

La méthode de Sarvas

Mise au point par Sarvas (1974) et reprise par Chuine *et al.*, (1998, 1999), il s'agit d'une méthode exponentielle se basant, originellement, sur la température moyenne (T dans l'équation) dont la formule est la suivante :

$$Sarvas = \frac{28,4}{1 + e^{-0,185(T-18,4)}}$$

En plus de la température moyenne, nous avons appliqué la formule sur les températures minimale et maximale. A l'origine, la valeur de l'unité de «chaleur» est considérée comme nulle si la température est inférieure à 0°C. Comme il s'agit d'une formule exponentielle, les valeurs sont toujours positives, quel que soit le signe des températures. En suivant l'hypothèse à l'origine de la mise au point de la méthode du Q_{10} , nous avons également appliqué cette méthode en prenant en compte toutes les températures, et non les seules valeurs positives.

2.2.2.3 Présentation d'autres méthodes de cumul d'unités de «chaleur»

Les paragraphes de cette section correspondent soit à un ensemble de méthodes proches par leur conception, soit à des méthodes assez caractéristiques qui sont alors présentées spécifiquement.

- Un certain nombre de méthodes, de conception simple, se rapprochent de celle de Lejoly-Gabriel (1978). Il y a, tout d'abord, celles qui se basent sur la température moyenne où l'on cumule, au-dessus d'un seuil donné, les écarts entre cette température et la valeur du seuil. Ces méthodes sont liées à des cumuls d'«unités de froid» (Cannell et Smith, 1983 ; Murray *et al.*, 1989 ; Kramer, 1994) ou non (Tabuenca et Herrero, 1965 ; Rickman *et al.*, 1983 ; Alcalá et Barranco, 1992 ; Yang *et al.*, 1995).

On retrouve le même type de processus, mais fondé sur la température maximale, dans l'article de Jato *et al.* (2002).

- Des versions proches des deux méthodes de GDD élaborées par Zalom *et al.* (1983) existent également. Ainsi Cesaraccio *et al.* (2004) proposent une version simplifiée de la méthode par triangulation où les degrés-jours, dans le cas où le seuil testé est compris entre la température minimale et maximale, correspondent à la moitié de l'écart entre la température maximale et la valeur du seuil testé. Orlandi *et al.* (2005) évoquent des méthodes qui ont très probablement inspiré Zalom *et al.*, (1983) pour l'élaboration de leur méthode sinusoïdale. On peut noter ainsi la méthode d'Allen (1976), celle de Parton et Logan (1981) – où l'on utilise une fonction sinusoïdale tronquée pendant le jour et une fonction exponentielle décroissante durant la nuit – et celle, postérieure, dite de Cesarracio *et al.* (2001) qui utilise deux fonctions sinusoïdales le jour et une fonction carrée décroissante durant la nuit. La méthode de De Wit *et al.* (1978) utilise deux fonctions cosinus, la première pour la période comprise entre la température minimale et la température maximale, la seconde entre cette dernière et la température minimale du jour suivant.
- **La méthode des GDH (*Growing Degree Hours*) d'Andersen (1991) :** Elle a été élaborée de telle sorte qu'elle soit utilisée graphiquement et que l'on puisse ainsi

estimer les GDH en fonction des températures horaires. De cette méthode, mise au point à Copenhague et valable – à priori – pour plusieurs espèces, ont découlé différents seuils d'action de la température utilisés dans la détermination graphique :

- 4°C qui correspond à la température de base sous laquelle il n'y a que peu ou pas de développement.
- 25°C qui est la valeur où le développement est maximal.
- 36°C qui correspond à la température critique au-delà de laquelle il y a peu ou pas de développement.

Ces seuils marquent la prise en compte de l'effet néfaste de températures trop élevées sur la plante, aspect inexistant dans les autres modèles, que l'on doit relier très probablement à un stress hydrique. Cependant le seuil de 36°C déterminé par Andersen doit être dépassé très rarement à Copenhague. La détermination de son rôle sur le développement de la plante ne se basant que sur de rares données, nous pouvons émettre un doute sur sa pertinence et sa robustesse à une telle latitude (55°N). De toute manière, ces seuils ne donnent pas de prévisions fiables pour des villes autres que Copenhague (Sutra *et al.*, 1992), suggérant la nécessité de développer des modèles propres au site étudié.

- **La méthode de Larsson (1993)** : Dite d'«activité cumulée», c'est une méthode exponentielle qui a la particularité d'utiliser trois paramètres, déterminés statistiquement, en plus de la température. Le premier paramètre a comme valeur de départ 0 et l'on teste toutes les valeurs supérieures de deux en deux. Pour les deux autres paramètres, les valeurs sont testées par palier de 0,2 à partir de la valeur 0,4 pour l'un, et 0 pour l'autre. L'auteur stipule, par ailleurs, que le démarrage du cumul doit s'effectuer après un certain nombre de jours où les températures dépassent un certain seuil ; et si ces températures descendent en dessous de ce seuil, pour un certain nombre de jours également, il considère que la plante perd toutes les unités de «chaleur» accumulées jusque-là, et que les cumuls doivent repartir de zéro. Cette méthode assez compliquée permet d'élaborer des modèles dont les prévisions divergent, en moyenne, de trois à six jours selon les taxons.

- **La méthode de Törnevik (1982)** : Elle se base, comme celle du Q_{10} , sur le principe de réaction chimique d'Arrhenius et prend en compte la température moyenne. Elle a été principalement élaborée pour la prévision du pollen de bouleau. Il s'agit d'une méthode exponentielle, complexe dans son élaboration, où l'on doit tester quatre paramètres, dépendant de l'espèce et du stade phénologique atteint par la plante, à côté de la température. Deux de ces paramètres sont, probablement, des coefficients de croissance, car leur utilisation dans la formule se rapproche de celle de Bidabe (1967) pour le Q_{10} ...en plus compliqué. L'un de ces paramètres, de par ses valeurs très faibles, pourrait être supprimé, sans préjudice, du contenu de la formule. Les deux paramètres restants sont des seuils de température placés en exposant, de manière similaire à la méthode du Q_{10} . La détermination statistique de ces quatre paramètres est particulièrement lourde en temps de calcul. De plus, il est nécessaire de connaître les normales journalières de température pour les stations étudiées et, surtout, la date de reprise d'activité de l'espèce. Cela complique la tâche si l'on ne dispose pas de séries phénologiques. Il est donc plus commode d'utiliser des modèles moins complexes, où l'on peut estimer la date de reprise d'activité statistiquement. Les prévisions de très bonne qualité issues de cette méthode appliquée au bouleau en France (Sutra *et al.*, 1992) ne justifient pas pleinement l'utilisation d'une méthode qui n'est reprise dans aucun article, excepté celui cité, traitant de la prévision de la date de début de pollinisation.

Suite à l'évocation des différentes méthodes rencontrées dans la littérature, il convient de préciser que l'Unité de Palynologie de l'Ecole Nationale Supérieure Agronomique de Montpellier (ENSA-M) a mis au point un logiciel de prévision des émissions polliniques, en temps réel, se basant sur des modèles phénologiques établis à partir de séries polliniques et météorologiques : le pollenoscope. Il est notamment cité dans le rapport de juillet 2006 du Plan Régional Santé-Environnement de la Préfecture de Languedoc-Roussillon (www.languedoc-roussillon.pref.gouv.fr).

2.2.3 Utilisation d'une méthode plus globale : la régression linéaire multiple

Le but de cette méthode est de prendre en compte, si cela s'avère nécessaire, d'autres paramètres météorologiques que la température dans la détermination de la date d'apparition d'un risque allergique d'exposition aux pollens. La régression multiple consiste, dans notre cas, à expliquer cette date par la conjonction des variations de plusieurs phénomènes (ici les variables météorologiques), en principe indépendants, appelés variables explicatives, variables indépendantes ou encore régresseurs (Chadule, 1997). L'équation de régression s'écrit de la manière suivante :

$$REAP = a_1x_1 + a_2x_2 + \dots + b$$

La variable REAP correspond ici au premier jour où survient un Risque d'Exposition Allergique aux Pollens, cette terminologie sera utilisée tout le long de la suite de ce travail. Les variables explicatives ou régresseurs sont symbolisés par x_1 et x_2 et les deux derniers paramètres correspondent au coefficient de régression (a) et à l'ordonnée à l'origine (b).

L'utilisation de cette méthode suppose que l'on considère que des paramètres tels que les précipitations, la durée d'insolation ou l'humidité relative peuvent jouer un rôle pendant la période qui précède l'apparition du risque ou la date de début de pollinisation pour une espèce donnée. En règle générale, les régresseurs retenus correspondent à des variables liées à différentes périodes des mois précédant le démarrage de la saison (par exemple : l'humidité relative minimale de la première «quinzaine» de mars et la température moyenne de la première «quinzaine» d'avril) étant donné les fortes autocorrélations existant entre les variables météorologiques concomitantes. Cependant il se peut que deux variables d'un même pas de temps (exemple : température maximale et précipitations de la première «quinzaine» de janvier) puissent être significativement corrélées avec les dates de début de pollinisation, sans pour autant être redondantes entre elles. Pour déterminer cela, nous allons utiliser la régression multiple pas à pas.

La démarche d'application de cette méthode consiste à recueillir un assez grand nombre de variables potentiellement explicatives et à choisir, de manière objective parmi elles, l'ensemble le plus explicatif. Ainsi après l'ajout d'un premier régresseur, on considère que les variables indépendantes incluses par la suite sont corrélées partiellement avec la variable dépendante. Par exemple, la température moyenne de la deuxième «quinzaine» d'avril est corrélée négativement avec la date d'apparition du risque lié aux Poacées ; plus la température

est élevée, plus l'apparition du risque est précoce : c'est le premier régresseur inclus. Or la régression multiple pas à pas met en avant une corrélation partielle positive entre l'apparition du risque et les cumuls de précipitations de cette «quinzaine» ; cela revient à dire qu'à température moyenne égale, les précipitations ont un effet retardateur sur l'apparition du risque. La température est, très certainement, le paramètre majeur influençant le développement et la croissance des plantes. Mais l'exemple évoqué précédemment montre l'intérêt d'utiliser cette méthode, qui permet de prendre en compte d'autres paramètres qui jouent un rôle équivalent à celui de la température (comme l'humidité relative qui y est très liée), modérant ou amplifiant celle-ci.

Par la régression multiple, nous avons voulu considérer l'impact de l'ensemble des paramètres météorologiques composant le temps «sensible» ressenti. Bien évidemment, du fait de redondances entre les paramètres, une à deux variables d'un même pas de temps sont généralement retenues. La significativité statistique, nécessaire à la validation des équations de régression, nous oblige à retenir un nombre assez réduit de variables explicatives.

Cette méthode a été privilégiée dans les articles d'auteurs francophones et notamment tous les articles, présentés en bibliographie, où M. Laaidi est l'auteur principal et qui se basent, le plus souvent, sur les travaux de Lejoly-Gabriel (1978) et Sutra *et al.*, (1992). Adams-Groom *et al.*, (2002) ont également utilisé cette méthode pour le bouleau, mais seulement à partir de la température et des précipitations d'après les résultats d'une étude britannique antérieure (Emberlin *et al.*, 1993). Enfin, cette méthode a été utilisée pour la prévision de la date de début de pollinisation des Poacées en Australie, en faisant ressortir l'influence des précipitations (Ong *et al.*, 1994,1997).

2.2.4 Les critères de choix des meilleurs modèles

Après avoir présenté les méthodes utilisées pour prévoir l'apparition d'un certain niveau de risque allergique, nous allons présenter, de manière générale, les critères qui vont nous permettre, de la manière la plus objective possible, de choisir les meilleurs modèles pour chaque méthode présentée.

2.2.4.1 Les critères destinés aux méthodes de cumul

Ces méthodes consistent à cumuler des températures, ou des unités de «chaleur», à partir d'une date supposée correspondre à la levée de dormance, et au-dessus d'un seuil donné. Le but est ici de déterminer la combinaison date/seuil de température (date/coefficient de développement pour la méthode du Q_{10} et date/type de température pour la méthode de Sarvas) qui minimise les écarts entre les valeurs prévues et observées sur l'échantillon d'apprentissage servant à l'élaboration du modèle. L'erreur moyenne absolue (*Mean Absolute Error*) se révèle être le meilleur critère dans ce cas-là. Cela consiste à calculer l'écart moyen, en valeur absolue, entre les valeurs prévues et observées. La formule est la suivante :

$$|MAE| = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i|$$

N correspondant au nombre d'observations, x_i à la valeur observée et $\hat{x}_i$ à la valeur prévue par le modèle. Cependant, du fait du grand nombre de combinaisons possibles, il se peut qu'il y en ait plusieurs présentant la même erreur moyenne absolue. Dans ce cas, nous nous basons sur l'erreur moyenne au carrée (*Root-Mean-Square-Error*) pour les «départager», en retenant la combinaison où cette erreur est minimisée. Il s'agit ici de calculer la moyenne des écarts, élevés au carré, entre les valeurs prévues et observées puis d'appliquer une racine carrée sur cette moyenne. Contrairement à l'erreur moyenne absolue, qui est un bon indicateur des écarts moyens de prévision, l'erreur moyenne au carré décrit aussi la distribution de l'amplitude des écarts, à travers une plus grande sensibilité aux valeurs extrêmes (Houzé, 2006). Ainsi, en cas d'égale valeur de l'erreur moyenne absolue, le fait de retenir la plus faible erreur moyenne au carré revient à retenir le modèle donnant, en théorie, de bonnes estimations de l'ensemble des dates de l'échantillon, plutôt que celui qui génère de «très» bonnes prévisions pour la grande majorité des observations et qui diverge totalement pour une ou deux années. Cependant, comme cela sera montré ultérieurement, un grand nombre de modèles issus de ces méthodes de cumul se caractérisent par une ou deux années où l'écart entre la valeur prévue et celle observée est important.

2.2.4.2 Les critères destinés à la régression linéaire multiple

Dans le cas présent, nous allons d'abord appliquer une régression multiple pas à pas ascendante sur les variables météorologiques d'une même période (exemple : application sur les variables de la première «quinzaine» de janvier, de la deuxième «quinzaine» de janvier et ainsi de suite). Cela permet de retenir les variables les plus significativement corrélées (pour $p < 0,05$) directement ou partiellement avec l'apparition du risque, sans qu'il y ait d'interférences entre les différentes périodes qui risqueraient de générer des corrélations partielles factices.

Les variables ainsi déterminées par la méthode pas à pas sont ensuite incluses dans une équation de régression linéaire multiple standard. Ces variables sont physiquement indépendantes mais, malheureusement, une multi-colinéarité purement statistique peut exister entre elles. Certains régresseurs vont donc apporter des informations redondantes et se caractérisent ainsi par des coefficients de corrélation partielle, calculés en fonction des autres variables explicatives, qui ne sont pas significatifs. Dans un souci de validation statistique des équations de régression, il convient alors de retirer les variables explicatives qui n'apportent qu'une information supplémentaire très minime. Pour choisir le nombre idéal de régresseurs à inclure dans l'équation, nous allons utiliser le test Af de performance en prévision réelle. Sa formule est la suivante :

$$Af = R^2 - 2[(M / N) + (1 - R^2)] / (1 - M + N)$$

La valeur de Af est une déclinaison du coefficient de détermination (R^2) qui prend en compte le nombre de régresseurs utilisés (M) et le nombre d'observations disponibles (N). Nous allons donc calculer la valeur de ce test pour l'équation d'origine, puis retirer la variable déterminée comme la plus redondante dans l'équation, et recalculer la valeur de Af une fois cette opération effectuée. En l'appliquant, nous avons pu constater que le fait de retirer une variable très redondante, faisant donc diminuer la valeur du R^2 très faiblement, entraînait une hausse de la valeur de Af, qui peut s'expliquer par la présence du nombre de régresseurs au dénominateur de la formule. Nous allons donc retirer, au fur et à mesure, les variables les plus redondantes jusqu'à atteindre la valeur optimale de Af, et l'équation qui en découle sera celle retenue pour les prévisions. Cependant, il faut préciser qu'avec ce critère, la redondance n'est pas toujours entièrement éliminée, et qu'il faudrait encore retirer des régresseurs afin de minimiser l'écart entre la valeur du R^2 et celle de Af, condition d'origine pour déterminer le meilleur modèle (Houzé, 2006) et pour totalement éliminer la multi-colinéarité. Mais cela

impose une variance expliquée encore plus faible et, par conséquent, une erreur moyenne absolue plus élevée du fait d'une diminution de l'information expliquée par le modèle. Il est donc préférable de retenir, comme critère de détermination, la valeur la plus élevée de A_f , quitte à garder des régresseurs apportant un peu de redondance, redondance qui n'est que purement statistique. Il faut alors juste s'assurer que le coefficient de corrélation multiple de l'équation est significatif (risque d'erreur inférieur à 5%) à travers le test F de Fisher, appliqué de manière automatique sur les équations de régression multiple.

2.2.5 Conclusion

Nous avons présenté, dans ce chapitre, les différentes méthodes de prévision de la date de début de pollinisation, ou d'apparition du risque allergique, présentes dans la littérature. Trois types de méthodes n'employant que la température ressortent : les méthodes linéaires (simples d'exécution), celles, proches des précédentes, employant des fonctions géométriques (sinus mais aussi cosinus) et enfin les méthodes dites exponentielles employant, le plus souvent, des coefficients de développement des plantes. Nous avons voulu retenir, au préalable, six méthodes qui englobent ces différents cas de figure et qui étaient présentées de manière claire dans la littérature. Quant à la régression linéaire multiple, elle a pour intérêt majeur de permettre la prise en compte de paramètres météorologiques autres que la température, reconnus pour leur influence sur les concentrations polliniques, et qui peuvent, en toute logique, avoir un rôle pendant la période précédant le début de la pollinisation. Les critères de choix des meilleurs modèles de prévision sont propres au type de méthode employé et seront utilisés de manière pratique dans la présentation des résultats.

CONCLUSION DE LA PARTIE

Nous avons évoqué, en premier lieu, les différents critères de détermination de la date de démarrage d'une saison pollinique nécessaires dans le cadre d'une prévision de l'apparition d'un risque d'exposition allergique aux pollens. Ces critères s'avérant tous valables, le choix doit s'effectuer selon l'objectif recherché. Parmi les méthodes de cumul d'unités de «chaleur» à destination de la prévision de l'apparition du risque présentes dans la littérature, dont plusieurs se basent sur la même méthodologie, nous en avons retenu préalablement six qui seront exploitées par la suite. Le calcul de la dormance, accompagnant souvent celui de la période de forçage des températures, n'a pas été retenu à cause de la détermination sujette à interprétations des limites de la période d'accumulation d'unités de «froid» et de la complexité des modèles engendrés, sans que cela puisse apporter, à coup sûr, une amélioration des prévisions. La régression linéaire multiple apparaît, quant à elle, comme une alternative multi-variables aux méthodes de cumul n'employant que la température qui reste, malgré tout, le paramètre le plus important dans le développement végétatif.

La troisième partie de ce travail consiste, ainsi, en une présentation des résultats issus de l'application des différentes méthodes retenues. Une étude sur les concentrations journalières de Poacées complète cette présentation des résultats.

PARTIE 3 :
PRESENTATION GENERALE DES RESULTATS

Cette partie va consister en une présentation des résultats issus des différentes méthodes évoquées dans la partie 2. Elles sont appliquées sur un certain nombre de villes afin de prévoir la date d'apparition d'un Risque d'Exposition Allergique aux Pollens (REAP) de frêne, de bouleau et de Poacées, ainsi que d'ambroisie pour Lyon. Avant de présenter, de manière générale, les résultats pour les sites ayant, au moins, 15 ans de données exploitables, nous allons tester la qualité des prévisions des différentes méthodes et, notamment, tenter de déterminer si l'une d'entre elles s'avère plus précise que les autres. Cette étude va porter sur les cinq sites de la moitié Nord de la France dont les séries démarrent avant 1990 : Paris, Rouen, Amiens, Nancy et Strasbourg. Auparavant, nous allons évoquer quelques opérations effectuées préalablement à l'analyse à proprement parler, et les variables météorologiques utilisées.

3.1 OPERATIONS SUR LES DONNEES POLLINIQUES ET PRESENTATION DES DONNEES METEOROLOGIQUES

3.1.1 La reconstitution des données polliniques manquantes

Ces opérations préalables consistent essentiellement à reconstituer les périodes de données manquantes quand celles-ci se situent en début de saison pollinique (à condition qu'elles ne soient pas trop longues) et risquent d'englober le démarrage de celle-ci ou, du moins, l'apparition du risque. Les concentrations à reconstituer dépendent d'un grand nombre de facteurs, si bien que l'opération est très délicate. Il serait plus judicieux d'éviter de le faire mais le démarrage trop tardif des capteurs certaines années – et donc une reconstitution totalement impossible – nécessite cette opération afin d'éviter de perdre trop d'années.

De par leur nature, les reconstitutions de ces valeurs ne peuvent être que sommaires et n'ont qu'un rôle indicatif. L'idéal serait une reconstitution des données d'une station à partir de celles d'une station voisine, comme c'est le cas pour les données météorologiques. Malheureusement, contrairement à ces dernières, la distance entre les sites de capture est trop importante pour effectuer cette opération, nous obligeant à nous rabattre sur d'autres procédés. Le fait d'employer des seuils bruts de risque ne nous contraint pas à effectuer des reconstitutions pour l'ensemble de la saison pollinique ; reconstitutions qui sont nécessaires si l'on emploie des seuils relatifs. Ainsi deux procédés ont été testés :

- **Une reconstitution à partir des données météorologiques par l'intermédiaire d'une régression linéaire multiple** : Les concentrations dépendant des conditions de précipitations, de température et d'humidité, il nous a semblé pertinent d'effectuer ces régressions. Le vent, qui joue un rôle majeur dans la dispersion des pollens, n'a pas été retenu car beaucoup trop complexe à mettre en équation. La distribution des concentrations n'étant pas gaussienne, il faut à priori effectuer une transformation logarithmique de ces données, sauf si l'on se réfère au théorème central limite qui stipule que l'augmentation de la taille d'un échantillon débouche sur une distribution approchant de la loi normale, même si la distribution de cette dernière ne l'est pas. Dans notre cas, pour des considérations pratiques, nous avons mis au point les modèles à partir des valeurs brutes de concentrations de la

première moitié de la saison pollinique du taxon étudié ; les concentrations de la fin de saison étant moins corrélées aux conditions météorologiques.

- **Une reconstitution en fonction des niveaux de risque :** L'utilisation de la régression linéaire multiple n'étant pas toujours pertinente (prévision de concentrations aberrantes par rapport à celles du reste de la saison), nous avons élaboré une méthode de reconstitution à partir des niveaux de risque et des données météorologiques. Il s'agit de calculer les conditions moyennes de température (minimale et maximale), d'humidité relative (minimale et maximale) et de précipitations de la période manquante, puis d'effectuer la même opération pour les jours associés à des niveaux de risque nul à faible, moyen, élevé et très élevé. Ensuite, grâce à une classification en nuées dynamiques, on cherche à déterminer la classe regroupant la période manquante et le niveau de risque dont les conditions météorologiques sont les plus proches statistiquement. Par exemple, si la période manquante présente des conditions météorologiques significativement équivalentes à celles enregistrées pour les jours à risque moyen d'une année donnée, les lacunes sont remplacées par la valeur-seuil associée à ce risque moyen. L'inconvénient de cette méthode tient en la répétition des mêmes valeurs, mais comme l'apparition du risque d'exposition allergique aux pollens s'effectue en fonction de celles-ci, cela ne pose pas de problème pour la détermination de la survenue du risque.

Les deux méthodes de reconstitution des données manquantes sont loin d'être idéales. Il est clair que l'existence actuelle d'un réseau plus dense de capteurs polliniques, si possible placés à des hauteurs équivalentes permettra d'élaborer, à la longue, des reconstitutions de meilleure qualité. A travers ces méthodes, et notamment celle se basant sur les seuils de risque, nous cherchons surtout à déterminer un ordre de grandeur. Les périodes manquantes reconstituées sont généralement courtes, si bien que la date estimée doit correspondre à celle qui aurait été effectivement déterminée s'il n'y avait pas eu de lacunes. Pour les périodes plus longues (de l'ordre d'une semaine), l'incertitude est plus grande et l'on n'utilise les valeurs reconstituées que si elles permettent de déterminer une date qui ne soit pas aberrante, notamment pour le bouleau dont la date de démarrage d'un site à l'autre ne présente pas une forte amplitude. Dans le cas contraire, on élimine l'année de l'analyse.

Dans le cas de la reconstitution d'une valeur manquante isolée (lorsque cela est nécessaire), ou si l'on souhaite déterminer les valeurs d'une période lacunaire ayant lieu pendant la deuxième moitié de la saison pollinique, l'utilisation d'une interpolation linéaire peut s'avérer suffisante. Cette technique est évidemment sujette à critique, mais les concentrations de fin de saison ne dépendent pas seulement des conditions météorologiques. Cela rend, par conséquent, l'application des deux autres méthodes difficile et risque de donner des concentrations aberrantes. Néanmoins si ces périodes sont trop longues, et/ou si elles sont liées à des concentrations potentiellement élevées, on n'effectue pas de reconstitutions de données.

D'une manière générale, il est impossible de fixer une seule méthode de reconstitution pour l'ensemble des lacunes. L'application se fait quasiment au cas par cas en fonction de la durée de la période manquante (critère qui influe peu sur la méthode employée), sa position par rapport à la période de pollinisation du taxon concerné, voire la durée de la saison pollinique. Dans le dernier cas, si la période manquante a lieu pendant la pleine saison et qu'elle est bornée par de faibles concentrations, la reconstitution est impossible. Nous vérifions, également, après coup si les concentrations prévues par une des méthodes sont réalistes par rapport à celles enregistrées avant et après la période lacunaire et, dans la négative, nous en employons une autre. Tout cela peut sembler un peu «artisanal» mais nous avons la volonté de tenter des reconstitutions qui ne soient pas seulement de simples interpolations linéaires, dans le but de «sauver» le plus d'années possibles.

Depuis peu le RNSA s'évertue à reconstituer les données manquantes d'une longueur de un à quatre jours. Cette reconstitution a été initiée à partir des travaux de C. Rieux (2006) et se décompose en trois catégories :

- Une interpolation linéaire simple à partir du jour précédent et du jour suivant la période manquante.
- Une interpolation linéaire en fonction des trois jours qui précèdent et suivent cette période lacunaire.
- Enfin, à l'aide d'un algorithme plus complexe, la reconstitution s'effectue à partir de la station géographiquement proche dont les résultats, avant et après l'interruption des données, sont les mieux corrélés avec les données de la station à

reconstituer. C'est, à priori, la meilleure méthode de reconstitution mais elle n'est exploitable que si le réseau de capteurs est dense, ce qui n'est le cas que dans la région lyonnaise avec les capteurs installés spécialement pour l'ambrosie et ne fonctionnant que pendant l'été.

3.1.2 Les données météorologiques utilisées

Ces données proviennent des stations météorologiques se situant, si possible, dans la même ville que les capteurs polliniques. Pour les méthodes de cumul, les températures minimales (**Tn**), maximales (**Tx**) et moyennes journalières (**Tm**) sont utilisées. La température moyenne est calculée en fonction de **Tn** et **Tx**.

Dans le cadre de l'utilisation de la régression linéaire multiple nous allons, en plus de la température de l'air, inclure d'autres variables météorologiques telles que :

- L'humidité relative minimale (**Hn**), maximale (**Hx**) et moyenne (**Hm**), cette dernière étant calculée à partir de **Hn** et **Hx**.
- Les précipitations (en mm) notées **P**.
- L'insolation (en minutes) notée **Insol**.
- La température mesurée dans le sol à – 10 cm (**T-10**) et – 20 cm (**T-20**). Les valeurs correspondent au relevé effectué à midi.
- La durée du gel (en min) notée **Gel**, qui correspond à la période où la température est inférieure ou égale à 0°C .

Ces trois dernières variables ne sont pas disponibles pour l'ensemble des villes retenues, surtout la durée du gel qui n'est utilisable que dans peu de cas. C'est pour cela que nous avons retenu le nombre de jours où la température minimale (**Tn0**), maximale (**Tx0**) et moyenne (**Tm0**) est inférieure à 0°C afin de tester l'influence de la durée d'une période de températures négatives sur la date d'apparition du risque. Certes, la définition par rapport à la valeur 0°C n'est pas la même entre les deux critères, mais les températures exactement égales à 0°C sont très rares. Cela a donc peu d'incidence, mais suppose d'employer le terme de températures négatives et non de gel.

Nous avons également déterminé le nombre total de jours de précipitations (**JP**), ceux où elles atteignent ou dépassent 1 mm (**JP1**) et 5 mm (**JP5**). Ces seuils sont ceux employés dans

l'Atlas climatique de la Côte-d'Or pour décrire les précipitations de ce département (Marceaux et Taboulot, 1994).

Comme cela est suggéré par l'emploi de ces deux derniers critères, nous ne nous basons pas sur les valeurs journalières pour élaborer les modèles de régression. Les articles utilisant cette méthode emploient des moyennes décadaires, mais nous avons préféré ici prendre un pas de temps bimensuel afin de limiter le temps de calcul et de nous baser sur des périodes plus longues et donc, à priori, plus représentatives des conditions pouvant influencer le développement de la plante jusqu'au démarrage de la saison pollinique.

Dans un but pratique, la durée de ces périodes bimensuelles diffère selon le mois considéré afin d'avoir des annotations lisibles dans les équations de régression et, surtout, pour pouvoir raisonner en terme de «quinzaine», expression couramment utilisée dans la vie quotidienne. Ainsi un mois de 30 jours est partagé en deux périodes de 15 jours ; la deuxième «quinzaine» d'un mois de 31 jours est réellement constituée de 16 jours. Le mois de février est partagé en deux périodes égales de 14 jours, excepté pour les années bissextiles où la deuxième moitié du mois comprend 15 jours.

Toutes les données météorologiques sont converties en moyennes. Pour les jours où la température est inférieure à 0°C, et pour ceux enregistrant des précipitations, nous avons préféré utiliser, pour une meilleure lisibilité, le total de ces jours. La seule exception concerne la deuxième «quinzaine» de février dont la durée est variable ; dans ce cas on n'utilise évidemment pas la somme, mais la fréquence d'apparition de ces paramètres.

Les périodes vont être identifiées en fonction du mois (**01** pour janvier, **10** pour octobre ...), **a** symbolise la première «quinzaine» et **b** la deuxième. Ainsi, par exemple, **T_x_04b** correspond à la moyenne de la température maximale de la deuxième «quinzaine» d'avril (16 au 30 avril).

Cette évocation globale des différents processus de reconstitution des données polliniques et des données météorologiques exploitées, dans le cadre d'une prévision de l'apparition du risque liée aux quatre taxons présentés dans le sous-chapitre 1.2.1, précède la présentation des résultats proprement dite.

3.2 PREVISION DU FRENE, DU BOULEAU, DES POACEES : LE CAS DE CINQ VILLES DE LA MOITIE NORD DE LA FRANCE

3.2.1 Choix des villes

Il s'est effectué non pas en fonction de considérations géographiques, mais en fonction de la disponibilité de séries de données relativement longues pour les trois taxons étudiés. Souhaitant retenir des sites ayant démarré avant 1990, seules cinq villes répondaient à ce critère. Deux d'entre elles, Amiens et Rouen, se caractérisent par un climat où l'influence océanique se fait ressentir – les deux villes se situent à une cinquantaine de kilomètres du littoral le plus proche – avec des hivers (décembre-janvier-février) relativement doux (des températures moyennes de 3,8 et 3,7°C respectivement sur la période 1961-1990 ; pour Amiens nous avons utilisé les normales d'Abbeville) et des étés (juin-juillet-août) assez frais (16 et 16,3°C respectivement). Nancy et Strasbourg présentent un caractère beaucoup plus continental avec des hivers plus rigoureux (2 et 1,8°C respectivement) et des valeurs de températures estivales supérieures aux deux villes précédentes à Nancy (17,2°C) et plus encore à Strasbourg (18,2°C). Enfin, Paris se caractérise par sa position géographique intermédiaire par rapport au littoral de la Manche, et par des températures (4,8°C en hiver et 18,7°C en été) influencées en partie par l'îlot de chaleur urbain. L'insolation annuelle dans ces villes est relativement faible allant de 1637,9 heures à Abbeville à 1687,2 heures à Rouen, seule Paris se distingue par un ensoleillement meilleur (1797,5 heures). Quant aux précipitations annuelles elles oscillent de 610,5 mm à Strasbourg à 785,2 mm à Rouen. On peut, notamment, expliquer la différence enregistrée entre Nancy (759,3 mm) et Strasbourg par l'effet d'abri généré par le massif des Vosges.

Pour information, nous précisons que les données météorologiques utilisées proviennent des sites de mesure suivants :

- Amiens-Glisly
- Rouen-Boos
- Paris-Montsouris
- Nancy-Essey
- Strasbourg-Entzheim

Nous allons comparer, taxon par taxon, les différentes méthodes de cumul d'une part, puis étudier les résultats issus de la régression linéaire multiple d'autre part.

Dans le cadre des différentes comparaisons nécessaires qui sont effectuées dans ce chapitre, nous employons le test t de Student pour un seuil de significativité de 95% aussi bien pour le frêne, le bouleau et les Poacées. En effet, la quasi-totalité des différentes séries testées, 193 sur 198, se caractérisent par une distribution normale selon les tests de Shapiro-Wilk et de Kolmogorov-Smirnov (pour un risque d'erreur de 5%), selon les modalités définies dans la section 1.1.5.3. Cela justifie le choix de s'en tenir à ce seul test de comparaison d'autant plus que celui non-paramétrique de Mann-Whitney, appliqué à titre expérimental, ne contredit pas les résultats du test de Student.

3.2.2 Prévision du frêne par les méthodes de cumul

Afin de mettre au point une comparaison précise, nous allons retenir les années communes aux cinq sites ; c'est-à-dire celles dont l'utilisation des données polliniques de frêne est réalisable. Les séries de données démarrant, au moins, en 1988, nous espérons pouvoir utiliser 17 à 18 ans de données. Malheureusement, le démarrage trop tardif des capteurs certaines années nous oblige à en retirer un certain nombre. Le site de Strasbourg est le plus touché avec les années 1988 à 1991 et 1996 qui sont concernées par ce problème. Les très faibles quantités enregistrées certaines années ne permettent pas toujours d'atteindre le seuil de risque escompté ; c'est notamment le cas d'Amiens en 1987 et 1988. De ce fait nous n'avons pu retenir que 11 années : 1992 à 1995, 1997, 1999 et 2001 à 2005.

3.2.2.1 Présentation des dates d'apparition et de l'application des méthodes

Il faut rappeler que la date d'apparition d'un risque d'exposition allergique au pollen de frêne a lieu dès la survenue le jour J d'au moins 20 grains/m³, avec un total cumulé d'au moins 60 grains de J à J+2.

Les dates moyennes d'apparition d'un risque allergique, classées par ordre croissant, sont les suivantes :

- 75,1^{ème} jour de l'année à Paris, soit le 16 mars pour une année non bissextile (écart-type de 7,9 jours).
- 79,2^{ème} jour de l'année à Strasbourg (écart-type de 7,7 jours).
- 80,6^{ème} jour de l'année à Amiens (écart-type de 7,9 jours).
- 82,5^{ème} jour de l'année à Nancy (écart-type de 9,5 jours).
- 82,9^{ème} jour de l'année à Rouen, soit le 24 mars pour une année non bissextile (écart-type de 10,4 jours).

Paris se caractérise par l'apparition moyenne du risque la plus précoce, ce qui semble être en phase avec les températures hivernales qui sont les plus douces des cinq villes. Les écarts-types traduisent une forte variabilité interannuelle des dates, l'amplitude maximale étant de 20 jours à Paris, 22 à Strasbourg, 25 à Amiens, 30 à Nancy et 33 à Rouen. La distribution de ces dates suit la loi normale d'après les résultats, concordants ici, des tests de Shapiro-Wilk et de Kolmogorov-Smirnov montrant, entre autres, que les valeurs extrêmes sont rares.

Une Analyse en Composantes Principales (ACP), appliquée sur les cinq séries de dates, permet de montrer que leur principale caractéristique (première composante) réside dans une variabilité interannuelle commune des dates d'apparition du risque. Deux années se démarquent par leur précocité : 1997 surtout (67,2^{ème} jour) et 2002 (72^{ème} jour). A l'opposé les années 1995 et 2005 se distinguent par les dates d'apparition globalement les plus tardives : 90,6^{ème} et 88,6^{ème} jours de l'année respectivement.

A partir de ces variables et des différentes méthodes de cumul présélectionnées, nous allons élaborer des modèles de prévision de l'apparition de ce risque. Nous allons, tout d'abord, rechercher les combinaisons entre les dates de démarrage de cumul et les seuils de température, les coefficients ou, selon les méthodes, le type de température qui minimisent l'erreur moyenne absolue. Précisons que les valeurs moyennes de cumul des unités de «chaleur», associées à ces modèles, se trouvent systématiquement à l'intersection de deux jours. De ce fait les dates prévues par les différents modèles correspondent aux jours précédant immédiatement ceux où le cumul moyen est dépassé. Ce choix d'application concernera tous les modèles de cumul présentés dans la suite de ce travail.

Pour les deux méthodes de *Growing Degree Days*, la méthode *Fraction-Time* et celle de Lejoly-Gabriel, nous testons les seuils d'action de température de 0 à 20°C, par pas de 0,5°C. Pour la méthode du Q₁₀, nous testons les coefficients de développement de 0,1 en 0,1, de 1 à 12. Enfin, pour la méthode de Sarvas, nous testons les températures minimale, maximale et

moyenne en remplaçant les unités de «chaleur» issues de températures négatives par 0 d'une part, et en les conservant telles quelles d'autre part. Parallèlement à cela, nous allons tester toutes les dates de début de cumul du 1^{er} octobre jusqu'au 28 février excepté à Paris où la date d'arrêt du test a lieu le 27 février du fait de l'apparition du risque le 28 février en 1997. Grâce à un programme élaboré sous Matlab® au Centre de Recherches de Climatologie de Dijon par Pascal Oettli, doctorant en climatologie et ingénieur informaticien, nous allons pouvoir, de manière quasi automatique, déterminer la combinaison optimale propre à chaque méthode. Avant d'exposer les résultats, nous allons présenter les sigles et abréviations affectés à chaque méthode :

- La méthode des GDD par triangulation : **GDD_1**.
- Celle employant une fonction sinusoïdale : **GDD_2**.
- La méthode de Lejoly-Gabriel à partir de la température maximale : **Lej_Tx**.
- La méthode FT : **FT**.
- La méthode du Q_{10} à partir du critère de Bidabe (1967) : **Q₁₀**.
- La méthode de Sarvas : **Sarvas**.

3.2.2.2 Présentation des meilleurs modèles

Nous allons présenter, pour chaque ville et dans l'ordre, la date de démarrage du cumul, le seuil ou le type de données retenu, l'erreur moyenne absolue (MAE), l'erreur moyenne au carré (RMSE) et le cumul moyen correspondant d'unités de «chaleur» à approcher le plus possible (cf. 3.2.2.1). La plus faible erreur moyenne absolue par ville est notée en gras, en cas d'ex aequo n'est surlignée que celle qui a la plus faible erreur moyenne au carré. Pour la méthode de Sarvas l'utilisation de la température minimale, maximale et moyenne, en tenant compte des unités issues des valeurs négatives, est symbolisée par Tn, Tx et Tm respectivement. Dans le cas de figure où l'on ne tient pas compte des unités issues de températures négatives les symboles sont Tn(0), Tx(0) et Tm(0), zéro étant la valeur affectée aux unités en question.

- **Amiens :**

- **GDD_1** : 27 février – 7,5°C – 3,09 jours – 4,05 jours – 37,30°D
- **GDD_2** : 26 février – 6,5°C – 3,73 jours – 5,18 jours – 55,23°D
- **Lej_Tx** : 3 février – 12,5°C – 3,91 jours – 5,27 jours – 172,99°C
- **FT** : 26 février – 10,5°C – **2,36 jours** – 3,16 jours – 5,46 unités
- **Q₁₀** : 3 février – 4,1 et 4,2 – 3,91 jours – 5,71 jours – 299,27 et 306,65 unités
- **Sarvas** : 28 février – Tn – 3,64 jours – 4,11 jours – 45,57 unités

- **Rouen :**

- **GDD_1** : 30 décembre – 11°C – 3,18 jours – 4,46 jours – 11,82°D
- **GDD_2** : 31 décembre et 1^{er} janvier – 11,5°C – 3,18 jours – 4,58 jours – 11,86°D
- **Lej_Tx** : 6 janvier – 11,5°C – 4,36 jours – 5,10 jours – 248,85°C
- **FT** : 6 janvier – 13°C – **3,09 jours** – 4,45 jours – 2,23 unités
- **Q₁₀** : 30 décembre – 12 – 5,45 jours – 6,76 jours – 1368,80 unités
- **Sarvas** : 8 janvier – Tn(0) – 6,45 jours – 10,54 jours – 119,62 unités

- **Paris :**

- **GDD_1** : 13 décembre – 10°C – 2,45 jours – 4,01 jours – 36,81°D
- **GDD_2** : 16 et 17 décembre – 10,5°C – 2,27 jours – 3,80 jours – 32,69 et 32,68°D
- **Lej_Tx** : 29 décembre – 12°C – **2,18 jours** – 2,86 jours – 264,69°C
- **FT** : 16 au 18 décembre – 11°C – 2,64 jours – 3,87 jours – 11,68 et 11,64 unités
- **Q₁₀** : 19 décembre – 10,3 – 4,00 jours – 5,69 jours – 1443,80 unités
- **Sarvas** : 27 février – Tm(0) – 4,36 jours – 5,49 jours – 82,12 unités

- **Nancy :**

- **GDD_1** : 13 février – 5°C – 3,73 jours – 5,09 jours – 88,55°D
- **GDD_2** : 17 février – 4°C – 3,91 jours – 5,50 jours – 111,33°D
- **Lej_Tx** : 27 janvier – 10°C – **3,73 jours** – 4,85 jours – 321,72°C
- **FT** : 20 février – 6,5°C – 3,91 jours – 5,39 jours – 15,21 unités
- **Q₁₀** : 27 janvier – 7,7 – 3,73 jours – 5,07 jours – 693,94 unités
- **Sarvas** : 20 février – Tm(0) – 4,00 jours – 5,43 jours – 101,63 unités

- **Strasbourg :**

- **GDD_1** : 14 décembre – 6,5°C – 2,09 jours – 3,61 jours – 93,92°D
- **GDD_2** : 14 décembre – 7°C – **2,00 jours** – 3,86 jours – 88,94°D
- **Lej_Tx** : 27 janvier – 11°C – 2,64 jours – 3,80 jours – 253,65°C
- **FT** : 13 décembre – 10°C – 2,36 jours – 4,26 jours – 11,39 unités
- **Q₁₀** : 25 janvier – 3,4-3,6-3,7 – 2,64 jours – 3,85 jours – 264,28-279,29-286,91 unités
- **Sarvas** : 28 et 29 janvier – Tm – 2,91 jours – 3,88 jours – 140,91 et 139,17 unités

Malgré l'emploi de l'erreur moyenne au carré destiné à départager d'éventuelles combinaisons présentant les mêmes erreurs moyennes absolues, on retrouve pour certaines méthodes et pour certaines villes plusieurs combinaisons ayant, à la fois, des valeurs de MAE et de RMSE identiques.

Nous pouvons remarquer que la qualité moyenne des modèles issus des six méthodes diffère selon les villes ; l'erreur moyenne absolue de l'ensemble des six villes est de 3,40 jours. Strasbourg et Paris présentent, en moyenne globale, les modèles les plus précis avec une erreur moyenne de 2,44 et 2,98 jours respectivement. Les modèles appliqués à Strasbourg se caractérisent d'ailleurs par une erreur moyenne significativement plus faible, d'après le test de Student, que celles des modèles d'Amiens (3,44 jours), Nancy (3,84 jours) et Rouen (4,29 jours). Les dates optimales de démarrage de cumul sont variables d'une ville à l'autre. Elles ont toutes lieu en février à Amiens, alors qu'à Paris cinq des six dates optimales de démarrage ont lieu en décembre. A Rouen et Strasbourg, elles se partagent entre décembre et

janvier et entre janvier (pour le Q_{10}) et février à Nancy. Quant aux seuils, ils diffèrent d'une ville à l'autre pour une méthode donnée et inversement. Nous pouvons cependant constater une certaine homogénéité des seuils de température déterminés à partir de la méthode de Lejoly-Gabriel qui oscillent de 10°C à Nancy à 12,5°C à Amiens, malgré la disparité des dates de début de cumul. Nous notons également que les seuils d'action et les dates de démarrage des modèles issus des méthodes GDD_1 et GDD_2 sont très proches, faisant ressortir leur genèse commune.

Les méthodes à l'origine des meilleurs modèles diffèrent d'une ville à l'autre. Ainsi à Amiens et Rouen c'est la méthode FT qui donne les meilleurs modèles, alors qu'à Paris et Nancy c'est celle de Lejoly-Gabriel. La méthode GDD_2 fournit le modèle avec l'erreur moyenne la plus faible à Strasbourg. Suite à ce constat de forte variabilité, nous allons déterminer statistiquement si une des méthodes est globalement meilleure que les autres.

3.2.2.3 Comparaisons inter-méthodes et intermensuelles

Nous allons, tout d'abord, présenter graphiquement l'erreur moyenne absolue liée à chaque méthode correspondant à la moyenne des erreurs des meilleures combinaisons des cinq villes de l'étude (figure 35).

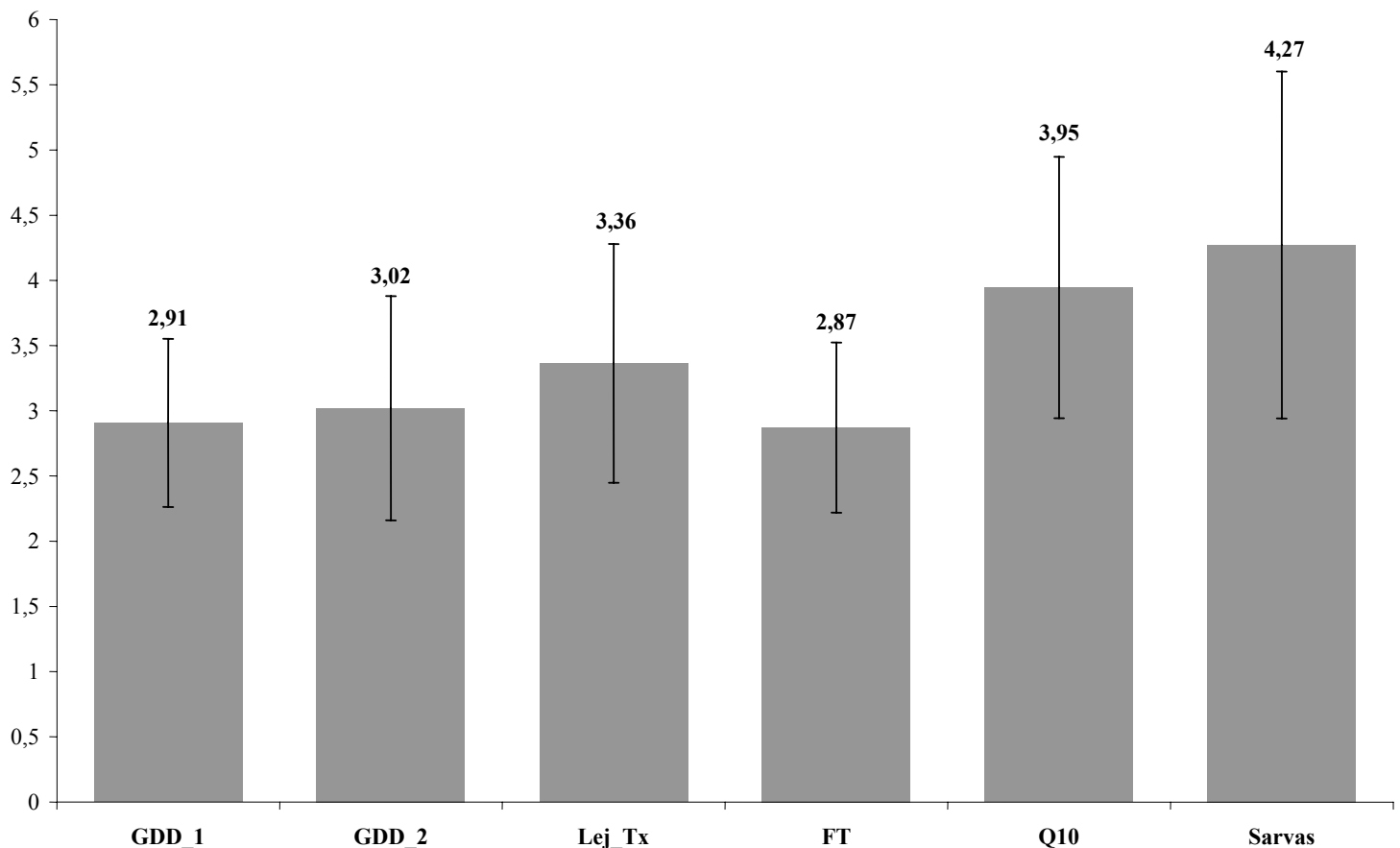


Figure 35 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour le frêne.

Les moyennes s'échelonnent de 2,87 jours pour la méthode FT à 4,27 jours pour celle de Sarvas. Malgré le fait que les deux méthodes «exponentielles» présentent une valeur moyenne assez nettement plus élevée que les autres, aucune différence significative n'est enregistrée entre les six modèles par l'intermédiaire du test de Student. Cela est dû au fait que ces deux méthodes présentent des modèles de précision équivalente aux quatre autres pour Amiens, Nancy et Strasbourg. Par ailleurs, nous pouvons noter que la méthode de Lejoly-Gabriel, la plus simple de conception, donne des résultats aussi fiables que les autres, ce qui semble montrer que la qualité d'une méthode ne réside pas dans sa complexité.

De ces six méthodes présélectionnées, nous n'allons en retenir que cinq pour la suite de l'analyse. En effet, les méthodes GDD_1 et GDD_2 étant très voisines en termes de processus d'élaboration, nous n'allons retenir que la plus simple : GDD_1. La justification de ce choix découle des résultats du test t de Student, et du fait que la complexité de la formule de la méthode GDD_2, quand le seuil testé est compris entre la température minimale et maximale, présente un risque important d'erreur de transcription. Notons seulement que dans ce dernier cas de figure, à températures identiques, les degrés-jours issus de GDD_2 sont très légèrement

plus élevés que ceux déterminés à partir de GDD_1. Ces degrés-jours étant liés à la surface comprise entre la courbe de température et le seuil testé, nous en déduisons que la surface associée à la courbe sinusoïdale est plus étendue que celle du triangle équilatéral de GDD_1.

Comme cela a été brièvement évoqué, nous avons constaté que les dates de démarrage sont très variables avec une bonne représentation des mois de décembre, janvier et février et, par conséquent, aucune date de démarrage en octobre et en novembre n'apparaît comme optimale pour aucune des méthodes.

Afin de déterminer si un mois se distingue particulièrement au niveau de la qualité des modèles qui en découlent, nous allons comparer les erreurs moyennes absolues de l'ensemble des méthodes ville par ville, mais aussi mois par mois (d'octobre à février), en fonction des combinaisons date-seuil minimisant les écarts entre valeurs observées et prévues pour chacun des mois en question. Nous faisons de même ensuite avec les erreurs moyennes de l'ensemble des villes méthode par méthode. Les résultats, ville par ville, sont présentés dans la figure 36 :

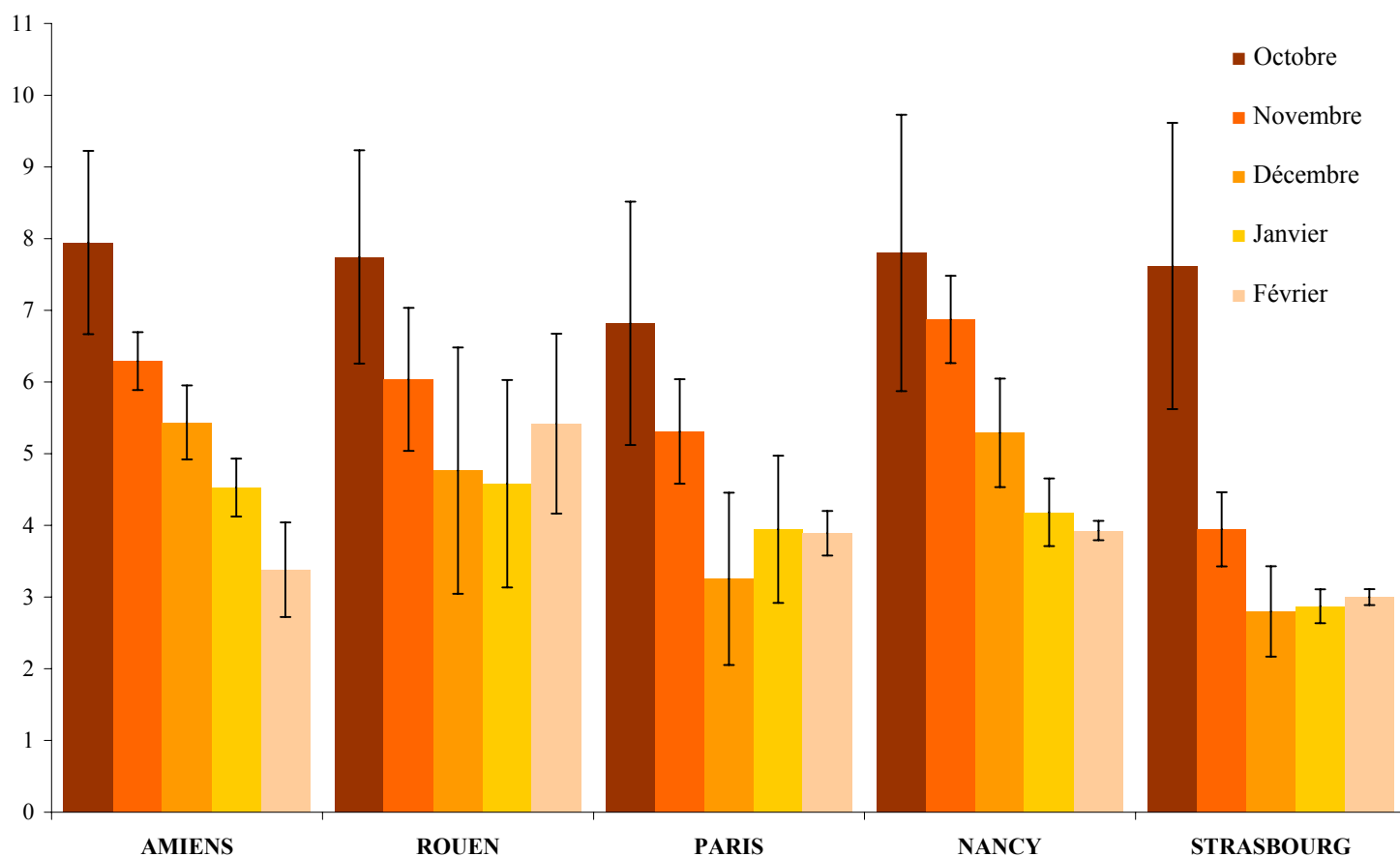


Figure 36 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour le frêne.

Nous allons analyser les résultats, ville par ville, à partir du test de Student appliqué selon un risque d'erreur de 5% :

- **Amiens** : L'erreur moyenne absolue décroît régulièrement en allant d'octobre à février en passant de 7,95 jours en octobre à 3,38 jours en février. De mois en mois, nous enregistrons une erreur moyenne absolue significativement plus faible selon le test t de Student. Ce résultat est donc conforme aux dates de début de cumul des modèles d'Amiens.
- **Rouen** : Pour cette ville, on peut remarquer une opposition entre les mois de décembre, janvier et février qui présentent des erreurs moyennes qui ne sont pas significativement différentes (4,76, 4,58 et 5,42 jours respectivement), et le mois d'octobre où la moyenne des erreurs absolues est significativement plus élevée. Le mois de novembre est particulier par son absence de différence significative avec les quatre autres mois. Cela est en partie dû à la valeur relativement élevée des écarts-types. Dans le cas présent, les dates de démarrage de cumul à l'origine des modèles les plus efficaces ont lieu en décembre et janvier, confirmant le constat effectué précédemment.
- **Paris** : Nous constatons quasiment le même cas de figure qu'à Rouen, à savoir que les erreurs moyennes absolues associées aux mois de décembre à février sont statistiquement équivalentes et qu'elles sont significativement plus faibles que pour octobre et novembre. Le mois de décembre présente la plus faible erreur des trois mois avec une valeur de 3,25 jours contre 3,95 en janvier et 3,89 en février.
- **Nancy** : Comme à Amiens, l'erreur moyenne absolue de l'ensemble des méthodes décroît au fur et à mesure que l'on se rapproche de la date de démarrage du frêne passant de 7,80 jours en octobre à 3,93 jours en février. Selon le test t de Student, pour un seuil de significativité de 95%, il y a des différences significatives entre tous les mois excepté entre octobre et novembre d'une part (à cause surtout de l'important écart-type enregistré en octobre) et janvier et février d'autre part. Il apparaît alors que pour cette ville il est plus pertinent de commencer les cumuls de température en janvier ou février, confirmant les résultats déterminés auparavant.

- **Strasbourg** : D'après le test t de Student, on enregistre un cas de figure strictement identique à celui de Paris où les mois de décembre à février (avec, respectivement des valeurs de 2,80, 2,87 et 3,00 jours) présentent des erreurs moyennes significativement plus faibles que les deux autres mois.

D'une manière générale, nous constatons que le démarrage des cumuls en octobre et novembre débouche sur des résultats médiocres. Le fait que ces deux mois correspondent, à priori, à la période de dormance du frêne peut expliquer ces résultats par le fait que l'action de la température n'est pas identique à celle de la période de développement précédant la pollinisation. Pour Rouen, Paris et Strasbourg, il n'y a pas de différences tangibles à propos de la qualité des modèles entre les mois de décembre, janvier et février. Ce qui n'est pas le cas pour Nancy et, plus encore, pour Amiens où les meilleurs modèles sont ceux dont le démarrage a lieu courant février. Pour Amiens, l'adéquation développement/température semble plus forte en ne tenant compte que des températures de février, qu'en faisant intervenir, en plus, celles des deux mois précédents. Par contre pour Rouen, Paris et Strasbourg, les modèles dont les cumuls sont initiés en décembre s'avèrent aussi bons que ceux initiés plus tard. La nette amélioration des résultats par rapport à novembre trouve, peut-être, son explication dans le fait que l'accomplissement de la dormance a lieu durant la première moitié de décembre. La qualité équivalente des modèles dont les cumuls sont initiés de décembre à février permet, dans un but préventif, de tester plusieurs modèles initiés à des dates différents au fur et à mesure de l'hiver.

Nous allons maintenant analyser les résultats, méthode par méthode mais tous sites confondus, avec, tout d'abord, la présentation graphique (figure 37) des erreurs moyennes absolues :

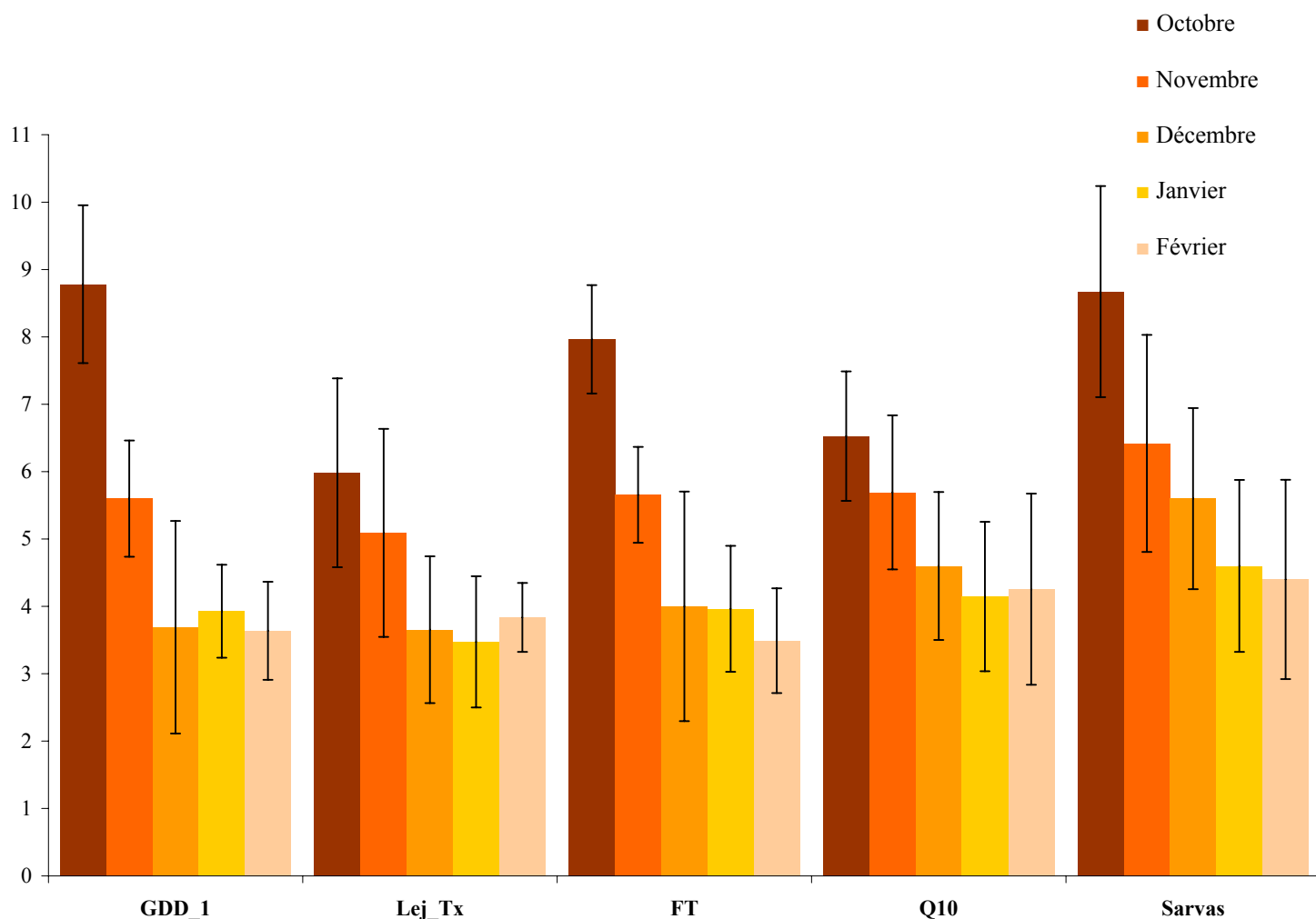


Figure 37 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour le frêne.

Nous allons analyser les résultats à partir du test de Student appliqué selon un risque d'erreur de 5% :

- **GDD_1** : Les erreurs moyennes absolues associées aux mois de décembre à février, statistiquement équivalentes, sont significativement plus faibles que celles enregistrées en octobre et novembre. Les valeurs moyennes de ces trois mois sont, dans l'ordre, de 3,69, 3,93 et 3,64 jours.
- **Lej_Tx** : Dans le cas de cette méthode, les mois de novembre à février conduisent à des erreurs moyennes absolues statistiquement équivalentes et la différence entre octobre et novembre n'est pas significative non plus. A l'opposé de la méthode GDD_1, l'erreur moyenne enregistrée en octobre est «relativement» faible (5,98 jours), mais l'amélioration au fil des mois se révèle être plus lente et débouche, au

final, sur des modèles qui ne sont pas plus précis (de 3,47 jours en janvier à 3,84 jours en février).

- **FT** : Les relations statistiques entre les différents mois sont très proches de celles de la méthode GDD_1. Les mois de décembre à février sont ceux qui fournissent, en moyenne, les meilleurs modèles. La différence la plus importante enregistrée par rapport à la méthode GDD_1 concerne l'erreur moyenne d'octobre un peu plus faible (7,96 jours).
- **Q₁₀** : Nous nous retrouvons, d'un point de vue statistique, dans le même cas de figure que pour la méthode de Lejoly-Gabriel, avec une amélioration lente de la qualité des modèles au fil des mois. D'une valeur comparativement basse en octobre (6,53 jours), on aboutit à une erreur minimale finalement plus élevée (4,15 jours en janvier) que celles des quatre méthodes précédentes.
- **Sarvas** : Les relations statistiques entre les différents mois sont strictement identiques à celles enregistrées pour les méthodes de Lejoly-Gabriel et du Q₁₀. Les erreurs moyennes, de 8,67 jours en octobre à 4,40 en février, y sont par contre plus élevées.

L'étude méthode par méthode confirme les résultats enregistrés ville par ville. Deux groupes se distinguent ici :

- Les méthodes GDD_1 et FT se caractérisent par une forte amplitude entre les erreurs moyennes d'octobre et celles de décembre à janvier.
- Les trois autres se distinguent par une amplitude maximale des erreurs mensuelles plus réduite (Lej_Tx et Q₁₀) et, pour la méthode de Sarvas, par les valeurs les plus élevées de novembre à février.

3.2.2.4 Bilan global

En guise de bilan, nous allons présenter les erreurs moyennes absolues mensuelles des modèles issus de l'ensemble des méthodes et des villes (figure 38) :

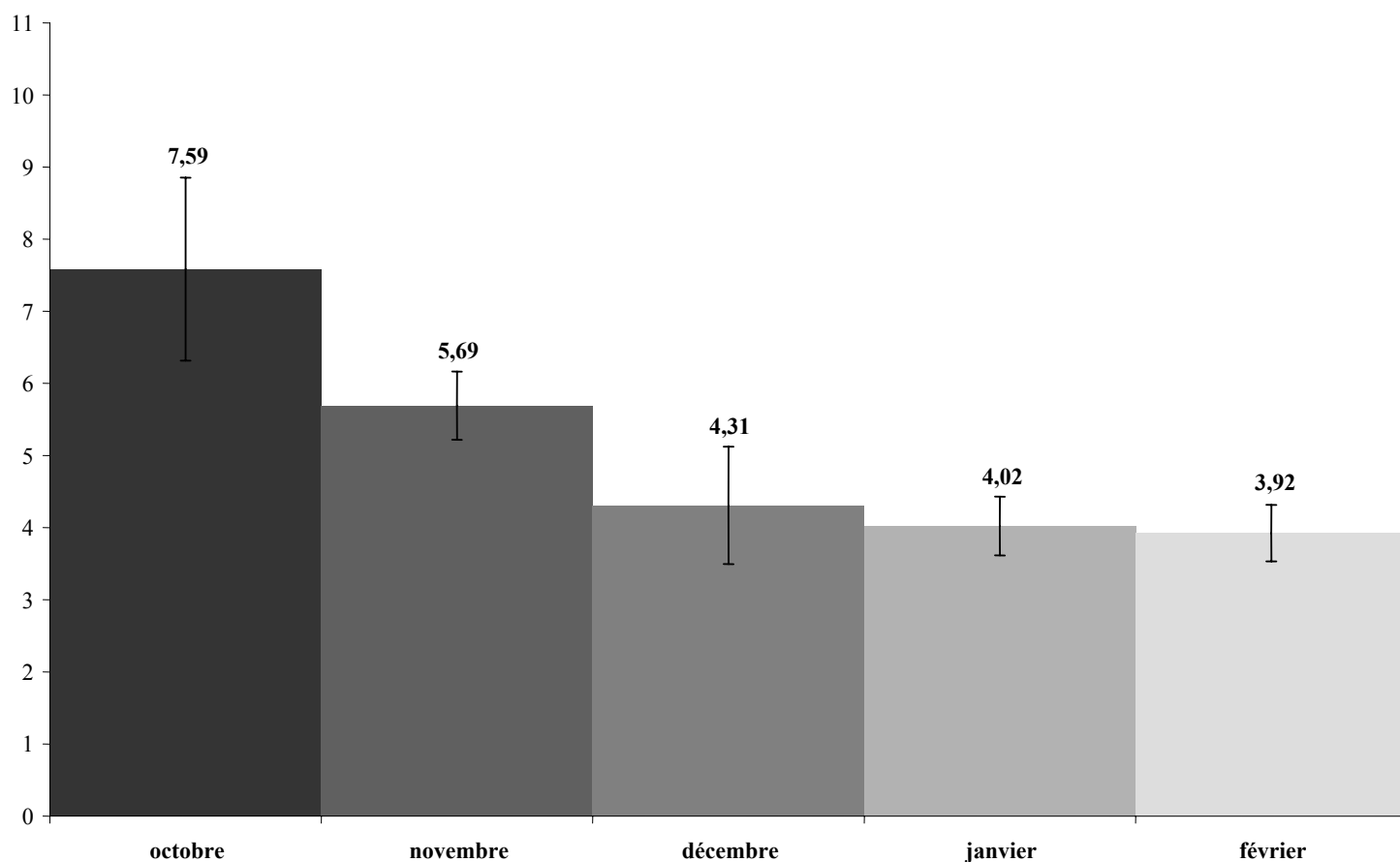


Figure 38 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour le frêne.

Sur l'ensemble des cinq villes, et d'après le test t de Student, les erreurs moyennes absolues associées aux mois de décembre, janvier et février sont équivalentes. Un démarrage assez précoce des cumuls débouche sur des modèles aussi fiables que ceux initiés plus tardivement, dont nous pouvons supposer que la courte période de cumul limiterait la divergence des prévisions. Cependant les séries d'Amiens et Nancy n'ont pas répondu à cette tendance. Par contre le démarrage des cumuls en novembre et, plus encore, en octobre s'est avéré systématiquement inapproprié avec une erreur moyenne de 7,59 jours pour ce dernier. Nous pouvons supposer que la période de dormance du frêne, accomplie à priori en décembre, a lieu pendant ces deux mois, et que le rôle de la température sur le développement de l'arbre

est différent de celui joué durant la période de forçage ou d'accumulation d'unités de «chaleur» (cf. 3.2.2.3).

3.2.2.5 Etude des résidus

La normalité de la distribution des résidus des différents modèles exposés ci-dessus (cf. 3.2.2.2) est testée à l'aide des tests de Shapiro-Wilk et Kolmogorov-Smirnov. Seuls les résidus des meilleurs modèles issus de la méthode de Sarvas à Rouen, des méthodes FT et de Lejoly-Gabriel à Nancy et des méthodes FT et GDD_1 à Strasbourg ne sont pas distribués normalement. Expliquer le caractère non gaussien de la distribution de ces résidus n'est pas évident, car la significativité ou la non-significativité de ces tests peut évoluer en ajoutant ou en retirant une observation, du fait de la faible taille de l'échantillon.

La précision des modèles liés au frêne s'avère assez bonne mais nous constatons, à travers les résidus, des années où la date d'apparition du risque est très mal estimée notamment pour les modèles issus des méthodes de Sarvas et du Q_{10} . En guise d'exemple, nous allons présenter, ville par ville et méthode par méthode, les années où la valeur absolue des résidus atteint ou dépasse deux semaines (14 jours). Les valeurs présentées correspondent à la relation suivante : **valeur prévue – valeur observée**. Les résidus sont généralement déterminés par la relation **valeur observée – valeur prévue** ; nous avons délibérément fait l'opposé pour que la lecture se fasse du point de vue de la date prévue. Cela permet de voir aisément si la prévision est trop tardive (résidu positif) ou trop précoce (résidu négatif).

- **Amiens :**
 - Q_{10} : – 14 jours (1995)

- **Rouen :**
 - Q_{10} : – 14 jours (1995)
 - **Sarvas** : – 30 jours (1995)

- **Nancy :**
 - **FT** : + 14 jours (2004)

Toutes les villes ne sont pas concernées par ce seuil de 14 jours mais toutes le sont par la présence d'au moins un résidu dont la valeur oscille de 10 à 13 jours. Nous allons tenter d'expliquer les causes de ces résidus.

- **Amiens** : Le meilleur modèle issu de la méthode du Q_{10} donne une prévision de 14 jours trop précoce en 1995. L'apparition du risque cette année-là a lieu le 5 avril, soit la date la plus tardive de l'échantillon de 11 années, alors que la date prévue tombe un 22 mars. Cela semble s'expliquer par les températures de février qui sont les plus douces, avec celles de 2002, et qui auraient dû être à l'origine d'un démarrage plus précoce.
- **Rouen** : Les modèles issus de la méthode du Q_{10} et de Sarvas, plutôt médiocres pour cette ville, fournissent une prévision beaucoup trop précoce de la date d'apparition du risque en 1995 (- 14 et -30 jours respectivement). Cette année se caractérise, comme à Amiens, par la date d'apparition du risque la plus tardive de la série (11 avril). La date prévue par le modèle issu de la méthode de Sarvas intervient le 12 mars, et le 28 mars pour celle du Q_{10} . L'absence d'adéquation entre les températures de février, douces, et la date d'apparition du risque tardive semble expliquer, en partie, la divergence entre les valeurs observées et prévues et notamment celle, conséquente, enregistrée pour la méthode de Sarvas. Cela confirme que l'utilisation de ces deux méthodes est peu judicieuse pour la prévision du frêne à Rouen.
- **Nancy** : La date d'apparition du risque allergique d'exposition au pollen de frêne en 2004 est trop tardivement prévue par le modèle issu de la méthode FT. La date d'apparition observée, 19 mars, est assez proche de la moyenne des 11 années de l'échantillon. Cela semble s'expliquer par les températures de la première «quinzaine» de mars et de la fin février, en dessous des valeurs moyennes des 11 années, à partir desquelles nous aurions dû nous attendre à une apparition effective du risque plus tardive.

A travers ces quelques exemples, les deux causes principales d'une mauvaise estimation des dates d'apparition du risque ressortent. Tout d'abord les dates d'apparition du risque mal

prévues par les méthodes de cumul font partie des plus tardives des échantillons étudiés, mais aussi des plus précoces avec des résidus atteignant, assez souvent, les 10 jours en valeur absolue. La prévision est trop précoce dans le cadre d'une apparition effective tardive et inversement ; cela est lié principalement à la non-adéquation théorique des dates d'apparition du risque aux conditions de température enregistrées auparavant.

Le cas de l'année 2004 à Nancy met en relief le phénomène opposé, à savoir qu'une année présentant une date d'apparition du risque «normale» peut être mal estimée si les conditions de températures la précédant sont remarquables (périodes anormalement froides ou chaudes), et cela pour les mêmes raisons que précédemment. Dans l'absolu, si l'on souhaite améliorer la qualité des modèles, il conviendrait de retirer ces années. Mais quel serait alors l'intérêt d'un modèle qui ne soit basé que sur des années «moyennes», alors que le but principal est qu'il puisse prévoir des années exceptionnelles ? De plus, les séries étant tellement courtes, il serait malvenu de se priver de l'information donnée par une année exceptionnelle.

3.2.3 Prévision du frêne par la régression linéaire multiple

Cette méthode a pour principale caractéristique de tenir compte d'autres variables que les seules températures de l'air. Toutes les variables détaillées dans le sous-chapitre 3.1.2 sont utilisées pour les sites de Rouen, Paris, Nancy et Strasbourg, excepté la durée du gel qui n'est exploitable qu'à Paris. La station météorologique d'Amiens, moins complète, ne mesure pas l'insolation, la température du sol et la durée du gel.

3.2.3.1 Etude des corrélations

Nous allons tout d'abord rechercher les coefficients de corrélation significatifs, pour un risque d'erreur de 5%, existant entre les valeurs bimensuelles des différentes variables météorologiques et les dates d'apparition du risque. Les périodes testées s'échelonnent de la première «quinzaine» d'octobre à la deuxième «quinzaine» de février (dont l'élaboration a été présentée dans le sous-chapitre 3.1.2) avec la prise en compte de la période allant du 15 au 27 février, particulière à Paris. Nous allons présenter, à travers le tableau 9, les résultats «quinzaine» par «quinzaine» afin de déterminer si plusieurs variables sont significativement

corrélées avec les dates d'apparition du risque de plusieurs villes. Les abréviations sont celles du sous-chapitre 3.1.2

	AMIENS	ROUEN	PARIS	NANCY	STRASBOURG
10a	Tn0 JP			Tm	
10b					
11a	JP			Hx Hm	
11b			JP5		
12a		T-20			
12b		Hn Hm	Tx0		
01a		T-20 Tx0 Tm0	Tn Tx Tm T-10 T-20 Tn0 Tx0 Tm0 Gel	Hn	Tx Hn Hm
01b	Hn Hm	Hx	JP5	T-10 T-20 Hn Hm	
02a				P JP5	P
02b			Tx	Tx Tx0	

Tableau 9 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique au pollen de frêne, sur les 11 années communes. En rouge = corrélation positive ; en bleu = corrélation négative ; en gras = variables communes à, au moins, deux villes

Nous avons également déterminé les corrélations partielles issues des régressions pas à pas ascendantes élaborées pour chaque période bimensuelle, la signification des couleurs est identique à celle du tableau 9 :

- **Amiens** : pas de corrélations
- **Rouen** : Tn_01a, T-10_01a
- **Paris** : pas de corrélations
- **Nancy** : T-20_10a, Hx_01a, Tx0_01a
- **Strasbourg** : Hm_02a

Nous pouvons constater la diversité des variables corrélées et, surtout, que la plupart de ces variables ne sont corrélées qu'avec les dates d'une seule ville. Cela est certainement dû à la faible longueur des séries qui est à l'origine d'un seuil élevé de significativité des corrélations, empêchant d'avoir un plus grand nombre de valeurs significatives. Cela montre, en tout cas, la nécessité d'élaborer les modèles ville par ville.

Au mieux, nous enregistrons des variables corrélées significativement avec les dates d'apparition du risque de deux villes et, ce qui est encourageant, le signe des corrélations est alors identique.

Les périodes où le plus grand nombre de corrélations est enregistré sont les deux «quinzaines» de janvier. Nous pouvons ainsi noter l'influence retardatrice des températures de la première «quinzaine» de janvier à Strasbourg, Paris et Rouen. Plus celles-ci sont élevées (ou plus le nombre de jours enregistrant des températures négatives est faible), plus l'apparition du risque est tardive. A partir de ces résultats, il semblerait, pour ces trois villes, que la température ne joue pas encore son rôle de «forçage» sur le développement du frêne...ce qui est contradictoire avec les résultats des méthodes de cumul montrant, dans plusieurs cas, que les meilleurs modèles sont initiés en décembre. L'humidité relative minimale et moyenne à Nancy, et moyenne à Strasbourg, joue un rôle parallèle à celui des températures, où des valeurs faibles sont liées à des dates d'apparition tardives. L'humidité relative minimale étant très dépendante de la température maximale, l'interprétation que l'on peut en faire est quasi identique.

Quant à la deuxième moitié du mois de janvier, nous enregistrons un nombre plus faible de corrélations significatives. Cependant l'interprétation à partir de l'humidité relative ou de la température du sol est identique à celle de la période précédente, à savoir que des conditions climatiques et/ou de sol clémentes ne favorisent pas une apparition précoce de concentrations de pollen de frêne «à risque». Cela confirme de nouveau le décalage entre les résultats des corrélations avec les variables de janvier et ceux des modèles issus des méthodes de cumul. Malgré tout, les modèles de cumul démarrant en février donnent d'aussi bons résultats que ceux démarrant plus tôt, un constat qui pourrait permettre de valider les corrélations enregistrées en janvier. De plus, dans le cas de la régression multiple, nous prenons en compte les valeurs brutes de température. Cela n'est pas le cas des méthodes de cumul utilisées, où l'on comptabilise les unités de chaleur à partir d'un certain seuil de température ou en fonction d'un coefficient de développement. Cette approche méthodologique différente peut expliquer la contradiction qui apparaît dans le cas du frêne en janvier.

3.2.3.2 Application de validations croisées

Du fait du petit nombre d'observations caractérisant les séries de dates de frêne, mais aussi celles des autres taxons, pour l'ensemble des stations de l'étude, nous avons voulu tester la robustesse des relations entre les différentes variables météorologiques et les dates d'apparition. Pour cela nous allons utiliser la méthode des validations croisées (*cross-validation*). Cela consiste à calculer les coefficients de corrélation sur l'ensemble de

l'échantillon d'apprentissage (11 ans dans le cas présent) moins une année qui sert, dans la pratique, de validation au modèle élaboré donc, dans notre cas, sur dix années. Nous allons donc répéter 11 fois l'opération en retenant, tout d'abord, dix années après avoir exclu 1992, puis le même nombre d'années une fois 1992 réintégrée, et 1993 retirée de l'analyse, et ainsi de suite.

L'objectif est de repérer prioritairement :

- Les variables corrélées significativement à la série de 11 dates grâce à une année exceptionnelle. En d'autres termes, si nous retirons cette année, les corrélations ne sont plus significatives.
- Les variables qui ne sont pas corrélées significativement à ces dates à cause d'une année particulière. En d'autres termes, sans cette année, ces variables sont corrélées significativement.

Dans le premier cas, toutes les validations croisées appliquées à une variable donnée vont déboucher sur des corrélations significatives sauf une : celle où est retirée l'année à l'origine de la relation significative.

Dans le deuxième cas, c'est la situation opposée. Toutes les validations croisées vont déboucher sur des corrélations non significatives excepté une : celle où est retirée l'année à l'origine de cette absence de significativité.

Nous allons tout d'abord nous intéresser au premier cas en présentant les variables où la moyenne des corrélations des validations croisées significatives est plus d'une fois et demi supérieure à la corrélation de la validation croisée non significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Paris :**
 - T-10_01a (+), T-20_01a (+) **(1997)**

Pour le frêne, seules les températures du sol relevées à -10 et -20 cm pendant la première «quinzaine» de janvier à Paris sont concernées. Cela signifie que les corrélations positives ne sont dues qu'à l'année 1997 qui, une fois retirée, entraîne la non-significativité du coefficient de corrélation entre ces variables et les dates d'apparition du risque. L'année 1997 se caractérise par l'apparition du risque la plus précoce de la série des 11 années (28 février).

Ainsi, ces corrélations sont dues à une année exceptionnelle qui est considérée statistiquement comme une valeur aberrante. Cependant, si l'on souhaite prévoir les dates particulièrement précoces et tardives, il est nécessaire de ne pas retirer cette année. En effet, du fait de la faible longueur des séries, nous ne pouvons nous permettre d'écarter une observation à cause de sa singularité, car ces séries courtes n'englobent pas tous les types d'années climatiquement exceptionnelles que l'on peut rencontrer. Retirer une année de l'analyse serait moins préjudiciable si nous disposions de séries au moins aussi longues que les périodes dites «normales» en météorologie (30 ans).

Nous allons nous intéresser au deuxième cas en présentant les variables où la moyenne des corrélations des validations croisées non significatives est plus d'une fois et demi inférieure à la corrélation de la validation croisée significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Amiens :**
 - JP_12a (-) **(1992)**
 - JP1_11b (-), JP5_01b (+) **(1993)**
 - JP1_02a (-) **(1995)**
 - Tx0_12b (+) **(1997)**
 - Tx_01b (+) **(2002)**
- **Rouen :**
 - Hx_10a (-), T-10_12a (+), P_01b (+), JP1_01b (+), JP5_01b (+) **(1992)**
 - Hx_12b (+), Tn0_01b (+), Tx0_01b (+), Tm0_01b (+), Insol_01b (+) **(1995)**
 - P_01a (-), Tn_01b (-) **(1997)**
 - Insol_12a (+) **(2002)**
 - T-10_12b (+) **(2003)**
- **Paris :**
 - Hx_01a (+), Hm_01b (-) **(1993)**
 - JP_01a (+) **(1994)**
 - Tx_12b (+) **(2003)**
- **Nancy :**
 - Hx_12a (+), Hx_01b (-) **(1992)**

- T-20_01a (+), Gel_01a (-) **(1994)**
- JP_02a (-), Tn_02b (-), Tm_02b (-) **(1995)**
- Tx0_12b (+), T-10_02a (-) **(1997)**
- **Strasbourg :**
 - JP5_02a (-) **(1992)**
 - Tx0_12a (+), Hx_02b (+) **(1993)**
 - Insol_11a (-), Insol_12b (-), Tn_01a (+), Tm_01a (+), Tn0_01a (-), Tm0_01a (-) **(1994)**
 - Hx_02a (+) **(1997)**
 - Insol_01a (+) **(2002)**
 - JP5_10a (-) **(2005)**

Ici un grand nombre de variables sont concernées, contrairement au cas précédent. Ces variables significativement corrélées aux dates d'apparition sur 10 années ne le sont plus si l'on inclut l'année indiquée entre parenthèses. Cela montre la fragilité des corrélations établies et certaines de celles déterminées sur 11 ans risquent de ne plus être significatives si l'on ajoute une année supplémentaire, surtout si cette dernière se caractérise par une date d'apparition du risque singulière par sa précocité ou son retard. En effet, plusieurs années présentées correspondent aux dates extrêmes d'apparition du risque propres à chaque série, à la fois pour les dates les plus précoces (1997 à Rouen, Nancy et Strasbourg ; 2002 à Nancy et Strasbourg ; 2003 à Paris) et les plus tardives (1993 à Nancy et Strasbourg ; 1995 à Amiens et Rouen). La propension de ces années exceptionnelles à faire et défaire les corrélations significatives montre l'importance de leur rôle dans l'élaboration des modèles de régression. Malgré l'instabilité qu'elles génèrent, il est nécessaire de les inclure dans les modèles afin d'obtenir, peut-être, des modèles plus robustes sur des séries plus longues où un large spectre de conditions météorologiques sera pris en compte.

3.2.3.3 Equations de régression et résidus

Nous présentons les équations de régression propres à chaque ville où les variables explicatives retenues sont liées à la valeur maximale du test Af. En plus de ce critère, ne sont retenues que les variables où le signe du coefficient de régression est identique au signe du

coefficient de corrélation. Cette dernière condition n'influe pas sur le résultat final, vu que les variables dont le signe du coefficient de régression est contraire à celui de la corrélation sont très redondantes et, par conséquent, éliminées d'office par l'application du test Af. De même, nous nous sommes abstenus d'inclure dans les modèles de régression certaines variables Tn0, Tx0 et Tm0 qui présentaient au moins 75% de valeurs nulles.

Conjointement aux équations, nous présentons le coefficient de détermination (R^2), la valeur du test Af (Af), l'erreur moyenne absolue (MAE), l'erreur moyenne au carré (RMSE) et le risque d'erreur du test F (p). Rappelons que le REAP correspond à la date d'apparition d'un Risque d'Exposition Allergique aux Pollens.

$$\text{REAP_Amiens} = -1,149 \cdot \text{JP_10a} - 1,326 \cdot \text{JP_11a} - 0,880 \cdot \text{Hn_01b} + 170,864$$

$$R^2 = 0,9356 ; \text{Af} = 0,8607 ; \text{MAE} = 1,40 \text{ jour} ; \text{RMSE} = 1,92 \text{ jour} ; p = 0,0002$$

$$\text{REAP_Rouen} = 3,752 \cdot \text{T-20_12a} + 1,166 \cdot \text{T-10_01a} - 3,000 \cdot \text{Hx_01b} + 343,662$$

$$R^2 = 0,8207 ; \text{Af} = 0,7202 ; \text{MAE} = 2,90 \text{ jours} ; \text{RMSE} = 4,20 \text{ jours} ; p = 0,0053$$

$$\text{REAP_Paris} = -1,372 \cdot \text{JP5_11b} - 0,981 \cdot \text{Tm0_01a} + 1,566 \cdot \text{JP5_01b} - 0,788 \cdot \text{Tx_02b} + 83,991$$

$$R^2 = 0,9087 ; \text{Af} = 0,7950 ; \text{MAE} = 1,65 \text{ jour} ; \text{RMSE} = 2,27 \text{ jours} ; p = 0,0028$$

$$\text{REAP_Nancy} = -2,266 \cdot \text{Hn_01a} + 2,856 \cdot \text{Hx_01a} + 1,006 \cdot \text{Tx0_01a} + 18,843 \cdot \text{Tx0_02b} - 22,316$$

$$R^2 = 0,9460 ; \text{Af} = 0,8416 ; \text{MAE} = 1,48 \text{ jour} ; \text{RMSE} = 2,10 \text{ jours} ; p = 0,0006$$

$$\text{REAP_Strasbourg} = -0,636 \cdot \text{Hn_01a} - 2,589 \cdot \text{P_02a} + 0,407 \cdot \text{Hm_02a} + 96,382$$

$$R^2 = 0,8133 ; \text{Af} = 0,7112 ; \text{MAE} = 2,78 \text{ jours} ; \text{RMSE} = 3,18 \text{ jours} ; p = 0,0060$$

Toutes les équations présentent un test F significatif, pour un risque d'erreur de 5%, indiquant que le coefficient de corrélation multiple l'est également. Nous pouvons remarquer la grande variété des régresseurs d'une équation à l'autre due, à la fois, à la présence de corrélations différentes d'une ville à l'autre et du jeu des corrélations partielles à l'intérieur de chaque modèle de régression. Si nous comparons l'erreur moyenne absolue des équations à celle des différents modèles issus des méthodes de cumul, nous constatons une plus grande précision de la part des équations avec, pour trois d'entre elles, une erreur inférieure à deux jours ; cas de

figure qui ne s'est jamais produit avec les méthodes de cumul. Cela montre donc l'intérêt, dans la prévision de l'apparition du risque d'exposition allergique lié au pollen de frêne, d'utiliser d'autres paramètres météorologiques que la seule température de l'air.

Quant aux résidus, seuls ceux des séries de Strasbourg et d'Amiens présentent une distribution suivant la loi normale selon les résultats des tests de Shapiro-Wilk et de Kolmogorov-Smirnov. A Nancy, Strasbourg, Amiens et Paris le résidu maximal en valeur absolue est inférieur à six jours. Nous sommes donc loin des écarts enregistrés certaines années avec les méthodes de cumul. Seule l'équation de Rouen présente un résidu dont la valeur absolue est supérieure à dix jours. Il s'agit de l'année 2004 où la prévision a été trop tardive de 11 jours (11,27 précisément) ; une année où l'apparition du risque, relativement précoce, ne compte pas parmi les valeurs exceptionnelles. De même, pour les autres villes, les résidus les plus importants ne concernent pas les dates extrêmes, ce qui est encourageant dans le cadre d'une prévision de bonne qualité de ces dernières.

3.2.4 Prévision du bouleau par les méthodes de cumul

Comme pour le frêne, nous allons retenir les années communes aux cinq sites ; c'est-à-dire celles pour lesquelles l'utilisation des données polliniques du bouleau est réalisable. A l'instar de ce qui a été fait pour ce dernier taxon, nous commençons par tenir compte de six méthodes. Du fait du démarrage plus tardif de la saison du bouleau, nous avons pu retenir un plus grand nombre d'années communes aux cinq sites. Quinze années sont ainsi retenues de 1989 à 2005 à l'exception de 1990 et 1996 à cause de démarrages des capteurs trop tardifs à Rouen (1990) et Strasbourg (1996), coïncidant avec la pleine saison de pollinisation du bouleau.

3.2.4.1 Présentation des dates d'apparition et de l'application des méthodes

Il faut rappeler que la date d'apparition d'un risque d'exposition allergique au pollen de bouleau a lieu dès la survenue le jour J d'au moins 20 grains/m³, avec un total cumulé d'au moins 60 grains de J à J+2.

Les dates moyennes d'apparition d'un risque allergique, classées par ordre croissant, sont les suivantes :

- 84,5^{ème} jour de l'année à Paris, soit le 25-26 mars pour une année non bissextile (écart-type de 6,4 jours).
- 86,5^{ème} jour de l'année à Strasbourg (écart-type de 5,8 jours).
- 87,6^{ème} jour de l'année à Rouen (écart-type de 5,7 jours).
- 88,8^{ème} jour de l'année à Amiens (écart-type de 5,2 jours).
- 89,7^{ème} jour de l'année à Nancy, soit le 31 mars pour une année non bissextile (écart-type de 4,9 jours).

Comme pour le frêne, Paris se caractérise par l'apparition du risque la plus précoce mais les écarts entre les moyennes d'une ville à l'autre sont plus faibles que pour le précédent taxon. Les écarts-types, s'échelonnant de 4,9 à 6,4 jours, sont également plus faibles et reflètent une variabilité interannuelle plus réduite. Cela se traduit par des amplitudes maximales sensiblement moins élevées : 25 jours à Paris, 20 à Strasbourg et Rouen, 21 à Amiens et 17 à Nancy. La distribution de ces dates suit la loi normale d'après les résultats des tests de Shapiro-Wilk et de Kolmogorov-Smirnov. Il faut noter la particularité de l'année 1992 où sont enregistrées les dates d'apparition du risque les plus tardives pour l'ensemble des cinq villes, du 4 avril à Paris et Strasbourg au 11 avril à Amiens.

A l'instar du frêne, une Analyse en Composantes Principales (ACP) est appliquée sur les cinq séries polliniques de bouleau. La première composante principale explique 83,91% de la variance totale et montre que, globalement, la variabilité interannuelle des dates d'apparition du risque est très similaire d'une ville à l'autre. Comme pour le frêne les années 1997 et 2002 se distinguent par leurs dates précoces : 78^{ème} et 80,4^{ème} jours de l'année respectivement. Les années 1992, 1995 et 2000 se démarquent nettement par leur caractère tardif, en intervenant les 98^{ème}, 93,4^{ème} et 93,2^{ème} jours de l'année. Pour 1992, cela confirme la constatation effectuée dans le paragraphe précédent.

A partir de ces variables et des différentes méthodes de cumul, nous allons élaborer des modèles de prévision de l'apparition de ce risque d'exposition allergique au pollen de bouleau. Nous allons procéder de la même manière que pour le frêne en recherchant les combinaisons entre les dates de démarrage de cumul et les seuils de température, les coefficients ou, selon les méthodes, le type de température qui minimisent l'erreur moyenne absolue. Les seuils d'action testés sont identiques à ceux du frêne (cf. 3.2.2.1) mais du fait d'une pollinisation moyenne plus tardive, nous avons souhaité tester toutes les dates de début

de cumul du 1^{er} octobre jusqu'au 10 mars, étant donné que la date d'apparition la plus précoce intervient le 11 mars 1997 à Paris. Les sigles et abréviations, affectés à chaque méthode, sont rappelés ci-après :

- La méthode des GDD par triangulation : **GDD_1**.
- Celle employant une fonction sinusoïdale : **GDD_2**.
- La méthode de Lejoly-Gabriel à partir de la température maximale : **Lej_Tx**.
- La méthode FT : **FT**.
- La méthode du Q_{10} à partir du critère de Bidabe (1967) : **Q₁₀**.
- La méthode de Sarvas : **Sarvas**.

3.2.4.2 Présentation des meilleurs modèles

A l'instar de ce qui a été fait pour le frêne, nous allons présenter, dans l'ordre, la date de démarrage du cumul, le seuil ou le type de données retenu, l'erreur moyenne absolue (MAE), l'erreur moyenne au carré (RMSE) et le cumul d'unités de «chaleur», de degrés-jours ou de température à approcher le plus possible (la date d'apparition du risque prévue ayant lieu la veille du jour où le cumul moyen est dépassé). La plus faible erreur moyenne absolue par ville est notée en gras, en cas d'ex aequo n'est surlignée que celle qui a la plus faible erreur moyenne au carré. Nous rappelons que pour la méthode de Sarvas l'utilisation de la température minimale, maximale et moyenne, en tenant compte des unités issues des valeurs négatives, est symbolisée par Tn, Tx et Tm respectivement ; dans le cas de figure où l'on n'en tient pas compte, les symboles sont Tn(0), Tx(0) et Tm(0) (cf. 3.2.2.2).

- **Amiens :**
 - **GDD_1** : 1^{er} janvier – 6,5°C – 1,80 jour – 3,00 jours – 126,07°D
 - **GDD_2** : 29 décembre – 7°C – **1,40 jour** – 2,46 jours – 117,16°D
 - **Lej_Tx** : 9 mars – 10°C – 2,13 jours – 2,80 jours – 232,27°C
 - **FT** : 10 mars – 8°C – 2,07 jours – 2,79 jours – 11,33 unités
 - **Q₁₀** : 29 décembre – 8,2 – 1,60 jour – 2,13 jours – 1110,50 unités
 - **Sarvas** : 3 mars – Tx et Tx(0) – 1,93 jour – 2,65 jours – 205,87 unités

- **Rouen :**

- **GDD_1** : 9 mars – 6,5°C – 1,80 jour – 2,46 jours – 50,58°D
- **GDD_2** : 2 mars – 7°C – 1,73 jour – 2,19 jours – 57,43°D
- **Lej_Tx** : 2 mars – 11°C – **1,67 jour** – 2,27 jours – 211,59°C
- **FT** : 9 mars – 9°C – 1,93 jour – 2,91 jours – 8,40 unités
- **Q₁₀** : 1^{er} mars – 8,7 – 2,20 jours – 2,93 jours – 550,45 unités
- **Sarvas** : 1^{er} mars – Tx et Tx(0) – 2,60 jours – 3,26 jours – 196,24 unités

- **Paris :**

- **GDD_1** : 26 février – 10,5°C – 2,07 jours – 3,49 jours – 25,85°D
- **GDD_2** : 28 février – 11°C – **2,00 jours** – 3,39 jours – 23,73°D
- **Lej_Tx** : 2 et 3 février – 14°C – 2,27 jours – 4,10 jours – 184,35 et 182,28°C
- **FT** : 28 février – 12,5°C – 2,47 jours – 3,68 jours – 4,96 unités
- **Q₁₀** : 7 mars – 6,5 – 2,80 jours – 3,90 jours – 364,34 unités
- **Sarvas** : 7 mars – Tx et Tx(0) – 2,87 jours – 4,31 jours – 167,72 unités

- **Nancy :**

- **GDD_1** : 1^{er} mars – 4°C – 2,07 jours – 3,62 jours – 118,41°D
- **GDD_2** : 28 février – 5°C – 2,00 jours – 3,33 jours – 103,97°D
- **Lej_Tx** : 3 mars – 7°C – 2,07 jours – 2,86 jours – 326,74°C
- **FT** : 10 mars – 6,5°C – 2,67 jours – 3,31 jours – 13,19 unités
- **Q₁₀** : 17 décembre – 4,9 – **1,27 jour** – 1,81 jour – 662,76 unités
- **Sarvas** : 29 décembre – Tx(0) – 1,53 jour – 2,27 jours – 437,58 unités

- **Strasbourg :**

- **GDD_1** : 11 février – 10°C – 2,07 jours – 3,57 jours – 25,78°D
- **GDD_2** : 16 février – 9,5°C – 1,80 jour – 2,98 jours – 35,42°D
- **Lej_Tx** : 26 et 27 décembre – 11°C – **1,60 jour** – 2,00 jours – 388,09 et 386,52°C
- **FT** : 11 et 12 février – 12°C – 2,07 jours – 3,45 jours – 5,05 et 5,03 unités
- **Q₁₀** : 20 février – 5,4 – 1,93 jour – 2,93 jours – 388,26 unités
- **Sarvas** : 20 février – Tx et Tx(0) – 2,20 jours – 3,64 jours – 253,61 et 253,35 unités

A l'instar des modèles appliqués au frêne, nous retrouvons pour certaines méthodes et pour certaines villes plusieurs combinaisons ayant, à la fois, des valeurs de MAE et de RMSE identiques. Nous pouvons remarquer que la qualité moyenne des modèles issus des six méthodes est très proche d'une ville à l'autre. Contrairement au frêne, les modèles présentent une bonne qualité prédictive caractérisée par une erreur moyenne absolue inférieure à trois jours. Quatre des cinq villes présentent, selon le test de Student, des erreurs moyennes absolues statistiquement équivalentes pour un risque d'erreur de 5% : Amiens (1,82 jour), Rouen (1,99 jour), Nancy (1,94 jour) et Strasbourg (1,95 jour). L'erreur moyenne absolue associée à Paris est plus élevée tout en s'avérant de très bonne facture (2,41 jours). Il semble ainsi plus aisé d'établir des modèles de prévision précis pour le bouleau que pour le frêne, l'erreur globale étant de 2,02 jours contre 3,40 jours pour le frêne mais, il est vrai, pour un nombre d'années différent. En plus de l'abondance et du caractère allergisant de ce pollen, le nombre relativement important d'articles traitant de la prévision de la date de début de pollinisation peut s'expliquer par la possibilité d'élaborer des modèles efficaces. Les dates de démarrage de cumul sont variables d'une méthode et d'une ville à l'autre, mais la grande majorité d'entre elles ont lieu en février et pendant la première décade de mars. Nous pouvons noter cependant des dates de début de cumul optimales en décembre (Q₁₀ à Amiens, Q₁₀ et Sarvas à Nancy, Lej_Tx à Strasbourg) et janvier (GDD_1 à Amiens). Les seuils d'action de température ou de développement varient d'une méthode à l'autre et d'une ville à l'autre. La seule exception concerne la méthode de Sarvas où les meilleurs modèles sont initiés à partir de la température maximale, en tenant compte des indices issus des températures négatives ou non. A l'instar du frêne, nous notons la grande ressemblance entre les seuils et les dates de démarrage des modèles issus de GDD_1 et GDD_2.

Il semble que toutes les méthodes débouchent sur des résultats équivalents ; celle qui donne la plus faible erreur moyenne absolue n'est pas la même d'une ville à l'autre. Ainsi à Nancy

c'est la méthode Q_{10} qui donne le meilleur modèle. A Paris et Amiens, il s'agit des modèles associés à la méthode GDD_2. A Rouen et Strasbourg, enfin, ce sont les modèles issus de la méthode de Lejoly-Gabriel qui conduisent aux erreurs les plus faibles. Face à ce constat, nous allons déterminer statistiquement, comme cela a été effectué pour le frêne, si une des méthodes est globalement meilleure que les autres.

3.2.4.3 Comparaisons inter-méthodes et intermensuelles

Nous allons, tout d'abord, présenter graphiquement l'erreur moyenne absolue liée à chaque méthode correspondant à la moyenne des erreurs des meilleures combinaisons des cinq villes de l'étude (figure 39).

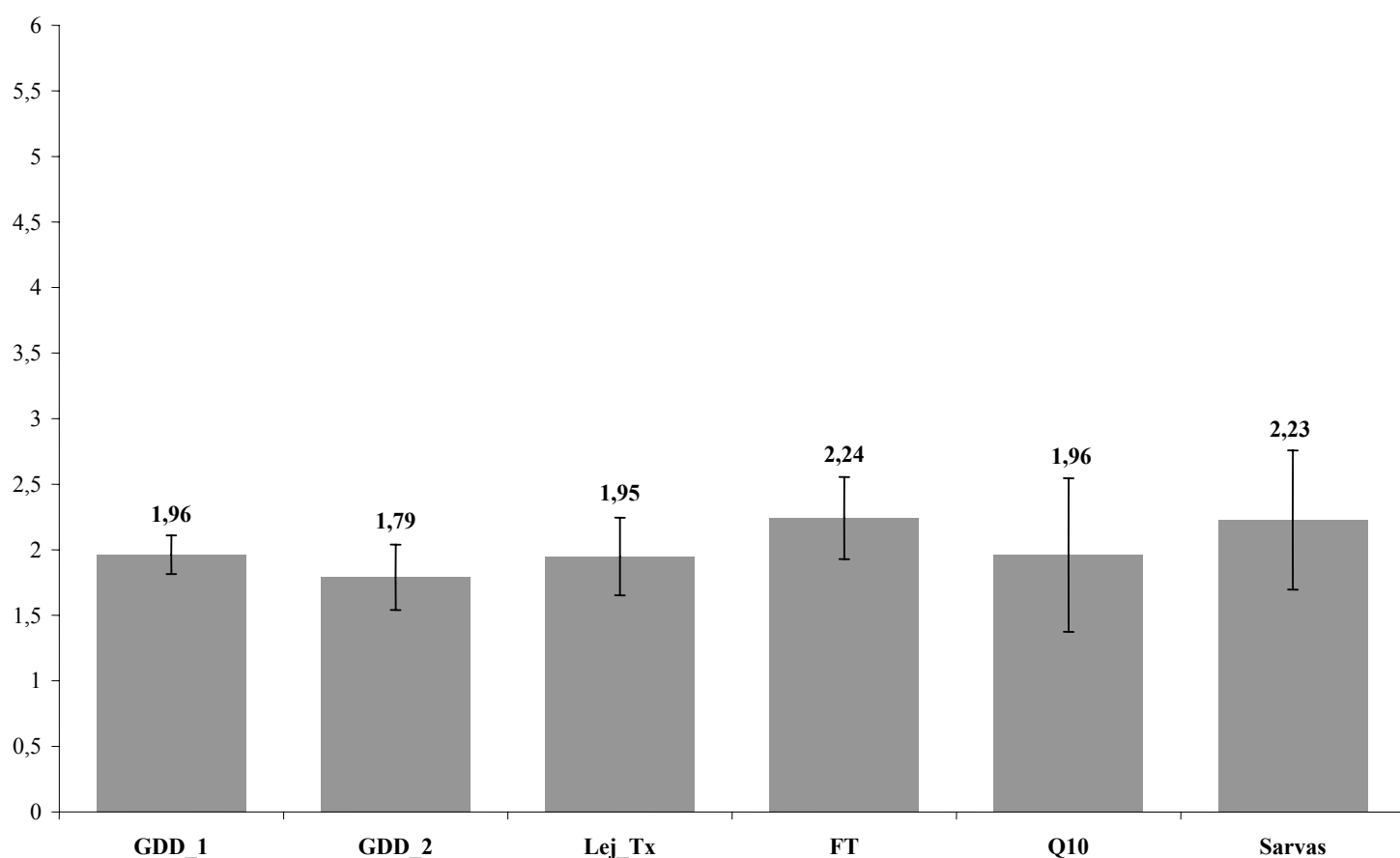


Figure 39 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour le bouleau.

Les moyennes s'échelonnent de 1,79 jour pour la méthode sinusoïdale des *Growing Degree Days* (GDD_2) à 2,24 jours pour celle du *Fraction-Time* (FT). L'utilisation du test de Student

révèle que les moyennes sont statistiquement équivalentes, confirmant l'impression laissée par la lecture du graphique. La seule différence significative enregistrée se place entre GDD_2 et FT mais elle n'est finalement que de 0,45 jour et, contrairement à ce qui avait été constaté graphiquement pour le frêne, les méthodes de Sarvas et du Q_{10} sont aussi fiables que les autres. Quant à la complexité des formules de certaines de ces six méthodes, elle n'est pas un gage d'amélioration de la performance des modèles. La méthode de Lejoly-Gabriel est tout aussi valable que GDD_2 qui n'apporte rien par rapport à GDD_1 et, à l'instar du frêne, sera retirée de la suite de l'analyse

Comme cela a été brièvement évoqué, nous avons constaté que les dates de démarrage les plus fréquemment rencontrées ont lieu en février et mars, c'est-à-dire relativement peu de temps avant la date d'apparition du risque lié au pollen de bouleau. Nous pouvons remarquer qu'aucune date de démarrage en octobre et en novembre n'apparaît comme optimale pour aucune des méthodes et des villes, et que les dates en décembre et janvier sont plus rares que pour le frêne.

Il semble donc que le démarrage des cumuls de température en février ou mars conduise à des modèles de prévision plus efficaces. Pour vérifier cette hypothèse, nous allons comparer les erreurs moyennes absolues mensuelles de l'ensemble des méthodes, ville par ville, puis celles de l'ensemble des villes, méthode par méthode. Nous testons les mois d'octobre à février et, du fait d'une pollinisation en moyenne plus tardive que le frêne, la première décade de mars. Nous déterminons, pour chaque cas de figure, les combinaisons date-seuil minimisant les écarts entre valeurs observées et prévues pour chacun des mois en question. Les résultats, ville par ville, sont présentés dans la figure 40 :

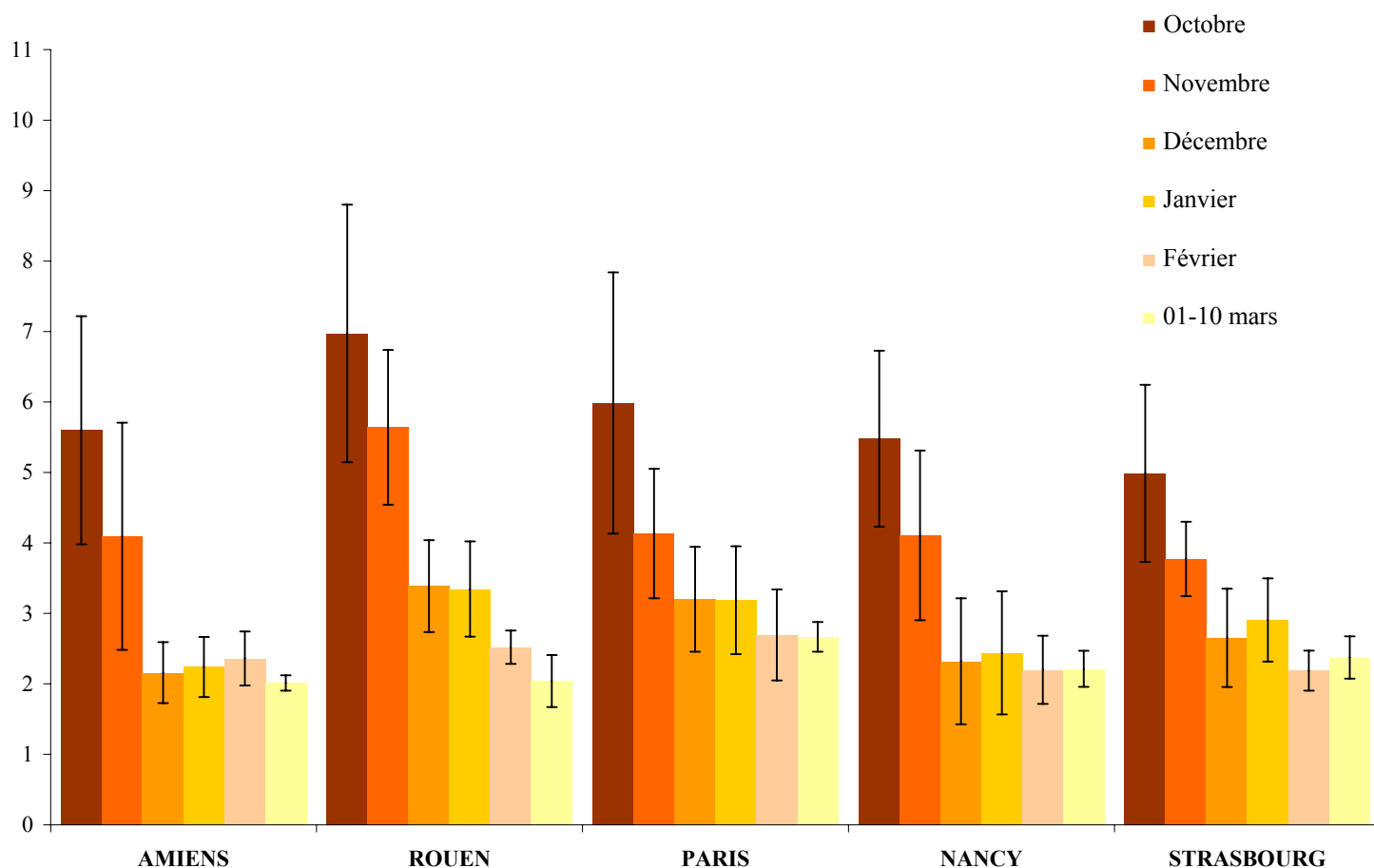


Figure 40 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour le bouleau.

Nous allons analyser les résultats ville par ville à partir du test t de Student :

- Amiens** : Ce test a permis de classer les mois en deux groupes significativement équivalents au niveau de l'erreur moyenne absolue ; entre ces groupes les différences de MAE sont statistiquement significatives pour un risque d'erreur de 5%. Les mois d'octobre et novembre forment un premier groupe avec des erreurs de 5,60 et 4,09 jours respectivement. Le deuxième groupe se compose des mois de décembre à février, ainsi que de la première décade de mars, où les erreurs sont statistiquement équivalentes avec des moyennes allant de 2,01 jours en mars à 2,36 jours en février. Ainsi, à Amiens, la mise en route des modèles de cumul dès le mois de décembre donne des résultats aussi fiables que pour des démarrages plus tardifs.

- **Rouen** : Pour cette ville, nous enregistrons une décroissance progressive des erreurs passant de 6,97 jours en octobre à 2,04 jours durant la première décade de mars. Les mois d'octobre et novembre forment un duo statistiquement équivalent, suivi par les mois de décembre et janvier caractérisés par des valeurs intermédiaires (3,39 et 3,35 jours). La première décade de mars présente, quant à elle, une erreur moyenne absolue significativement plus faible que celle de tous les autres mois, février compris. Ce résultat était grandement suggéré par le fait que les meilleures dates de démarrage aient systématiquement lieu en mars pour les meilleurs modèles de chaque méthode. Dans le cas de Rouen, l'influence des températures précédant immédiatement l'apparition du risque semble plus prépondérante que celle des mois précédents, différant sensiblement de ce qui est constaté à Amiens.
- **Paris** : Nous constatons quasiment le même cas de figure qu'à Amiens avec des erreurs moyennes qui ne sont pas significativement différentes du mois de décembre jusqu'à la première décade de mars (de 3,20 à 2,67 jours respectivement), même si une sorte de palier se dessine à l'instar du cas de figure de Rouen. Les erreurs en octobre et novembre sont significativement plus importantes mais nous constatons, cependant, que l'erreur moyenne absolue de novembre n'est pas significativement supérieure à celles de décembre et janvier. Malgré tout, dans les faits, cela représente tout de même une amélioration de la précision des modèles de l'ordre de un jour. Ainsi, le démarrage des cumuls en décembre donne des prévisions de qualité équivalente, en moyenne, à celles de modèles initiés plus tardivement. Cela nous laisse supposer que la dormance du bouleau est accomplie avant la mi-décembre et que les températures, à partir de cette date-là, jouent un rôle bénéfique dans son développement.
- **Nancy** : La distribution des erreurs moyennes absolues est quasi identique à celle de Paris à l'exception du mois de novembre, dont l'erreur est statistiquement supérieure à celles de décembre et janvier. Ainsi, nous constatons, de nouveau, une équivalence des moyennes de décembre à la première décade de mars, les valeurs s'échelonnant de 2,20 jours en février à 2,44 jours en janvier.

- **Strasbourg** : Nous enregistrons le même schéma qu'auparavant à l'exception de la présence de différences significatives entre janvier et février. Les erreurs moyennes sont ainsi de 2,65 jours (décembre), 2,91 jours (janvier), 2,19 jours (février) et 2,37 jours (1^{er} au 10 mars).

D'une manière générale, nous constatons que le démarrage des cumuls en octobre et novembre débouche sur des résultats médiocres, même s'ils le sont moins que ceux enregistrés pour le frêne. La période de dormance du bouleau correspondant probablement à ces deux mois, cela explique la moins bonne qualité des modèles. A l'exception de Rouen, les modèles initiés en décembre donnent des résultats aussi fiables que ceux démarrant de janvier à début mars, ce qui rejoint le constat effectué pour le frêne. Même si en valeur absolue la plupart des meilleurs modèles sont initiés en février (ou début mars) contrairement au frêne, l'utilisation de modèles démarrant plus précocement est tout à fait adéquate et permet de conserver une assez longue période de cumul. Dans un but préventif, utiliser différents modèles de cumul initiés à différentes dates peut se révéler une option intéressante, afin de voir si les dates prévues proposées sont proches ou non.

L'utilisation de la première décade de mars comme point de départ des cumuls n'apporte pas d'amélioration significative par rapport aux modèles initiés de décembre à février, à l'exception notable de Rouen.

Ainsi, d'un point de vue phénologique, ces résultats rejoignent ceux du frêne avec un accomplissement de la dormance du bouleau qui a lieu, à priori, durant la première moitié du mois de décembre, période qui se caractérise par les jours les plus courts de l'année.

Nous allons nous intéresser à présent aux résultats méthode par méthode par l'intermédiaire, notamment, de la figure 41 :

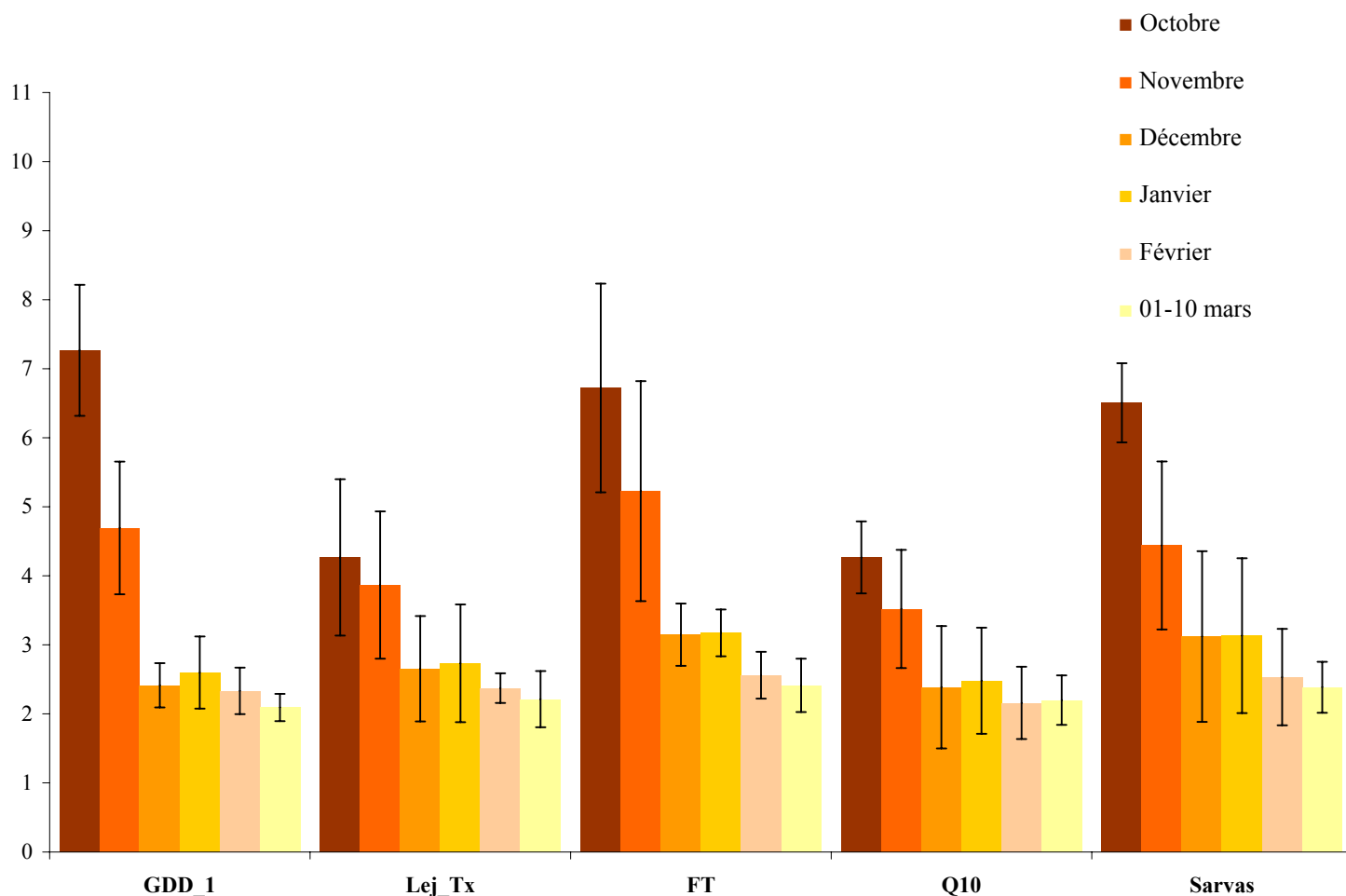


Figure 41 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour le bouleau.

Nous allons analyser les résultats, méthode par méthode, à partir du test t de Student appliqué selon un risque d'erreur de 5% :

- GDD_1** : Les erreurs moyennes absolues associées à chaque période comprise entre le mois de décembre et la première décade de mars, statistiquement équivalentes, sont significativement plus faibles que celles enregistrées en octobre et novembre. Les valeurs moyennes de ces quatre périodes sont, dans l'ordre, de 2,41, 2,60, 2,33 et 2,09 jours.
- Lej_Tx** : Dans le cas de cette méthode, les mois de novembre à mars présentent des erreurs moyennes absolues statistiquement équivalentes, à l'exception des combinaisons novembre-février et novembre-mars. La différence entre octobre et novembre n'est pas significative non plus. A l'opposé de la méthode GDD_1, l'erreur moyenne enregistrée en octobre est «relativement» faible (4,27 jours contre

7,27 et 8,47 jours). Mais les erreurs moyennes associées aux mois de décembre à mars sont quasi identiques à celles des méthodes précédentes, s'échelonnant de 2,21 jours en mars à 2,73 en janvier.

- **FT** : Pour cette méthode, nous n'enregistrons pas de différences significatives pour trois combinaisons de mois : octobre-novembre, décembre-janvier et février-mars. Contrairement aux trois méthodes précédentes, nous relevons une évolution, par palier de deux mois, de la précision des modèles issus de cette méthode. Les erreurs moyennes absolues les plus faibles sont enregistrées en février (2,56 jours) et mars (2,41 jours).
- **Q₁₀** : Nous nous retrouvons, d'un point de vue statistique, dans le même cas de figure que pour la méthode de Lejoly-Gabriel, avec une amélioration lente de la qualité des modèles au fil des mois. D'une valeur comparativement basse en octobre (4,27 jours), on aboutit à une erreur minimale équivalente à celle des quatre méthodes précédentes (2,16 jours en février).
- **Sarvas** : Les relations statistiques entre les différents mois sont strictement identiques à celles enregistrées pour les méthodes de Lejoly-Gabriel et du Q₁₀, à l'exception des mois d'octobre et novembre dont les erreurs moyennes sont statistiquement équivalentes. Les valeurs vont de 6,51 jours en octobre à 2,39 jours durant la première décade de mars.

L'étude par méthode confirme les résultats enregistrés ville par ville avec, principalement, l'absence de différences significatives entre les erreurs moyennes absolues des modèles initiés de décembre à mars, à l'exception de la méthode FT qui se distingue par des modèles, plus précis, mis en route en février et mars. Un même constat peut être établi pour la méthode de Sarvas, par la lecture de la figure 41. Cependant les moyennes ne sont pas significativement différentes de décembre à mars, selon le test de Student.

3.2.4.4 Bilan global

Nous allons d'abord présenter les erreurs moyennes absolues mensuelles des modèles issus de l'ensemble des méthodes et des villes (figure 42) :

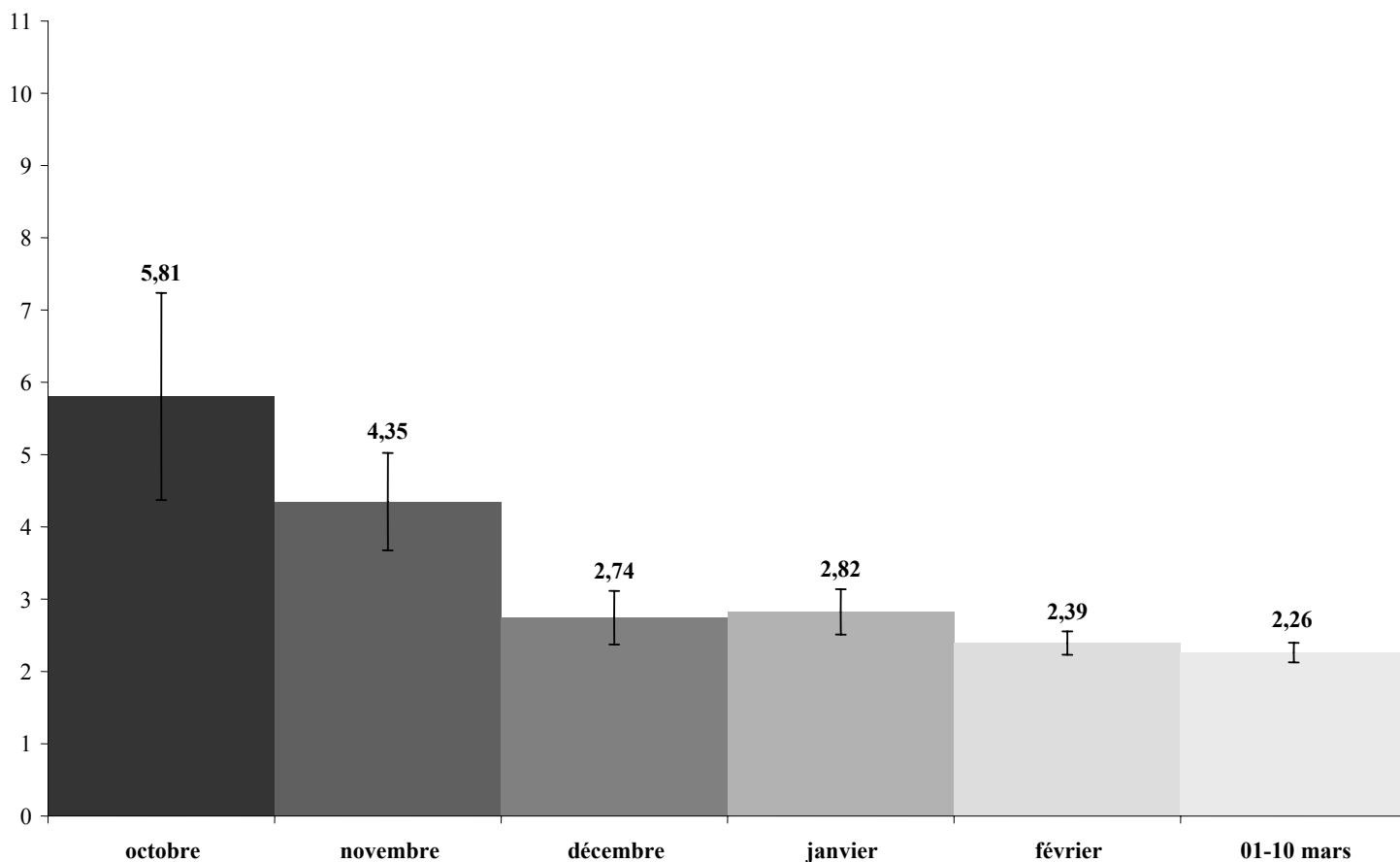


Figure 42 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour le bouleau.

De manière globale, et d'après le test t de Student, nous enregistrons quatre cas de figure où les erreurs moyennes ne sont pas significativement différentes : entre octobre et novembre, décembre et janvier, décembre et février ainsi qu'entre février et mars. La première décade de mars est celle fournissant la plus faible erreur moyenne qui s'avère, par ailleurs, être significativement plus faible qu'en décembre et janvier. Ce résultat est influencé grandement par le cas particulier de Rouen, seule ville où l'on constate des différences significatives pour ces deux duos. De ce fait, nous pouvons considérer qu'à l'instar du frêne le démarrage des modèles de cumul peut s'effectuer de décembre à février avec, peut-être, une préférence pour ce dernier mois. La qualité des modèles y est très proche et montre bien l'influence des

températures de ces trois mois sur le développement du bouleau. Comme pour le frêne, on peut émettre l'hypothèse que l'accomplissement de la dormance a lieu en décembre. Utilisée à titre expérimental, la première décade de mars n'est guère exploitable dans un cadre préventif, à cause de périodes de cumul pouvant se réduire à moins d'une semaine en cas d'apparition précoce du risque d'exposition allergique aux pollens. Il est clair que dans un but opérationnel les cumuls doivent être, au moins, entamés mi-février afin d'avoir une période d'accumulation qui ne soit pas réduite à une portion congrue de quelques jours.

3.2.4.5 Etude des résidus

La normalité de la distribution des résidus des différents modèles exposés ci-dessus est testée, comme à l'accoutumée, grâce aux tests de Shapiro-Wilk et de Kolmogorov-Smirnov. Près de 40% d'entre eux présentent des distributions non gaussiennes, la ville la plus touchée étant Strasbourg (quatre des cinq modèles). A l'opposé quatre des cinq séries de résidus d'Amiens, Nancy et Rouen présentent une distribution normale. Comme pour le frêne, l'explication du caractère non gaussien de la distribution de ces résidus n'est pas évidente car cela peut évoluer avec l'ajout d'une variable du fait du nombre réduit d'observations (15). De plus aucune méthode et aucune ville n'est systématiquement touchée ou épargnée par ce cas de figure.

La précision des modèles liés au bouleau est bonne, néanmoins la date d'apparition du risque de certaines années est mal estimée. Contrairement au frêne, aucun écart n'atteint 14 jours en valeur absolue. Nous avons alors décidé de présenter les années où la valeur absolue des résidus dépasse de cinq jours et plus celle des autres années, à condition que cette valeur atteigne au moins 10 jours. Les valeurs présentées correspondent à la relation suivante : **valeur prévue – valeur observée.**

- **Amiens :**
 - **GDD_1** : – 10 jours (2002)
- **Paris :**
 - **GDD_1** : – 11 jours (1991)
 - **Lej_Tx** : – 12 jours (1991)
 - **FT** : – 11 jours (1991)

- **Nancy** :
 - **GDD_1** : – 12 jours (1991)

Trois villes sont concernées par ce cas de figure et seulement deux années : 1991 et 2002.

- **Amiens** : Le modèle issu de la méthode GDD_1 donne une prévision dix jours trop précoce par rapport à la date réelle d'apparition du risque en 2002. Cela est d'autant plus surprenant que la date d'apparition la plus précoce de la série a lieu cette année-là (22 mars). Les températures de la deuxième «quinzaine» de janvier et de la première de février, les plus douces de la série, semblent expliquer, en grande partie, cette propension à une prévision très précoce (12 mars).
- **Paris** : Les modèles issus des méthodes GDD_1, Lej_Tx et FT débouchent sur des prévisions trop précoces de 11 et 12 jours pour l'année 1991. La date d'apparition est proche de la moyenne (24 mars) alors que les dates prévues sont particulièrement précoces (12 et 13 mars). Les cumuls d'unités de chaleur pour GDD_1 et FT sont entamés les 26 et 28 février, et la première quinzaine de mars 1991 est la plus douce (avec 1997) de la série. Les conditions de température exceptionnelles expliquent ici la prévision précoce. Quant à la méthode de Lejoly-Gabriel, les cumuls sont initiés début février, période rigoureuse en 1991. Mais le seuil d'action, élevé (14°C), va tenir compte des températures maximales de la fin février jusqu'à la date prévue qui sont quasiment exclusivement supérieures à ce seuil, ce qui doit expliquer la précocité de la prévision.
- **Nancy** : Le modèle issu de la méthode GDD_1 débouche sur une prévision trop précoce de 12 jours pour l'année 1991. Comme à Paris, la date d'apparition du risque cette année (1^{er} avril) est proche de la moyenne des 15 années (31 mars). Ce modèle est initié le 1^{er} mars et ne prend en compte que les températures de la première moitié du mois de mars qui sont, comme à Paris, les plus douces de la série avec 1997.

A travers ces quelques exemples, nous retrouvons les mêmes causes, que pour le frêne, à l'origine de résidus importants. Une année présentant une date d'apparition du risque

«normale» peut être mal estimée si les conditions de température la précédant sont remarquables, à l’instar de la grande douceur de la première «quinzaine» du mois de mars 1991. Le phénomène contraire se retrouve aussi avec, dans le cas de Nancy, une date précoce dont la prévision, qui lui est associée, l’est encore plus.

3.2.5 Prévision du bouleau par la régression linéaire multiple

Rappelons que toutes les variables présentées dans le sous-chapitre 3.1.2 sont utilisées pour les sites de Rouen, Paris, Nancy et Strasbourg, excepté la durée du gel qui n’est exploitable qu’à Paris. La station météorologique d’Amiens, moins complète, ne mesure pas l’insolation, la température du sol et la durée du gel.

3.2.5.1 Etude des corrélations

Nous allons tout d’abord rechercher les coefficients de corrélation significatifs, pour un risque d’erreur de 5%, existant entre les valeurs bimensuelles des différentes variables météorologiques et les dates d’apparition du risque. Les périodes testées s’échelonnent de la première «quinzaine» d’octobre à la première décade de mars. Les dates les plus précoces ayant généralement lieu entre le 10 et le 15 mars, nous n’avons pas voulu nous priver des conditions des dix premiers jours du mois dans l’analyse. Parallèlement à l’étude des résultats issus des méthodes de cumul, nous avons exceptionnellement inclus une période décadaire parmi les périodes bimensuelles. Nous allons présenter les résultats «quinzaine» par «quinzaine», plus la première décade de mars, à travers le tableau 10 en reprenant les abréviations du sous-chapitre 3.1.2. Le but est, ici encore, de déterminer si plusieurs variables sont significativement corrélées avec les dates d’apparition du risque de plusieurs villes.

	AMIENS	ROUEN	PARIS	NANCY	STRASBOURG
10a				Hx JP JP5	
10b		Tx			
11a				Hx	
11b	P JP5			Hx	P JP1 JP5
12a		Hx			
12b			Tn0 Tx0 Tm0 Gel		
01a		Insol	Tn T-10 T-20 Tn0 Tx0 Tm0 Gel Insol	Hn	Tx0 Tm0
01b	Tx0	Hn Hx Hm	Tm0 Hn Hm	Hn Hx Hm	
02a					
02b					
03a		Hx		T-10	Tx T-10

Tableau 10 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique au pollen de bouleau, sur les 15 années communes. En rouge = corrélation positive ; en bleu = corrélation négative ; en gras = variables communes à, au moins, deux villes

Nous avons également déterminé les corrélations partielles issues des régressions pas à pas ascendantes élaborées pour chaque période bimensuelle :

- **Amiens** : Hn_01b
- **Rouen** : Tx0_01b, JP1_03a
- **Paris** : JP_01a, JP_01b
- **Nancy** : JP1_11a
- **Strasbourg** : Tx_11b, Insol_11b, JP1_11b, Tn0_01a

A l'instar du frêne, nous pouvons constater la diversité des variables corrélées et, surtout, que la plupart d'entre elles ne le sont qu'avec les dates d'une seule ville. Cela montre, de nouveau, la nécessité d'élaborer des équations propres à chaque site.

Cependant, nous constatons la présence de variables communes à deux et trois villes où l'on enregistre des coefficients de corrélation significatifs. A l'instar du frêne, le signe des corrélations est alors identique. Nous remarquons par ailleurs l'absence de corrélations entre les dates d'apparition et les variables météorologiques de février ; résultat qui s'avère être en opposition avec celui des méthodes de cumul où un certain nombre de dates de démarrage, déterminées comme les plus pertinentes, ont lieu durant ce mois.

Le plus grand nombre de corrélations est enregistré durant les deux «quinzaines» de janvier, résultat similaire à celui constaté pour le frêne. Nous pouvons ainsi noter l'influence retardatrice des températures de la première «quinzaine» de janvier à Strasbourg et Paris. Plus

celles-ci sont élevées (ou plus le nombre de jours enregistrant des températures négatives est faible), plus l'apparition du risque est tardive. Il semblerait pour ces deux villes que la température de cette période ne joue pas encore son rôle de «forçage» sur le développement du bouleau. Comme cela a été évoqué pour le frêne, il y a une certaine contradiction avec les résultats découlant des méthodes de cumul. Mais la méthodologie n'étant pas la même, il est difficile d'interpréter ces divergences.

Quant à la deuxième moitié du mois de janvier, à travers les corrélations liées aux températures d'Amiens et Paris, nous constatons le rôle «bénéfique» des valeurs de températures et, dans le cas présent, du faible nombre de jours avec des températures négatives sur l'apparition précoce du risque. Cela se confirme, pour Nancy et Strasbourg du moins, durant la première décade de mars. L'humidité relative se cantonne pour sa part, à l'exception des mois d'octobre et novembre, dans un rôle quelque peu différent que celui joué pour le frêne. En effet, quand les corrélations sont significatives, une forte humidité relative est liée à une date d'apparition du risque précoce, notamment au cours de la deuxième «quinzaine» de janvier. Durant cette période et celle de mars, l'influence de ce paramètre semble être en contradiction avec les températures. Les deux étant liés, nous aurions dû nous attendre (surtout pour l'humidité minimale et moyenne) à des corrélations positives. L'humidité relative joue-t-elle un rôle particulier ? Il est difficile de l'affirmer sans disposer d'un plus grand nombre de variables corrélées significativement.

3.2.5.2 Application de validations croisées

A travers cette méthode, nous avons voulu tester la robustesse des relations entre les différentes variables météorologiques et les dates d'apparition. Pour cela nous allons utiliser la méthode des validations croisées (*cross-validation*) exposée dans la section 3.2.3.2. A l'instar de ce qui a été fait pour le frêne, l'objectif est de repérer prioritairement :

- Les variables corrélées significativement à la série de 15 dates grâce à une année exceptionnelle. En d'autres termes, si nous retirons cette année, les corrélations ne sont plus significatives.

- Les variables qui ne sont pas corrélées significativement à ces dates à cause d'une année particulière. En d'autres termes, sans cette année, ces variables sont corrélées significativement.

Dans le premier cas, toutes les validations croisées appliquées à une variable donnée vont déboucher sur des corrélations significatives sauf une : celle où est retirée l'année à l'origine de la relation significative.

Dans le deuxième cas, c'est la situation opposée. Toutes les validations croisées vont déboucher sur des corrélations non significatives excepté une : celle où est retirée l'année à l'origine de cette absence de significativité.

Nous allons tout d'abord nous intéresser au premier cas en présentant les variables où la moyenne des corrélations des validations croisées significatives est plus d'une fois et demi supérieure à la corrélation de la validation croisée non significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Paris :**
 - Tx0_12b (-), Tm0_12b (-), Gel_12b (-), Tn0_01a (-), Tx0_01a (-), Tm0_01a (-), Gel_01a (-), T-10_01a (+) **(1997)**
- **Strasbourg :**
 - Tx0_01a (-) **(1997)**

Les conditions de température, et des variables découlant de celle-ci, de mi-décembre 1996 à mi-janvier 1997 à Paris se distinguent, ainsi que le nombre de jours avec des températures constamment négatives à Strasbourg. Cela signifie que les corrélations ne sont dues qu'à l'année 1997 qui, une fois retirée, entraîne une non-significativité du coefficient de corrélation entre ces variables et les dates d'apparition du risque. Cette année se caractérise, à la fois, par l'apparition du risque la plus précoce de la série de ces deux villes (11 et 16 mars respectivement), et par la période allant de mi-décembre à mi-janvier la plus froide des années de la série. Cela explique donc les corrélations négatives entre le nombre de jours où les températures sont négatives et la durée du gel d'un côté, et la date d'apparition du risque de l'autre. Durant ces périodes, plus la durée de gel et plus le nombre de jours concernés par des valeurs négatives sont importants, plus l'apparition du risque est précoce.

L'année 1997 se singularise donc par deux aspects : un début d'hiver très rigoureux et une date d'apparition du risque très précoce. Les corrélations ne sont dues qu'à une année exceptionnelle, mais faut-il cependant la retirer ? Comme pour le frêne, n'ayant qu'un faible nombre d'années, cela n'est pas judicieux surtout si cette année sort de l'ordinaire. Il se peut que l'ajout progressif de nouvelles années dans les modèles débouche sur une corrélation non significative de ces variables, mais il n'est pas exclu que le même phénomène se retrouve dans le cas d'une période exceptionnellement froide en décembre et janvier. Cependant, les températures de la mi-février au 10 mars, les plus douces à Paris et Strasbourg sur les 15 ans de l'étude, ont, très certainement, beaucoup contribué à l'apparition précoce du risque. Partant de ce dernier constat, nous n'avons donc pas de raisons de retirer cette année de l'analyse.

Nous allons nous intéresser au deuxième cas en présentant les variables où la moyenne des corrélations des validations croisées non significatives est plus d'une fois et demi inférieure à la corrélation de la validation croisée significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Amiens :**
 - Tm0_01b (+), Hx_02b (+), Hm_02b (+) **(1991)**
 - Tm_10a (-), Tn_12a (+), Tx_12a (+), Tm_12a (+), Tn0_12a (-), JP_01a (+) **(1992)**
- **Rouen :**
 - JP_10a (-) **(1991)**
 - P_12a (+), Tn_12a (+), Tx_12a (+), Tm_12a (+), Tn0_12a (-), Tm0_12a (-), T-10_12a (+), JP_12a (+), JP1_12a (+), P_01b (+), JP5_01b (+) **(1992)**
 - Hn_12a (-), Insol_12a (+), Tn_02a (+), Tx_02a (+), Tm_02a (+), T-10_02a (+) **(2002)**
 - Hx_02b (-) **(2003)**
- **Paris :**
 - Insol_01b (+) **(1995)**
 - Tn_01b (-), T-10_01b (-), T-20_01b (-) **(1997)**
 - Hn_10b (+) **(1998)**
 - Insol_12a (+) **(2002)**

- **Nancy :**
 - JP1_12b (-) **(1989)**
 - Hx_12a (+), Hx_12b (+) **(1992)**
 - Tx_10b (-), Tm_10b (-), JP_10b (-), JP1_11b (-) **(1994)**
 - Insol_12a (+) **(2002)**

- **Strasbourg :**
 - T-20_01a (+) **(1991)**
 - JP5_02a (-) **(1992)**
 - Hx_12b (+) **(1995)**
 - Tx_11a (+), Tm_11a (+), Tn0_11a (-), Hx_02a (+) **(1997)**

Ici un grand nombre de variables sont concernées, contrairement au cas précédent. Ces variables significativement corrélées aux dates d'apparition sur 14 années ne le sont plus si l'on inclut l'année indiquée entre parenthèses. Cela montre une nouvelle fois la fragilité des corrélations établies et la nécessité de mettre à jour les équations année après année. Trois années sont, au moins, présentes dans trois villes et leur singularité est à l'origine de la non-significativité de certaines corrélations :

- **1991 :** Pour Amiens, Rouen et Strasbourg, cette année se caractérise par une apparition du risque assez précoce ne semblant plus correspondre, à Amiens, aux relations établies entre l'humidité relative de la deuxième «quinzaine» de février et les dates des 14 autres années. Cependant, les variables concernées diffèrent d'une ville à l'autre, rendant difficile une constatation générale.

- **1992 :** Seule la série de Paris n'est pas concernée par cette année. Elle se caractérise par la date d'apparition du risque qui se révèle la plus tardive des cinq séries étudiées. Les variables qui reviennent le plus fréquemment sont celles de la première «quinzaine» de décembre 1991 dont les conditions de température assez rigoureuses et l'absence de précipitations, notamment, auraient du conduire vers une apparition plus précoce. L'influence de ces dates très tardives est pour beaucoup dans la remise en cause de ces relations.

- **2002** : Rouen, Paris et Nancy sont concernées par cette année où l'apparition du risque s'est effectuée tôt dans la saison, sans pour autant atteindre les dates les plus précoces de la série. Nous retrouvons, pour ces trois villes, la durée d'insolation de la première «quinzaine» de décembre 2001 qui s'avère être corrélée positivement avec les dates d'apparition du risque, hors l'année 2002. En d'autres termes, nous aurions dû nous attendre à une date tardive. L'insolation de la première moitié du mois de décembre 2001 est la plus élevée de la période à Rouen et Nancy, et se situe juste derrière l'année 1991 à Paris. Or les dates d'apparition du risque se sont révélées précoces, ce qui a rendu la corrélation non significative.

Ainsi l'ajout d'une année présentant des conditions climatiques hivernales «moyennes», mais se caractérisant par une date d'apparition du risque parmi les plus tardives ou les plus précoces, risque de chambouler une partie des corrélations établies. De même, si des conditions climatiques remarquables ont lieu durant l'hiver et qu'elles débouchent sur une date d'apparition du risque «moyenne», nous aurons les mêmes effets.

La propension de ces années climatiquement et/ou phénologiquement exceptionnelles à faire et défaire les corrélations significatives montre l'importance de leur rôle dans l'élaboration des modèles de régression. Cela souligne, en tout cas, l'absolue nécessité de pérenniser le réseau de capteurs afin d'avoir un historique plus complet des conditions climatiques extrêmes pouvant survenir durant les mois précédant la pollinisation, ainsi que des dates d'apparition du risque, extrêmes ou non, qui y sont liées. C'est la seule condition permettant de valider (ou de réfuter) les corrélations établies sur des séries qui pèchent par leur faible longueur.

3.2.5.3 Equations de régression et résidus

Comme pour le frêne, nous présentons les équations de régression propres à chaque ville où les variables explicatives retenues sont liées à la valeur maximale du test Af. En plus de ce critère, ne sont retenues que les variables où le signe du coefficient de régression est identique au signe du coefficient de corrélation. De plus, nous nous sommes de nouveau abstenus d'inclure dans les modèles de régression certaines variables Tn0, Tx0 et Tm0 qui présentaient au moins 75% de valeurs nulles.

Conjointement aux équations, nous indiquons le coefficient de détermination (R^2), la valeur du test Af (Af), l'erreur moyenne absolue (MAE), l'erreur moyenne au carré (RMSE) et le risque d'erreur du test F (p).

$$\text{REAP_Amiens} = -0,606 \cdot \text{JP5_11b} - 0,444 \cdot \text{Hn_01b} + 6,342 \cdot \text{Tx0_01b} + 123,142$$

$R^2 = 0,6073$; Af = 0,5161 ; MAE = 2,59 jours ; RMSE = 3,17 jours ; $p = 0,0135$

$$\text{REAP_Rouen} = -0,038 \cdot \text{Insol_01a} - 1,177 \cdot \text{Hx_01b} - 1,125 \cdot \text{Hx_03a} + 0,939 \cdot \text{JP1_03a} + 307,955$$

$R^2 = 0,8833$; Af = 0,8194 ; MAE = 1,65 jour ; RMSE = 1,89 jour ; $p = 0,0001$

$$\text{REAP_Paris} = -1,049 \cdot \text{Tn0_12b} - 0,027 \cdot \text{Insol_01a} + 3,162 \cdot \text{Tm0_01b} + 88,522$$

$R^2 = 0,6349$; Af = 0,5480 ; MAE = 2,59 jours ; RMSE = 3,70 jours ; $p = 0,0092$

$$\text{REAP_Nancy} = -0,244 \cdot \text{JP_10a} + 0,779 \cdot \text{Hx_11a} + 0,518 \cdot \text{Hx_11b} - 0,917 \cdot \text{Hx_01b} + 53,260$$

$R^2 = 0,8399$; Af = 0,7688 ; MAE = 1,49 jour ; RMSE = 1,89 jour ; $p = 0,0005$

$$\text{REAP_Strasbourg} = 1,080 \cdot \text{Tx_11b} + 0,034 \cdot \text{Insol_11b} - 2,013 \cdot \text{JP1_11b} + 1,017 \cdot \text{Tn0_01a} - 1,173 \cdot \text{Tx0_01a} + 82,331$$

$R^2 = 0,9071$; Af = 0,8296 ; MAE = 1,53 jour ; RMSE = 1,72 jour ; $p = 0,0002$

Toutes les équations présentent un test F significatif, pour un risque d'erreur de 5%, indiquant que le coefficient de corrélation multiple l'est également. Comme pour le frene, nous pouvons remarquer la grande variété des régresseurs d'une équation à l'autre due, à la fois, à la présence de corrélations différentes d'une ville à l'autre et au jeu des corrélations partielles à l'intérieur de chaque modèle de régression. Si nous comparons l'erreur moyenne absolue des équations à celle des différents modèles issus des méthodes de cumul, nous constatons une précision équivalente, avec une erreur moyenne absolue de 2,04 jours pour l'ensemble des cinq villes. De ce fait, l'intérêt de la régression linéaire multiple est confirmé dans la prévision de l'apparition du risque d'exposition allergique.

Quant aux résidus, ils sont distribués normalement pour l'ensemble des séries, selon les tests de Shapiro-Wilk et de Kolmogorov-Smirnov. A Rouen, Nancy et Strasbourg, la valeur

absolue des résidus est inférieure à cinq jours. A Amiens nous enregistrons, au pire, une prévision de huit jours (7,97 exactement) trop tardive en 2002. A Paris, une prévision trop précoce de dix jours (10,42 exactement) est obtenue en 1995. Nous constatons, par ailleurs, que les années connaissant les dates d'apparition les plus tardives (1992) et les plus précoces (1997 en moyenne) ne sont pas celles où l'on constate les plus forts résidus.

3.2.6 Prévision des Poacées par les méthodes de cumul

Comme pour les deux taxons précédents, nous allons retenir les années communes aux cinq sites, dont l'utilisation est réalisable pour la prévision des Poacées. Nous avons cette fois pu retenir la période 1988-2005 complète, soit 18 années.

3.2.6.1 Présentation des dates d'apparition et de l'application des méthodes

Il faut rappeler que la date d'apparition d'un risque d'exposition allergique au pollen de Poacées a lieu dès l'apparition le jour J d'au moins 10 grains/m³, avec un total cumulé d'au moins 50 grains de J à J+4, à condition que l'on enregistre 10 grains ou plus pour trois des cinq jours en question.

Les dates moyennes d'apparition d'un risque allergique, classées par ordre croissant, sont les suivantes :

- 129,8^{ème} jour de l'année à Nancy, soit le 10 mai pour une année non bissextile (écart-type de 6 jours).
- 132,5^{ème} jour de l'année à Strasbourg (écart-type de 5,2 jours).
- 134^{ème} jour de l'année à Amiens (écart-type de 5,8 jours).
- 138^{ème} jour de l'année à Paris (écart-type de 7,8 jours).
- 138,8^{ème} jour de l'année à Rouen, soit le 19 mai pour une année non bissextile (écart-type de 8,2 jours).

Nancy et Strasbourg se caractérisent par l'apparition du risque la plus précoce. Les écarts-types s'échelonnent de 5,2 à 8,2 jours, ils sont plus élevés à Paris et Rouen où les dates moyennes sont les plus tardives. Cela se traduit dans les amplitudes maximales : 31 jours à Paris, 36 à Rouen contre «seulement» 21 jours à Nancy, 18 à Strasbourg et 23 à Amiens. La

distribution de l'ensemble des séries suit la loi normale, d'après les résultats des deux tests de normalité employés. Il faut noter la particularité de l'année 1991 où sont enregistrées les dates d'apparition du risque les plus tardives pour l'ensemble des cinq villes, à l'exception de Strasbourg où elle intervient un jour avant celle de 1996. Paris et Rouen enregistrent même, pour cette année 1991, des dates d'apparition du risque en juin, les 9 et 8 respectivement.

Une Analyse en Composantes Principales (ACP) est utilisée sur les cinq séries polliniques de Poacées. La première composante principale permet de confirmer la constatation précédente, à savoir le caractère très tardif de l'apparition du risque en 1991 (150,6^{ème} jour de l'année) et, dans une moindre mesure, de celle de 1996 (144,6^{ème} jour) pour les cinq villes. Les années 1998 à 2000 se distinguent, au contraire par leurs dates précoces comprises entre les 127^{ème} et 129^{ème} jours de l'année.

A partir de ces variables et des différentes méthodes de cumul, nous allons élaborer des modèles de prévision de l'apparition de ce risque d'exposition allergique au pollen de Poacées. Nous allons procéder de la même manière que pour le frêne et le bouleau en recherchant les combinaisons entre les dates de démarrage de cumul et les seuils de température, les coefficients ou le type de température – selon les méthodes – qui minimisent l'erreur moyenne absolue. Nous allons tester, à côté des seuils d'action présentés dans la section 3.2.2.1, toutes les dates de début de cumul du 1^{er} janvier, date de démarrage de cumul la plus lointaine rencontrée dans la littérature (Smith et Emberlin, 2005) jusqu'au 30 avril, étant donné que la date d'apparition la plus précoce intervient les 2 mai 1999 et 2003 à Nancy. Les sigles et abréviations, affectés à chaque méthode, sont rappelés ci-après :

- La méthode des GDD par triangulation : **GDD_1**.
- Celle employant une fonction sinusoïdale : **GDD_2**.
- La méthode de Lejoly-Gabriel à partir de la température maximale : **Lej_Tx**.
- La méthode FT : **FT**.
- La méthode du Q₁₀ à partir du critère de Bidabe (1967) : **Q₁₀**.
- La méthode de Sarvas : **Sarvas**.

3.2.6.2 Présentation des meilleurs modèles

Ils sont présentés de la même manière que pour les deux autres taxons à partir des six méthodes retenues au préalable. La plus faible erreur moyenne absolue par ville est notée en gras, en cas d'ex aequo n'est surlignée que celle qui a la plus faible erreur moyenne au carré.

Nous rappelons que pour la méthode de Sarvas l'utilisation de la température minimale, maximale et moyenne, en tenant compte des unités issues des valeurs négatives, est symbolisée par Tn, Tx et Tm respectivement ; dans le cas de figure où l'on n'en tient pas compte, les symboles sont Tn(0), Tx(0) et Tm(0) (cf. 3.2.2.2 et 3.2.4.2).

- **Amiens :**

- **GDD_1** : 21 avril – 7°C – 2,06 jours – 2,95 jours – 116,39°D
- **GDD_2** : 19 avril – 4,5°C – 2,06 jours – 2,97 jours – 181,25°D
- **Lej_Tx** : 20 avril – 11°C – 2,28 jours – 3,32 jours – 387,66°C
- **FT** : 24 avril – 10°C – 1,83 jour – 2,72 jours – 13,62 unités
- **Q₁₀** : 26 avril – 4,5 – 2,39 jours – 3,60 jours – 345,10 unités
- **Sarvas** : 23 avril – Tn(0) – **1,72 jour** – 2,46 jours – 73,42 unités

- **Rouen :**

- **GDD_1** : 20 avril – 7°C – 3,39 jours – 5,17 jours – 139,93°D
- **GDD_2** : 20 avril – 7°C – 3,50 jours – 5,22 jours – 142,75°D
- **Lej_Tx** : 27 avril – 14°C – 3,67 jours – 5,62 jours – 288,38°C
- **FT** : 27 avril – 13,5°C – **3,33 jours** – 5,15 jours – 7,19 unités
- **Q₁₀** : 27 avril – 3,1 – 3,72 jours – 5,72 jours – 216,93 unités
- **Sarvas** : 28 avril – Tm et Tm(0) – 3,44 jours – 5,60 jours – 154,62 unités

- **Paris :**

- **GDD_1** : 19 avril – 11,5°C – 3,78 jours – 5,45 jours – 88,93°D
- **GDD_2** : 19 avril – 13°C – **3,67 jours** – 5,52 jours – 66,18°D

- **Lej_Tx** : 19 avril – 16°C – 4,11 jours – 5,48 jours – 399,24°C
 - **FT** : 19 avril – 15,5°C – 3,78 jours – 5,54 jours – 9,29 unités
 - **Q₁₀** : 18 avril – 6 et 6,1 – 4,11 jours – 6,11 jours – 1240,70 et 1282,20 unités
 - **Sarvas** : 16 avril – Tn et Tn(0) – 3,72 jours – 5,03 jours – 160,44 unités
-
- **Nancy** :
 - **GDD_1** : 25 avril – 7,5°C – 2,50 jours – 3,34 jours – 77,62°D
 - **GDD_2** : 24 avril – 8°C – 2,61 jours – 3,39 jours – 77,30°D
 - **Lej_Tx** : 29 avril – 16,5°C – 2,56 jours – 3,38 jours – 135,57°C
 - **FT** : 25 avril – 13°C – 2,39 jour – 3,26 jours – 6,27 unités
 - **Q₁₀** : 4 janvier – 3,2 – 2,33 jours – 3,25 jours – 734,83 unités
 - **Sarvas** : 2 janvier – Tm(0) – **1,94 jour** – 3,06 jours – 445,19 unités
-
- **Strasbourg** :
 - **GDD_1** : 30 avril – 8,5°C – 2,39 jours – 3,34 jours – 73,36°D
 - **GDD_2** : 30 avril – 9,5°C – 2,39 jours – 3,34 jours – 63,56°D
 - **Lej_Tx** : 30 avril – 17°C – **1,94 jour** – 2,66 jours – 183,94°C
 - **FT** : 29 avril – 14,5°C – 2,39 jours – 3,50 jours – 5,92 unités
 - **Q₁₀** : 30 avril – 3,6 – 2,28 jours – 3,63 jours – 224,51 unités
 - **Sarvas** : 22 et 30 avril – Tx-Tx(0) et Tm-Tm(0) – 2,56 jours – 3,33 jours – 314,44 et 120,96 unités

A l’instar du bouleau, nous pouvons remarquer que la qualité moyenne des modèles issus des six méthodes, pour une ville donnée, est très proche. Elle n'en est pas moins significativement différente, selon le test t de Student, entre chaque ville. Nous n’enregistrons ainsi aucune différence significative entre les erreurs moyennes absolues d’Amiens et Strasbourg, et celle de cette dernière et de Nancy. Les erreurs de ces trois villes sont les suivantes : 2,06 jours à Amiens, 2,33 à Strasbourg et 2,39 à Nancy. Elles sont significativement inférieures à celles de Paris (3,86 jours) et Rouen (3,51 jours), valeurs symptomatiques de modèles moins précis. La moyenne globale est, quant à elle, de 2,83 jours soit intermédiaire entre celles

enregistrées, sur des périodes plus courtes, pour le frêne et le bouleau. Les dates de démarrage de cumul sont toutes comprises sur une période allant du 15 au 30 avril, à l'exception des meilleurs modèles issus des méthodes du Q_{10} et de Sarvas, où les dates optimales ont lieu début janvier. Cette homogénéité se distingue du caractère plus varié des dates de démarrage de cumul associées au frêne et au bouleau. Quant aux seuils d'action liés à chaque méthode, ils diffèrent d'une ville à l'autre. Nous pouvons cependant voir que la plupart des seuils de température sont supérieurs ou égaux à 10°C et qu'un certain nombre sont proches de 15°C. Cela semble lié aux périodes de cumul qui englobent les températures plus douces d'avril et mai, qui ne sont pas évidemment pas prises en compte pour le frêne et le bouleau. Une fois de plus, les combinaisons date-seuil issues des modèles élaborés à partir des méthodes GDD_1 et GDD_2 sont très proches et mêmes identiques dans le cas de Rouen. Toutes les méthodes semblent conduire à des résultats équivalents, et celle qui donne la plus faible erreur moyenne absolue n'est pas la même d'une ville à l'autre. Trois méthodes se distinguent ici : Sarvas qui procure les erreurs les plus faibles à Amiens et Nancy, GDD_2, FT et Lejoly-Gabriel qui fournissent respectivement les meilleurs modèles à Paris, Rouen et Strasbourg. Devant ce constat, nous allons, une nouvelle fois, déterminer statistiquement si une des méthodes est globalement meilleure que les autres.

3.2.6.3 Comparaisons inter-méthodes et intermensuelles

Nous allons, tout d'abord, présenter graphiquement l'erreur moyenne absolue liée à chaque méthode correspondant à la moyenne des erreurs des meilleures combinaisons des cinq villes de l'étude (figure 43).

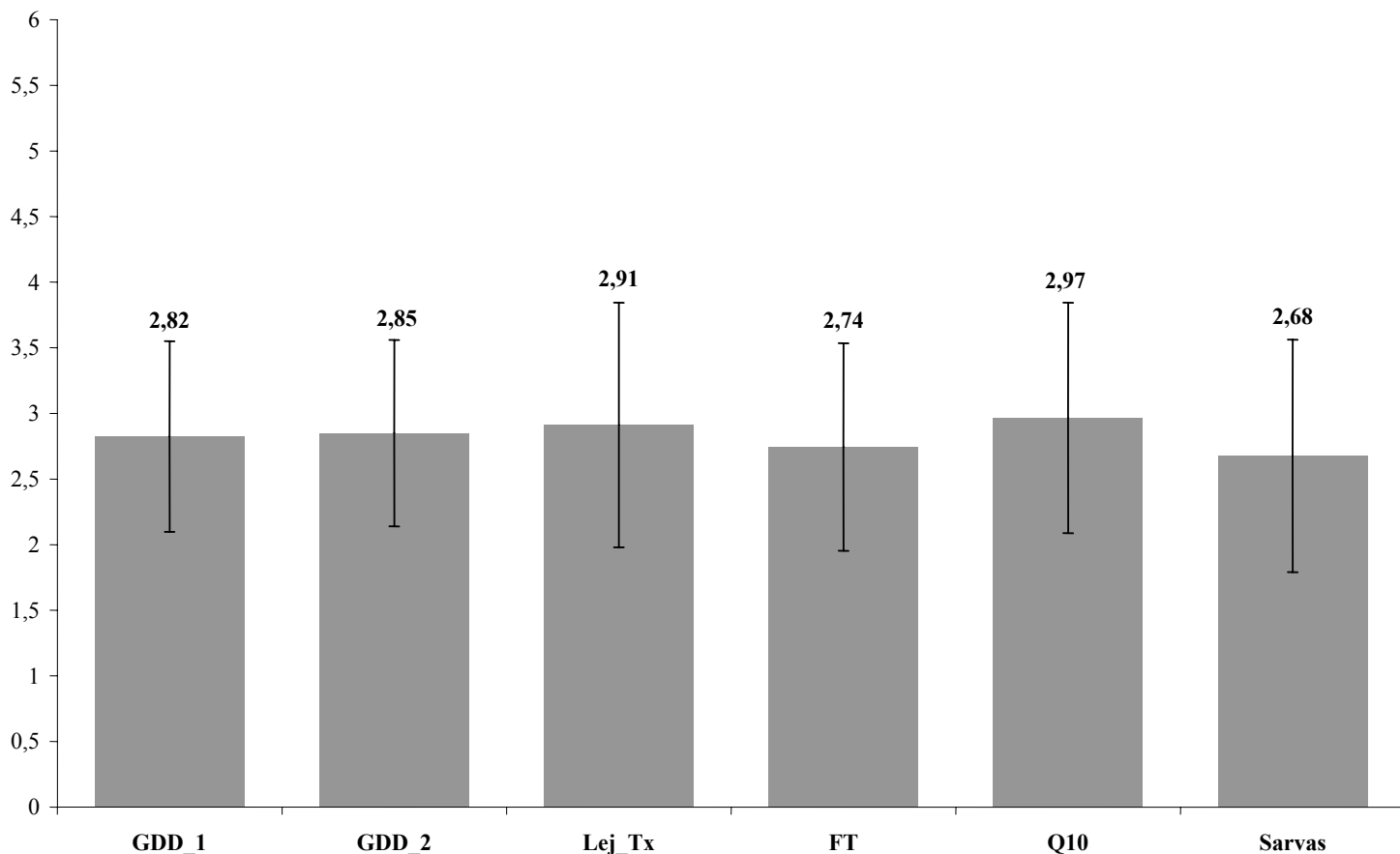


Figure 43 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour les Poacées.

Les moyennes s'échelonnent de 2,68 jours pour la méthode de Sarvas à 2,97 jours pour celle du Q₁₀. Le test t de Student révèle que les moyennes sont statistiquement équivalentes, confirmant l'impression laissée par l'analyse du graphique. Une fois de plus la méthode de Lejoly-Gabriel, dont la méthodologie est la plus simple, donne des modèles aussi efficaces que ceux issus des méthodes plus élaborées. De nouveau, la méthode GDD_2 n'apporte rien de plus par rapport à GDD_1. De ce fait, elle sera également retirée de la suite de l'analyse.

Toutes les dates de démarrage, à deux exceptions près, ont lieu en avril, mois à partir duquel les modèles les plus efficaces semblent être initiés. Pour vérifier cette hypothèse, nous allons comparer les erreurs moyennes absolues de l'ensemble des méthodes ville par ville, mais aussi mois par mois (de janvier à avril), en fonction des combinaisons date-seuil minimisant les écarts entre valeurs observées et prévues pour chacun des mois en question. Nous faisons de même ensuite avec les erreurs moyennes de l'ensemble des villes méthode par méthode. Les résultats, ville par ville, sont présentés dans la figure 44 :

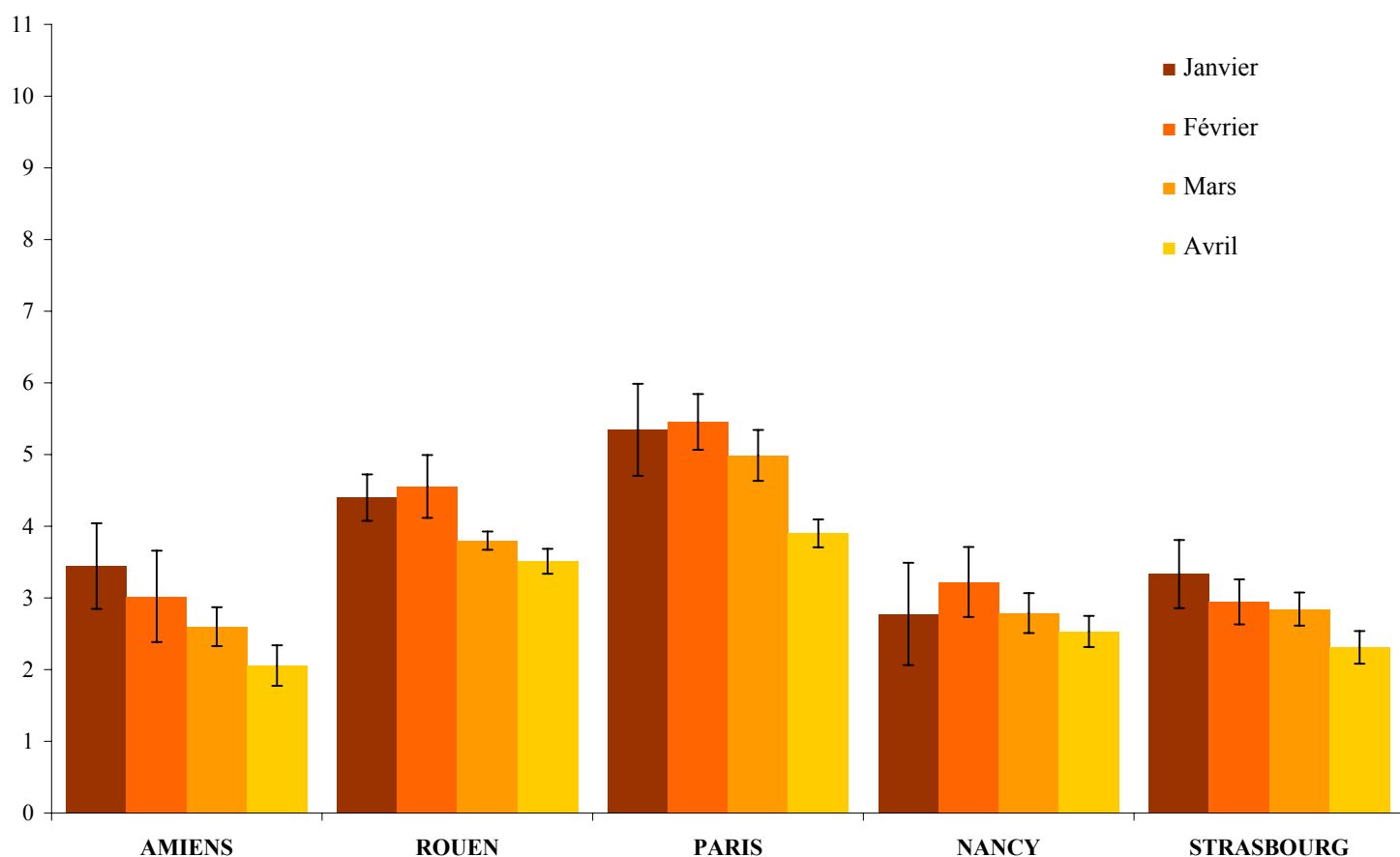


Figure 44 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour les Poacées.

Nous allons analyser les résultats à l'aide du test t de Student pour un seuil de significativité de 95% :

- **Amiens** : Nous constatons une décroissance régulière de l'erreur moyenne absolue au fur et à mesure que la date d'apparition du risque se rapproche : 3,44 jours en janvier, 3,02 en février, 2,60 en mars et 2,06 en avril. Nous enregistrons ainsi des différences significatives entre toutes les combinaisons d'erreurs, excepté entre janvier et février d'une part, et février et mars d'autre part. Ces résultats confirment le caractère optimal des modèles dont les débuts de cumul ont lieu en avril.
- **Rouen** : Seuls les mois de janvier et février enregistrent des erreurs qui ne sont pas significativement différentes. Nous constatons que l'erreur de février est plus élevée que celle de janvier (4,56 contre 4,40 jours). Le démarrage des cumuls en avril débouche sur les modèles les plus fiables comme à Amiens, mais l'erreur

moyenne associée est plus élevée (3,51 jours). Le mois de mars, d'un point de vue statistique, occupe une position intermédiaire au niveau de l'erreur moyenne (3,80 jours).

- **Paris** : L'erreur moyenne absolue associée au mois de janvier n'est pas significativement différente de celles de février et de mars. Les valeurs moyennes de ces trois mois sont élevées, s'échelonnant de 4,99 jours en mars à 5,46 en février. A l'opposé, l'erreur moyenne du mois d'avril est significativement plus faible. Cela démontre l'amélioration de la précision des modèles, malgré une erreur proche de quatre jours (3,90 jours).
- **Nancy** : Selon le test t de Student, seule la différence entre l'erreur moyenne absolue de février et celle d'avril est significativement différente. Mis à part ce cas de figure, toutes les autres combinaisons mensuelles ne présentent pas d'erreurs significativement différentes. Celles-ci s'échelonnent sur une assez faible amplitude, allant de 3,22 jours en février à 2,53 jours en avril.
- **Strasbourg** : Les combinaisons janvier-février, janvier-mars et février-mars ne présentent pas d'erreurs significativement différentes. Le mois d'avril présente l'erreur moyenne la plus faible (2,31 jours) et janvier la plus élevée (3,33 jours). Les démarrages initiés en avril débouchent donc, ici aussi, sur les modèles les plus précis.

D'une manière générale, nous constatons que les démarrages de cumul en janvier et février sont statistiquement équivalents ; puis nous avons une gradation avec une amélioration significative en mars et, encore plus, en avril. Seule la ville de Nancy ne répond pas à cette description. Contrairement à ce qui a été démontré pour le bouleau et le frêne, il semble bien que ce sont les conditions de température des quelques semaines précédant la date d'apparition du risque qui jouent un rôle majeur dans sa survenue. Cela ne remet pas en cause l'influence des températures des mois précédents, mais leur inclusion a tendance à diminuer la précision des modèles. Les dates optimales de démarrage sont systématiquement égales ou postérieures au 15 avril, dont une bonne partie se situe dans les derniers jours d'avril. Cela risque de déboucher, pour des années précoces, à des périodes de cumul de deux semaines à seulement quelques jours. Dans une optique opérationnelle, cela est trop court. Il conviendrait

de se fixer le 15 avril comme date ultime de début de cumul, même si cela ne permet pas de prendre en compte les modèles optimaux de chaque méthode. Cela devient particulièrement pertinent dans le cadre d'une année exceptionnelle, telle l'année 2007, où le risque est apparu dès la fin du mois d'avril pour toutes les villes à l'exception de Rouen. En effet, le premier jour répondant au critère de survenue du risque a été le 25 avril à Nancy, le 26 à Amiens et le 28 à Paris et Strasbourg. Ces caractéristiques exceptionnelles étaient liées aux conditions climatiques inédites du mois d'avril, présentant un ensoleillement et des températures record, dignes d'un mois de juin.

Nous allons nous intéresser à présent aux résultats méthode par méthode par l'intermédiaire, notamment, de la figure 45 :

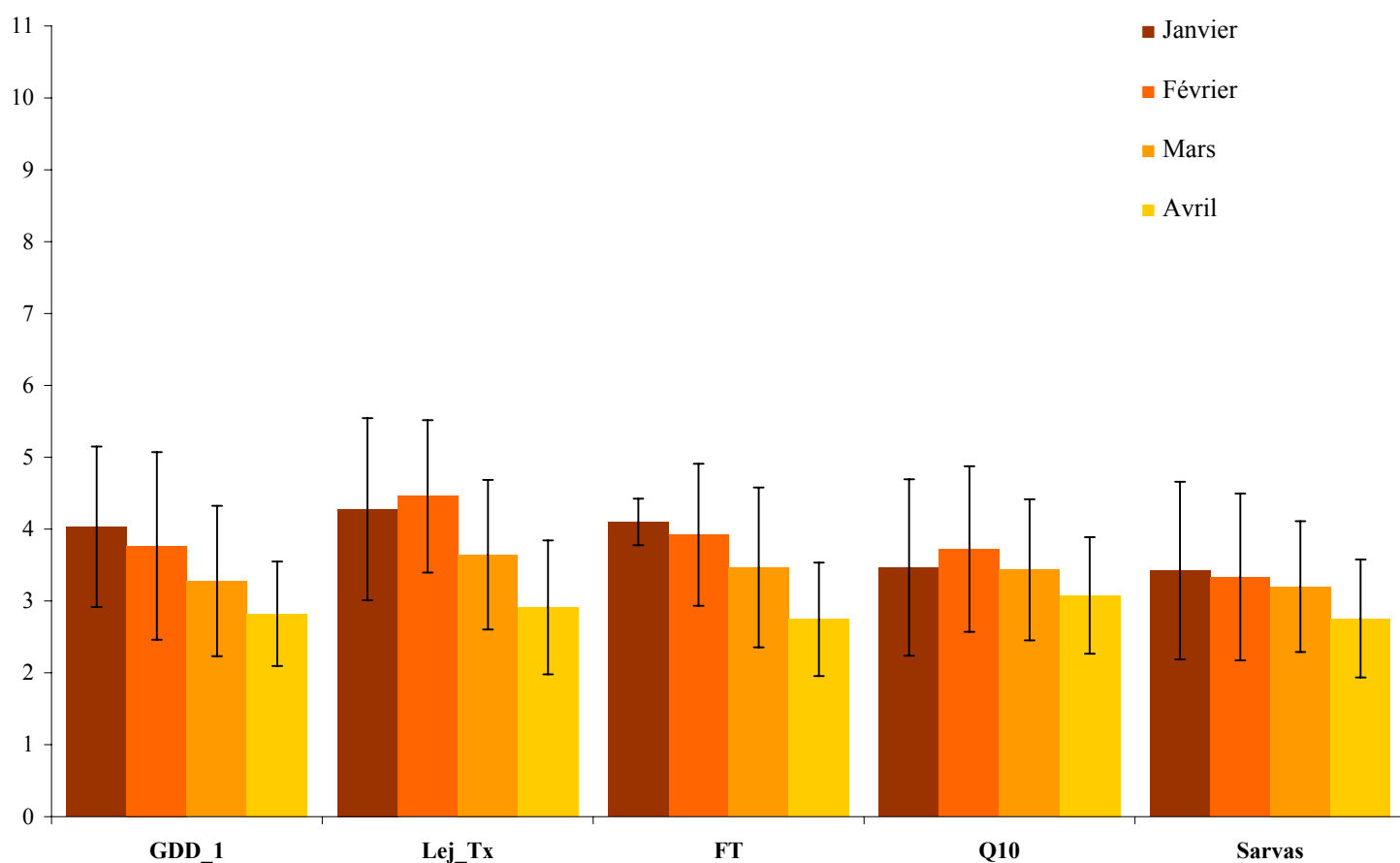


Figure 45 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour les Poacées.

Nous allons analyser les résultats, méthode par méthode, à partir du test t de Student appliqué selon un risque d'erreur de 5% :

- **GDD_1** : Aucune différence significative n'est enregistrée entre les erreurs moyennes des différents mois, à cause, essentiellement, d'écart-types relativement élevés au vu de la valeur brute des erreurs. Ils sont dus, en grande partie, à l'écart constaté entre les erreurs moyennes de Rouen et Paris, élevées, et celles des trois autres villes, plus faibles. Graphiquement nous constatons, cependant, une amélioration progressive de la qualité des modèles, mise en avant par les erreurs moyennes qui s'échelonnent de 4,03 jours en janvier à 2,82 jours en avril.
- **Lej_Tx** : Dans le cas de cette méthode, seuls les mois de février et avril présentent des erreurs moyennes absolues significativement différentes. Le mois de février présente, en effet, l'erreur moyenne absolue la plus élevée (4,46 jours), et avril la plus faible avec 2,91 jours. A l'instar des deux méthodes précédentes, les forts écart-types expliquent essentiellement l'absence de différence significative entre les différentes combinaisons mensuelles, à l'exception d'une.
- **FT** : Aucune différence significative n'est enregistrée à l'exception de la combinaison janvier-avril. Les erreurs moyennes s'échelonnent de 4,10 jours en janvier à 2,74 jours en avril. Nous pouvons constater, à la différence de ce qui a été vu jusqu'à présent, le faible écart-type enregistré en janvier (0,32 jour).
- **Q₁₀** : Comme pour la méthode GDD_1, aucune différence significative n'est enregistrée. L'amplitude entre la moyenne la plus forte et la plus faible est cependant plus réduite que pour les quatre méthodes précédentes, car l'on passe de 3,72 jours en février à 3,08 jours en avril.
- **Sarvas** : Les relations statistiques entre les différents mois sont strictement identiques à celles enregistrées pour les méthodes GDD_1 et du Q₁₀. A l'instar de cette dernière méthode, l'amplitude maximale est plus réduite que pour les quatre méthodes non exponentielles. Ainsi, les erreurs s'échelonnent de 3,42 jours en janvier à 2,76 jours en avril.

Malgré l'absence de différences significatives, due en grande partie aux écarts d'erreur moyenne absolue intra mensuels, nous constatons que le démarrage des cumuls en avril est à

l'origine des meilleurs modèles pour l'ensemble des méthodes. Cela confirme donc ce qui a été vu au niveau de l'étude ville par ville.

3.2.6.4 Bilan global

Nous allons présenter les erreurs moyennes absolues mensuelles des modèles issus de l'ensemble des méthodes pour toutes les villes (figure 46) :

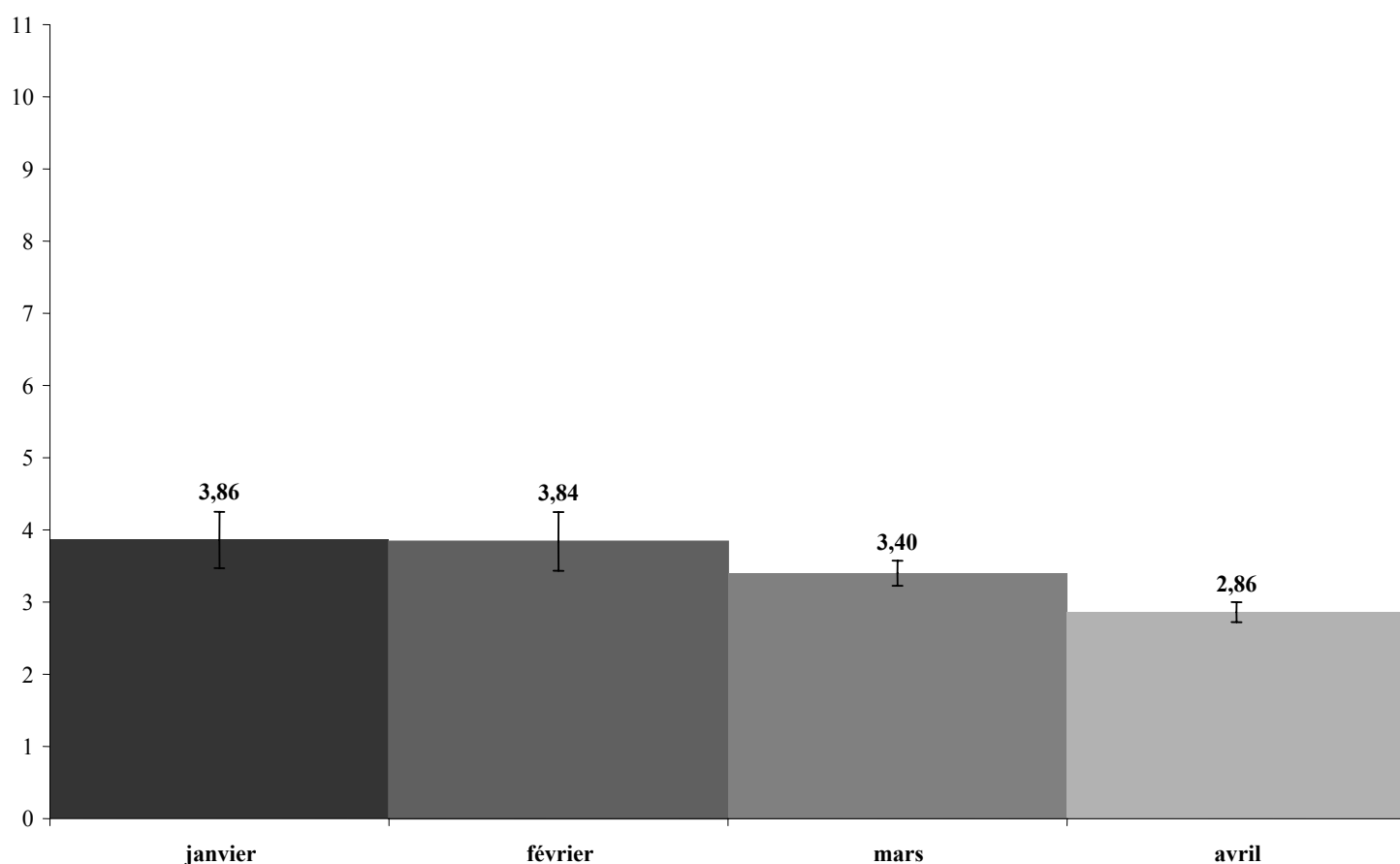


Figure 46 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour les Poacées.

D'après le test t de Student, seules les différences entre les erreurs moyennes de janvier et février, et entre celles de ce dernier mois et mars, ne sont pas significatives. Le mois d'avril apparaît comme la période la plus adéquate pour initier des modèles efficaces. Mais nous pouvons constater que la différence entre les valeurs de janvier et avril est relativement faible (1 jour), montrant que l'utilisation de modèles dont les cumuls sont entamés plus précocement

peut être envisageable. En effet, cela ne débouche pas sur des valeurs aussi médiocres que celles enregistrées pour le bouleau et le frêne, à partir de modèles initiés en octobre et novembre. Mais ici le dernier mois précédant le début de la pollinisation des Poacées se distingue statistiquement des autres, cas de figure qui n'a pas été constaté de manière globale pour le frêne et le bouleau.

3.2.6.5 Etude des résidus

La normalité de la distribution des résidus des différents modèles exposés ci-dessus est déterminée par l'intermédiaire des tests de Shapiro-Wilk et Kolmogorov-Smirnov. Onze modèles présentent des résidus qui ne sont pas distribués de manière normale, et des différences sont enregistrées entre les villes. Alors que seuls les résidus du modèle issu de la méthode du Q_{10} de Strasbourg n'ont pas une distribution gaussienne, quatre des cinq modèles de Paris ne présentent pas de résidus distribués selon la loi normale.

Au vu du faible nombre d'observations (18), les valeurs des deux tests peuvent devenir significatives ou non par le simple fait d'ajouter ou de retirer une variable. Nous pouvons supposer que sur une période plus longue (30-40 ans), les résidus tendront plus souvent vers une distribution gaussienne.

La précision des modèles liés aux Poacées est bonne à l'instar des résultats du bouleau, mais, comme précédemment, la date d'apparition du risque est mal appréhendée certaines années. Nous allons présenter celles où la valeur absolue des résidus dépasse de cinq jours et plus celle des autres années. Les valeurs présentées correspondent à la relation suivante : **valeur prévue – valeur observée**.

- **Rouen :**
 - **GDD_1** : – 16 jours (1994)
 - **Lej_Tx** : – 17 jours (1994)
 - **FT** : – 18 jours (1994)
- **Paris :**
 - **GDD_1** : – 16 jours (1990)
 - **FT** : – 16 jours (1990)
 - **Q₁₀** : – 17 jours (1990)

- **Nancy :**
 - **Q₁₀** : – 10 jours (1990)
 - **Sarvas** : – 10 jours (1990)
- **Strasbourg :**
 - **Q₁₀** : – 11 jours (2005)

Les années répondant au critère évoqué précédemment concernent quatre villes : Rouen, Paris, Nancy et Strasbourg.

- **Rouen** : Les modèles issus de quatre méthodes débouchent sur une prévision systématiquement trop précoce de la date d'apparition du risque de l'année 1994, avec un écart assez considérable de 16 à 18 jours selon les méthodes. La date d'apparition de cette année est la deuxième plus tardive (29 mai) de la série, spécialement si nous la comparons à celles des quatre autres villes (8 au 11 mai). Les modèles ont ici prévu une date oscillant du 11 au 13 mai, et les températures de la deuxième «quinzaine» d'avril et de la première moitié du mois de mai sont légèrement supérieures à la moyenne 1988-2005. Le décalage entre ces dates tardives, et les conditions plutôt favorables à une apparition plus précoce, semblent expliquer l'ampleur des résidus liés à cette année.
- **Paris** : Les modèles issus de quatre méthodes donnent une prévision de 16 à 17 jours trop en avance pour l'année 1990. La date d'apparition (23 mai) se situe plus tardivement que la moyenne 1988-2005 (18 mai) sans pour autant être excessive. Les dates prévues, beaucoup plus précoces (6 au 7 mai), interviennent même plus tôt que la date la plus précoce de la série, le 9 mai, qui a été observée en 1998. Les températures moyennes de la première «quinzaine» de mai sont, avec celles de 2000, les plus élevées de la période, ce qui pourrait expliquer cette précocité de la prévision.
- **Nancy** : Pour les modèles issus des méthodes du Q₁₀ et de Sarvas, la prévision de la date d'apparition du risque en 1990 est trop précoce de dix jours. Cette année se caractérise par une date d'apparition (13 mai) assez proche de la moyenne (10 mai)

et la première «quinzaine» de mai par des températures plus élevées que la moyenne 1988-2005 ; ce qui n'est pas le cas de la deuxième moitié du mois d'avril, plus froide que la moyenne. Au bout de compte l'explication de la précocité de cette date n'est pas évidente ou, du moins, n'est pas caractérisée par une date d'apparition exceptionnelle ou des conditions de température qui le seraient également.

- **Strasbourg** : Le modèle issu de la méthode du Q_{10} donne une prévision de 11 jours trop précoce en 2005. La date prévue par le modèle est le 9 mai contre une apparition réelle du risque le 20 mai. L'année 2005 fait partie des trois années les plus tardives au niveau de l'apparition du risque. Cependant les températures moyennes de la première «quinzaine» de mai, la date optimale de début de cumul étant le 30 avril, sont inférieures à la moyenne 1988-2005. Cela aurait normalement dû présager d'une prévision plus tardive, ce qui n'est pas le cas ici.

A l'instar du bouleau et du frêne, les exemples de mauvaises prévisions exposés ici sont liés à des dates d'apparition du risque qui se distinguent par leur caractère tardif (le cas de Rouen en 1994 et de Strasbourg en 2005) ou précoce, alors que les conditions de température ne sont pas exceptionnelles. Elles sont liées également à des dates d'apparition se situant dans la moyenne, et qui sont mal prévues à cause de conditions de température particulières (grande fraîcheur ou douceur), comme c'est le cas à Paris et Nancy en 1990 où la première «quinzaine» de mai se situe parmi les plus douces de la période d'étude.

3.2.7 Prévision des Poacées par la régression linéaire multiple

3.2.7.1 Etude des corrélations

A l'instar du frêne et du bouleau, nous étudions, tout d'abord, les coefficients de corrélation significatifs, pour un risque d'erreur de 5%, existant entre les valeurs bimensuelles des différentes variables météorologiques et les dates d'apparition du risque. Les périodes testées s'échelonnent de la première «quinzaine» de janvier à la dernière «quinzaine» d'avril. Nous allons présenter les résultats «quinzaine» par «quinzaine», à travers le tableau 11, afin de

déterminer si plusieurs variables sont significativement corrélées avec les dates d'apparition du risque de plusieurs villes.

	AMIENS	ROUEN	PARIS	NANCY	STRASBOURG
01a	Hn Hx Hm				
01b	Tn Tx Tm Tn0 Tx0 Hn Hm JP	Tx0		Tn Tx Tm T-10 T-20 Tn0 Tx0 Tm0 Hn P JP JP1 JP5	Tx Tm Tn0 Tx0 Tm0 JP JP1
02a	Tx Tm Tx0 Tm0	Tx Tm Tx0 Tm0	Tn0 Tm0 Gel	T-10 T-20 Tx0 Tm0	
02b	Hn Hm	T-10 T-20 Tm0			Tx0
03a	Hn Hm				Tn Tm0
03b			Hx		
04a	P JP1	T-20	Tx Tm Insol Hn Hx Hm P JP JP1 JP5		
04b	Tn Tn0		Hx Hm	Tn	

Tableau 11 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique au pollen de Poacées, sur les 18 années communes. En rouge = corrélation positive ; en bleu = corrélation négative ; en gras = variables communes à, au moins, deux villes

Nous avons également déterminé les corrélations partielles issues des régressions pas à pas ascendantes élaborées pour chaque période bimensuelle :

- **Amiens** : JP_03a
- **Rouen** : pas de corrélations
- **Paris** : Tm_04b
- **Nancy** : pas de corrélations
- **Strasbourg** : pas de corrélations

A l'instar du frêne et du bouleau, nous pouvons constater la diversité des variables corrélées et, surtout, qu'une grande partie d'entre elles ne le sont qu'avec les dates d'une seule ville. Cela est certainement dû encore à la faible longueur des séries et montre, de nouveau, la nécessité d'élaborer les modèles ville par ville.

Cependant, nous enregistrons des variables corrélées significativement avec les dates d'apparition du risque pour trois à quatre villes durant la deuxième «quinzaine» de janvier et la première moitié du mois de février. De même, toutes les périodes bimensuelles sont représentées à, au moins, une reprise.

Le plus grand nombre de corrélations est obtenu durant les deux «quinzaines» évoquées précédemment. Nous pouvons ainsi noter l'influence «positive» des températures, ou du

nombre de jours concernés par des valeurs négatives, de la deuxième «quinzaine» de janvier pour l'ensemble des villes, à l'exception de Paris. Plus ces températures sont élevées (ou plus le nombre de jours enregistrant des températures négatives est faible), plus l'apparition du risque est précoce. Ce cas de figure concerne toutes les périodes, ce qui semble supposer l'inexistence d'une période de dormance durant ces quatre mois, période qui a peut-être lieu avant le mois de janvier. Nous aboutissons aux mêmes conclusions avec la première «quinzaine» de février, ainsi qu'avec les autres variables thermiques prises à d'autres périodes. Paris se distingue pour sa part par des corrélations significativement positives entre les températures maximale et moyenne de la première moitié du mois d'avril et les dates d'apparition du risque.

L'humidité relative se caractérise par un rôle opposé à celui des températures, montrant la forte auto-corrélation existant entre les deux paramètres. Ainsi, le fait qu'elle soit significativement corrélée avec l'apparition du risque et que la température ne le soit pas, pour certaines périodes, n'est pas problématique. Du fait de la forte relation entre les deux paramètres, notamment l'humidité relative minimale avec la température maximale, l'interprétation des résultats est très proche. Paris, de ce point de vue, se distingue de nouveau à partir de la mi-mars par des corrélations négatives (logiques au vu du signe des corrélations avec les températures) alors qu'elles sont positives pour toutes les autres villes.

Nous pouvons noter, enfin, que les variables de précipitations corrélées significativement le sont systématiquement négativement. Que ce soit en termes de cumul ou de nombre de jours concernés par ce phénomène, des périodes pluvieuses semblent contribuer à une précocité de l'apparition du risque. Les précipitations semblent ainsi avoir un rôle bénéfique sur le développement des différentes espèces de Poacées, en plus des températures clémentes.

3.2.7.2 Application de validations croisées

Nous rappelons ici que l'objectif des validations croisées est de repérer prioritairement :

- Les variables corrélées significativement à la série de 18 dates grâce à une année exceptionnelle. En d'autres termes, si nous retirons cette année, les corrélations ne sont plus significatives.

- Les variables qui ne sont pas corrélées significativement à ces dates à cause d'une année particulière. En d'autres termes, sans cette année, ces variables sont corrélées significativement.

Dans le premier cas, toutes les validations croisées appliquées à une variable donnée vont déboucher sur des corrélations significatives sauf une : celle où est retirée l'année à l'origine de la relation significative.

Dans le deuxième cas, c'est la situation opposée. Toutes les validations croisées vont déboucher sur des corrélations non significatives excepté une : celle où est retirée l'année à l'origine de cette absence de significativité.

Nous allons tout d'abord nous intéresser au premier cas en présentant les variables où la moyenne des corrélations des validations croisées significatives est plus d'une fois et demi supérieure à la corrélation de la validation croisée non significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Amiens :**
 - Hn_01b (+), Hm_01b (+), Tx0_02a (+), Tm0_02a (+), Hn_03a (+), Tn_04b (-) **(1991)**
- **Rouen :**
 - Tx_02a (-), Tx0_02a (+), Tm0_02a (+) **(1991)**
- **Paris :**
 - Tn0_02a (+), Tm0_02a (+), Gel_02a (+) **(1991)**
- **Nancy :**
 - Tx0_02a (+), Tn_04b (-) **(1991)**

Les conditions de température et/ou d'humidité relative de 1991 de quatre des cinq villes se distinguent. Cela signifie que les corrélations ne sont dues qu'à l'année 1991 qui, une fois retirée, entraîne une non-significativité du coefficient de corrélation entre ces variables et les dates d'apparition du risque. L'année 1991 se caractérise, à la fois, par l'apparition du risque d'exposition allergique le plus tardif, et par des températures très nettement inférieures à la

moyenne de mi-janvier à mi-février et durant la deuxième «quinzaine» d'avril. A l'instar de ce qui a été évoqué pour les deux autres taxons, les corrélations ne sont donc dues qu'à une année exceptionnelle, qu'il convient de conserver dans l'analyse. Comme cela a été montré pour le frêne et le bouleau, les situations exceptionnelles doivent être prises en compte dans l'élaboration des modèles de régression, afin de pouvoir prévoir ces situations et, surtout, inclure, au fur et à mesure, un panel plus large de paramètres «hors-normes», qu'ils soient climatiques ou aérobiologiques.

Nous allons nous intéresser au deuxième cas en présentant les variables où la moyenne des corrélations des validations croisées non significatives est plus d'une fois et demi inférieure à la corrélation de la validation croisée significative ; l'année de cette dernière est indiquée. Le signe du coefficient de corrélation est précisé entre parenthèses :

- **Amiens :**
 - JP_04a (-) **(1991)**
 - JP_02a (-), JP_03a (-) **(1993)**
 - Hx_03a (+), Tm_04b (-) **(1996)**
- **Rouen :**
 - Tn_04a (-), T-10_04a (-), JP_04b (-), JP1_04b (-) **(1991)**
 - JP5_04a (-) **(1994)**
- **Paris :**
 - Tn_01a (-), Tx_01a (-), Tm_01a (-), Tn0_01a (+), JP_01a (-), Hn_02b (-), Hx_02b (-), Hm_02b (-), Insol_03b (+) **(1991)**
 - Hn_01b (+), Hm_01b (+) **(1993)**
 - T-10_02a (-), T-20_02a (-) **(1998)**
- **Nancy :**
 - T-10_04b (-), T-20_04b (-) **(1996)**
 - JP_02b (-), JP1_02b (-), Tn0_04a (+), Tx_04b (-), Tm_04b (-) **(2003)**
- **Strasbourg :**
 - Tn0_03a (+) **(1991)**
 - T-20_01b (-) **(1993)**

- Tx_04a (+) **(1996)**
- P_02b (-), Tn_02b (-), Tn0_02b (+), Tm0_02b (+), T-10_02b (-), T-20_02b (-) **(2003)**

Ces variables significativement corrélées aux dates d'apparition sur 17 années ne le sont plus si l'on inclut l'année indiquée entre parenthèses. La fragilité des corrélations établies est de nouveau mise à jour, mettant en avant la nécessité, comme pour les deux autres taxons, d'actualiser les équations année après année.

Trois années sont, au moins, présentes dans trois villes et leur singularité est à l'origine de la non-significativité de certaines corrélations:

- **1991** : Pour Amiens, Rouen, Paris et Strasbourg, la cause de ce phénomène vient du caractère très tardif de l'apparition du risque dans ces quatre villes. Paris est la ville la plus concernée, avec notamment des corrélations négatives entre les dates d'apparition et les valeurs de températures de la première moitié du mois de janvier sur l'ensemble de la série, moins l'année 1991. Des températures supérieures aux normales saisonnières durant cette période devaient conduire à une apparition précoce du risque. La première «quinzaine» de janvier 1991 s'est caractérisée par des températures supérieures à la moyenne 1988-2005. Mais la date d'apparition du risque très tardive (9 juin) a entraîné la non-significativité des corrélations entre ces variables et les dates d'apparition du risque. L'humidité relative de la deuxième «quinzaine» de février ne s'est plus retrouvée corrélée significativement, et négativement, avec l'inclusion de l'année 1991, car l'humidité relative nettement supérieure à la moyenne 1988-2005 (81,46% d'humidité relative moyenne contre une moyenne de 73,88%) n'a pas correspondu à une apparition précoce du risque.
- **1993** : Les séries d'Amiens, Paris et Strasbourg sont concernées. L'apparition du risque compte parmi les plus précoces à Amiens (8 mai) et Paris (10 mai). Ce caractère particulier est à l'origine de l'absence de corrélations significatives, une fois incluse cette année. Ainsi à Amiens, un nombre important de jours pluvieux durant les premières «quinzaines» de février et mars étaient liés une apparition précoce du risque. Cela n'a pas été le cas en 1993 où ces deux «quinzaines» ont été sèches (2 et 1 jours de précipitations respectivement), n'empêchant pas une survenue précoce du risque. La date enregistrée à Strasbourg (15 mai) se situe dans

la moyenne de la série (12-13 mai). L'absence de corrélation significative enregistrée, en incluant cette année, pour la température moyenne du sol à -20 cm du 16 au 31 janvier, semble être un cas isolé car toutes les températures de l'air de cette période sont corrélées significativement avec les dates d'apparition du risque, 1993 incluse.

- **1996** : Amiens, Nancy et Strasbourg sont concernées par cette année dont l'apparition du risque est la plus tardive des différentes séries avec l'année 1991. Les variables concernées sont différentes d'une ville à l'autre. Nous pouvons seulement constater que les températures de la deuxième «quinzaine» d'avril sont concernées à Amiens et Nancy. Sans l'année 1996, des températures de sol ou de l'air élevées auraient dû déboucher sur une date précoce. Les températures supérieures à la moyenne 1988-2005 de l'année 1996 n'ont pas été suivies d'une telle configuration, rendant la relation non significative.

Comme cela a été évoqué pour le bouleau et le frêne, il convient de tenir compte des années exceptionnelles d'un point de vue météorologique ou aérobiologique, dans l'optique de la mise au point de modèles à visée préventive. Les élaborer à partir, exclusivement, de paramètres «moyens» n'a, soulignons-le de nouveau, aucun intérêt.

3.2.7.3 Equations de régression et résidus

Nous présentons les équations de régression propres à chaque ville où les variables explicatives retenues sont liées à la valeur maximale du test Af. Comme pour le frêne et le bouleau, en plus de ce critère, ne sont retenues que les variables où le signe du coefficient de régression est identique au signe du coefficient de corrélation. De plus, nous nous sommes de nouveau abstenus d'inclure dans les modèles de régression certaines variables Tn0, Tx0 et Tm0 qui présentaient au moins 75% de valeurs nulles.

Conjointement aux équations, nous présentons le coefficient de détermination (R^2), la valeur du test Af (Af), l'erreur moyenne absolue (MAE), l'erreur moyenne au carré (RMSE) et le risque d'erreur du test F (p).

$$\text{REAP_Amiens} = -0,384 \cdot \text{Hm_01a} + 0,180 \cdot \text{Hn_01b} + 6,342 \cdot \text{Tm0_02a} + 0,194 \cdot \text{Hn_02b} - 0,600 \cdot \text{JP1_04a} + 75,426$$

$$R^2 = 0,8433 ; Af = 0,7812 ; MAE = 1,85 \text{ jour} ; RMSE = 2,23 \text{ jours} ; p = 0,0002$$

$$\text{REAP_Rouen} = 3,723 \cdot \text{Tx0_01b} + 7,398 \cdot \text{Tm0_02b} - 0,854 \cdot \text{Hx_04b} + 216,360$$

$$R^2 = 0,6715 ; Af = 0,6096 ; MAE = 3,28 \text{ jours} ; RMSE = 4,56 \text{ jours} ; p = 0,0011$$

$$\text{REAP_Paris} = 1,679 \cdot \text{Tm0_02a} + 0,037 \cdot \text{Insol_04a} - 0,259 \cdot \text{Hx_04b} + 147,488$$

$$R^2 = 0,8265 ; Af = 0,7840 ; MAE = 2,63 \text{ jours} ; RMSE = 3,14 \text{ jours} ; p < 0,0001$$

$$\text{REAP_Nancy} = 1,059 \cdot \text{Tm0_01b} - 1,308 \cdot \text{Tn_04b} + 131,622$$

$$R^2 = 0,7724 ; Af = 0,7326 ; MAE = 2,28 \text{ jours} ; RMSE = 2,77 \text{ jours} ; p < 0,0001$$

$$\text{REAP_Strasbourg} = 0,798 \cdot \text{Tm0_01b} + 1,120 \cdot \text{Tm0_03a} + 127,790$$

$$R^2 = 0,5794 ; Af = 0,5168 ; MAE = 2,82 \text{ jours} ; RMSE = 3,28 \text{ jours} ; p = 0,0015$$

Toutes les équations présentent un test F de Fisher significatif, pour un risque d'erreur de 5%, indiquant que le coefficient de corrélation multiple l'est également. Comme pour le frêne et le bouleau, nous pouvons remarquer la grande variété des régresseurs d'une équation à l'autre due, à la fois, à la présence de corrélations différentes d'une ville à l'autre et au jeu des corrélations partielles à l'intérieur de chaque modèle de régression. L'équation de Strasbourg se compose de variables explicatives dont les coefficients de régression sont positifs. Cela pose problème car l'on ne peut prévoir de dates avant le 8 mai. Il est donc préférable de remplacer le nombre de jours où la température moyenne est inférieure à 0°C de la première «quinzaine» de mars par la température minimale. L'équation est alors la suivante :

$$\text{REAP_Strasbourg} = 0,873 \cdot \text{Tm0_01b} - 0,802 \cdot \text{Tn_03a} + 130,154$$

La présence d'un coefficient de régression négatif permet, à priori, la prévision de n'importe quelle date d'apparition du risque. L'erreur moyenne absolue associée à cette équation est quasiment identique à celle de l'équation précédente avec une valeur de 2,86 jours.

Quant aux résidus, ils sont distribués normalement pour l'ensemble des séries, selon les résultats des tests de Shapiro-Wilk et Kolmogorov-Smirnov. A Amiens, Paris, Nancy et

Strasbourg, la valeur absolue des résidus est inférieure à six jours. A Rouen nous enregistrons, au pire, une prévision de dix jours (10,41 exactement) trop précoce en 1994 et une de 11,63 jours trop tardive en 1999. L'année 1994 se caractérise pour cette ville par une apparition tardive du risque et 1999 par une apparition particulièrement précoce (3 mai).

3.2.8 Conclusion

A travers l'exemple de ces cinq villes, nous avons d'abord montré que les six méthodes de cumul fournissent des résultats statistiquement équivalents, que ce soit pour la prévision de la date d'apparition du risque allergique d'exposition aux pollens de frêne, de bouleau ou de Poacées. La méthode GDD_2 n'ayant rien apporté de plus par rapport à GDD_1, elle a été retirée de l'analyse. Ainsi, si nous devons retenir une seule méthode, nous choisirions celle de Lejoly-Gabriel, qui ne consiste qu'à cumuler les températures maximales au-dessus d'un seuil donné, car c'est méthodologiquement la plus simple de toutes. Elle s'est avérée aussi efficace que les autres méthodes employant des formules plus complexes. Cela démontre que la simplicité d'une formule n'est pas en contradiction avec la recherche de modèles de prévision précis. Nous avons démontré également que la période optimale de démarrage des cumuls, pour le frêne et le bouleau, était entamée en décembre, semblant indiquer que la fin de la dormance de ces deux genres intervient durant ce mois. Cependant, les meilleurs modèles ont été le plus fréquemment initiés en février. Dans le cas des Poacées, le mois d'avril s'est distingué des autres par l'erreur moyenne absolue significativement plus faible des modèles associés. La régression linéaire multiple, permettant l'inclusion d'autres variables que la température de l'air, est une alternative intéressante par la proposition d'équations dont les erreurs moyennes absolues sont proches de celles des meilleurs modèles issus des différentes méthodes de cumul.

Par cette étude, nous avons pu constater le poids des années exceptionnelles dans l'élaboration des modèles de régression, que ce soit par le caractère précoce ou tardif de l'apparition du risque ou par des conditions climatiques sortant de l'ordinaire. Cette influence a tendance à générer une instabilité des corrélations d'une année sur l'autre du fait, en très grande partie, de la faible longueur des séries. Mais il est essentiel d'inclure ces années si nous voulons les prévoir le plus précisément possible. Plus généralement, à propos des méthodes de cumul, nous constatons de mauvaises prévisions quand la date d'apparition, trop précoce ou trop tardive, n'est pas en phase avec les températures qui la précèdent. Dans le

même ordre d'idées, si nous enregistrons des températures exceptionnellement fraîches, ou douces, et que la date d'apparition du risque est «moyenne», cette dernière est également mal estimée.

Le dernier point, lié à celui précédemment évoqué, est la difficulté d'élaborer des modèles globaux du fait de la faible longueur des séries. Celles-ci ne démarrant au plus tôt qu'en 1987, il est difficile de retrouver les mêmes variables météorologiques corrélées significativement, dans le cas de la régression multiple, d'une ville à l'autre. La constitution progressive de séries plus longues permettra, théoriquement, de résoudre ce problème. De même, en ce qui concerne les méthodes de cumul, des séries plus longues pourraient déboucher sur des seuils d'action proches d'une ville à l'autre ou corrélés aux caractéristiques géographiques et météorologiques des villes concernées.

Suite à cette étude, nous allons présenter les résultats des modèles issus des méthodes de cumul et de la régression linéaire multiple pour l'ensemble des années disponibles d'un certain nombre de villes françaises.

3.3 PRESENTATION DES RESULTATS POUR LES SERIES POLLINIQUES LES PLUS LONGUES

Nous allons, dans ce chapitre, évoquer les résultats des différentes méthodes de cumul (dont les sigles correspondent à ceux utilisés dans le chapitre 3.2) et de la régression linéaire multiple pour les stations présentant des séries de 15 ans et plus. Il s'agit d'effectuer une revue exhaustive des meilleures combinaisons date-seuil et équations de régression, taxon par taxon. Pour chaque modèle, nous allons présenter la date de début de cumul, le seuil d'action, l'erreur moyenne absolue et le cumul moyen à approcher le plus possible. Rappelons que les dates prévues par les méthodes de cumul correspondent aux jours précédant immédiatement ceux où les cumuls moyens sont dépassés, et que les seuils d'action testés sont ceux présentés dans la section 3.2.2.1. Ainsi, pour ces dernières méthodes, cet inventaire a pour but de faire ressortir les dates de démarrage des accumulations d'unités de «chaleur» et les seuils d'action qui sont à l'origine des erreurs moyennes absolues les plus faibles. De ce fait, dans le cadre d'une année précoce, des dates de démarrage peu ou pas exploitables, car initiées trop tardivement, seront susceptibles d'être présentées. Dans ce dernier cas de figure, une avancée de deux à trois semaines de la date-butoir, si la pollinisation est précoce, serait nécessaire dans le cadre de modèles opérationnels et, comme cela a été évoqué auparavant (cf. 3.2.2.4, 3.2.4.4 et 3.2.6.4), il serait intéressant de tester plusieurs modèles initiés à des dates différentes. Mais ici nous nous bornerons à une présentation purement statistique des résultats, avec une détermination prospective de la date d'apparition du risque en 2006 à partir des modèles ne présentant pas d'erreurs moyennes trop élevées (supérieures à six ou sept jours). Notons, par ailleurs, que seules les méthodes de Sarvas, du Q_{10} et de Lejoly-Gabriel sont retenues pour certaines stations. Ce choix s'est justifié par rapport aux résultats du chapitre 3.2 (ces trois méthodes sont statistiquement équivalentes aux autres en terme d'erreur moyenne absolue) et par leur plus grande facilité de mise au point. Pour les sites de Rouen, Paris, Amiens, Nancy et Strasbourg, un parallèle sera établi entre les seuils, dates de démarrage de cumul et erreurs moyennes absolues issues des modèles élaborés sur les années communes aux cinq villes – qui seront présentés entre parenthèses – et ceux déterminés à partir de toutes les années disponibles. Enfin, il convient de signaler que la normalité de la distribution de toutes les séries de dates utilisées, dans ce chapitre, a été testée par l'intermédiaire des tests de Shapiro-

Wilk et Kolmogorov-Smirnov, pour un risque d'erreur de 5%. D'après les résultats découlant de ces applications, toutes les séries suivent une distribution gaussienne.

Nous allons entamer la présentation par le frêne.

3.3.1 Résultats pour le frêne

Les résultats sont exposés ville par ville avec, tout d'abord, une visualisation graphique des dates d'apparition du risque. Il s'ensuit, pour les méthodes de cumul, une présentation des combinaisons date-seuil issues des modèles où l'erreur moyenne absolue est minimisée. Sont ensuite abordées, dans le cadre de la régression linéaire multiple, les équations en découlant. Enfin, nous présentons les prévisions pour 2006 à partir des différents modèles. Notons que chaque date prévue pour cette année, à partir des méthodes de cumul et des équations de régression respectivement, correspond :

- A la veille du jour où les cumuls moyens associés à chaque modèle sont dépassés.
- Au $x^{\text{ième}}$ jour de l'année déterminé par chaque équation, dont la valeur est arrondie à l'unité près.

Nous procéderons de la même manière pour le bouleau, les Poacées et l'ambroisie.

Certaines méthodes de cumul sont à l'origine de modèles présentant plusieurs seuils, ou plusieurs dates de démarrage, différents. Les dates prévues pour 2006 sont toujours identiques sauf, quelquefois, dans le cas du Q_{10} où un décalage de un jour peut être constaté quand plusieurs seuils sont retenus. Dans ce cas nous choisissons de présenter la date déterminée à partir du plus faible de ces seuils. Ce choix concernera également les modèles de prévision du bouleau, des Poacées et de l'ambroisie.

3.3.1.1 Rouen

Dix-sept années sont disponibles pour cette ville de 1988 à 2005, l'année 1990 étant retirée. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 47.

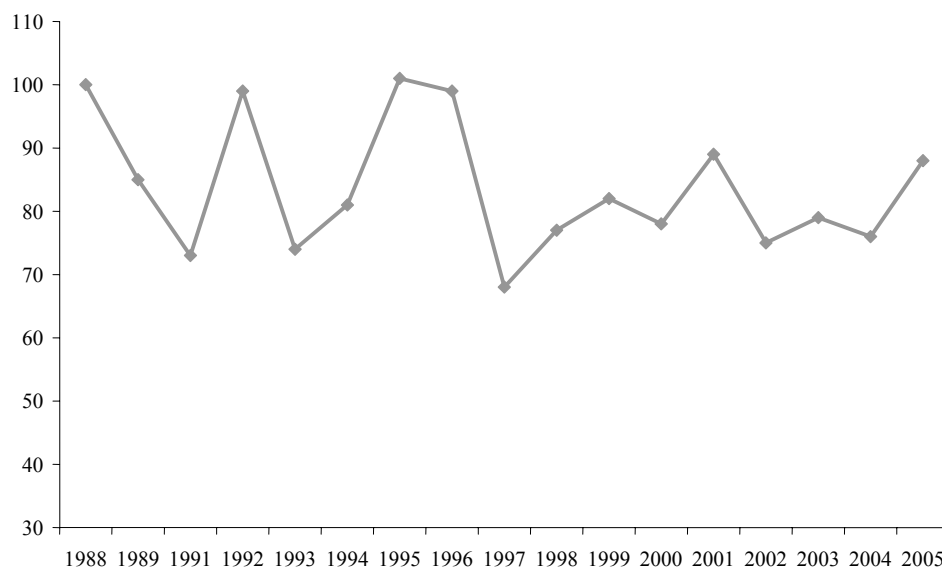


Figure 47 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Rouen, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 83,8^{ème} jour de l'année, soit le 25 mars pour une année non bissextile. Les dates s'échelonnent du 9 mars, en 1997, au 11 avril en 1995, soit une amplitude maximale de 33 jours. Nous remarquons qu'à partir de 1997 la variation interannuelle de l'apparition du risque est plus faible qu'au début de la période, où des dates particulièrement tardives ont été enregistrées pour plusieurs années.

Rappelons, par ailleurs, que nous testons les dates de démarrage du 1^{er} octobre au 28 février.

- **GDD_1** : 21 et 22 février (30 décembre) – 12 °C (11°C) – 4,59 jours (3,18 jours) – 4,55°D
- **Lej_Tx** : 16 février (6 janvier) – 9°C (11,5°C) – 5,06 jours (4,36 jours) – 272,91°C
- **FT** : 5 février (6 janvier) – 13°C (13°C) – 5,12 jours (3,09 jours) – 1,74 unité
- **Q₁₀** : 16 février (30 décembre) – 8,8 à 9,5 (12) – 5,12 jours (5,45 jours) – 531,62 à 580,86 unités
- **Sarvas** : 18 février (8 janvier) – Tm(0) (Tn(0)) – 4,88 jours (6,45 jours) – 113,93 unités

L'ajout de six années entraîne des dates de démarrage de cumul optimales différentes, qui ont toutes lieu en février contrairement aux modèles élaborés sur onze années. Cela met en avant l'instabilité certaine affectant les modèles élaborés sur de courtes séries temporelles. La

moyenne des erreurs absolues sur les cinq méthodes est plus élevée en tenant compte de 17 années (5,07 jours) que de 11 (4,51 jours).

L'équation issue de l'application de la régression linéaire multiple est la suivante, où sont repris les sigles présentés dans le sous-chapitre 3.1.2. Le coefficient de détermination, le résultat du test Af, l'erreur moyenne absolue et le risque d'erreur du test F s'établissent comme suit :

$$\text{REAP_Rouen} = -1,180 \cdot \text{JP1_11b} + 2,079 \cdot \text{T-20_11b} + 0,381 \cdot \text{Hn_12b} - 0,621 \cdot \text{Tm0_01a} + 1,096 \cdot \text{JP1_02a} + 39,862$$

$$R^2 = 0,7179 ; \text{Af} = 0,6293 ; \text{MAE} = 4,61 \text{ jours} ; p = 0,0083$$

Les régresseurs sont complètement différents de ceux de l'équation établie sur 11 années. Les variables JP1_11b, T-20_11b (corrélations partielles) et JP1_02a, présentes dans l'équation ci-dessus, ne sont pas significativement corrélées pour un risque d'erreur de 5% avec les dates d'apparition du risque de la série de 11 ans. Par contre, Hn_12b et Tm0_01a sont corrélées dans les deux cas de figure, mais elles n'ont pas été retenues dans l'équation sur les 11 années par le jeu des corrélations internes existant entre les différentes variables météorologiques. A l'instar des modèles issus des méthodes de cumul, l'équation élaborée sur 17 années est moins précise que celle mise au point sur 11 ans.

En 2006, l'apparition d'un risque allergique d'exposition au pollen de frêne a eu lieu le 1^{er} avril. Les modèles issus des méthodes GDD_1 (28 mars), FT (29 mars), Q₁₀ et Sarvas (1^{er} avril pour ces dernières) ont donné une prévision exacte ou légèrement précoce. La prévision par les modèles issus de la méthode Lej_Tx s'est révélée moins précise et, surtout, trop tardive (7 avril). Quant à la date prévue par l'équation, elle est tombée le 21 mars et s'est avérée, par conséquent, beaucoup trop précoce.

3.3.1.2 Amiens

Nous disposons conjointement des variables météorologiques et des données polliniques sur la période 1989-2005. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 48.

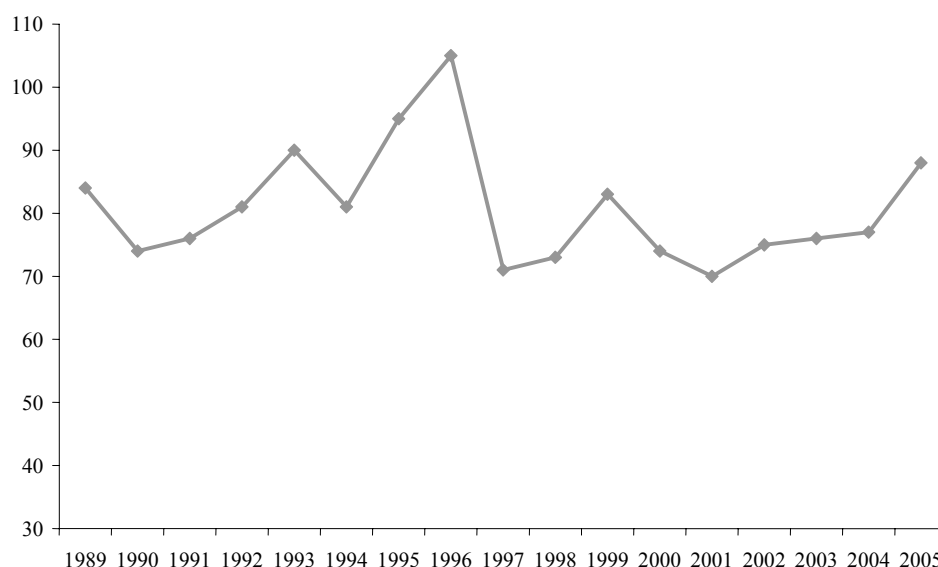


Figure 48 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Amiens, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 80,8^{ème} jour de l'année, soit le 22 mars pour une année non bissextile. Les dates s'échelonnent du 11 mars, en 2001, au 14 avril en 1996, soit une amplitude maximale de 35 jours. Il convient de noter qu'aucune tendance significative à la précocité (ou au caractère plus tardif) des dates d'apparition n'est enregistrée sur cette période.

Les dates de démarrage testées s'échelonnent du 1^{er} octobre au 28 février.

- **GDD_1** : 24 février (27 février) – 6,5°C (7,5°C) – 3,59 jours (3,09 jours) – 53,71°D
- **Lej_Tx** : 8 février (3 février) – 11°C (12,5°C) – 4,29 jours (3,91 jours) – 239,20°C
- **FT** : 27 février (26 février) – 10°C (10,5°C) – 3,35 jours (2,36 jours) – 6,15 unités
- **Q₁₀** : 3 février (3 février) – 11,7 (4,1 et 4,2) – 3,94 jours (3,91 jours) – 997,78 unités
- **Sarvas** : 26 février (28 février) – Tn (Tn) – 3,76 jours (3,64 jours) – 48,69 unités

Contrairement à Rouen, la variation des dates de démarrage de cumul et des seuils est très faible, à l'exception du modèle issu du coefficient de Q₁₀, entre les modèles élaborés sur 17 et 11 années. A l'instar de Rouen, la précision globale des modèles élaborés sur 17 années est moindre (erreur moyenne de 3,79 jours) que celle élaborée sur les 11 années (3,38 jours).

Quant à l'équation de régression, elle se décline de la manière suivante :

$$\text{REAP_Amiens} = -1,641 * \text{JP_10a} - 1,526 * \text{JP1_11a} + 103,087$$

$$R^2 = 0,6753 ; Af = 0,6200 ; MAE = 4,28 \text{ jours} ; p = 0,0004$$

Il faut noter que seules trois variables sont significativement corrélées ($p = 0,05$) avec les 17 dates : les deux régresseurs ainsi que JP_11a. JP_10a et JP_11a étaient déjà significativement corrélées sur 11 ans, mais toutes les autres variables corrélées sur ce dernier pas de temps (cf. 3.2.3) ne le sont plus ici. L'ajout d'années a dégradé la qualité de la «meilleure» régression, ce qui montre, une nouvelle fois, la fragilité des corrélations.

En 2006, l'apparition d'un risque allergique d'exposition au pollen de frêne a eu lieu le 2 avril. Les modèles issus des méthodes GDD_1, FT et Sarvas ont donné des dates proches s'échelonnant du 28 au 30 mars. En ce qui concerne le Q₁₀ et Lej_Tx, les dates prévues ont été un peu trop tardives (4 et 7 avril respectivement). Les modèles, issus de ces deux dernières méthodes, se singularisent des autres par une date de démarrage de cumul début février. Il semble que le moment de la date de démarrage a dû quelque peu influencer le caractère un peu trop précoce ou tardif de la prévision. Quant à la date estimée à partir de l'équation, elle a eu lieu le 30 mars et, étant donné le laps de temps entre les dates d'apparition et les périodes des deux régresseurs, ce résultat est à utiliser avec précaution, malgré la bonne prévision.

3.3.1.3 Paris

Nous disposons des années 1988 à 2005, à l'exception de 1998 et 2000, soit un total de 16 ans. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 49.

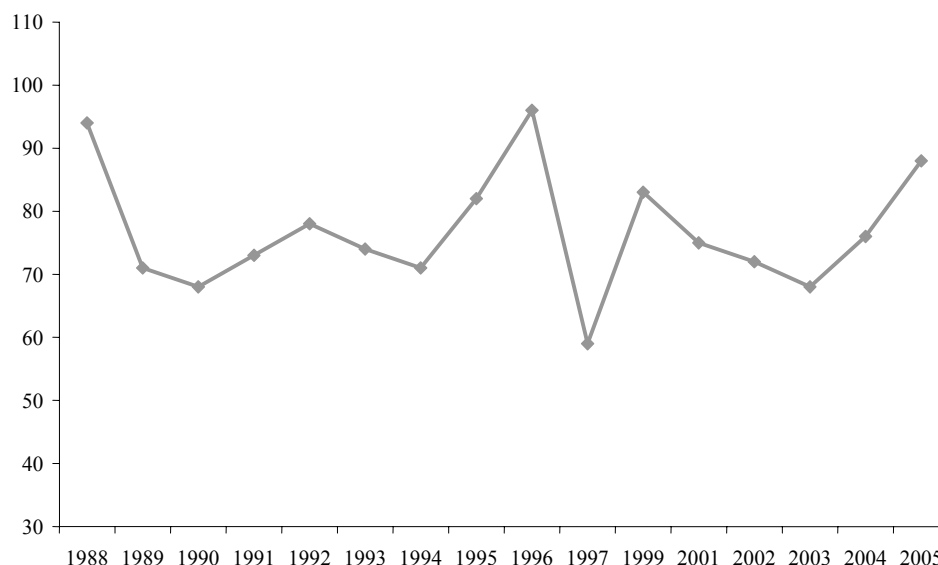


Figure 49 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Paris, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 76,8^{ème} jour de l'année, soit le 18 mars pour une année non bissextile. Les dates s'échelonnent du 28 février, en 1997, au 5 avril en 1996, soit une amplitude maximale de 37 jours à un an d'intervalle.

Les dates de démarrage testées s'étendent du 1^{er} octobre au 27 février.

- **GDD_1** : 18 décembre (13 décembre) – 9°C (10°C) – 3,25 jours (2,45 jours) – 57,97°D
- **Lej_Tx** : 28 décembre (29 décembre) – 12,5°C (12°C) – 3,06 jours (2,18 jours) – 244,41°C
- **FT** : 18 décembre (16 au 18 décembre) – 11°C (11°C) – 3,44 jours (2,64 jours) – 11,74 unités
- **Q₁₀** : 21 février (19 décembre) – 10,2 à 12 (10,3) – 3,88 jours (4,00 jours) – 664,43 à 837,47 unités
- **Sarvas** : 3 février (27 février) – Tx(0) (Tm(0)) – 4,56 jours (4,36 jours) – 261,75 unités

Comme remarque majeure, nous pouvons noter un grand décalage, entre les dates de démarrage optimales pour les modèles issus du Q₁₀. Nous avons ainsi trois modèles démarrant en décembre, et les deux autres en février, à partir des 16 années disponibles. La moyenne des erreurs absolues est plus élevée sur 16 années (3,64 jours) que sur 11 (3,13 jours). Cependant,

nous constatons que cela ne concerne pas les méthodes du Q₁₀ et de Sarvas où les valeurs sont très proches entre les deux pas de temps. Cela rejoint les constatations que nous pouvons faire sur les modèles issus de ces deux méthodes pour Amiens et Rouen. Dans cette dernière ville les erreurs moyennes, pour les séries de 17 ans, sont plus faibles que celles constatées sur celles de 11 ans.

L'équation issue de la régression linéaire multiple est la suivante :

$$\text{REAP_Paris} = 1,967 \cdot \text{Tm0_12a} + 0,919 \cdot \text{Tn_01a} + 2,141 \cdot \text{JP5_01b} + 6,920 \cdot \text{Tm0_01b} + 1,306 \cdot \text{T-20_01b} + 58,683$$

$$R^2 = 0,9045 ; Af = 0,8365 ; MAE = 2,07 \text{ jours} ; p = 0,0001$$

L'erreur moyenne absolue est relativement proche de celle de l'équation établie sur 11 ans (1,65 jour). Les deux équations, composées de quatre régresseurs chacune, n'ont en commun que la variable JP5_01b. Les variables Tm0_12a, Tm0_01b et T-20_01b sont spécifiquement corrélées avec la série de 16 ans, les deux dernières étant issues de corrélations partielles. Cela montre une nouvelle fois la variabilité des corrélations quand est ajouté un certain nombre d'années. Dans le cas de Paris cela n'empêche cependant pas l'obtention de deux équations expliquant, chacune, plus de 90% de la variance.

En 2006, l'apparition d'un risque allergique d'exposition au pollen de frêne a eu lieu le 31 mars. Les modèles issus du Q₁₀ et de la méthode de Sarvas ont donné d'excellentes prévisions avec, respectivement le 30 et le 31 mars. A l'opposé, les trois derniers modèles ont donné des dates beaucoup trop tardives (11 et 12 avril) et ainsi amplifié le caractère déjà tardif de cette année 2006. Nous nous rendons rapidement compte que les «bons» modèles de cumul ont tous été initiés en février, et les trois autres en décembre. Un démarrage durant ce dernier mois ne s'est pas avéré adéquat en 2006, sans pour autant affirmer que ce sera systématiquement le cas pour d'autres années. Le caractère très tardif de la prévision s'explique par les conditions de l'hiver 2005-2006, parmi les plus rigoureuses de ces 15 dernières années. La date effective d'apparition du risque a été plus tardive de 13 jours par rapport à la moyenne, mais pas autant que ce qu'auraient laissé présager les conditions de température. Cela montre en tout cas que la date optimale de démarrage a joué un rôle important sur la qualité d'une prévision, et confirme qu'il est essentiel de mettre à jour les modèles, année après année. Néanmoins, le caractère tardif de l'apparition du risque en 2006 a été systématiquement repéré. Avec une date prévue tombant le 29 mars, l'équation de

régression est à l'origine d'une très bonne estimation, qui s'est avérée plus précoce que la date réelle de deux jours.

3.3.1.4 Nancy

A l'instar de Rouen, nous disposons des années 1988 à 2005, à l'exception de 1990, soit une série de 17 années. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 50.

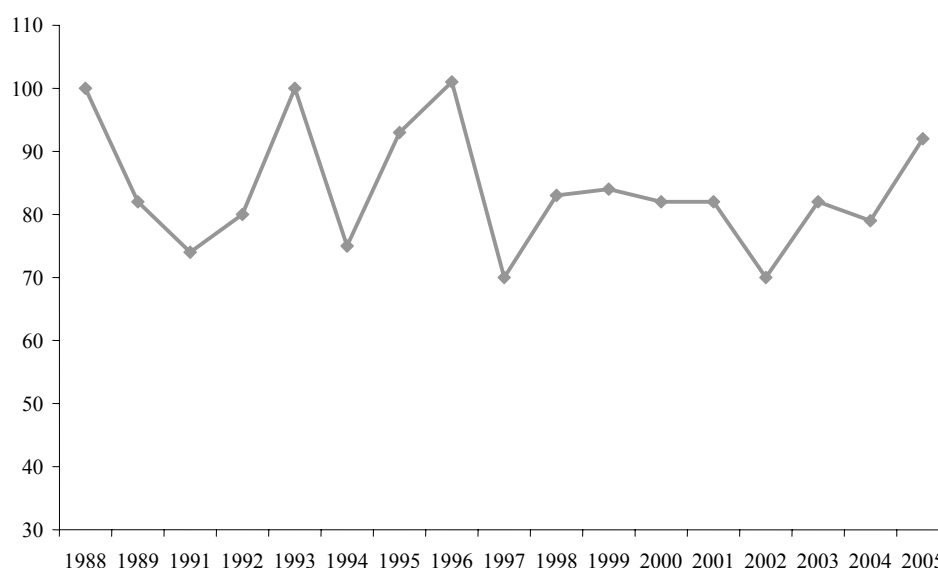


Figure 50 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Nancy, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition du risque a lieu le 84,1^{ème} jour de l'année, soit le 25 mars dans le cadre d'une année non bissextile. Les dates s'échelonnent du 11 mars, en 1997 et 2002, au 10 avril en 1996, soit une amplitude maximale de 31 jours.

A l'instar de Rouen et d'Amiens, la borne supérieure des dates de début de cumul testées est le 28 février.

- **GDD_1** : 20 février (13 février) – 4°C (5°C) – 2,94 jours (3,73 jours) – 104,64°D
- **Lej_Tx** : 27 janvier (27 janvier) – 10,5°C (10°C) – 3,41 jours (3,73 jours) – 288,36°C
- **FT** : 19 février (20 février) – 8,5°C (6,5°C) – 2,94 jours (3,91 jours) – 10,61 unités

- **Q₁₀** : 29 janvier (27 janvier) – 6,8 et 6,9 (7,7) – 3,35 jours (3,73 jours) – 564,85 à 574,75 unités
- **Sarvas** : 21 février (20 février) – T_m(0) (T_m(0)) – 3,47 jours (4,00 jours) – 102,68 unités

Les variations de dates et de seuils sont plutôt faibles pour l'ensemble des méthodes. Contrairement aux trois villes précédentes, la moyenne des erreurs absolues des modèles élaborés sur 17 années est plus faible (3,22 jours) que celle calculée à partir des modèles sur 11 ans (3,82 jours).

L'équation de régression est la suivante :

$$\text{REAP_Nancy} = 0,893 \cdot \text{Hx_12b} + 1,879 \cdot \text{Tx0_01b} + 4,623 \cdot \text{T-10_01b} + 35,307 \cdot \text{Tx0_02b} - 19,026$$

$$R^2 = 0,6988 ; \text{Af} = 0,6222 ; \text{MAE} = 4,30 \text{ jours} ; p = 0,0039$$

Les quatre variables explicatives sont complètement différentes des quatre de l'équation élaborée sur 11 ans et l'erreur moyenne absolue y est nettement supérieure (4,30 contre 1,49 jour). A l'instar d'Amiens, il n'y a pas d'amélioration de la qualité des modèles en tenant compte d'années supplémentaires dans l'élaboration de modèles. Nous pouvons également remarquer que les variables Hx_12b et Tx0_01b, cette dernière étant issue d'une corrélation partielle, sont spécifiquement corrélées à la série de 17 ans.

En 2006, la date d'apparition du risque est intervenue tard dans la saison (8 avril). Seul le modèle issu de la méthode de Lejoly-Gabriel (initié fin janvier) a fourni une très bonne prévision qui s'est avérée exacte. Les quatre autres ont donné des dates trop précoces oscillant du 31 mars au 3 avril qui se rapprochent des dates observées à Rouen, Amiens et Paris. L'équation de régression a débouché, quant à elle, sur une date de prévision complètement erronée (19 mars), rejoignant, au niveau «qualitatif», la prévision établie pour Rouen.

Après avoir présenté ces quatre stations qui ont fait l'objet des sous-chapitres 3.2.2 et 3.2.3, les séries polliniques de Strasbourg ne comptant que 13 années exploitables n'étant pas évoquées, nous allons présenter les résultats pour d'autres villes fournissant au moins une série de 15 années où les dates d'apparition du risque lié au frêne sont exploitables. Ne sont concernées ici que Lyon et Bordeaux qui sont traitées conjointement. Montluçon, qui présentait le nombre d'années requis, n'a pas été retenu faute de données météorologiques

locales fiables. Pour Bordeaux, seules les méthodes du Q₁₀, de Sarvas et de Lejoly-Gabriel ont été exploitées pour les raisons évoquées dans l'introduction du chapitre 3.3. Etant donné que quatre taxons ont été étudiés à Lyon, nous avons employé les cinq méthodes de cumul.

3.3.1.5 Lyon et Bordeaux

Quinze années sont exploitables à Lyon de 1988 à 2005 inclus, 1997, 2000 et 2002 étant retirées de l'analyse. Il s'agit du même nombre d'années à Bordeaux mais réparties en deux périodes : 1988-1992 et 1996-2005. Ces séries sont mises en parallèle avec les données météorologiques, des années correspondantes, de Lyon-Bron et Bordeaux-Mérignac. Les conditions météorologiques de la période 1961-1990 sont brièvement évoquées à l'instar de ce qui a été fait dans le sous-chapitre 3.2.1. La température moyenne hivernale (décembre à février) est ainsi plus douce à Bordeaux (6,4°C) qu'à Lyon (3,4°C) alors que la température moyenne estivale (juin à août) est équivalente entre les villes (19,3 et 19,6°C respectivement). Les précipitations annuelles, assez abondantes, sont du même acabit (923 mm à Bordeaux et 824,5 à Lyon) ainsi que l'insolation, supérieure à la moitié Nord de la France, qui atteint 2083,7 heures par an à Bordeaux et 1755,7 heures à Lyon. Les dates d'apparition du risque d'exposition allergique à Lyon (figure 51) et Bordeaux (figure 52) sont présentées ci-après.

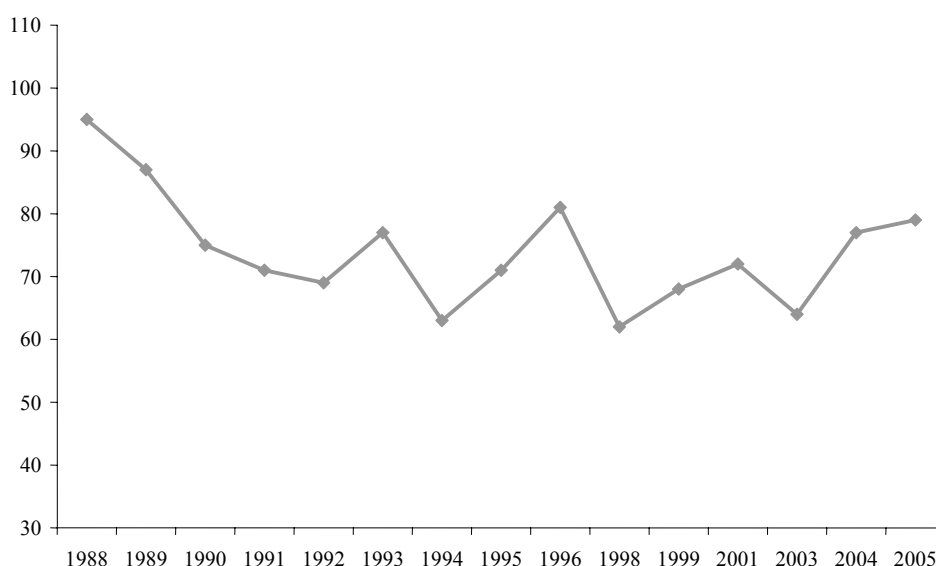


Figure 51 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Lyon, en nombre de jours à partir du 1^{er} janvier.

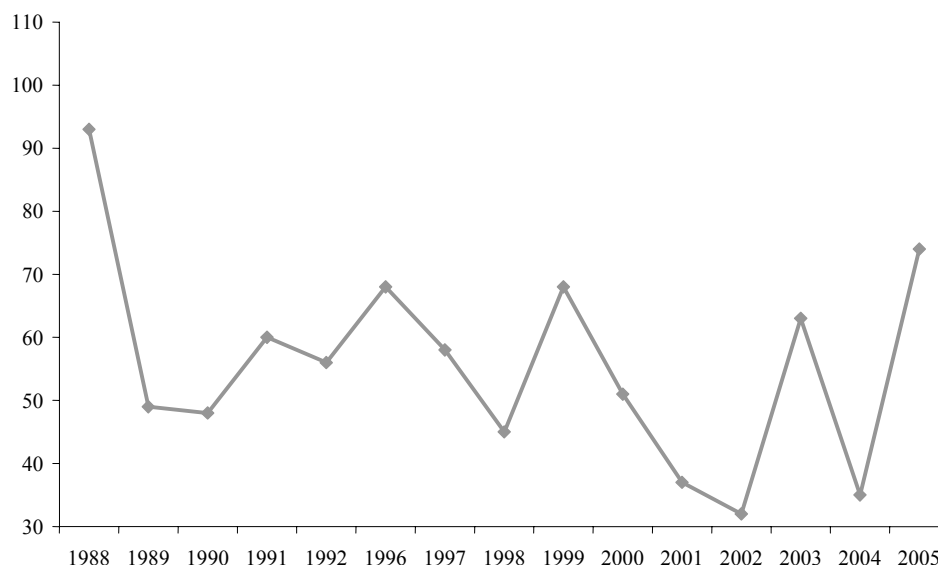


Figure 52 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.

A Lyon, la date moyenne d'apparition a lieu le 74,1^{ème} jour de l'année, soit le 15 mars dans le cadre d'une année non bissextile. Sur la figure 51, nous constatons une apparition progressivement plus précoce de 1988 à 1992, avant une relative stagnation. Cette date moyenne est nettement plus précoce à Bordeaux intervenant le 55,8^{ème} jour de l'année soit le 25 février (année non bissextile). Les dates d'apparition à Lyon oscillent du 3 mars, en 1998, au 4 avril en 1988, soit une amplitude maximale de 33 jours. Cette dernière est autrement plus élevée à Bordeaux atteignant 61 jours du 1^{er} février (2002) au 2 avril (1988). Cette variabilité des dates d'apparition du risque à Bordeaux est probablement liée, en partie, aux concentrations enregistrées assez fluctuantes d'une année à l'autre, puisqu'elles oscillent de 464 grains comptabilisés en 1988 à 2 131 en 1991.

Ainsi les dates de début de cumul testées s'échelonnent du 1^{er} octobre au 28 février à Lyon, et jusqu'au 31 janvier à Bordeaux.

- **Lyon :**

- **GDD_1** : 25 février – 1,5°C – 4,60 jours – 115,25°D
- **Lej_Tx** : 25 février – 8°C – 4,40 jours – 200,73°C
- **FT** : 25 février – 8°C – 4,53 jours – 8,43 unités
- **Q₁₀** : 9 février – 2,5 – 4,80 jours – 147,13 unités
- **Sarvas** : 23 février – Tm(0) – 4,73 jours – 81,19 unités

- **Bordeaux :**

- **Lej_Tx** : 23 janvier – 17°C – 2,40 jours – 42,1°C
- **Q₁₀** : 31 janvier – 12 – 5,13 jours – 826,03 unités
- **Sarvas** : 28 janvier – Tx et Tx(0) – 7,53 jours – 208,36 et 208,32 unités

Il faut tout d'abord noter que les dates optimales de démarrage des deux villes sont extrêmement proches des dates d'apparition les plus précoces des deux séries. Dans un cadre réel d'exploitation à visée préventive, il faudrait se borner au 15 février à Lyon, et au 15 janvier à Bordeaux maximum. Nous constatons que la précision des modèles est équivalente à Lyon et qu'elle est assez quelconque avec une moyenne des erreurs absolues de 4,61 jours. Par contre, à Bordeaux, des différences inédites sont enregistrées entre les trois modèles. Le modèle issu de la méthode Lej_Tx jouit d'une bonne précision alors que les deux modèles issus des méthodes exponentielles sont médiocres, particulièrement celui lié à la méthode de Sarvas. Les équations de régression sont, quant à elles, les suivantes :

$$\text{REAP_Lyon} = -2,070 \cdot \text{JP_11a} + 2,757 \cdot \text{T-20_01b} + 77,501$$

$$R^2 = 0,5457 ; \text{Af} = 0,4618 ; \text{MAE} = 4,39 \text{ jours} ; p = 0,0088$$

$$\text{REAP_Bordeaux} = 1,131 \cdot \text{Hn_12a} + 1,051 \cdot \text{Hn_01b} - 88,878$$

$$R^2 = 0,4121 ; \text{Af} = 0,3091 ; \text{MAE} = 10,47 \text{ jours} ; p = 0,0413$$

Les erreurs moyennes absolues sont à l'image de celles des modèles issus des méthodes de cumul : quelconque à Lyon et encore plus médiocre que le modèle associé à la méthode de Sarvas à Bordeaux.

En 2006, l'apparition du risque d'exposition allergique est survenue le 23 mars à Lyon et le 15 du même mois à Bordeaux. A Lyon les prévisions issues des cinq modèles ont été particulièrement bonnes, oscillant du 20 au 24 mars, la palme revenant aux modèles issus des méthodes FT et Sarvas qui ont donné la date exacte. Quant à l'équation de régression, elle a fourni une prévision trop précoce de 12 jours.

A Bordeaux, ne sont présentés que les prévisions établies à partir de Lej_Tx et Q₁₀ du fait de la trop mauvaise qualité du modèle de cumul et de l'équation restants. D'une manière assez inattendue, la prévision découlant de la méthode du Q₁₀ est tombée la veille de la date observée alors que celle issue de Lej_Tx est trop tardive de huit jours. Dans tous les cas, la

très grande variabilité des dates d'apparition du risque à Bordeaux, qui risquent d'être décorréliées de la phénologie du frêne, nécessitent une très grande prudence quant à l'utilisation des modèles de prévision.

3.3.2 Résultats pour le bouleau

Comme pour le frêne, les résultats des modèles de cumul sont exposés ville par ville avec, tout d'abord, une présentation graphique des dates d'apparition du risque. Rappelons que pour Rouen, Amiens, Paris, Nancy et Strasbourg, un parallèle est établi avec les résultats issus des modèles élaborés sur 15 ans. Sont ensuite présentées les équations de régression multiple. Enfin, nous présentons les prévisions pour 2006 à partir des différents modèles, à condition que la qualité de ces derniers soit relativement bonne.

3.3.2.1 Rouen

Dix-sept années sont disponibles pour cette ville de 1988 à 2005, l'année 1990 étant retirée. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 53.

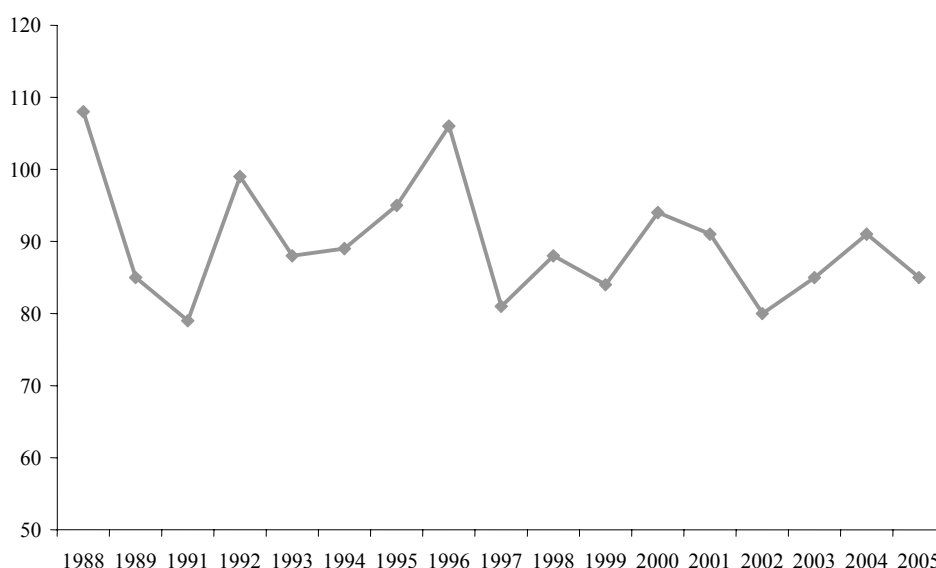


Figure 53 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Rouen, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 89,9^{ème} jour de l'année, soit le 31 mars pour une année non bissextile c'est-à-dire, en moyenne, six jours après le frêne sur une période identique. Les dates s'échelonnent ainsi du 20 mars, en 1991, au 17 avril en 1988, soit une amplitude maximale de 29 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars. Rappelons que les sigles des méthodes sont identiques à ceux présentés et utilisés dans les sous-chapitres 3.2.2, 3.2.4 et 3.2.6.

- **GDD_1** : 3 mars (9 mars) – 7°C (6,5°C) – 1,82 jour (1,80 jour) – 52,94°D
- **Lej_Tx** : 1^{er} mars (2 mars) – 11°C (11°C) – 1,71 jour (1,67 jour) – 216,62°C
- **FT** : 10 mars (9 mars) – 9,5°C (9°C) – 2,12 jours (1,93 jour) – 7,22 unités
- **Q₁₀** : 24 février (1^{er} mars) – 11,4 à 12 (8,7) – 2,12 jours (2,20 jours) – 876,45 à 938,29 unités
- **Sarvas** : 2 mars (1^{er} mars) – Tx et Tx(0) (Tx et Tx(0)) – 3,12 jours (2,60 jours) – 200,18 unités

Les variations des dates de démarrage de cumul et des seuils sont très faibles en passant d'un échantillon de 15 à 17 ans. Seul le modèle associé au Q₁₀ présente un décalage d'une semaine quant à la date de démarrage de cumul, et un coefficient de développement nettement différent. La moyenne des erreurs absolues est très proche : 2,18 jours avec 17 années contre 2,04 jours avec 15. Le fait d'ajouter seulement deux années pourrait expliquer la faible évolution des dates de démarrage et des seuils.

Quant à l'équation de régression, elle se décline de la manière suivante :

$$\text{REAP_Rouen} = -1,524 * \text{Hx_12a} - 2,424 * \text{Tx_03a} - 2,682 * \text{Tm0_03a} + 262,659$$

$$R^2 = 0,6496 ; \text{Af} = 0,5794 ; \text{MAE} = 3,64 \text{ jours} ; p = 0,0028$$

Le critère 03a correspond, ici, à la première décade de mars. Aucun des régresseurs de cette équation n'est commun à ceux de l'équation élaborée sur les 15 années communes à Rouen, Amiens, Paris, Nancy et Strasbourg. Il faut souligner que seules quatre variables sont significativement corrélées aux dates d'apparition du risque : les trois de l'équation (dont Tm0_03a issue d'une corrélation partielle) plus Hm_12a. Seule la variable Hx_12a est à la fois corrélée aux dates d'apparition du risque sur 15 et 17 ans. Ainsi, à l'instar de ce qui a été constaté pour le frêne, l'inconstance des variables corrélées, après l'inclusion de quelques

années supplémentaires, est mise en exergue. Dans le cas présent, cela a débouché sur une équation plus médiocre que celle présentée dans le sous-chapitre 3.2.5.

En 2006, l'apparition du risque d'exposition allergique lié au pollen de bouleau est survenue tard dans la saison (15 avril). La prévision issue du modèle associé au Q_{10} s'est révélée bonne (deux jours en avance) et celle de Lej_Tx de cinq jours trop précoce. Par contre, les trois autres modèles ont donné des prévisions trop précoces de huit jours (Sarvas) à 15 jours (FT). Ainsi, contrairement à ce qui a été constaté pour le frêne, des modèles avec des dates de démarrage proches ont donné des prévisions très divergentes, même si celle du modèle issu du Q_{10} intervient une semaine avant les autres. Nous pouvons supposer ici que la qualité de la prévision a plus dépendu de la méthode retenue que de la date d'initiation des modèles. La prévision issue de l'équation s'est révélée aussi trop précoce, avec une estimation tombant le 6 avril.

3.3.2.2 Amiens

A l'instar du frêne, 17 années de données polliniques et météorologiques conjointes sont disponibles de 1989 à 2005. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 54.

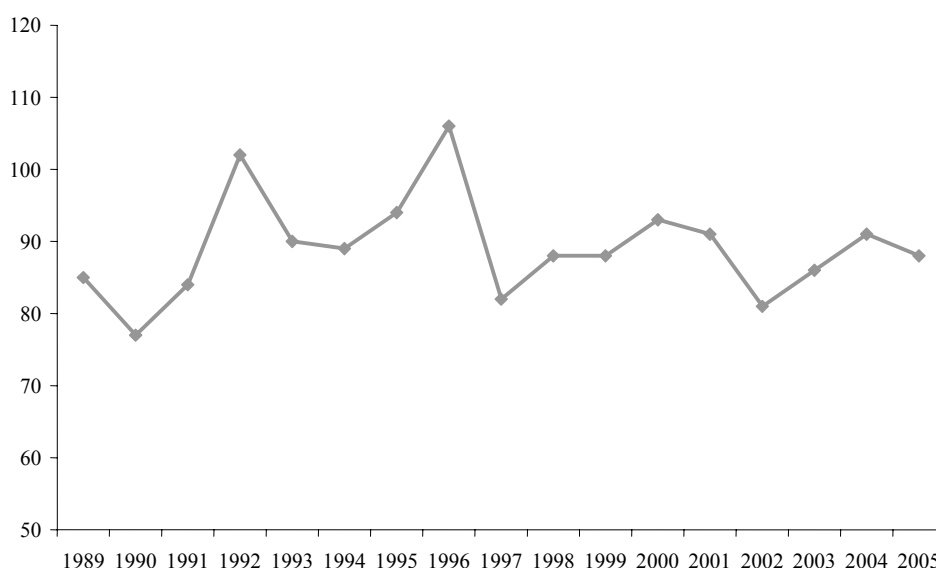


Figure 54 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Amiens, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 89,1^{ème} jour de l'année, soit le 30 mars pour une année non bissextile c'est-à-dire, en moyenne, huit jours après le frêne sur une période identique. Les dates s'échelonnent du 18 mars, en 1990, au 15 avril en 1996, soit une amplitude maximale de 29 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars.

- **GDD_1** : 6 mars (1^{er} janvier) – 5,5 °C (6,5°C) – 2,24 jours (1,80 jour) – 80,25°D
- **Lej_Tx** : 7 mars (9 mars) – 10,5°C (10°C) – 2,29 jours (2,13 jours) – 234,63°C
- **FT** : 10 mars (10 mars) – 9°C (8°C) – 2,29 jours (2,07 jours) – 8,88 unités
- **Q₁₀** : 16 février (29 décembre) – 7,4 à 8 (8,2) – 1,76 jour (1,60 jour) – 616,67 à 680,53 unités
- **Sarvas** : 15 février (3 mars) – Tx(0) (Tx et Tx(0)) – 2,18 jours (1,93 jour) – 279,44 unités

La variation des dates et des seuils, en passant de 15 à 17 ans de données, est assez faible, à l'exception des dates de début de cumul associées aux modèles issus de GDD_1 et, dans une moindre mesure, de Sarvas. La précision des modèles élaborés sur 17 années est très légèrement inférieure à celle de ceux mis au point sur 15 ans, avec une moyenne des erreurs de 2,15 et 1,91 jours respectivement.

L'équation de régression, quant à elle, se présente ainsi :

$$\text{REAP_Amiens} = 1,598 \cdot \text{JP1_01b} + 2,922 \cdot \text{Tm0_01b} + 74,613$$

$$R^2 = 0,5102 ; \text{Af} = 0,4343 ; \text{MAE} = 3,96 \text{ jours} ; p = 0,0068$$

Il convient de rappeler qu'à partir de la série de 15 ans, seules trois variables étaient corrélées significativement, plus une variable partiellement corrélée. Dans le cas présent, il s'agit de cinq variables et d'une autre corrélée partiellement (JP1_01b), se situant toutes durant la deuxième «quinzaine» de janvier. Cet état de fait montre la difficulté de déterminer, pour cette ville, des variables significativement corrélées aux dates d'apparition du risque, caractérisée par une mauvaise représentativité des différentes périodes bimensuelles incluses dans l'analyse. L'erreur moyenne absolue de l'équation, plus faible que celles des modèles de cumul, est supérieure à celle élaborée sur l'échantillon de 15 années.

Comme à Rouen, le 15 avril 2006 est marqué par l'apparition du risque lié au bouleau. Les deux modèles issus des méthodes exponentielles de cumul (Q₁₀ et Sarvas) ont donné une

bonne prévision : 12 avril. Par contre, les trois autres modèles ont estimé une date d'apparition du risque beaucoup trop précoce : 7 avril pour Lej_Tx, 1^{er} avril pour GDD_1 et FT. Ces trois derniers modèles ont été initiés début mars, alors que les deux modèles issus des méthodes exponentielles (Q₁₀ et Sarvas) l'ont été mi-février. Outre le fait que démarrer des cumuls en mars n'est pas très pertinent dans un but préventif réellement opérationnel, si la saison pollinique du bouleau est précoce, il s'avère que la prise en compte, seule, des dernières semaines avant l'apparition du risque ne s'est pas révélée adéquate pour 2006. Les modèles initiés plus tôt ont été plus performants. A l'instar de ce qui a été constaté pour le frêne à Paris, nous remarquons l'influence de la date de démarrage de cumul sur la qualité d'une prévision pour une année donnée. Cependant, nous pouvons constater, qu'à date de démarrage équivalente, le type de méthode peut aussi influencer sur la qualité de la méthode. C'est ce que nous constatons pour le modèle issu de la méthode de Lejoly-Gabriel par rapport à ceux déterminés à partir des méthodes GDD_1 et FT. Comme à Rouen, l'équation de régression a donné une date d'apparition du risque beaucoup trop précoce (1^{er} avril), équivalente à celles fournies par les «mauvais» modèles.

3.3.2.3 Paris

Nous disposons de la série 1987-2005 complète, soit 19 ans. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 55.

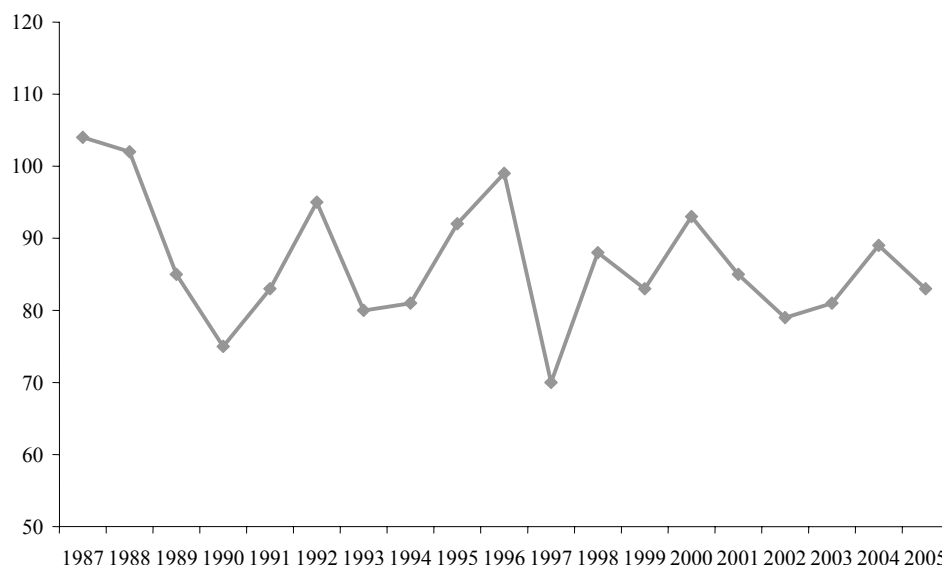


Figure 55 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Paris, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 86,7^{ème} jour de l'année, soit le 28 mars pour une année non bissextile. Les dates s'échelonnent du 11 mars, en 1997, au 14 avril en 1987, soit une amplitude maximale de 34 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars.

- **GDD_1** : 28 février (26 février) – 10,5°C (10,5°C) – 1,89 jour (2,07 jours) – 24,12°D
- **Lej_Tx** : 1^{er} mars (2 et 3 février) – 13 °C (14°C) – 2,47 jours (2,27 jours) – 184,89°C
- **FT** : 28 février (28 février) – 13°C (12,5°C) – 2,21 jours (2,47 jours) – 3,91 unités
- **Q₁₀** : 18 février (7 mars) – 9,4 (6,5) – 2,63 jours (2,80 jours) – 914,86 unités
- **Sarvas** : 15 février (7 mars) – Tm(0) (Tx et Tx(0)) – 3,74 jours (2,87 jours) – 173,83 unités

Les variations de seuil sont ici faibles à l'exception du modèle issu du Q₁₀. Les dates optimales de démarrage pour les deux séries se situent toutes en février et mars. La moyenne des erreurs absolues est quasi identique : 2,59 jours pour la série de 19 ans et 2,50 jours pour celle de 15 ans. L'équation de régression est, quant à elle, la suivante :

$$\text{REAP_Paris} = 0,385 \cdot \text{Hn_10b} + 3,653 \cdot \text{P_01b} + 0,084 \cdot \text{Insol_01b} + 2,715 \cdot \text{Tm0_01b} - 0,826 \cdot \text{Tx_02b} + 50,348$$

$$R^2 = 0,8386 ; Af = 0,7820 ; MAE = 2,73 \text{ jours} ; p = 0,0001$$

Nous pouvons noter que seule la variable Tm0_01b est commune aux deux équations établies l'une sur 15 ans de données et l'autre, ci-dessus, sur 19 ans. Les variables P_01b et Insol_01b sont corrélées partiellement à l'apparition du risque, compte tenu de la variable Tm0_01b. L'erreur moyenne est proche de celle de l'équation établie sur 15 ans (2,59 jours) avec, cependant, deux régresseurs en plus.

En 2006, l'apparition d'un risque a eu lieu le 12 avril, une date mal prévue par l'ensemble des modèles de cumul. En effet, ils ont tous fourni une prévision trop précoce allant de cinq jours, au mieux, avec le modèle issu du Q₁₀ (et six jours avec Lej_Tx) jusqu'à 13 jours avec le modèle élaboré à partir de la méthode FT et 12 à partir de la méthode GDD_1. Une initiation des cumuls à cette période ne s'est pas avérée adéquate pour cette année, un démarrage plus précoce aurait, peut-être, donné de meilleurs résultats. Ainsi, nous avons testé toutes les dates de démarrage du 1^{er} octobre au 31 décembre afin de déterminer les meilleurs modèles de chaque méthode dans ce laps de temps. Sans entrer dans les détails, toutes les dates de démarrages sont comprises entre le 15 et le 31 décembre et les modèles associés présentent une erreur moyenne absolue globale de 3,58 jours, soit, tout de même, un jour de plus que pour les modèles testés ci-dessus. Ici, contrairement à précédemment, les prévisions ont été trop tardives, allant du 13 avril (Lej_Tx) au 17 avril (Q₁₀), mais meilleures.

Le 7 avril a été la date estimée par l'équation de régression, soit cinq jours en avance par rapport à la date réelle d'apparition du risque.

3.3.2.4 Nancy

Nous disposons de la série 1988-2005 complète, soit 18 ans. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 56.

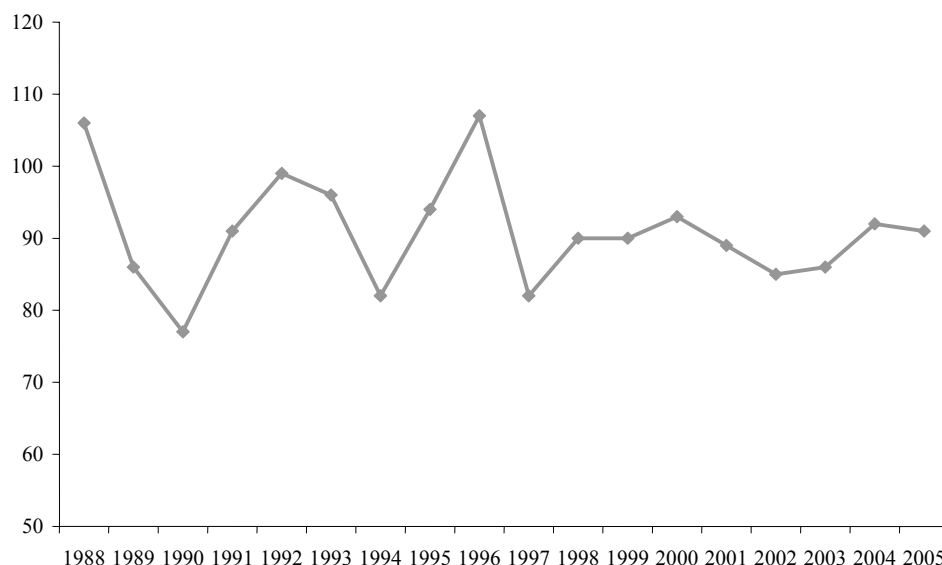


Figure 56 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Nancy, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 90,9^{ème} jour de l'année, soit le 1^{er} avril pour une année non bissextile. Les dates s'échelonnent du 18 mars, en 1990, au 16 avril en 1996, soit une amplitude maximale de 30 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars.

- **GDD_1** : 28 février (1^{er} mars) – 5,5°C (4°C) – 2,22 jours (2,07 jours) – 86,98°D
- **Lej_Tx** : 4 mars (3 mars) – 7,5°C (7°C) – 2,61 jours (2,07 jours) – 313,96°C
- **FT** : 2 mars (10 mars) – 9°C (6,5°C) – 2,72 jours (2,67 jours) – 10,37 unités
- **Q₁₀** : 29 décembre (17 décembre) – 4,4 (4,9) – 1,67 jour (1,27 jour) – 548,20 unités
- **Sarvas** : 1^{er} janvier (29 décembre) – Tx (Tx(0)) – 1,78 jour (1,53 jour) – 439,50 unités

Les variations de dates optimales et de seuils sont relativement faibles à l'exception des seuils liés à la méthode FT qui passent de 6,5 à 9°C. Nous remarquons que les deux modèles issus des deux méthodes exponentielles (Q₁₀ et Sarvas) ont pour point commun un démarrage des cumuls beaucoup plus précoce que les autres modèles. L'erreur moyenne de l'ensemble des modèles est globalement équivalente d'un pas de temps à l'autre. Elle est un peu meilleure sur 15 ans de données (erreur moyenne absolue de 1,92 jour) que sur 18 ans (2,20 jours).

Nous passons maintenant à l'équation de régression :

$$\text{REAP_Nancy} = 2,148 * \text{Hx_10b} - 0,043 * \text{Insol_11b} - 1,190 * \text{Tx_03a} - 99,333$$

$$R^2 = 0,6892 ; \text{Af} = 0,6295 ; \text{MAE} = 3,51 \text{ jours} ; p = 0,0008$$

Le critère 03a correspond à la première décade de mars. A l'instar de plusieurs équations présentées auparavant, les variables explicatives sont totalement différentes de celles de l'équation élaborée sur 15 ans. L'erreur moyenne absolue est ici nettement plus importante que celle de l'équation présentée dans le sous-chapitre 3.2.5 qui débouchait sur une valeur très faible (1,49 jour).

En 2006, l'apparition du risque d'exposition allergique est survenue le 15 avril, soit parmi les dates les plus tardives depuis 1988. Ici nous remarquons une opposition entre les deux modèles dont les cumuls sont initiés les 29 décembre et 1^{er} janvier et les quatre autres. En ce qui concerne les deux premiers cas, les prévisions ont été bonnes mais un peu trop tardives, 17 avril avec le Q₁₀ et 19 avril avec Sarvas, alors qu'elles ont été trop précoces avec les trois autres s'échelonnant, au mieux, de huit jours (Lej_Tx) à 13 jours (FT). Même si des disparités existent entre les modèles initiés après la mi-février, nous constatons que la date de démarrage joue, de nouveau, un rôle majeur dans la qualité des prévisions. A l'instar de ce que nous avons constaté à Paris, un démarrage des cumuls fin décembre – début janvier s'avère être la moins mauvaise des solutions. Cependant, nous remarquons une opposition, quant à la qualité de la prévision, entre les deux méthodes où sont prises en compte toutes les températures de la période d'accumulation et les trois autres.

L'équation de régression a donné, quant à elle, une prévision trop précoce de dix jours (5 avril), cas de figure similaire à celui des trois stations précédentes.

3.3.2.5 Strasbourg

Dix-sept années sont exploitables de 1988 à 2005, l'année 1996 étant écartée de l'analyse. Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 57.

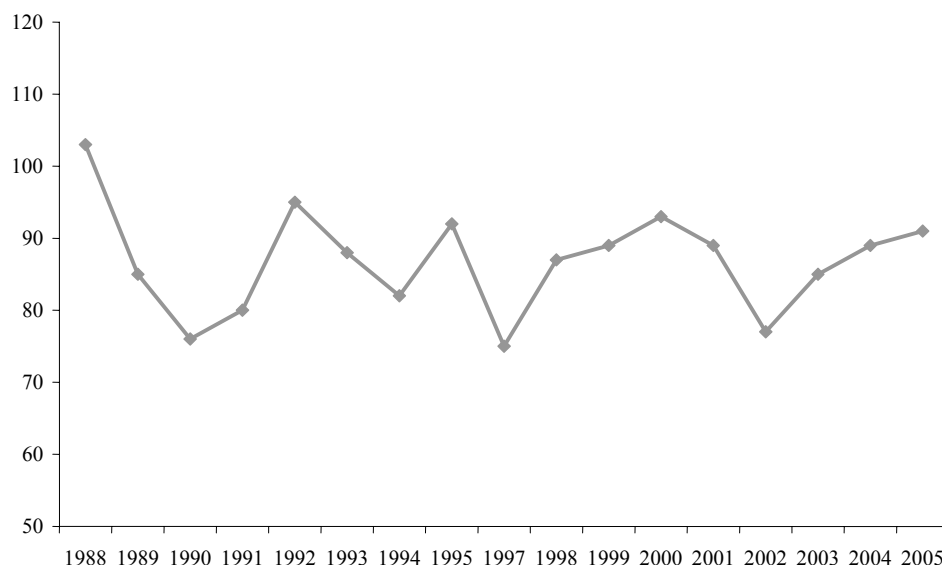


Figure 57 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Strasbourg, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition a lieu le 86,8^{ème} jour de l'année, soit le 28 mars pour une année non bissextile. Les dates s'échelonnent du 16 mars, en 1997, au 12 avril en 1988, soit une amplitude maximale de 28 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars.

- **GDD_1** : 1^{er} mars (11 février) – 8°C (10°C) – 2,12 jours (2,07 jours) – 42,12°D
- **Lej_Tx** : 28 décembre (26 et 27 décembre) – 11°C (11°C) – 2,12 jours (1,60 jour) – 388,12°C
- **FT** : 12 février (11 et 12 février) – 12°C (12°C) – 2,29 jours (2,07 jours) – 5,10 unités
- **Q₁₀** : 21 février (20 février) – 5,3 (5,4) – 1,82 jour (1,93 jour) – 375,15 unités
- **Sarvas** : 20 février (20 février) – Tx(0) (Tx et Tx(0)) – 2,24 jours (2,20 jours) – 254,71 unités

Les seuils et les dates de démarrage évoluent très peu en passant de 15 à 17 ans, à l'exception du modèle issu de la méthode GDD_1 qui connaît des variations un peu plus importantes. La moyenne des erreurs absolues, de ce fait, est très proche : 2,12 jours sur 17 ans de données contre 1,97 jour sur 15 ans.

L'équation de régression, quant à elle, se présente de la manière suivante :

$$\text{REAP_Strasbourg} = -0,532 \cdot \text{Tm0_01a} + 1,098 \cdot \text{Tn0_02b} - 1,072 \cdot \text{Tx_03a} + 99,884$$

$$R^2 = 0,6836 ; \text{Af} = 0,6179 ; \text{MAE} = 3,11 \text{ jours} ; p = 0,0015$$

Tx_03a correspond ici à la température maximale de la première décade de mars. Aucun des trois régresseurs n'est retrouvé dans l'équation établie sur 15 ans. Cependant le nombre de jours où la température moyenne est inférieure à 0°C durant la première «quinzaine» de janvier est, par nature, proche des variables Tn0_01a et Tx0_01a présentes dans l'équation sur 15 ans. L'évolution de l'erreur moyenne absolue de 15 vers 17 ans de données est similaire à celle de Nancy. La valeur de cette erreur passe ainsi de 1,53 à 3,11 jours.

En 2006, l'apparition du risque est intervenue le 8 avril, ce qui s'est avéré plus précoce que pour les autres sites présentés de la partie Nord de la France. Les modèles issus des méthodes du Q₁₀ et de Sarvas ont donné d'excellentes prévisions : 6 avril pour la première et 7 avril pour la seconde. Ces deux modèles ont pour particularité d'avoir été initiés à des dates concomitantes (20 et 21 février). Le modèle issu de la méthode de Lejoly-Gabriel, initié en décembre, a débouché sur une prévision beaucoup trop divergente : 19 avril. La date d'apparition du risque, certes tardive, ne l'a pas autant été que ce que aurait laissé envisager cet hiver parmi les plus rigoureux des 20 dernières années. Quant aux modèles FT et GDD, initiés à 17 jours d'intervalle à des dates englobant celles des modèles associés au Q₁₀ et Sarvas, ils ont donné des prévisions trop précoces de sept (FT) et huit jours (GDD_1), ce qui est assez singulier au vu des dates prévues par les deux méthodes exponentielles. L'influence de la date de démarrage de cumul sur la prévision semble, tout du moins dans le cas de la méthode de Lejoly-Gabriel, être la cause la plus visible. Cependant, les écarts de prévision constatés par ailleurs sont probablement liés à la fois aux dates de démarrage de cumul et au type de méthode employé. Comme à Nancy et Amiens, ce sont les deux modèles issus des méthodes du Q₁₀ et de Sarvas qui donnent les meilleures prévisions pour 2006, avec des dates de démarrage de cumul nettement plus précoces à Nancy que dans les deux autres villes. Quant à l'équation de régression, elle a donné une prévision neuf jours trop en avance par rapport à la date réelle.

Après avoir présenté ces cinq stations qui se caractérisent par des modèles de bonne ou très bonne facture, débouchant sur des prévisions de qualité variable pour 2006, nous allons développer les résultats pour d'autres villes présentant au moins une série de 15 années où les dates d'apparition du risque lié au bouleau sont exploitables. Quatre villes répondent à ce

critère : Lyon, Bordeaux, Clermont-Ferrand et Montluçon. Pour les raisons évoquées à la fin de la section 3.3.1.4, les résultats de Montluçon ne sont pas présentés.

3.3.2.6 Les sites du Centre-Est : Lyon et Clermont-Ferrand

Une série de 19 années est exploitable à Lyon de 1987 à 2005. A Clermont-Ferrand, toutes les années depuis l'installation du capteur, en 1991, sont utilisables, soit 15 années ; les données météorologiques utilisées sont celles de la station homonyme. En comparaison des conditions météorologiques de Lyon, Clermont-Ferrand (normale 1961-1990) se caractérise par des températures équivalentes en hiver (3,6°C) mais plus fraîches en été (18,3°C contre 19,6 à Lyon). Les précipitations annuelles ne sont pas très élevées (590,6 mm) et revêtent une distribution digne d'un climat continental, avec des valeurs plus faibles du mois de novembre au mois de mars (155,3 mm) que de mai à septembre (338,8 mm). Enfin, l'insolation annuelle se révèle légèrement plus faible qu'à Lyon (1907,3 heures contre 1975,7). Les dates d'apparition du risque d'exposition allergique à Lyon (figure 58) et Clermont-Ferrand (figure 59) sont présentées ci-après.

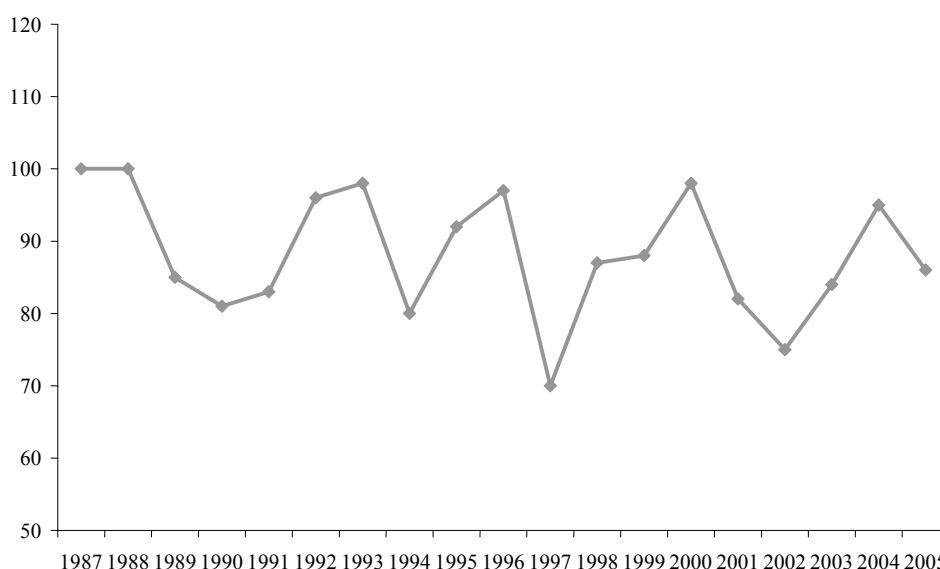


Figure 58 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Lyon, en nombre de jours à partir du 1^{er} janvier.

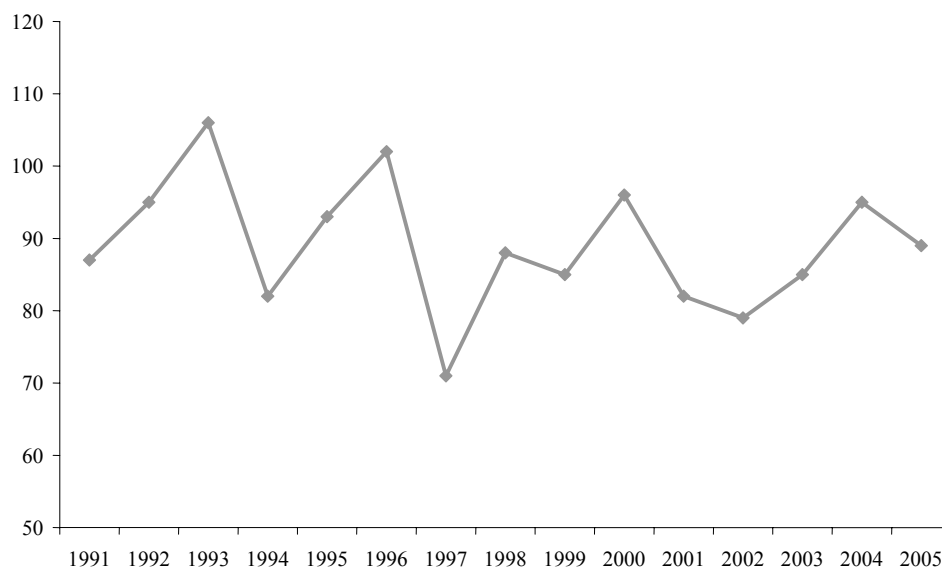


Figure 59 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Clermont-Ferrand, en nombre de jours à partir du 1^{er} janvier.

A Lyon, la date moyenne d'apparition a lieu le 88,3^{ème} jour de l'année, soit le 29 mars dans le cadre d'une année non bissextile. Elle intervient, en moyenne, le 89^{ème} jour de l'année à Clermont-Ferrand soit, en arrondissant, un jour après Lyon. Les dates d'apparition à Lyon oscillent du 11 mars, en 1997, au 10 avril (1987) et 9 avril (1988) qui correspondent au 100^{ème} jour des deux années en question. L'amplitude maximale est ainsi de 30 jours. En ce qui concerne Clermont-Ferrand, les dates d'apparition s'échelonnent du 12 mars, en 1997, au 16 avril, en 1993, soit une amplitude maximale de 35 jours. Les dates de début de cumul testées s'étendent du 1^{er} octobre au 10 mars pour les deux villes.

- **Lyon :**

- **GDD_1** : 9 février – 8,5°C – 2,74 jours – 70,46°D
- **Lej_Tx** : 5 mars – 8,5°C – 3,37 jours – 315,11°C
- **FT** : 12 février – 11,5°C – 2,95 jours – 9,68 unités
- **Q₁₀** : 9 février – 6,2 et 6,3 – 2,21 jours – 721,55 et 738,29 unités
- **Sarvas** : 21 février – Tm – 2,58 jours – 159,12 unités

- **Clermont-Ferrand :**

- **Lej_Tx** : 2 mars – 13°C – 3,53 jours – 274,12°C
- **Q₁₀** : 2 février – 5,7 à 5,9 – 2,40 jours – 784,65 à 826,42 unités
- **Sarvas** : 20 février – T_m – 2,87 jours – 164,35 unités

Pour Clermont-Ferrand, nous n'avons testé que trois des cinq méthodes pour les raisons évoquées dans l'introduction du chapitre 3.3. Nous constatons que les dates de démarrage de cumul ont lieu en février et mars pour les deux villes et que les erreurs moyennes absolues s'échelonnent de 2,21 à 3,53 jours, indiquant des modèles de bonne qualité avec cependant une erreur un peu plus élevée des modèles issus de la méthode Lej_Tx.

Les équations de régression sont les suivantes :

$$\text{REAP_Lyon} = -1,731 \cdot \text{Tx0_12b} - 0,816 \cdot \text{Tm_02b} - 0,955 \cdot \text{Tm_03a} + 101,375$$

$$R^2 = 0,5291 ; \text{Af} = 0,4551 ; \text{MAE} = 4,54 \text{ jours} ; p = 0,0087$$

$$\text{REAP_Clermont} = 2,296 \cdot \text{T-20_10a} + 0,935 \cdot \text{Hx_02a} + 4,365 \cdot \text{Tn0_03a} - 5,430 \cdot \text{Tm0_03a} - 38,483$$

$$R^2 = 0,8741 ; \text{Af} = 0,8087 ; \text{MAE} = 2,52 \text{ jours} ; p = 0,0002$$

Il est important de préciser ici que les variables T_m_03a, T_{n0}_03a et T_{m0}_03a correspondent à la première décade de mars. L'équation de cette dernière ville présente une erreur moyenne absolue assez faible, alors que celle liée à l'équation de Lyon présente une valeur assez quelconque. Nous pouvons noter que la température, sous différentes formes, de la première décade de mars est retenue dans chacune des équations. Cela montre bien l'importance de son impact peu avant l'apparition du risque.

En 2006, l'apparition de ce dernier est survenue le 9 avril à Lyon et, de manière surprenante, le 20 mars à Clermont-Ferrand. A Lyon les prévisions issues des cinq modèles de cumul ont été homogènes mais trop précoces, oscillant du 30 mars au 1^{er} avril. La date prévue par l'équation de régression a été encore plus précoce : 27 mars.

A Clermont-Ferrand, la date précoce d'apparition n'a été correctement estimée par aucun modèle de cumul. Ces derniers ont systématiquement donné des dates beaucoup trop tardives : 29 mars (Lej_Tx), 1^{er} avril (Sarvas) et 9 avril (Q₁₀). Quant à la date déterminée par l'équation de régression, elle l'a été également (8 avril). Cependant, la date d'apparition du risque lié au

bouleau a été en total décalage avec celles observées à Lyon et pour les stations relativement proches de Montluçon (12 avril) et Saint-Etienne (8 avril). Cela est d'autant plus étrange que la date d'apparition du risque lié au frêne (30 mars) a été quasi identique à celles de Montluçon (29 mars) et Saint-Etienne (31 mars). Nous pouvons donc craindre qu'une confusion ait eu lieu entre le pollen de bouleau et, à priori, celui de noisetier (également triporé) à la fin du mois de mars, même si, en général, ce dernier genre a accompli sa pollinisation à cette période.

3.3.2.7 Bordeaux

Seize années sont exploitables de 1989 à 2005, l'année 1993 étant retirée de l'analyse. Les dates d'apparition du risque d'exposition allergique sont présentées dans la figure 60.

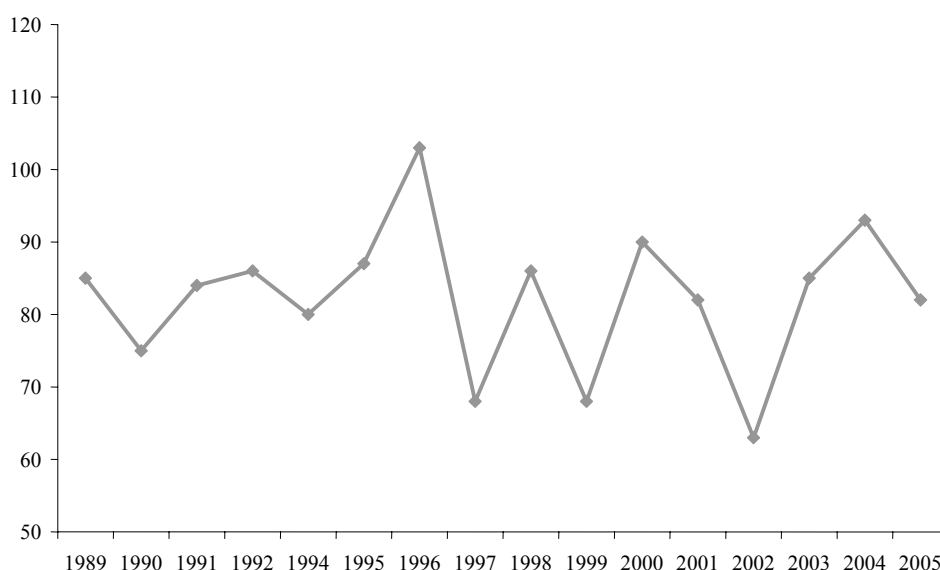


Figure 60 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.

La date moyenne d'apparition a lieu le 82,3^{ème} jour de l'année, soit le 23 mars dans le cadre d'une année non bissextile pour ces 16 années. Les dates d'apparition s'échelonnent du 4 mars, en 2002, jusqu'au 12 avril, en 1996, soit une amplitude maximale de 40 jours. Il faut noter cependant que les quantités de pollen de bouleau comptabilisées à Bordeaux sont plus faibles que dans les autres sites, avec cinq années sur les 16 enregistrant moins de 1 000 grains. A cela il faut ajouter les années 1988 et 1993 comptant chacune moins de 400 grains, d'où leur

exclusion de l'analyse, et les années 2006 et 2007 où 718 et 858 grains ont été enregistrés respectivement. Nous présentons tout de même les résultats du bouleau pour cette ville, tout en sachant que le risque associé n'est fondamentalement pas très élevé. Sur la période 1997-2005, il n'est que le 10^{ème} taxon par ordre d'importance, ne représentant que 3,08% du total de tous les pollens comptabilisés. Précisons également que les dates de début de cumul testées s'étendent du 1^{er} octobre au 28 février.

- **Lej_Tx** : 27 février – 14°C – 6,31 jours – 267,03°C
- **Q₁₀** : 22 février – 4 et 4,1 – 5,38 jours – 357,86 et 371,78 unités
- **Sarvas** : 19 février – Tn(0) – 4,88 jours – 88,47 unités

Du fait, peut-être, des faibles concentrations en pollen de bouleau enregistrées certaines années, la qualité des meilleurs modèles est très quelconque, voire médiocre pour Lej_Tx. Les dates optimales de démarrage ont toutes lieu en février quelques jours avant les dates les plus précoces d'apparition du risque.

L'équation de régression, quant à elle, est la suivante :

$$\text{REAP_Bordeaux} = -0,031 * \text{Gel_12b} - 1,568 * \text{Tn_02b} + 92,600$$

$$R^2 = 0,5546 ; \text{Af} = 0,4785 ; \text{MAE} = 4,97 \text{ jours} ; p = 0,0052$$

L'erreur moyenne absolue de l'équation est équivalente à celles des modèles de cumul. Nous pouvons noter que celle-ci contient la température minimale de la dernière période précédant les dates d'apparition du risque les plus précoces, à l'instar des équations de Lyon et Clermont-Ferrand.

En 2006, l'apparition du risque a eu lieu précocement par rapport aux autres sites étudiés : le 1^{er} avril. Les modèles issus des méthodes du Q₁₀ et de Lejoly-Gabriel, malgré leur précision «perfectible», ont donné des dates proches de la date réelle : 28 et 29 mars respectivement. La prévision du modèle issu de la méthode de Sarvas se révèle, quant à elle, trop précoce de sept jours. Le type de modèle semble ici être le critère le plus influent de la qualité de la prévision en 2006. Quant à l'équation de régression elle a, malheureusement, donné une date complètement divergente en prévoyant une apparition du risque dès le 7 mars.

3.3.3 Résultats pour les Poacées

A l'instar des deux taxons précédents, les résultats des meilleurs modèles de cumul sont exposés ville par ville avec une présentation graphique des dates d'apparition du risque. Sont ensuite détaillées les équations de régression linéaire multiple. Enfin, nous présentons les prévisions pour 2006 à partir des différents modèles, dont les erreurs moyennes absolues ne sont pas trop élevées.

3.3.3.1 Rouen, Amiens, Nancy et Strasbourg

Ces quatre villes ont en commun la période 1988-2005, soit 18 ans de données disponibles, période qui a été présentée dans les sous-chapitres 3.2.6 et 3.2.7. Les résultats des modèles issus des méthodes de cumul et de la régression linéaire multiple y sont développés, en rappelant que les dates de début de cumul testées s'échelonnent du 1^{er} janvier au 30 avril. Nous allons donc rappeler, graphiquement, les dates d'apparition du risque allergique d'exposition aux pollens de Poacées ainsi que les prévisions pour 2006. Les dates d'apparition du risque des quatre villes sont représentées dans la figure 61.

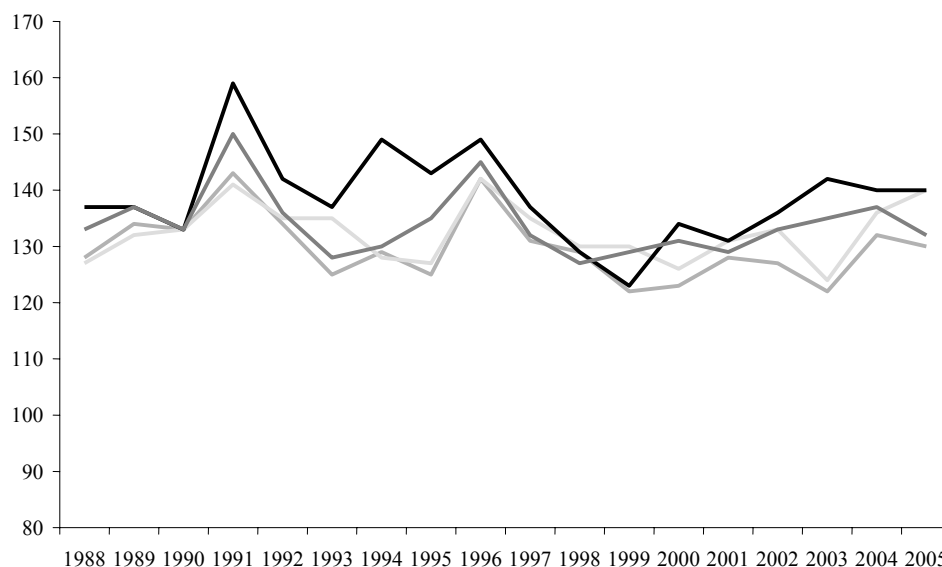


Figure 61 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, en nombre de jours à partir du 1^{er} janvier, à Rouen (noir), Amiens (gris foncé), Nancy (gris moyen) et Strasbourg (gris clair).

Les dates moyennes d'apparition surviennent le 138,8^{ème} jour à Rouen, le 134^{ème} à Amiens, le 129,8^{ème} à Nancy et le 132,5^{ème} à Strasbourg. Dans le cadre d'années non bissextiles, cela correspond au 19 mai à Rouen, au 14 à Amiens, au 10 à Nancy et aux 12-13 mai à Strasbourg. Les dates d'apparition s'échelonnent du 3 mai (1999) au 8 juin (1991) à Rouen, du 7 mai (1998) au 30 mai (1991) à Amiens, du 2 mai (1999 et 2003) au 23 mai (1991) à Nancy et du 4 mai (2003) au 21 mai (1996) à Strasbourg. Cela débouche sur les amplitudes maximales suivantes : 36 jours à Rouen, 23 à Amiens, 21 à Nancy et 18 à Strasbourg.

Cette rapide présentation faite, nous allons évoquer les prévisions de la date d'apparition du risque d'exposition allergique en 2006 en précisant que nous ne tenons pas compte des résultats des modèles issus de la méthode GDD_2.

Pour cette année, le risque est survenu à Rouen le 17 mai, soit une date proche de la moyenne 1988-2005. Les meilleurs modèles de cumul sont tous initiés après le 19 avril, ne laissant qu'une période d'accumulation relativement courte. Les prévisions ont été cependant de très bonne qualité. Trop précoce de quatre jours pour le modèle issu de la méthode GDD_1, qui a donné la prévision la plus divergente, elles se sont révélées quasiment exactes pour les modèles issus des méthodes du Q₁₀ et de Sarvas (16 mai) ; les deux derniers modèles donnant des dates intermédiaires (14 mai pour FT et 15 mai pour Lej_Tx). La seule prise en compte des conditions de température des dernières semaines s'est avérée tout à fait appropriée pour la prévision de l'année 2006. Par contre l'équation de régression a fourni une date beaucoup trop tardive : le 29 mai.

A Amiens, l'apparition du risque est survenue le 18 mai, soit un peu plus tard que la moyenne 1988-2005 de cette ville. Tous les modèles de cumul ont prévu la survenue du risque le 10 ou 11 mai, prévision qui s'est avérée donc trop précoce. Les dates de démarrage de cumul très rapprochées, du 20 au 26 avril (cf. 3.2.6.2), ont probablement été à l'origine de ces dates quasi identiques, malgré la variété de la conception des différentes méthodes et des seuils associés. A l'instar de Rouen, la date d'initiation des modèles est le facteur le plus visible à l'origine de ces prévisions très proches.

Quant à l'équation de régression, développée dans le sous-chapitre 3.2.7, elle donne une excellente prévision pour 2006, la veille de la date réelle.

A Nancy, l'apparition du risque est intervenue le 15 mai, plus tard que la date moyenne. Les modèles issus des méthodes du Q₁₀ et de Sarvas, initiés début janvier (cf. 3.2.6.2), se sont avérés performants avec une prévision à un jour près pour Sarvas (14 mai) et deux jours pour l'autre (13 mai). Les autres modèles, initiés après le 24 avril, ont donné des prévisions de dix (FT, GDD_1) et neuf jours (Lej_Tx) trop précoces. La qualité de la prévision pour 2006 a

dépendu, à priori, essentiellement de la date de démarrage de cumul. Contrairement à Rouen, la prise en compte d'une longue période d'accumulation s'est avérée la plus pertinente pour déterminer une date d'apparition du risque très précise en 2006. Cependant, le fait que ce soit les deux méthodes exponentielles qui ressortent n'est, peut-être, pas étranger à la bonne qualité de la prévision sur cette ville. La date prévue par l'équation de régression, le 11 mai, s'est située à mi-chemin de celles prévues par les deux groupes de modèles singularisés par leurs dates de démarrage.

A Strasbourg, en 2006, le risque est apparu un peu plus tôt que pour les trois villes précédentes, à une date correspondant à la moyenne 1988-2005 de cette ville : le 12 mai. A l'instar de Rouen et Amiens, tous les modèles, initiés dans les derniers jours d'avril (cf. 3.2.6.2), ont donné de très bonnes prévisions s'échelonnant du 9 mai (Lej_Tx) au 11 mai (FT et Sarvas). Les autres modèles ayant prévu l'apparition du risque le 10 mai, nous pouvons constater l'influence de dates de démarrages de cumul rapprochées (22 au 30 avril selon les modèles) dans l'homogénéité des prévisions. L'équation de régression a, quant à elle, fourni une prévision trop tardive de huit jours. Paris, dernière des cinq stations faisant l'objet des sous-chapitres 3.2.6 et 3.2.7, est traitée ici à part car 19 années, de 1987 à 2005, y sont disponibles.

3.3.3.2 Paris

Les dates d'apparition du risque d'exposition allergique aux pollens sont présentées dans la figure 62.

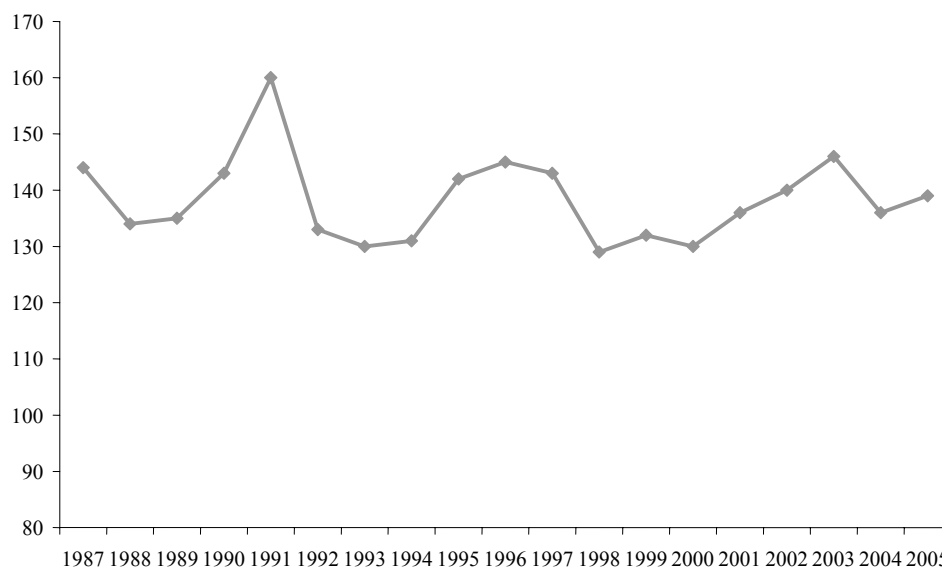


Figure 62 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Paris, en nombre de jours à partir du 1^{er} janvier.

Sur cette période, la date moyenne d'apparition se situe le 138,3^{ème} jour de l'année, soit le 18 mai dans le cadre d'une année non bissextile. Les dates s'échelonnent du 9 mai, en 1998, au 9 juin en 1991, soit une amplitude maximale de 31 jours. Les dates de début de cumul testées s'étendent du 1^{er} janvier au 30 avril. Entre parenthèses sont rappelés les résultats déterminés à partir de la série de 18 années présentée dans le chapitre 3.2.

- **GDD_1** : 20 avril (19 avril) – 12°C (11,5°C) – 3,79 jours (3,78 jours) – 78,24°D
- **Lej_Tx** : 26 avril (19 avril) – 18,5°C (16°C) – 4 jours (4,11 jours) – 226,59°C
- **FT** : 20 avril (19 avril) – 15,5°C (15,5°C) – 3,84 jours (3,78 jours) – 9,21 unités
- **Q₁₀** : 20 avril (18 avril) – 4 (6 et 6,1) – 4,26 jours (4,11 jours) – 548,95 unités
- **Sarvas** : 19 avril (16 avril) – Tn et Tn(0) (Tn et Tn(0)) – 3,95 jours (3,72 jours) – 151,05 unités

Les variations de seuils et de dates de démarrage sont, globalement, relativement faibles. Les dates optimales, pour les deux séries, se placent toutes en avril. La moyenne des erreurs absolues est quasi identique : 3,97 jours pour la série de 19 ans et 3,90 jours pour celle de 18 ans. L'équation de régression est, quant à elle, la suivante :

$$\text{REAP}_{\text{Paris}} = 1,664 \cdot \text{Tm0_02a} + 0,035 \cdot \text{Insol_04a} - 0,316 \cdot \text{Hx_04b} + 153,378$$

$$R^2 = 0,8051 ; Af = 0,7636 ; MAE = 2,81 \text{ jours} ; p < 0,0001$$

Les régresseurs sont strictement identiques à ceux de l'équation élaborée sur 18 ans (cf. 3.2.7), l'ajout d'une seule année n'a pas modifié le contenu des équations, entraînant la stabilité de l'erreur moyenne absolue (2,81 contre 2,63 jours).

En 2006, l'apparition d'un risque est intervenue le 15 mai. Du fait de la proximité des dates de démarrage de cumul, les prévisions, de très bonne facture, ont été très proches allant du 11 mai (GDD_1 et FT) jusqu'au 15 mai exactement pour le modèle issu de la méthode de Sarvas. Quant à l'équation de régression, elle s'est avérée moins précise avec une prévision trop tardive de sept jours.

Après avoir présenté ces cinq stations qui se caractérisent par des modèles présentant des erreurs moyennes absolues peu élevées (cf. 3.2.6 et 3.2.7), nous allons jeter un regard sur les sites présentant au moins 15 ans de données de Poacées exploitables.

3.3.3.3 Les sites du Centre-Est : Lyon, Clermont-Ferrand et Saint-Etienne

Une série de 19 années est exploitable à Lyon de 1987 à 2005. A Clermont-Ferrand et Saint-Etienne, toutes les années depuis l'installation du capteur, en 1991, sont utilisables, soit un total de 15. Notons que les données météorologiques associées à cette dernière ville sont celles de la station dénommée Saint-Etienne-Bouthéon. Sur la période 1961-1990 les températures moyennes hivernale (3,1°C) et estivale (18°C) de cette dernière sont similaires à celles de Clermont-Ferrand, ainsi que l'insolation annuelle (1920 heures). Quant aux précipitations annuelles, bien qu'un peu plus abondantes (708,1 mm), elles présentent la même distribution mensuelle qu'à Clermont-Ferrand (saison «froide» moins arrosée que la saison «chaude»). Les résultats de Montluçon n'ont pas été présentés une nouvelle fois du fait de l'absence de données météorologiques locales fiables. Les dates d'apparition du risque d'exposition allergique à Lyon (figure 63), Clermont-Ferrand et Saint-Etienne (figure 64) sont présentées ci-après.

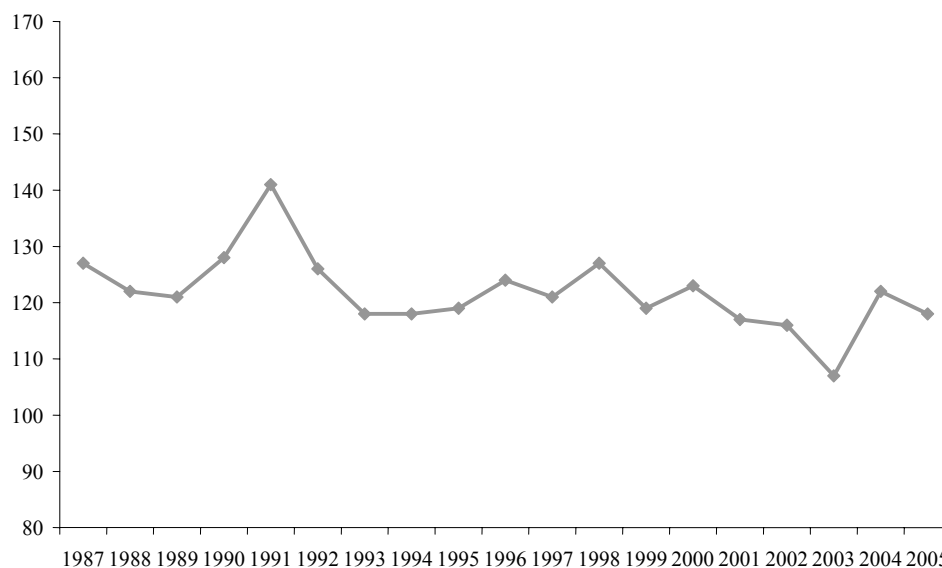


Figure 63 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Lyon, en nombre de jours à partir du 1^{er} janvier.

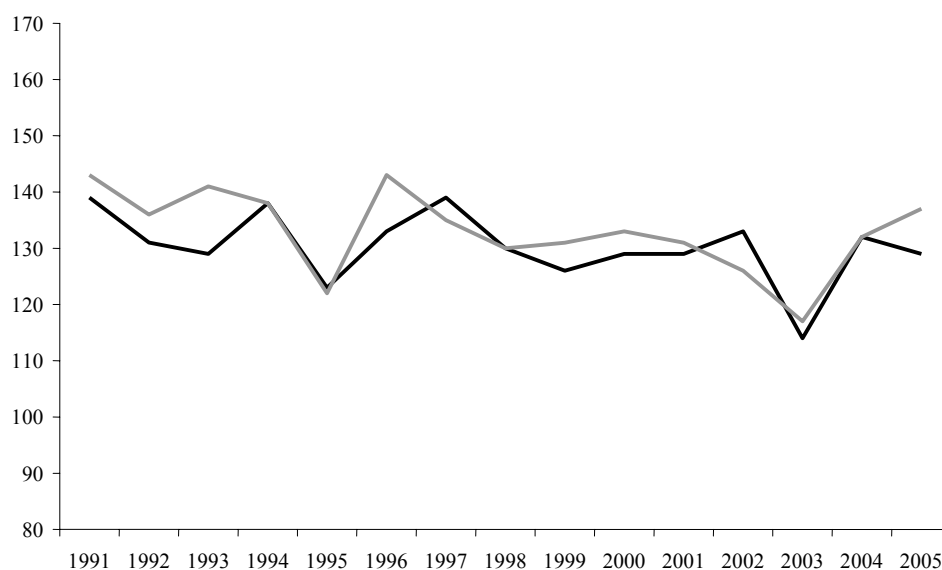


Figure 64 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, en nombre de jours à partir du 1^{er} janvier, à Clermont-Ferrand (noir) et Saint-Etienne (gris).

A Lyon, la date moyenne d'apparition a lieu le 121,8^{ème} jour de l'année, soit le 2 mai dans le cadre d'une année non bissextile. Elle intervient, en moyenne, le 130,3^{ème} jour de l'année à Clermont-Ferrand, soit le 10 mai pour une année non bissextile, et le 133^{ème} jour de l'année à Saint-Etienne ce qui correspond au 13 mai d'une année non bissextile. Les dates d'apparition à Lyon oscillent du 17 avril, en 2003, au 21 mai, en 1991, soit une amplitude maximale de 34 jours. En ce qui concerne Clermont-Ferrand, elles s'échelonnent du 24 avril, en 2003, au 19

mai, en 1991 et 1997, soit une amplitude maximale de 25 jours. Enfin, à Saint-Etienne, l'année 2003 se caractérise aussi par la date d'apparition la plus précoce de la série (27 avril), les plus tardives survenant les 23 mai 1991 et 22 mai 1996 correspondant au 143^{ème} jour de l'année. L'amplitude maximale est ici de 26 jours.

Remarquons qu'au contraire des stations plus septentrionales, des dates d'apparition du risque antérieures au 1^{er} mai apparaissent dans les trois stations. Cela ne concerne que 2003 pour Clermont-Ferrand et Saint-Etienne, mais huit années sur 19 à Lyon.

Notons que les bornes supérieures des dates de début de cumul testées correspondent à la veille de l'apparition du risque la plus précoce. Rappelons que pour les raisons évoquées au début du chapitre 3.3 (équivalence des méthodes et facilité d'élaboration), seules trois méthodes sont utilisées pour Clermont-Ferrand et Saint-Etienne.

- **Lyon :**

- **GDD_1** : 9 mars – 7°C – 2,84 jours – 219,00°D
- **Lej_Tx** : 10 mars – 6 et 6,5°C – 2,89 jours – 819,38 et 816,40°C
- **FT** : 18 mars – 8,5°C – 3,16 jours – 30,83 unités
- **Q₁₀** : 7 mars – 3,1 – 2,84 jours – 471,98 unités
- **Sarvas** : 8 et 9 mars – Tm et Tm(0) – 2,74 jours – 327,20 et 323,14 unités

- **Clermont-Ferrand :**

- **Lej_Tx** : 28 mars – 7,5°C – 3,27 jours – 694,22°C
- **Q₁₀** : 30 mars – 2,4 – 3,27 jours – 259,29 unités
- **Sarvas** : 31 mars et 1^{er} avril – Tm et Tm(0) – 3,33 jours – 255,98 et 250,86 unités

- **Saint-Etienne :**

- **Lej_Tx** : 14 janvier – 0°C – 4,20 jours – 1470,60°C
- **Q₁₀** : 18 et 19 janvier – 2,9 – 3,60 jours – 713,16 et 709,76 unités
- **Sarvas** : 12 janvier – Tm(0) – 3,53 jours – 515,85 unités

Nous constatons que les dates de démarrage de cumul des meilleurs modèles ont lieu en mars à Lyon et Clermont-Ferrand. Pour chacune des deux villes, ces dates sont relativement proches. C'est également le cas à Saint-Etienne, sauf que ces dates ont lieu aux alentours de la mi-janvier. Pour les trois méthodes communes à ces sites, nous enregistrons une erreur moyenne absolue de 2,82 jours à Lyon, 3,29 à Clermont-Ferrand et 3,78 à Saint-Etienne. Cela montre la bonne qualité prévisionnelle des modèles, qui diffère peu d'un type de méthode à l'autre. Quant aux équations de régression, elles se décomposent de la manière suivante :

$$\text{REAP_Lyon} = -1,201 \cdot \text{Tn_01b} + 0,516 \cdot \text{JP5_02a} + 0,949 \cdot \text{Tm0_02a} + 0,588 \cdot \text{Hx_02b} + 0,238 \cdot \text{Hn_03b} + 56,587$$

$$R^2 = 0,7998 ; \text{Af} = 0,7380 ; \text{MAE} = 2,38 \text{ jours} ; p = 0,0004$$

$$\text{REAP_Clermont} = 0,779 \cdot \text{Hn_03b} - 1,134 \cdot \text{JP_03b} + 4,008 \cdot \text{T-20_03b} + 65,892$$

$$R^2 = 0,6992 ; \text{Af} = 0,6222 ; \text{MAE} = 2,55 \text{ jours} ; p = 0,0033$$

$$\text{REAP_St-Etienne} = -0,007 \cdot \text{Gel_01a} - 0,943 \cdot \text{JP1_01b} + 0,567 \cdot \text{Hx_03b} + 88,749$$

$$R^2 = 0,5162 ; \text{Af} = 0,4110 ; \text{MAE} = 4,58 \text{ jours} ; p = 0,0399$$

Nous pouvons noter qu'apparaissent, dans chaque équation, des variables météorologiques de la deuxième moitié du mois de mars. Cette période semble jouer un rôle important dans l'apparition des premières concentrations susceptibles de gêner les allergiques et montre, de nouveau, l'importance des conditions climatiques durant les dernières semaines précédant l'apparition du risque. A noter également que l'équation stéphanoise donne une erreur moyenne absolue plus élevée que les deux autres.

En 2006, l'apparition du risque d'exposition allergique au pollen de la famille des Poacées est survenue le 2 mai à Lyon, le 15 à Clermont-Ferrand et le 14 à Saint-Etienne. A Lyon, le modèle issu de la méthode de Lejoly-Gabriel a donné une date très proche de celle observée (1^{er} mai), tandis que les autres ont donné des prévisions trop précoces allant du 25 au 28 avril, soit sept à quatre jours trop tôt. Les modèles ont ainsi bien pris en compte la spécificité de Lyon se traduisant par une apparition moyenne du risque plus précoce que dans les autres sites et, notamment, en 2006. L'équation de régression, quant à elle, a donné une prévision trop tardive pour cette année : le 8 mai.

A Clermont-Ferrand, dont la date d'apparition du risque est similaire aux stations plus septentrionales, les prévisions des modèles issus des méthodes de cumul ne se sont pas

avérées bonnes car trop précoces de neuf et dix jours. La période de démarrage des cumuls des meilleurs modèles (fin mars) n'est donc pas adéquate pour 2006. La prévision donnée par l'équation de régression a été pire encore en étant en avance de 13 jours par rapport à la date réelle.

Par contre les différents modèles ont très correctement estimé la date d'apparition du risque à Saint-Etienne. Les prévisions s'échelonnent du 11 mai (équation de régression) au 16 mai (modèle issu de Lej_Tx) encadrant les prévisions fournies par les modèles issus du Q₁₀ et de Sarvas qui se sont avérées exactes. En ce qui concerne les modèles de cumul, à Clermont-Ferrand comme à Saint-Etienne, nous constatons de nouveau que des dates de démarrage de cumul équivalentes débouchent sur des dates prévues très rapprochées.

3.3.3.4 Les sites du Sud-Ouest : Bordeaux et Toulouse

Les années 1988 à 2005 sont exploitables à Bordeaux, soit une série de 18 ans. A Toulouse, 16 années sont utilisables pour l'analyse de 1989 à 2005, l'année 1990 étant écartée. Les données météorologiques, affectées à cette dernière, sont celles de la station de Toulouse-Blagnac. Sur la période 1961-1990, la température moyenne hivernale est proche de celle de Bordeaux (6°C), mais l'été y est plus chaud (20,2°C contre 19,3 à Bordeaux). Du fait de l'éloignement par rapport à l'Atlantique, les précipitations annuelles sont modérées (655,4 mm) et plus faibles qu'à Bordeaux. Par contre la durée d'insolation est similaire : 2047,3 heures par an à Toulouse contre 2083,7 à Bordeaux. Les dates d'apparition du risque d'exposition allergique des deux villes sont présentées dans les figures 65 et 66.

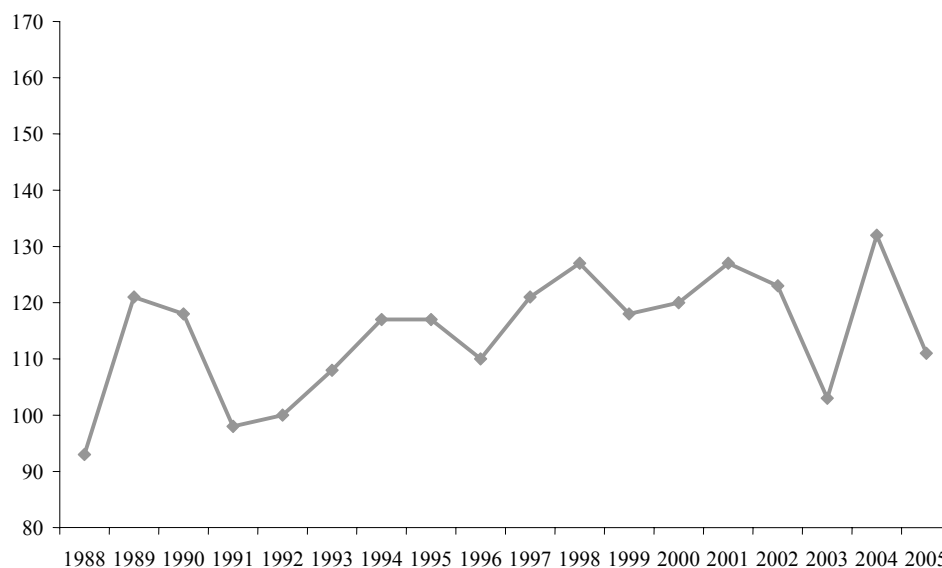


Figure 65 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.

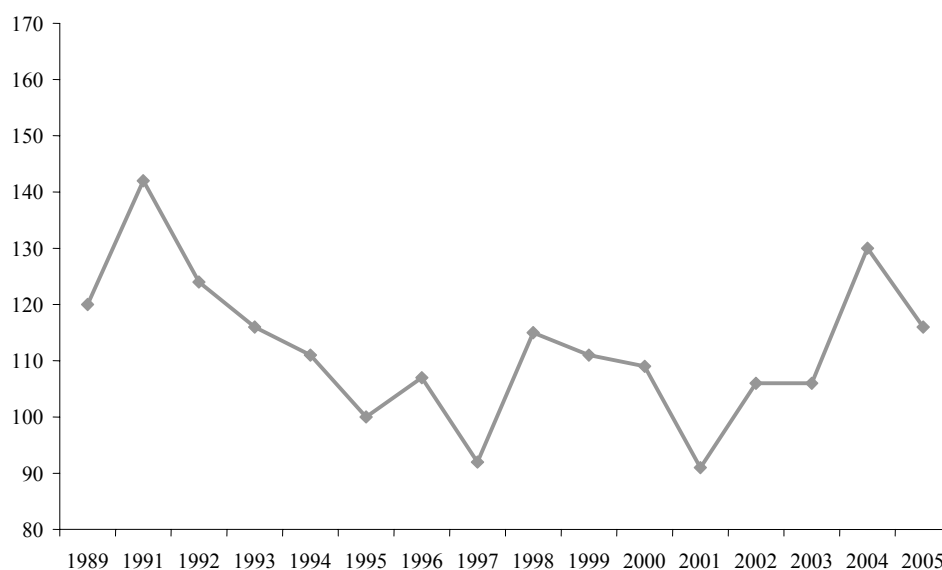


Figure 66 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Toulouse, en nombre de jours à partir du 1^{er} janvier.

A Bordeaux, la date moyenne d'apparition a lieu le 114,7^{ème} jour de l'année, soit le 25 avril dans le cadre d'une année non bissextile. Elle intervient, en moyenne, le 112,3^{ème} jour de l'année à Toulouse soit le 22 avril pour une année non bissextile. Les dates d'apparition à Bordeaux oscillent du 2 avril, en 1988, au 11 mai, en 2004, soit une amplitude maximale de 39 jours. En ce qui concerne Toulouse, elles s'échelonnent du 1^{er} avril, en 2001, au 22 mai, en 1991, soit une amplitude maximale de 50 jours. Les dates moyennes sont donc nettement plus

précoces que pour les sites étudiés jusqu'alors. Cependant, elles doivent être considérées avec précaution à Bordeaux. En effet, pour plusieurs années, des concentrations supérieures à 10 grains/m³ sont enregistrées sur quelques jours d'affilée dès le début du mois de mars, et sont entrecoupées de périodes de plusieurs jours de concentrations très faibles. La détermination de l'apparition du risque n'en est rendue, alors, que plus compliquée. Nous avons ainsi décidé de n'étudier que les périodes où le total cumulé sur cinq jours atteint ou dépasse 50 grains/m³, démarrant à partir du 1^{er} avril. Cela dans le but de limiter la variabilité interannuelle et de se rapprocher de la pleine saison pollinique des Poacées qui démarre généralement à partir de la mi-mai. A Toulouse, par contre, nous ne retrouvons pas de telles concentrations dès le mois de mars. Le démarrage de la saison y étant plus marqué, la détermination du risque s'en trouve d'autant plus aisée.

Notons que les dates de début de cumul testées s'échelonnent du 1^{er} janvier à la veille de l'apparition du risque la plus précoce pour chacune des deux villes.

- **Bordeaux :**

- **Lej_Tx** : 1^{er} avril – 19°C – 7,67 jours – 114,00°C
- **Q₁₀** : 18 mars – 2,4 – 8,39 jours – 243,70 unités
- **Sarvas** : 16 mars – Tx et Tx(0) – 8,22 jours – 484,98 unités

- **Toulouse :**

- **GDD_1** : 21 mars – 14°C – 4,38 jours – 18,34°D
- **Lej_Tx** : 21 mars – 19,5°C – 3,56 jours – 98,51°C
- **FT** : 21 mars – 16,5°C – 3,94 jours – 2,91 unités
- **Q₁₀** : 19 janvier – 10,1 – 5,06 jours – 3885,60 unités
- **Sarvas** : 14 janvier – Tm et Tm(0) – 5,31 jours – 472,35 et 471,25 unités

A Bordeaux, nous n'avons testé que trois des cinq méthodes pour les raisons évoquées au début du chapitre 3.3. A Toulouse, les cinq méthodes avaient été testées dans l'optique préalable d'une étude sur les Cupressacées-Taxacées qui a été abandonnée par la suite. Nous constatons la qualité médiocre des modèles de Bordeaux, probablement due à la présence de ces petites séries de concentrations oscillant entre 10 et 30 grains qui semblent décorrélées

avec la phénologie de cette famille. Au vu de ce contexte, nous avons déterminé, en guise de test, comme date d'apparition du risque le premier jour (J) atteignant 30 grains/m³ avec, au moins, un total cumulé de 150 grains de J à J+4 ; à condition que trois des cinq jours enregistrent des concentrations supérieures ou égales à 30 grains. Ce critère se basant sur un seuil journalier plus élevé que celui retenu (10 grains), les dates d'apparition du risque sont en moyenne plus tardives, 13 mai pour une année non bissextile, et surtout plus faciles à déterminer. Les modèles issus des méthodes de cumul, appliqués sur ces nouvelles dates, ne sont pour autant guère meilleurs que ceux appliqués sur les dates issues des niveaux de risque initiaux. Avec des erreurs moyennes absolues allant de 5,83 (Lej_Tx) à 6,11 jours (Sarvas), ces modèles ne sont pas particulièrement performants.

A Toulouse, où la qualité des modèles est meilleure qu'à Bordeaux mais assez quelconque, nous constatons une opposition entre les modèles issus des méthodes du Q₁₀ et de Sarvas, et les autres au niveau de la période de démarrage des cumuls : janvier pour les deux premiers cités, mars pour le reste. A noter également que les erreurs moyennes des modèles issus du Q₁₀ et de la méthode de Sarvas dépassent cinq jours.

Les équations de régression sont, quant à elles, les suivantes :

$$\text{REAP_Bordeaux} = 2,267 \cdot \text{Tx_02a} + 19,938 \cdot \text{JP_02b} + 1,156 \cdot \text{JP_03b} - 2,779 \cdot \text{JP5_03b} + 74,722$$

$$R^2 = 0,8360 ; \text{Af} = 0,7845 ; \text{MAE} = 3,51 \text{ jours} ; p < 0,0001$$

$$\text{REAP_Toulouse} = - 2,637 \cdot \text{Tm_01b} - 11,016 \cdot \text{Tn_02a} - 6,239 \cdot \text{Tn0_02a} + 0,486 \cdot \text{Hn_03b} + 159,868$$

$$R^2 = 0,9494 ; \text{Af} = 0,9032 ; \text{MAE} = 2,48 \text{ jours} ; p < 0,0001$$

La précision des équations de régression est meilleure que celle des modèles de cumul, symbolisée en cela par des erreurs moyennes absolues nettement plus faibles.

En 2006, l'apparition du risque d'exposition allergique a eu lieu le 22 avril à Bordeaux et le 17 du même mois à Toulouse.

Du fait de la mauvaise qualité des modèles de cumul, seule la prévision issue de l'équation est présentée à Bordeaux. Celle-ci tombant le 14 avril, elle s'est avérée trop précoce de huit jours.

A Toulouse, les meilleures prévisions ont été celles issues des modèles découlant des méthodes du Q₁₀ (21 avril) et de Sarvas (22 avril) ; ce qui représente quatre et cinq jours d'écart avec les valeurs observées. Ici encore les dates de démarrage de cumul, proches et

précoces (janvier), ainsi que le type de méthode employé (exponentiel), ont influencé les dates prévues. Les trois autres méthodes de cumul, initiées beaucoup plus tard, ont donné des prévisions beaucoup trop précoces (31 mars et 1^{er} avril), alors que celle fournie par l'équation de régression a été beaucoup trop tardive (7 mai).

3.3.3.5 Nantes

Seize années sont exploitables de 1989 à 2005, l'année 1992 étant retirée. Les données météorologiques proviennent de la station de Nantes-Bouguenais. Les températures moyennes hivernale et estivale (période 1961-1990) sont clémentes avec des valeurs de 5,9 et 18,2°C respectivement. Bien que relativement proche de l'océan, la ville enregistre des précipitations annuelles qui ne sont pas excessives (788,5 mm). L'insolation annuelle, quant à elle, s'établit à 1956,3 heures. Les dates d'apparition du risque d'exposition allergique sont présentées dans la figure 67.

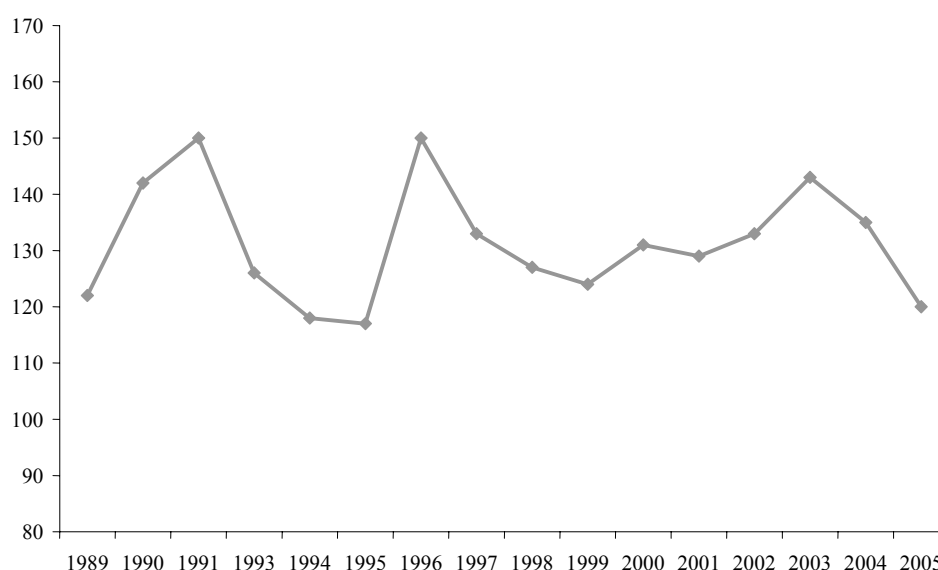


Figure 67 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Nantes, en nombre de jours à partir du 1^{er} janvier.

La date moyenne d'apparition du risque a lieu le 131,3^{ème} jour de l'année, soit le 11 mai dans le cadre d'une année non bissextile. Les dates d'apparition s'échelonnent du 27 avril, en 1995, jusqu'aux 30 mai 1991 et 29 mai 1996, deux dates qui correspondent au 150^{ème} jour de l'année. L'amplitude maximale qui en résulte atteint 33 jours. Les dates de début de cumul testées s'échelonnent du 1^{er} janvier au 26 avril.

- **Lej_Tx** : 26 avril – 19°C – 5,69 jours – 109,01°C
- **Q₁₀** : 26 avril – 11,8 à 12 – 5,19 jours – 2269,70 à 2351,40 unités
- **Sarvas** : 26 avril – T_m et T_m(0) – 7,06 jours – 137,28 unités

La qualité des modèles est globalement médiocre, surtout au regard de l'erreur moyenne absolue associée au modèle issu de la méthode de Sarvas. Statistiquement, la date de démarrage de cumul à l'origine des «meilleurs» modèles a lieu le 26 avril, soit un jour avant la date d'apparition la plus précoce de la série. Cette date n'est évidemment pas adaptée dans le cadre préventif, ce qui permet de rappeler que nous nous attachons à présenter, dans cette section, la meilleure combinaison date-seuil d'un point de vue statistique. Pour que les différents modèles soient opérationnels, si l'année se révèle précoce, il faudrait se fixer le 10 avril comme date-butoir de démarrage de cumul. En tout cas, cela montre que les modèles initiés plus tôt présentent des erreurs moyennes absolues encore plus élevées.

L'équation de régression, quant à elle, est la suivante :

$$\text{REAP_Nantes} = 2,965 * \text{Tm0_02a} + 129,211$$

$$R^2 = 0,3128 ; \text{Af} = 0,2191 ; \text{MAE} = 7,07 \text{ jours} ; p = 0,0243$$

Il s'agit de la seule variable significativement corrélée, pour un risque d'erreur de 5%, aux dates d'apparition du risque. De ce fait cette équation n'est guère exploitable dans un but préventif.

En 2006, l'apparition du risque a eu lieu le 14 mai, soit une date aux alentours de la moyenne des 16 années de la série. Nous présentons seulement les dates prévues à partir des modèles issus des méthodes Lej_Tx et du Q₁₀ qui sont les moins mauvais de la liste. Paradoxalement, compte tenu de leur précision relativement médiocre, ils ont donné tous deux de bonnes prévisions : le 10 mai pour Lej_Tx et le 13 mai pour le Q₁₀.

3.3.3.6 Marseille

Dix-huit années sont exploitables de 1988 à 2005 auxquelles sont associées les données météorologiques de la station de Marignane. Le climat est méditerranéen avec, sur la période 1961-1990, une température moyenne hivernale douce (7,3°C), un été chaud (22,6°C) et,

surtout, une insolation annuelle moyenne nettement supérieure aux autres stations de l'étude (2835,5 heures). Les précipitations sont variables d'une saison à l'autre, les plus forts cumuls étant enregistrés à l'automne et les plus faibles durant l'été. Le total annuel de 544,4 mm se révèle modéré, et fait de Marseille la station la moins pluvieuse des villes de l'étude. Les dates d'apparition du risque d'exposition allergique sont présentées dans la figure 68.

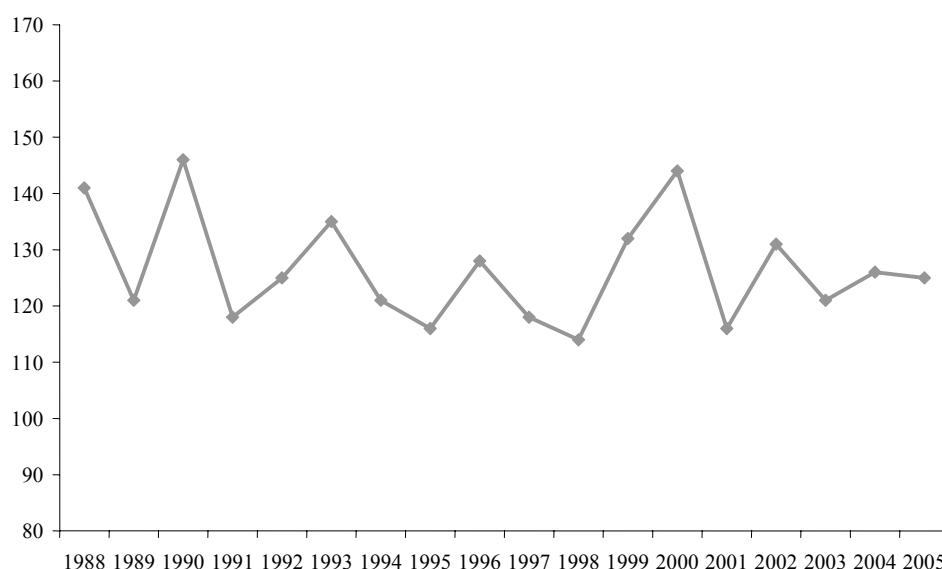


Figure 68 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Marseille, en nombre de jours à partir du 1^{er} janvier.

La date moyenne d'apparition du risque a lieu le 126,6^{ème} jour de l'année, soit le 7 mai dans le cadre d'une année non bissextile. Les dates s'échelonnent du 24 avril, en 1998, au 26 mai, en 1990, soit une amplitude maximale de 32 jours. A première vue, ces dates apparaissent tardives par rapport à ce qui est enregistré à Lyon où la date moyenne, sur la période 1988-2005, survient le 2 mai malgré des températures plus basses de janvier à avril. Les causes de ce relatif caractère tardif sont à chercher dans les concentrations mêmes. Sur la période 1997-2005, on comptabilise 2 à 3,5 fois moins de pollens de Poacées à Marseille que dans les villes précédemment étudiées. De ce fait les concentrations journalières nécessaires pour atteindre le niveau de risque retenu pour ce taxon ne peuvent être atteintes aussi tôt que pour une ville comme Bordeaux, qui présente des conditions hivernales de température assez proches, mais où le total global de pollens de Poacées est 2,6 fois supérieur à celui enregistré à Marseille. Les dates de début de cumul testées s'échelonnent, quant à elles, du 1^{er} janvier au 23 avril.

- **GDD_1** : 28 mars – 0°C – 7,94 jours – 573,29°D
- **Lej_Tx** : 28 février – 9,5°C – 7,33 jours – 1210,20°C
- **FT** : 22 février – 8°C – 7,22 jours – 63,53 unités
- **Q₁₀** : 2 mars – 1,1 – 7,61 jours – 150,56 unités
- **Sarvas** : 23 mars – Tx et Tx(0) – 7,78 jours – 664,04 unités

Les cinq méthodes avaient été utilisées pour l'étude des Cupressacées-Taxacées abandonnée à cause d'un trop grand nombre d'années où le capteur ne fonctionna pas de manière continue.

La qualité globale des modèles de cumul est médiocre avec des erreurs moyennes absolues comprises entre sept et huit jours. Dans ces conditions, ils ne sont guère utilisables dans une optique préventive.

L'équation de régression, quant à elle, est la suivante :

$$\text{REAP_Marseille} = -25,488 \cdot \text{JP_02b} + 3,088 \cdot \text{P_03b} + 1,844 \cdot \text{P_04a} + 125,975$$

$$R^2 = 0,5553 ; \text{Af} = 0,4789 ; \text{MAE} = 5,15 \text{ jours} ; p = 0,0084$$

L'équation de régression présente une erreur moyenne absolue plus faible que celles des différents modèles de cumul, mais qui dépasse tout de même les cinq jours. Nous pouvons noter que seules des variables de précipitations sont incluses dans l'équation. Elles semblent ainsi jouer un rôle important dans l'apparition du risque. C'est la seule équation, avec celle de Saint-Etienne, à ne pas inclure au moins un paramètre thermique.

En 2006, l'apparition du risque a eu lieu le 26 avril, soit la deuxième date la plus précoce depuis 1988, égalant les années 1995 et 2001. Nous présentons seulement la date prévue à partir de l'équation de régression à cause de la trop médiocre qualité des modèles de cumul. Cette date est tombée le 30 avril, ce qui constitue plutôt une bonne prévision, de quatre jours trop tardive.

Nous allons ensuite nous attacher au cas de l'ambrosie à Lyon, seule ville où les concentrations de ce taxon sont réellement exploitables. Pour cela nous allons développer ce sous-chapitre sur le même mode que ceux du chapitre 3.2, en plus simplifié.

3.3.4 Résultats pour l'ambroisie : le cas de Lyon

Dix-neuf années sont exploitables de 1987 à 2005. Les dates d'apparition du risque d'exposition allergique sont présentées dans la figure 69.

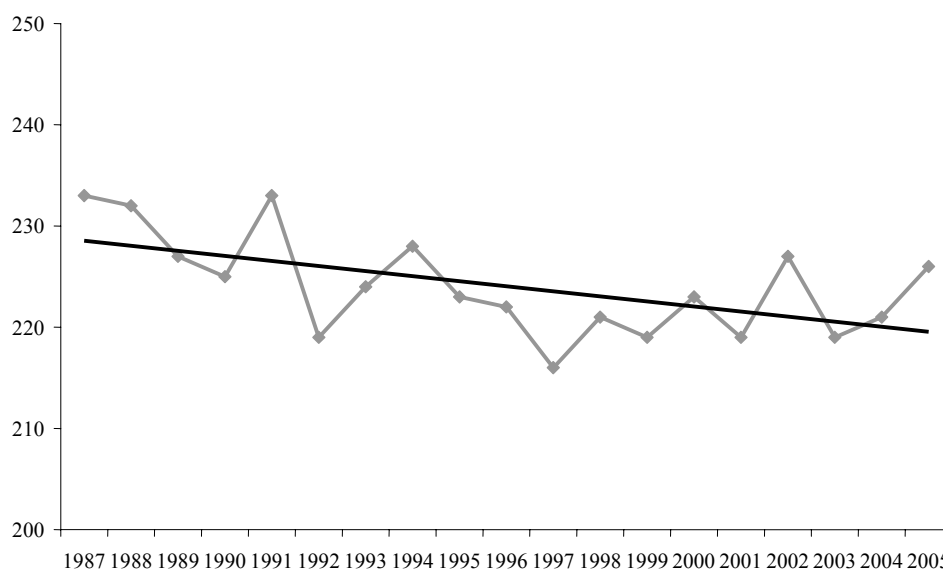


Figure 69 : Dates d'apparition du risque d'exposition allergique au pollen d'ambroisie, à Lyon, en nombre de jours à partir du 1^{er} janvier et droite de tendance linéaire.

La date moyenne d'apparition du risque correspond au 224,1^{ème} jour de l'année, soit le 12 août dans le cadre d'une année non bissextile. Les dates s'échelonnent du 4 août, en 1997, au 21 août, en 1987 et 1991, soit une amplitude maximale de 17 jours. Contrairement aux graphiques précédents, nous avons ajouté la droite de tendance linéaire car nous constatons une corrélation négative significative ($-0,5608$), pour un risque d'erreur de 5%, entre les dates d'apparition du risque et les années. Il s'agit, toutes villes et tous taxons confondus du coefficient le plus élevé. Une tendance globale à la précocité de l'apparition du risque semble se dessiner même si elle n'explique que 31% de l'évolution interannuelle de ces dates. Mais l'apparition du risque ayant eu lieu le 15 août en 2006 et le 17 août en 2007, cette précocité devient beaucoup moins visible à cause d'une inversion de tendance qui se confirme depuis 2003.

Nous allons maintenant présenter les résultats liés aux cinq méthodes de cumul et ceux issus de la régression linéaire multiple.

3.3.4.1 Résultats des méthodes de cumul

A l'instar de ce qui a été effectué dans le chapitre 3.2, nous souhaitons déterminer si un mois se distingue, en particulier au niveau de la qualité des modèles qui en découlent. Nous allons comparer les erreurs moyennes absolues de l'ensemble des cinq méthodes de cumul, mois par mois. En lien avec la germination de la plante qui s'effectue dès le mois d'avril et courant mai (Thibaudon *et al.*, 2004), nous allons ainsi tester les dates de démarrage de cumul du 1^{er} mars au 31 juillet, et cela pour les seuils d'actions présentés dans la section 3.2.2.1. La seule exception concerne les méthodes FT, GDD_1 et Lej_Tx où le seuil maximal de température testé est 25°C, et non 20°C comme pour les autres taxons, du fait de la période d'accumulation des unités de «chaleur» qui comprend une bonne partie de l'été. Les résultats sont présentés dans la figure 70.

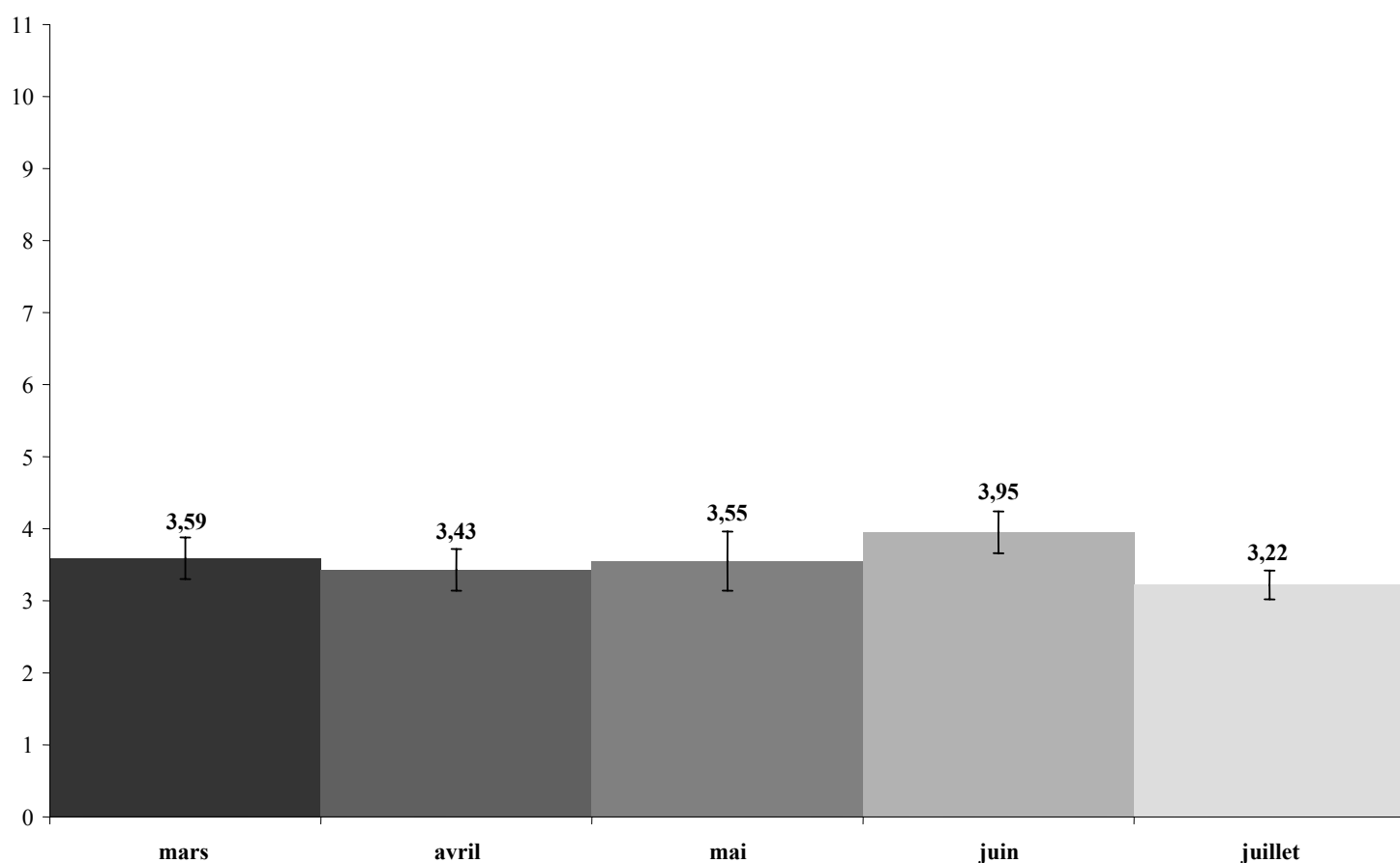


Figure 70 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles, mois par mois, pour l'ambrosie à Lyon.

Des cinq séries d'erreur moyenne absolue testées seule celle de mai présente une distribution non gaussienne, selon les tests de normalité de Shapiro-Wilk et Kolmogorov-Smirnov. Pour les comparaisons incluant cette série, nous devrions utiliser un test de Mann-Whitney. Cependant, les résultats étant identiques à ceux fournis par le test de Student, ce dernier est le seul à être retenu pour l'ensemble de l'étude sur l'ambroisie. De manière générale, et d'après le test de Student, l'erreur moyenne absolue de juin est significativement supérieure à celles d'avril et de juillet, qui présentent toutes deux les valeurs les plus faibles. Nous notons également une différence significative entre mars et juillet. Malgré tout, les écarts entre les valeurs étant faibles, 0,73 jour maximum, nous pouvons constater que des démarrages de cumul précoces débouchent sur des modèles à priori aussi fiables que ceux initiés plus tardivement. Il semble donc que la température joue un rôle bénéfique sur le développement de l'ambroisie dès le mois de mars. Dans un but préventif, il serait intéressant d'initier, pour chaque méthode, au moins cinq modèles (c'est-à-dire un par mois) ; le modèle initié, de préférence, entre le 1^{er} et le 15 juillet apparaissant incontournable. Mais ici, nous nous bornons à présenter les combinaisons date-seuil minimisant l'erreur moyenne absolue.

Suite à cela, nous présentons les meilleurs modèles issus des cinq méthodes de cumul. Sont évoqués dans l'ordre la date de démarrage, le seuil ou le type de données retenu, l'erreur moyenne absolue, l'erreur moyenne au carré et le cumul d'unités de «chaleur», de degrés-jours ou de température à approcher le plus possible :

- **GDD_1** : 17 juillet – 7,5°C – 3,53 jours – 4,41 jours – 405,09°D
- **Lej_Tx** : 29 juillet – 24,5°C – 3,21 jours – 4,33 jours – 359,68°C
- **FT** : 15 mai – 13°C – 2,84 jours – 3,64 jours – 83,65 unités
- **Q₁₀** : 27 juillet – 2,2 – 3,26 jours – 4,19 jours – 231,38 unités
- **Sarvas** : 28 juillet – T_n et T_n(0) – 3 jours – 4,05 jours – 194,59 unités

La qualité globale des modèles de cumul est équivalente, le modèle issu de la méthode FT se caractérise par une erreur moyenne absolue inférieure à trois jours. C'est, par ailleurs, le seul modèle initié en mai.

3.3.4.2 Résultats de la régression linéaire multiple et prévisions pour 2006

Les variables bimensuelles significativement corrélées aux dates d'apparition du risque, pour un seuil de significativité de 95%, sont tout d'abord présentées «quinzaine» par «quinzaine».

Le signe du coefficient de corrélation est indiqué entre parenthèses :

- **2^{ème} «quinzaine» de mars (03b) :**

- **JP (+)**

- **1^{ère} «quinzaine» d'avril (04a) :**

- **Hn (+)**

- **Hx (+)**

- **Hm (+)**

- **2^{ème} «quinzaine» d'avril (04b) :**

- **Hx (+)**

- **1^{ère} «quinzaine» de mai (05a) :**

- **Tx (-)**

- **Tm (-)**

- **Hm (+)**

- **T-10 (-)**

- **T-20 (-)**

- **2^{ème} «quinzaine» de mai (05b) :**

- **Tn (-)**

- **Tx (-)**

- **Tm (-)**

- **T-10 (-)**

- T-20 (-)
- 2^{ème} «quinzaine» de juin (06b) :
 - Hn (+)
 - Hm (+)
- 2^{ème} «quinzaine» de juillet (07b) :
 - JP (+)

A ces variables il faut ajouter cinq autres partiellement corrélées avec les dates d'apparition du risque : JP1_04a (la seule corrélée négativement), Insol_04b, Hm_05b, Insol_05b et Insol_06b.

Nous constatons que l'humidité relative et la température sont les paramètres qui ressortent le plus fréquemment. Plus l'humidité est faible et plus la température est élevée, plus l'apparition du risque sera précoce. De ce point de vue, nous pouvons constater l'importance du rôle joué par les températures du mois de mai sur l'apparition de ce risque, période durant laquelle, rappelons-le, se déroule la germination de la plante.

L'équation de régression, qui découle de ces corrélations, est ensuite présentée. Il s'agit bien évidemment de celle qui maximise la valeur du test Af.

$$\text{REAP_Lyon} = 0,732 \cdot \text{Hx_04b} + 0,022 \cdot \text{Insol_04b} - 1,859 \cdot \text{Tn_05b} + 0,358 \cdot \text{Hm_05b} + 0,017 \cdot \text{Insol_05b} + 141,167$$

$$R^2 = 0,8983 ; \text{Af} = 0,8497 ; \text{MAE} = 1,20 \text{ jour} ; \text{RMSE} = 1,56 \text{ jour} ; p < 0,0001$$

L'équation de régression présente une erreur moyenne absolue très faible et se caractérise par l'absence de variables des mois de juin et juillet.

En 2006, l'apparition du risque est survenue le 15 août. Les différents modèles de cumul ont estimé des dates assez différentes. Les modèle issu de la méthodes GDD_1 a fourni une prévision trop précoce : le 7 août. Cette précocité est à mettre en lien avec les températures très élevées de la deuxième moitié du mois de juillet. Les modèles issus des méthodes FT et de Sarvas ont abouti à une prévision plus précise : le 11 août. Le modèle découlant de la

méthode du Q_{10} a été le plus précis (14 août) alors que celui issu de la méthode de Lejoly-Gabriel a déterminé le 18 août comme date d'apparition du risque. Dans ce dernier cas, le seuil d'action très élevé ($24,5^{\circ}\text{C}$) a été dépassé moins fréquemment qu'à l'accoutumée, du fait de températures maximales inférieures aux normales saisonnières en août, expliquant le caractère légèrement tardif de la prévision. Enfin l'équation de régression a donné une prévision de quatre jours trop précoce (11 août). Malgré la bonne qualité générale des modèles, la grande variété des dates obtenues se révèle quelque peu gênante dans le cadre d'une prévision efficace. Le type de méthode semble être ici la cause principale de ces divergences. Au bout du compte la régression linéaire multiple, pour ce taxon particulier, se distingue par l'aisance de son application.

3.3.5 Discussion et conclusion

Les différents modèles de cumul et les équations de régression ont donc été présentés pour les séries les plus longues à disposition. La comparaison entre les résultats des villes de Rouen, Amiens, Paris, Nancy et Strasbourg, exposés dans le chapitre 3.2, et ceux rapportés ci-dessus ont été très instructifs dans le sens où ils ont permis de comparer l'évolution des modèles avec celle de la longueur des séries. Nous avons pu ainsi voir que les prédicteurs retenus, dans les équations de régression, varient de manière importante avec l'ajout de nouvelles années servant à l'élaboration des modèles. La faible longueur des séries est à l'origine de cette inconstance qui aura, sans doute, tendance à s'atténuer dès que les séries seront plus longues (30 ans et plus). Le poids d'une année nouvelle allant déclinant avec l'augmentation de la durée de la série à laquelle elle est ajoutée, nous pouvons espérer, à la longue, déterminer plusieurs variables robustes qui seront à l'origine d'équations de contenu beaucoup plus stable.

En ce qui concerne les modèles issus des méthodes de cumul, nous avons noté que les seuils d'action et les dates de démarrage de cumul étaient moins fluctuants. Pour plusieurs modèles, l'ajout d'une ou plusieurs années ne faisait quasiment pas varier ces deux paramètres. Cette plus grande constance est probablement à relier au fait que l'on n'utilise qu'un seul paramètre : la température. Les intercorrélations existant entre les différents paramètres météorologiques, dans le cadre de la régression linéaire multiple, génèrent l'instabilité constatée du fait de la grande diversité des variables mises en commun ; cela n'est pas le cas

pour les modèles de cumul. Nous avons enfin constaté que l'ajout d'années n'était pas toujours un gage d'amélioration de la précision des modèles.

Par ailleurs nous avons eu confirmation que, globalement, les méthodes de cumul étaient statistiquement équivalentes, notamment les deux méthodes de *Growing Degree Days* développées par Zalom *et al.* (1983). Cela explique le fait que nous ayons seulement retenu trois méthodes pour certaines des villes présentées. Nous avons remarqué, essentiellement pour les Poacées, que la qualité des modèles élaborés sur les villes de la moitié Sud de la France (Bordeaux, Toulouse et Marseille) était moins bonne que pour les villes plus septentrionales, à l'exception de Nantes. La difficulté de détermination de l'apparition du risque, du fait de petites quantités apparaissant tôt dans l'année (Bordeaux) ou de concentrations deux à trois fois plus faibles que dans la moitié Nord (Marseille), semble être à l'origine de ce phénomène. Les meilleurs modèles ont été déterminés pour des villes présentant à la fois de fortes concentrations et une pollinisation progressive, non entrecoupée (ou rarement) de longues périodes de faibles concentrations, idéale pour une détermination aisée de la date d'apparition du risque d'exposition allergique.

L'utilisation prospective des modèles pour l'année 2006 a été également très instructive. Cette année s'est révélée intéressante, car elle a permis de valider les modèles par la prévision de dates d'apparition tardives (frêne et bouleau) et se situant dans la moyenne (Poacées). Nous avons pu ainsi constater que les équations de régression n'ont, dans la plupart des cas, pas fourni de bonnes prévisions. De ce point de vue, les modèles de cumul ont été globalement plus performants avec, à plusieurs reprises, des prévisions exactes ou très proches des dates observées. Ce dernier constat nous a permis de mettre en avant l'influence de différentes dates de démarrage éloignées sur la qualité de la prévision (le frêne à Paris et les Poacées à Nancy notamment). Des dates d'initiation des modèles très proches ont, par ailleurs, été à l'origine de prévisions qui l'ont également été : c'est particulièrement net pour les Poacées. Cependant, nous avons aussi constaté, dans plusieurs cas, des variations de la prévision en lien avec d'assez faibles différences dans les dates de démarrage, comme pour le bouleau à Strasbourg et l'ambroisie à Lyon. Ici le type de méthode employé, et les seuils d'action qui en découlent, semblent primer sur les dates de début de cumul pour expliquer les écarts de prévision. Notons enfin, qu'à plusieurs reprises, les méthodes exponentielles du Q_{10} et de Sarvas se sont distinguées par la qualité de leur prévision et par les dates de démarrage de cumul très différentes des autres méthodes. Cela a été le cas à Nancy (bouleau et Poacées), Amiens (bouleau), Strasbourg (bouleau) et Toulouse (Poacées).

Ainsi, les différences enregistrées dépendent à la fois de la date début de cumul, du type de méthode et du seuil associé. Le premier paramètre cité joue un rôle important dans plusieurs cas et, de ce fait, peut à nouveau poser le problème de l'intérêt de la détermination de la dormance. En effet, le calcul de cette dernière permettrait de déterminer une date de début de cumul des unités de «chaleur» propre à chaque année, qui ne concernerait d'ailleurs que le frêne et le bouleau. Cela permettrait d'améliorer, en théorie, la précision globale des modèles ou, tout du moins, réduire les résidus des années servant à leur élaboration où les dates sont mal estimées.

Cependant cela est loin d'être certain. Une accumulation des unités de «froid» a été réalisée, à titre d'exemple sur quelques stations, à partir d'une des équations développées par Aron et Gat (1991). Sans trop entrer dans les détails, nous nous sommes d'abord appuyés sur les modèles présentés dans ce chapitre, en considérant, pour chacun d'eux, que la date de démarrage de cumul (par exemple le 20 février) correspondait à la fin de la dormance pour les années où les écarts entre dates prévues et observées étaient inférieures à trois jours. A partir de ces années nous avons calculé le cumul moyen de «chilling» en choisissant arbitrairement le premier jour, après l'équinoxe d'automne, où la température moyenne est inférieure à 12,5°C et la date de début de cumul des unités de «chaleur» (le 20 février par exemple) comme bornes de détermination de la période de dormance. Ainsi, au lieu d'avoir une date fixe, nous avons démarré les accumulations d'unités de «chaleur» (en conservant les seuils d'action initiaux, mais en adaptant les cumuls moyens) à des dates différentes, dont la variabilité maximale atteignait un à deux mois. Les résultats n'ont pas été probants et se sont caractérisés par une détérioration de la qualité des modèles. De plus, beaucoup de ceux initiés en février ont donné de bonnes prévisions en 2006, période où l'on peut raisonnablement penser que la dormance a été accomplie depuis un petit moment. A cela il faut ajouter que le fait de s'appuyer sur une date de démarrage de cumul tardive (durant la deuxième moitié de février par exemple) pour déterminer les unités de «froid» a conduit, pour des années à hiver doux, à des dates de fin de dormance ultérieures à l'apparition du risque lié au frêne ou au bouleau. Au vu des différents cas présentés auparavant, la détermination de la dormance n'est d'aucune utilité si la date de début de cumul s'effectue tardivement. De plus, les dates définies ne s'accompagnent pas d'une amélioration de la qualité des modèles, confirmant notre choix d'écarter le calcul de la dormance de l'analyse.

Nous pouvons noter, par ailleurs, que les modèles de cumul à l'origine des meilleures dates de prévision en 2006, pour un taxon donné, ne sont pas initiés à la même période ; c'est notamment le cas des Poacées. Pour ce dernier taxon, les dates de démarrage de cumul à

l'origine des prévisions les plus performantes se situent pendant la dernière décade d'avril à Amiens, Strasbourg, Rouen et Paris ; alors qu'elles ont lieu au mois de janvier à Toulouse, Saint-Etienne et Nancy. Ainsi, pour une même année, la date de démarrage optimale peut différer grandement d'une ville à l'autre, et cela même entre deux villes assez proches d'un point de vue thermique (Nancy et Strasbourg).

Par conséquent, dans un cadre préventif et opérationnel, et au vu des constations effectuées ci-dessus, il semble pertinent, pour une méthode donnée, de disposer d'un jeu de modèles initiés à des mois différents en conservant, au moins, deux à trois semaines d'accumulation d'unités de «chaleur». De plus il convient de tester l'ensemble des cinq méthodes utilisées au regard de l'influence que peut avoir leur conception, et les seuils associés, sur la qualité de la prévision d'une année donnée. Mais si nous étions forcés de choisir une méthode et une période de début de cumul, au vu des résultats précédents et de ceux du chapitre 3.2, la méthode de Lejoly-Gabriel et les mois précédant immédiatement celui de la pollinisation la plus précoce seraient retenus.

Face à ce constat, l'apport d'observations phénologiques serait d'une très grande utilité durant les dernières semaines précédant l'apparition du risque. Cependant si l'on se réfère à l'étude d'Estrella *et al.* (2006), à l'échelle de l'Allemagne, des écarts considérables ont été enregistrés entre les dates de floraison et de début de pollinisation du bouleau, pouvant atteindre plusieurs dizaines de jours. L'apport de pollens exogènes à la zone de recouvrement du capteur, le type de critère de détermination du début de la saison, ou une mauvaise identification des dates de floraison sont autant de raisons expliquant les réserves émises, par les auteurs, à propos de l'utilisation de la phénologie dans les prévisions polliniques. Malgré tout, cela pourrait apporter des informations précieuses, même si elles ne sont pas très précises, sur le stade de développement du bouleau et du frêne, car cela semble plus difficile pour les Poacées du fait de la multitude d'espèces composant cette famille. Ainsi un stade phénologique donné pourrait être utilisé comme point de départ des modèles de cumul et, dans le cas des équations de régression linéaire multiple, les dates des différents stades seraient considérées comme des régresseurs conjointement aux variables météorologiques.

3.4 ETUDES COMPARATIVES COMPLEMENTAIRES

Elles concernent les cinq méthodes de cumul (Q_{10} , FT, GDD_1, Lej_Tx et Sarvas) et les modèles qui en sont issus sous différents angles. Il va s'agir, tout d'abord, d'évoquer les relations éventuelles entre les seuils d'action associés aux quatre premières et différents paramètres géographiques et météorologiques. Nous tenterons ensuite d'évaluer la différence de précision des modèles, à travers l'erreur moyenne absolue, selon le type de critère de détermination de la date d'apparition du risque d'exposition allergique. Le critère de Nilsson et Persson (1981), qui stipule que l'apparition du risque (ou la date de début de pollinisation) intervient le premier jour où est atteint 5% du total annuel en fréquence cumulée, est ici utilisé du fait de sa fréquence d'apparition dans la littérature.

3.4.1 Détermination de relations entre seuils d'action et caractéristiques physiques : le cas d'Amiens, Rouen, Paris, Nancy et Strasbourg

Précédemment, nous avons pu voir que les dates de démarrage de cumul, propres à chaque méthode, pouvaient fluctuer grandement. Afin de déterminer les relations existant entre les seuils d'actions et les caractéristiques géographiques et climatiques, il est nécessaire de fixer les dates de début de cumul, à l'instar de ce qui a été fait pour l'olivier (Orlandi *et al.*, 2005). Malheureusement, la configuration du programme utilisé nous a obligé à tester deux dates, le 1^{er} et le 2 de chaque mois où nous avons retenu, à chaque fois, celle qui débouche sur la combinaison minimisant l'erreur moyenne absolue et, en cas d'égalité, celle où l'erreur moyenne au carré est la plus faible. Mais, au bout du compte, les seuils d'action associés ne diffèrent pas, ou très légèrement, entre le 1^{er} et le 2 du mois considéré. La prise en compte de deux dates a donc une influence quasi nulle sur l'angle d'analyse de notre étude. Nous étudions ici le cas des cinq villes présentées dans le chapitre 3.2, pour le frêne, le bouleau et les Poacées, et cela sur les années communes aux cinq sites, taxon par taxon.

Il convient de préciser que la quasi-totalité des séries utilisées présentent une distribution gaussienne, selon les résultats des tests de normalité de Shapiro-Wilk et de Kolmogorov-Smirnov. Ainsi, par l'intermédiaire de coefficients de corrélation linéaire, nous allons mettre en relation, dans un premier temps, les seuils d'action, méthode par méthode à l'exception de

celle de Sarvas, avec la latitude, la longitude, l'altitude et la distance à la mer des villes en question. Ces paramètres sont présentés dans le tableau 12.

	AMIENS	ROUEN	PARIS	NANCY	STRASBOURG
latitude Nord (en ° décimaux)	49,9	49,4	48,9	48,7	48,6
longitude Est (en ° décimaux)	2,3	1,1	2,4	6,2	7,8
altitude (en m)	24	17	48	207	140
distance à la mer (en km)	60	55	150	360	435

Tableau 12 : Caractéristiques géographiques des cinq villes de l'étude.

A côté de ces critères géographiques, nous allons également comparer ces seuils aux températures minimale, maximale et moyenne du mois, complet, qui précède l'apparition la plus précoce du risque. Cela correspond donc aux mois de février (frêne et bouleau) et avril (Poacées). Ces températures sont calculées sur les années communes disponibles et sont donc propres à chaque taxon. Elles sont présentées dans le tableau 13.

		AMIENS	ROUEN	PARIS	NANCY	STRASBOURG
frêne	Tn_02	2,0	2,0	3,6	0,3	0,4
	Tx_02	8,0	7,9	8,9	7,1	7,1
	Tm_02	5,0	4,9	6,2	3,7	3,7
bouleau	Tn_02	1,7	1,7	3,4	-0,1	0,1
	Tx_02	8,0	7,9	9,0	7,1	7,2
	Tm_02	4,9	4,8	6,2	3,5	3,6
Poacées	Tn_04	4,6	4,7	7,3	4,1	5,2
	Tx_04	14,1	13,5	15,4	14,6	15,6
	Tm_04	9,4	9,0	11,3	9,4	10,4

Tableau 13 : Températures minimales (Tn), maximales (Tx) et moyennes (Tm) des mois de février (02) et avril (04) des années communes à chaque taxon. Frêne = 11 ans ; bouleau = 15 ans ; Poacées = 18 ans.

Dans le chapitre 3.2 nous avons démontré que les modèles initiés en décembre présentaient une erreur moyenne absolue équivalente, globalement, à celle des modèles initiés en janvier et février pour le frêne et le bouleau. De ce fait nous avons examiné les seuils d'action associés à des démarrages ayant lieu les 1^{er} ou 2 décembre, 1^{er} ou 2 janvier, 1^{er} ou 2 février et également 1^{er} ou 2 mars pour le bouleau. Quant aux Poacées, seuls les modèles initiés le 1^{er} ou 2 avril ont été testés. Peu de corrélations s'avèrent significatives au seuil de 95% pour le frêne et les

Poacées. Par contre elles se révèlent être en nombre assez important pour le bouleau, justifiant que l'on ne s'attarde que sur ce seul taxon.

L'étude ne portant que sur cinq variables, le seuil de significativité de 95% est très élevé, ne mettant ainsi en avant que des corrélations très fortes. De ce fait nous présentons, paramètre par paramètre, les valeurs des corrélations significatives propres aux méthodes symbolisées par les abréviations décrites dans le chapitre 3.2 et, entre parenthèses, le mois correspondant.

- **Latitude :**
 - **Lej_Tx** (décembre) = - 0,9042
- **Longitude :**
 - **Q₁₀** (décembre) = - 0,9852
 - **Q₁₀** (janvier) = - 0,9511
 - **Q₁₀** (février) = - 0,9323
- **Altitude :**
 - **Q₁₀** (décembre) = - 0,8901
 - **Q₁₀** (janvier) = - 0,9651
 - **Q₁₀** (février) = - 0,9252
 - **Q₁₀** (mars) = - 0,8890
- **Distance à la mer :**
 - **Q₁₀** (janvier) = - 0,9390
 - **Q₁₀** (février) = - 0,8870
- **Tn_02 :**
 - **GDD_1** (décembre) = 0,9445
 - **Q₁₀** (janvier) = 0,9104
 - **GDD_1** (février) = 0,8870
- **Tx_02 :**
 - **GDD_1** (décembre) = 0,9597
 - **GDD_1** (février) = 0,8804

- **Tm_02 :**
 - **GDD_1** (décembre) = 0,9519
 - **Q₁₀** (janvier) = 0,8859
 - **GDD_1** (février) = 0,8834

Malgré le fait que nous ne retrouvons pas l'ensemble des associations méthode-date de début de cumul pour chaque variable, nous constatons une constance dans le signe des corrélations paramètre par paramètre. Ainsi, les corrélations sont systématiquement négatives avec les quatre variables géographiques, et positives avec les valeurs de température. Les corrélations négatives significatives concernent quasi exclusivement la méthode du Q₁₀, semblant montrer que les coefficients de développement de la plante diminuent avec l'altitude, la latitude et la continentalité, cette dernière étant symbolisée à la fois par la distance à la mer et la longitude dans le cas présent.

Au niveau des températures les méthodes GDD_1 et Q₁₀ sont mises en évidence. De par le signe positif des corrélations, elles mettent en avant le critère de continentalité, à savoir que les seuils d'action diminuent avec la température mensuelle de février. Il semble donc, pour le bouleau, que les seuils d'action sont liés à la position géographique des villes dont dépend, en grande partie, les températures. Ainsi pour les stations les plus froides – Nancy et Strasbourg par conséquent – nous enregistrons des seuils de température et des coefficients de développement plus faibles que pour les autres villes présentant des conditions plus clémentes. Cela est très net pour le Q₁₀ où la moyenne des coefficients pour Nancy et Strasbourg (5,5 en décembre, 4,9 en janvier et 3,7 en février) est plus faible que la moyenne des trois autres villes (11,7 en décembre, 9,8 en janvier et 9,1 en février). Expliquer d'un point de vue phénologique ces résultats n'est pas chose aisée. Cependant, nous constatons que le signe des coefficients de corrélation associés à l'altitude et à la distance à la mer sont identiques à ceux présentés pour l'olivier par Orlandi *et al.* (2005). Cela semble indiquer une constance quant à l'effet de la continentalité sur les valeurs des seuils d'action pour des espèces ou genres pollinisant au début du printemps. Mais peut-on affirmer pour autant, d'un point de vue phénologique, que le bouleau nécessite des températures plus élevées pour assurer son besoin en «chaleur» dans les stations où les températures sont les plus clémentes ? Il semble difficile de l'affirmer, d'autant plus que cela ne se retrouve pas pour tous les cas de figure et notamment avec les modèles issus de la méthode FT, qui ne sont concernés par aucune corrélation significative. A l'heure actuelle, les résultats présentés n'ont qu'une valeur indicative et n'apparaissent guère exploitables en vue d'une estimation des seuils d'action

d'une ville lambda, pour une date de démarrage de cumul fixe, à partir de ses caractéristiques géographiques et thermiques. Ce type de détermination des seuils d'action sera probablement plus réalisable à partir d'un réseau au moins aussi dense que l'actuel, notamment dans la moitié Nord de la France, dont les sites les plus récents disposeront de séries au moins aussi longues que celles utilisées dans ce travail.

3.4.2 Précision des modèles de prévision : comparaison en fonction de deux critères de détermination de l'apparition du risque

A travers les villes présentées dans le chapitre 3.3, nous allons essayer de déterminer l'influence du critère de détermination de l'apparition d'un risque d'exposition allergique aux pollens (REAP), sur la qualité des modèles qui en sont issus. De ce fait, nous allons comparer les erreurs moyennes absolues présentées dans le chapitre précédent avec celles issues des modèles élaborés pour prévoir l'apparition du premier jour où est atteint 5% du total annuel cumulé. Outre le fait que ce critère est le plus communément employé dans la littérature, il permet la comparaison entre des seuils bruts, adaptés d'un point de vue sanitaire, et un seuil relatif dont l'utilisation est plus adéquate en agronomie qu'en aérobiologie.

La présence de données manquantes a nécessité des reconstitutions. A côté de celles déjà effectuées dans le cadre de la prévision de l'apparition du risque, nous avons reconstitué, à l'aide de la seule interpolation linéaire, les autres séries lacunaires présentes plus en avant dans la saison. Malheureusement, certaines périodes manquantes n'ont pu être reconstituées à cause de leur longueur et/ou par le fait qu'elles intervenaient aux alentours du pic de pollinisation. Ainsi, par rapport aux séries présentées dans le chapitre 3.3, les années suivantes ont dû être retirées :

- **Frêne :**
 - **Rouen** : 1999
 - **Paris** : 1996 et 2001
 - **Nancy** : 2000, 2002 et 2004
 - **Bordeaux** : 1997

- **Bouleau :**
 - **Rouen** : 1999
 - **Paris** : 1996 et 2001
 - **Nancy** : 1999, 2000 et 2004

- **Poacées :**
 - **Rouen** : 1993, 2004 et 2005
 - **Nancy** : 1997
 - **Strasbourg** : 1993, 1997 et 2001
 - **Clermont-Ferrand** : 2000
 - **Saint-Etienne** : 1997, 1998 et 2000
 - **Bordeaux** : 1988, 1993 et 1994
 - **Nantes** : 1993 et 2001

Nous allons présenter les résultats taxon par taxon en comparant, tout d'abord, les dates moyennes issues des deux critères, puis les erreurs moyennes des modèles à proprement parler. Les écarts entre les valeurs moyennes de ces dernières vont être testées grâce au test de Student, choisi en fonction de la distribution gaussienne de la quasi-totalité des séries utilisées, selon les tests de Shapiro-Wilk et de Kolmogorov-Smirnov.

3.4.2.1 Le cas du frêne

Les dates moyennes d'apparition du risque sont présentées dans le tableau 14.

	REAP	5% cumulé
ROUEN	25 mars	18 mars
AMIENS	22 mars	20 mars
PARIS	17 mars	9 mars
NANCY	27 mars	24 mars
LYON	15 mars	12 mars
BORDEAUX	25 février	9 février

Tableau 14 : Dates moyennes d'apparition du risque allergique au pollen de frêne, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.

Il convient de préciser que les moyennes sont calculées, pour chaque ville, sur les années où l'on a pu déterminer des dates d'apparition à partir des deux critères et que les périodes diffèrent d'une ville à l'autre. Les dates moyennes sont arrondies et correspondent à l'apparition du risque lors d'une année non bissextile.

Nous constatons ici que les dates moyennes déterminées avec le critère de 5% sont toujours plus précoces que celles issues du REAP, l'écart allant de deux jours (Amiens) à 16 (Bordeaux). Cela montre que la date d'apparition du risque, intervenant le premier jour (J) atteignant 20 grains/m³ avec un total cumulé d'au moins 60 grains de J à J+2, correspond à une fréquence cumulée en moyenne supérieure à 5%. Nous évoquons à présent les erreurs moyennes absolues des modèles propres à chaque critère. Nous nous bornons simplement à présenter la moyenne des erreurs des meilleurs modèles issus des cinq méthodes ou, à défaut, des trois méthodes de cumul. Il convient de préciser que les modèles présentés pour chaque critère ont été déterminés à partir d'un pas de temps allant du 1^{er} octobre à la veille de l'apparition du risque la plus précoce, pour chaque ville. Cependant, nous avons fixé le 28 février, comme date-limite d'initiation des modèles, pour les sites dont l'apparition la plus précoce du risque a eu lieu après cette date. Le même processus est appliqué au bouleau. Les erreurs moyennes absolues sont présentées dans le tableau 15.

	REAP	5% cumulé
ROUEN	5,13	5,79
AMIENS	3,79	3,62
PARIS	3,93	7,43
NANCY	2,50	3,80
LYON	5,05	4,56
BORDEAUX	5,43	9,48

Tableau 15 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour le frêne.

Il convient de rappeler que la moyenne calculée pour Bordeaux ne correspond qu'aux modèles de trois méthodes de cumul, chiffre que nous avons considéré trop faible pour appliquer un test statistique. Pour les cinq autres villes nous avons comparé les valeurs grâce au test t de Student. Ainsi, à l'exception d'Amiens et Rouen, nous enregistrons des différences significatives entre les moyennes pour un risque d'erreur de 5%. La précision des modèles liés au seuil relatif est, de fait, significativement moins bonne à Nancy et Paris que celle des modèles liés au REAP. Le même cas de figure se retrouve à Bordeaux alors qu'à Lyon c'est l'opposé, mais avec une différence de 0,49 jour. L'erreur moyenne absolue associée au seuil de 5% à Paris, particulièrement élevée, semble trouver son explication dans la précocité des dates liées aux années 1995 (13 février) et 2004 (6 février). Avec, respectivement, 38 et 39 jours d'avance sur les dates déterminées à partir du REAP, elles contribuent très probablement à dégrader la qualité des modèles ; cet état de fait semblant également expliquer le cas de Bordeaux.

D'une manière générale, dans le cas du frêne, la prise en compte du critère relatif, censé mieux retranscrire la phénologie de la plante, n'est pas un gage d'obtention de meilleurs modèles en comparaison de ceux basés sur le seuil brut.

3.4.2.2 Le cas du bouleau

A l'instar du frêne, nous commençons par présenter les dates moyennes d'apparition du risque liées au bouleau (tableau 16) calculées, rappelons-le, sur des périodes propres à chaque ville.

	REAP	5% cumulé
ROUEN	31 mars	3 avril
AMIENS	30 mars	1 ^{er} avril
PARIS	27 mars	28 mars
NANCY	1 ^{er} avril	2 avril
STRASBOURG	28 mars	30 mars
LYON	29 mars	28 mars
CLERMONT-FERRAND	30 mars	31 mars
BORDEAUX	23 mars	18 mars

Tableau 16 : Dates moyennes d'apparition du risque allergique au pollen de bouleau, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.

Nous constatons ici que les dates moyennes, pour une ville donnée, sont très proches. Les écarts oscillent d'un jour (Nancy, Paris, Lyon et Clermont-Ferrand) à cinq (Bordeaux). Le démarrage relativement brutal de la pollinisation du bouleau explique ces écarts réduits. Nous évoquons à présent les erreurs moyennes absolues des modèles propres à chaque critère. Ces dernières sont présentées dans le tableau 17.

	REAP	5% cumulé
ROUEN	2,41	2,04
AMIENS	2,39	2,20
PARIS	2,68	2,31
NANCY	2,21	2,25
STRASBOURG	2,21	2,02
LYON	2,79	2,24
CLERMONT-FERRAND	2,98	2,58
BORDEAUX	5,81	6,25

Tableau 17 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour le bouleau.

Il faut noter que les moyennes calculées pour Bordeaux et Clermont-Ferrand ne correspondent qu'aux modèles de trois méthodes de cumul. Pour les six autres villes nous avons comparé les valeurs grâce au test t de Student. Ce dernier permet de mettre en évidence qu'aucune différence significative n'est enregistrée entre les erreurs issues des modèles liés au seuil de 5% et ceux liés au REAP. Nous aboutissons à la même constatation pour Bordeaux et Clermont-Ferrand. Cette proximité des valeurs semble logique car les dates d'apparition moyennes du risque définies par les deux critères sont très proches. Cela confirme en tout cas que l'apparition rapide de concentrations polliniques assez importantes permet d'enregistrer des différences de dates faibles entre les deux critères de détermination. Dans l'absolu, l'utilisation d'un seuil brut est aussi pertinente que celle d'un seuil relatif pour la prévision du bouleau.

3.4.2.3 Le cas des Poacées

Nous commençons par la présentation des dates moyennes d'apparition du risque liées aux Poacées (tableau 18).

	REAP	5% cumulé
ROUEN	19 mai	26 mai
AMIENS	14 mai	20 mai
PARIS	18 mai	22 mai
NANCY	10 mai	18 mai
STRASBOURG	12 mai	18 mai
LYON	2 mai	8 mai
CLERMONT-FERRAND	10 mai	23 mai
SAINT-ETIENNE	13 mai	22 mai
BORDEAUX	26 avril	9 avril
TOULOUSE	22 avril	26 avril
NANTES	12 mai	15 mai
MARSEILLE	7 mai	2 mai

Tableau 18 : Dates moyennes d'apparition du risque allergique au pollen de Poacées, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.

Contrairement au bouleau, nous constatons une assez nette différence entre les dates moyennes déterminées à partir des deux critères. A l'exception de Marseille et Bordeaux, les dates correspondant au REAP sont en moyenne plus précoces que celles issues de la fréquence cumulée. La combinaison d'une pollinisation abondante et progressive débouche sur des valeurs brutes, associées au seuil de 5%, élevées (300 grains pour une année à 6 000 grains par exemple). En comparaison, les dates déterminées à partir de quantités aussi faibles que 10 grains ne peuvent, dans ce contexte, qu'être plus précoces. Quant à Marseille et Bordeaux, nous trouvons la configuration opposée qui s'explique, dans le premier cas, par des concentrations nettement plus faibles que dans les autres villes. La présence de longues séries de petites concentrations dès le mois de mars, que l'on ne retrouve pas ailleurs, semble expliquer les valeurs bordelaises précoces. Précisons que les modèles présentés ont été déterminés à partir d'une période allant du 1^{er} janvier à la veille de l'apparition du risque la plus précoce, pour chaque ville. Cependant pour les sites où les dates de survenue du risque ont toujours eu lieu en mai, nous avons fixé le 30 avril comme date-limite d'initiation des modèles. Nous présentons maintenant les erreurs moyennes absolues liées à chaque critère. Elles sont regroupées dans le tableau 19.

	REAP	5% cumulé
ROUEN	4,03	2,37
AMIENS	2,06	1,89
PARIS	3,97	1,62
NANCY	2,38	1,68
STRASBOURG	2,31	2,75
LYON	2,89	3,02
CLERMONT-FERRAND	3,43	2,48
SAINT-ETIENNE	3,50	2,06
BORDEAUX	7,24	6,13
TOULOUSE	4,45	2,49
NANTES	7,26	6,64
MARSEILLE	7,58	3,99

Tableau 19 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour les Poacées.

Il faut noter que les moyennes calculées pour Bordeaux, Clermont-Ferrand, Saint-Etienne et Nantes ne correspondent qu'aux modèles de trois méthodes de cumul. Pour les huit autres villes nous avons, de nouveau, comparé les valeurs grâce au test t de Student. A l'exception d'Amiens et Lyon, les différences enregistrées sont significativement différentes. Ainsi, en dehors de Strasbourg, les modèles associés au critère de 5% sont de meilleure qualité que les autres. Cette caractéristique semble s'expliquer à travers les dates elles-mêmes. Généralement plus tardives, ces dates sont à priori plus en osmose avec la phénologie des principales espèces de Poacées que celles déterminées à partir du REAP, dont la valeur-seuil associée est faible. La raison de la meilleure qualité des modèles est, en tout cas, à rechercher dans cette direction. Cela se confirme par ailleurs pour les quatre villes sur lesquelles n'avaient été testées que trois méthodes de cumul.

A travers l'exemple de ces trois taxons, il semble bien que les dates issues des différents critères de détermination influencent l'élaboration des modèles et, surtout, leur précision. Cela est particulièrement net en analysant conjointement les résultats liés au bouleau et aux Poacées : plus l'écart entre les dates est élevé, plus les différences entre les erreurs moyennes absolues sont importantes.

Par cette étude nous avons montré, pour le frêne et le bouleau, qu'en termes de qualité les modèles élaborés à partir du REAP n'avaient rien à envier à ceux construits à partir du critère le plus couramment rencontré dans la littérature. Il en est tout autrement avec les Poacées, où l'on se rend compte que le seuil brut retenu ne favorise pas particulièrement la qualité des

modèles. D'un point de vue strictement statistique, l'emploi de la fréquence cumulée serait plus judicieux à condition que le but de notre travail ne soit pas à visée préventive. En définitive, cette étude ne remet nullement en cause le choix des seuils effectué dans le sous-chapitre 2.1.5.

Dans le dernier chapitre de cette troisième partie, nous allons traiter d'un sujet qui n'a pas encore été abordé dans ce travail : l'étude des relations entre conditions climatiques et concentrations polliniques.

3.5 ETUDE SUR LES CONCENTRATIONS POLLINIQUES : LES POACEES A NANTES ET LA ROCHE-SUR-YON

Les articles traitant de la prévision des concentrations polliniques ou, seulement, des relations existant entre les conditions météorologiques et ces concentrations, sont nombreux et variés. Cet aspect étant secondaire dans notre travail de thèse, nous ne nous y sommes pas attardé.

Les Poacées ont été choisies du fait de la relation, mainte fois démontrée, existant entre les concentrations polliniques et différentes variables météorologiques tels la température de l'air, les précipitations, l'humidité relative, l'ensoleillement ou la vitesse du vent (Moseholm *et al.*, 1987 ; Norris-Hill et Emberlin, 1991 ; Bricchi *et al.*, 1992 ; Galán *et al.*, 1995 ; Norris-Hill, 1997 ; Laaidi M, 1997 ; Sánchez-Mesa *et al.*, 2003). Nous avons voulu, ainsi, étudier l'influence de ces variables sur les concentrations en pollen de Poacées, à travers les séries de Nantes et de La Roche-sur-Yon. Le choix de ces deux villes a été conditionné par la part très importante occupée par ce taxon, par rapport à l'index pollinique moyen de la période 1997-2005 (cf. 1.2.2.2).

Le travail se base sur les différents niveaux de Risque d'Exposition Allergique aux Pollens (REAP) liés aux Poacées.

3.5.1 Matériel et méthodes

Les données polliniques de l'étude sont celles de Nantes de 1989 à 2005 (moins l'année 1992) et celles de la Roche-sur-Yon de 1993 à 2005, sans reconstitution des données manquantes. Les données météorologiques quotidiennes exploitées sont la température (minimale, maximale et moyenne), l'humidité relative (minimale, maximale et moyenne), le cumul et le nombre de jours de précipitations. Selon les tests de Shapiro-Wilk et Kolmogorov-Smirnov, la distribution des différentes séries étudiées par la suite ne suit pas la loi normale ; le test non-paramétrique de Mann-Whitney est donc utilisé. Ces niveaux de risque sont également associés à des classes de précipitations, des types de temps créés à partir d'une classification en nuées dynamiques, ainsi que des combinaisons de précipitations à J et J-1. Les relations entre ces variables qualitatives et les niveaux de REAP sont mesurées à l'aide d'un test du χ^2 et d'une méthode exploratoire qui en découle : l'analyse des correspondances multiple (ACM).

Toute l'étude va donc se baser sur le REAP que le RNSA a développé pour chaque pollen (Ickovic *et al.*, 1988 ; Thibaudon, 2003). Nous rappelons brièvement que cet indice comprend six niveaux allant du risque 0 (nul) à 5 (très élevé). Nous avons décidé ici de regrouper les risques 0 à 2 (faible) qui correspondent à des concentrations polliniques très basses. Les différents niveaux, et les concentrations correspondantes, sont rappelés dans le tableau 20.

niveau		concentration (gr/m ³ /j)
0-2	nul à faible	< 10
3	moyen	[10 ; 34]
4	élevé	[35 ; 99]
5	très élevé	> 99

Tableau 20 : Niveaux de REAP associés aux Poacées et concentrations correspondantes.

L'étude des relations entre les différents niveaux de risque et les conditions météorologiques va porter sur la période comprise entre le premier jour atteignant un risque 4 et le dernier où ce niveau est atteint. Cela permet de se centrer, en terme de risque d'exposition, sur la pleine saison pollinique, celle qui affecte plus particulièrement les allergiques à ce pollen.

3.5.2 Relations entre REAP et conditions météorologiques

3.5.2.1 Description des données de Poacées

La durée moyenne de la pleine saison pollinique est significativement plus élevée ($p = 0,0409$ avec un test U de Mann-Whitney) à La Roche-sur-Yon (77,5 jours) qu'à Nantes (64,1 jours). Si on ne prend en compte que la période 1993-2005, la durée moyenne de la saison à Nantes est de 66,5 jours (différence non significative pour $p = 0,05$). La plus faible longueur des séries à Nantes s'explique, pour une bonne partie, par une proportion plus forte qu'à La Roche-sur-Yon de données manquantes émaillant la pleine saison des Poacées.

Le tableau 21 présente la distribution de chaque niveau de risque pour chaque station :

REAP	NANTES	LA ROCHE-SUR-YON
0-2	10,05%	8,23%
3	27,02%	28,57%
4	33,76%	31,75%
5	29,17%	31,45%

Tableau 21 : Distribution des niveaux de REAP pour les périodes d'étude respectives de chaque ville.

Les proportions entre les deux villes sont quasiment identiques. Le test z de comparaison des fréquences, qui se base sur la probabilité P d'apparition du caractère étudié en fonction de la taille de la série et de la probabilité complémentaire (1-P) (Chadule, 1997), n'enregistre aucune différence significative, pour un risque d'erreur de 5% et un même niveau de risque, entre les deux villes. On remarque que plus de la moitié des jours de cette période sont caractérisés par un risque élevé et très élevé : 62,93% à Nantes et 63,20% à La Roche-sur-Yon.

3.5.2.2 Conditions moyennes de température, d'humidité relative et de précipitations associées à chaque niveau

Par l'intermédiaire d'un test U de Mann-Whitney qui – rappelons-le – est l'alternative non-paramétrique à un test t de Student pour comparer deux échantillons indépendants, nous allons étudier les valeurs des différentes variables météorologiques associées à chaque niveau de REAP. Les graphiques 71 à 76 présentent la température, l'humidité relative et le nombre de jours de précipitations en fonction du REAP pour chacune des deux villes :

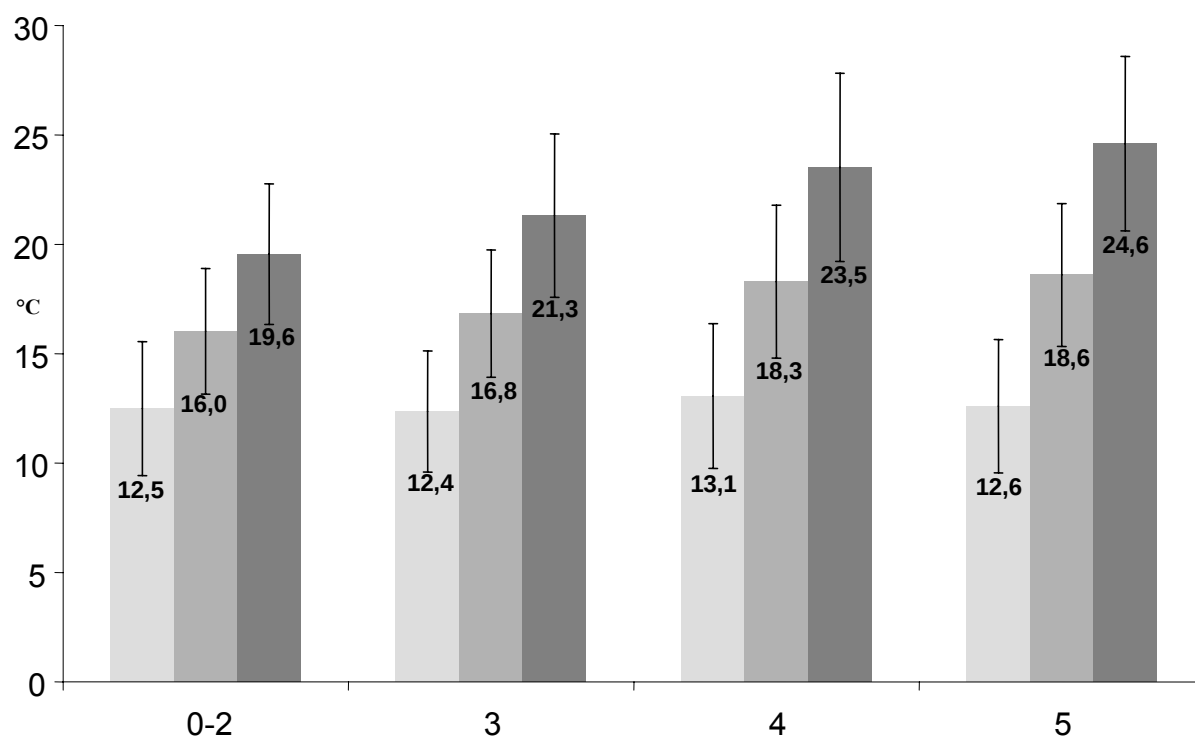


Figure 71 : Valeurs moyennes de température minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé) selon le niveau de REAP à Nantes.

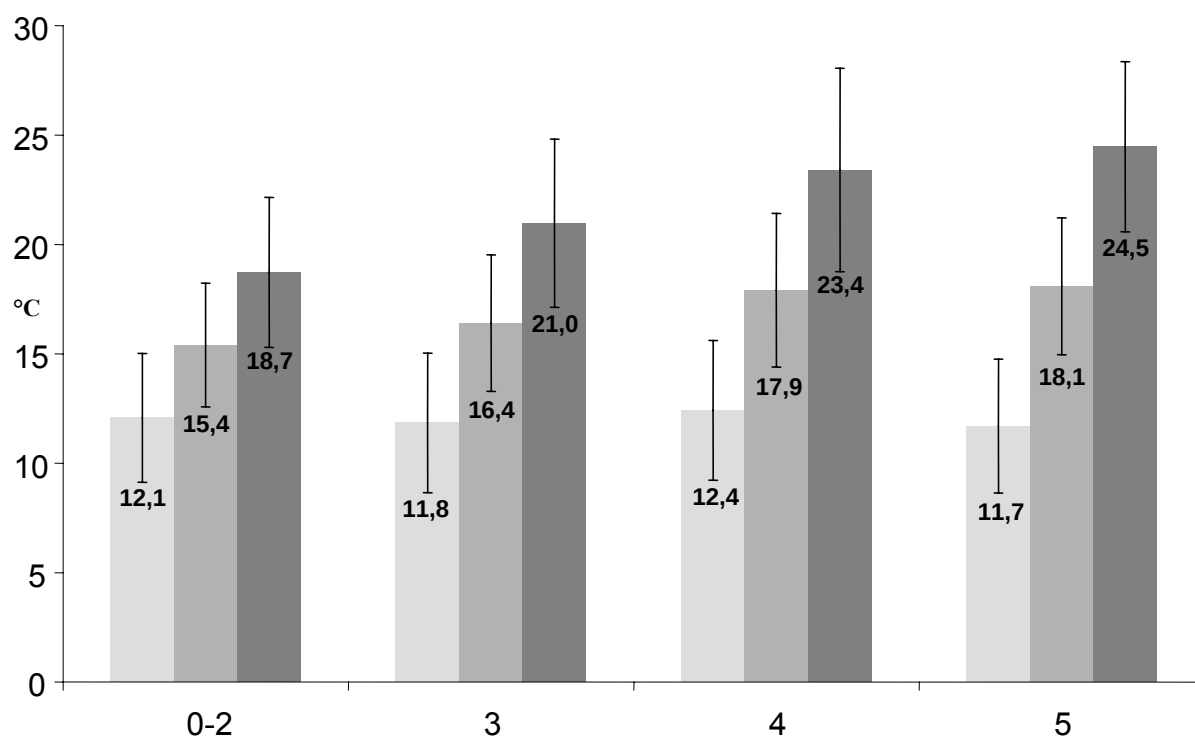


Figure 72 : Valeurs moyennes de température minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé) selon le niveau de REAP à La Roche-sur-Yon.

D'après les résultats issus du test de Mann-Whitney, on enregistre des différences significatives ($p < 0,05$) entre les valeurs moyennes de température maximale associées à chaque niveau de risque. Plus celles-ci sont élevées, plus le niveau de risque l'est aussi. Cette gradation régulière ne se retrouve pas pour les températures minimales statistiquement équivalentes d'un niveau de risque à l'autre.

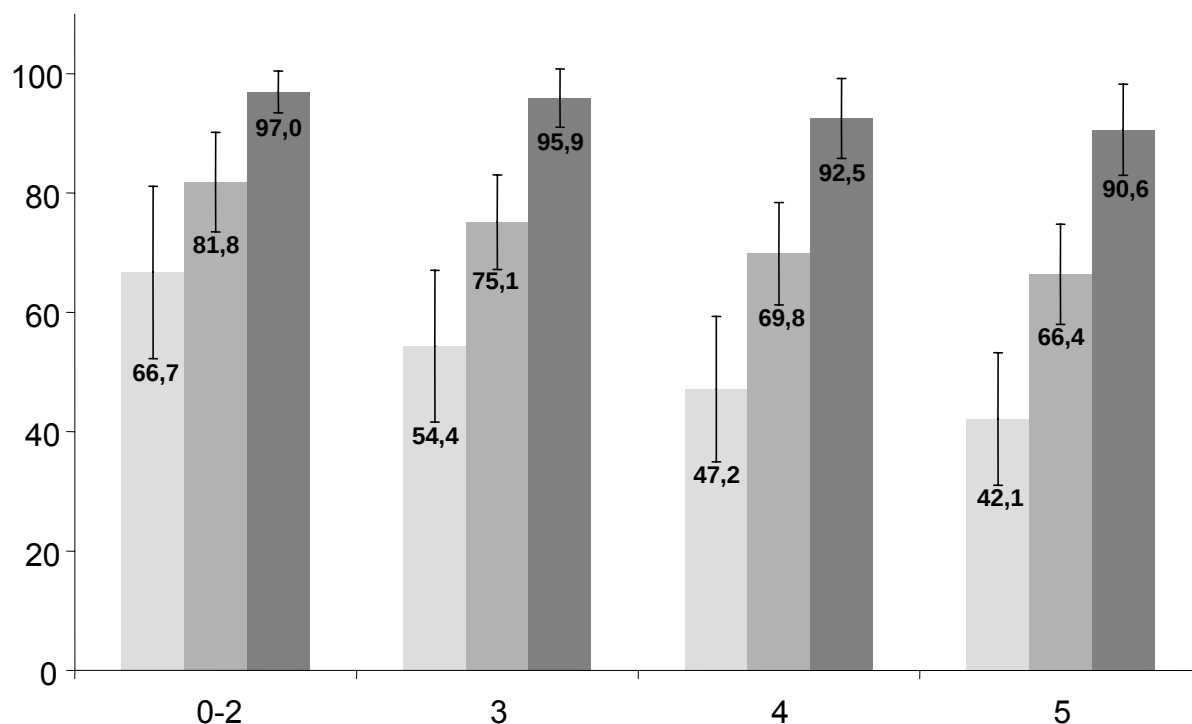


Figure 73 : Valeurs moyennes d'humidité relative minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé), en %, selon le niveau de REAP à Nantes.

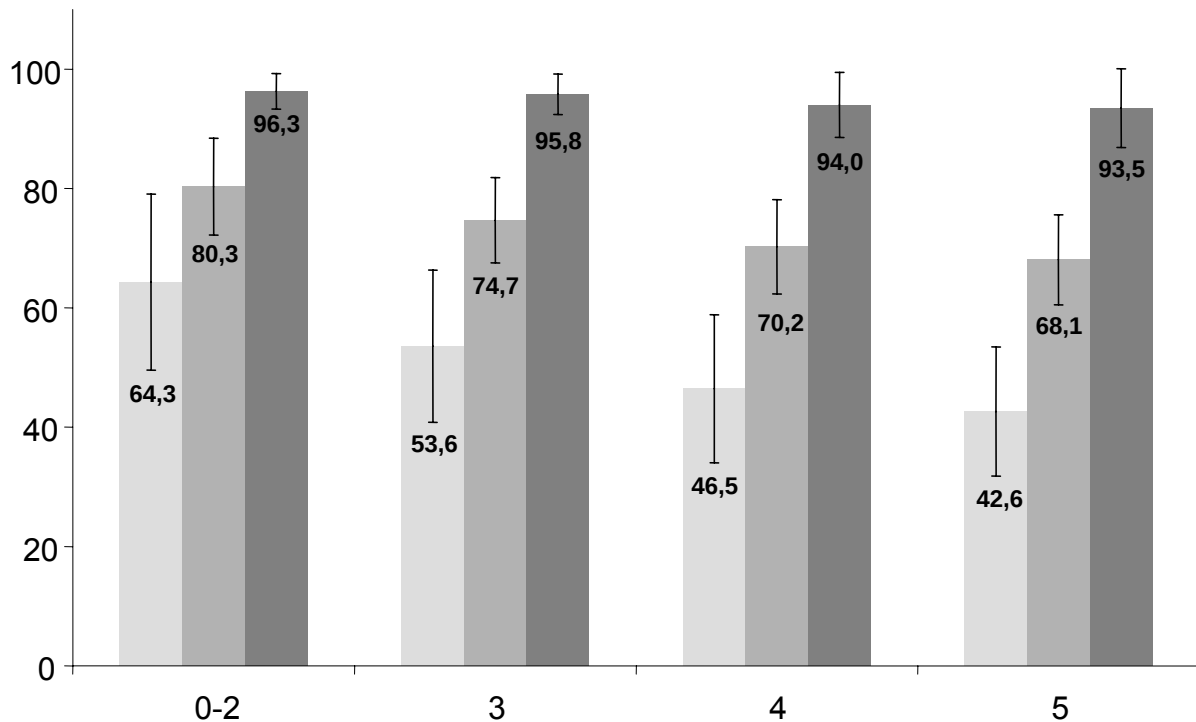


Figure 74 : Valeurs moyennes d'humidité relative minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé), en %, selon le niveau de REAP à La Roche-sur-Yon.

L'humidité relative minimale étant fortement liée à la température maximale, on enregistre logiquement des différences significatives ($p < 0,05$) entre les valeurs moyennes en passant d'un niveau à l'autre. Ainsi plus l'humidité relative est faible, plus le niveau de risque est élevé. Il en est de même pour l'humidité relative moyenne. Quant à l'humidité relative maximale, les différences sont minimales d'un niveau à l'autre. Néanmoins, à Nantes, ces différences sont significatives selon le test de Mann-Whitney, et suivent la tendance des valeurs minimales et moyennes.

Nous allons, maintenant, nous intéresser à la proportion de jours pluvieux associés à chaque niveau de REAP (figures 75 et 76). Le test z de comparaison des fréquences est donc approprié dans ce cas précis.

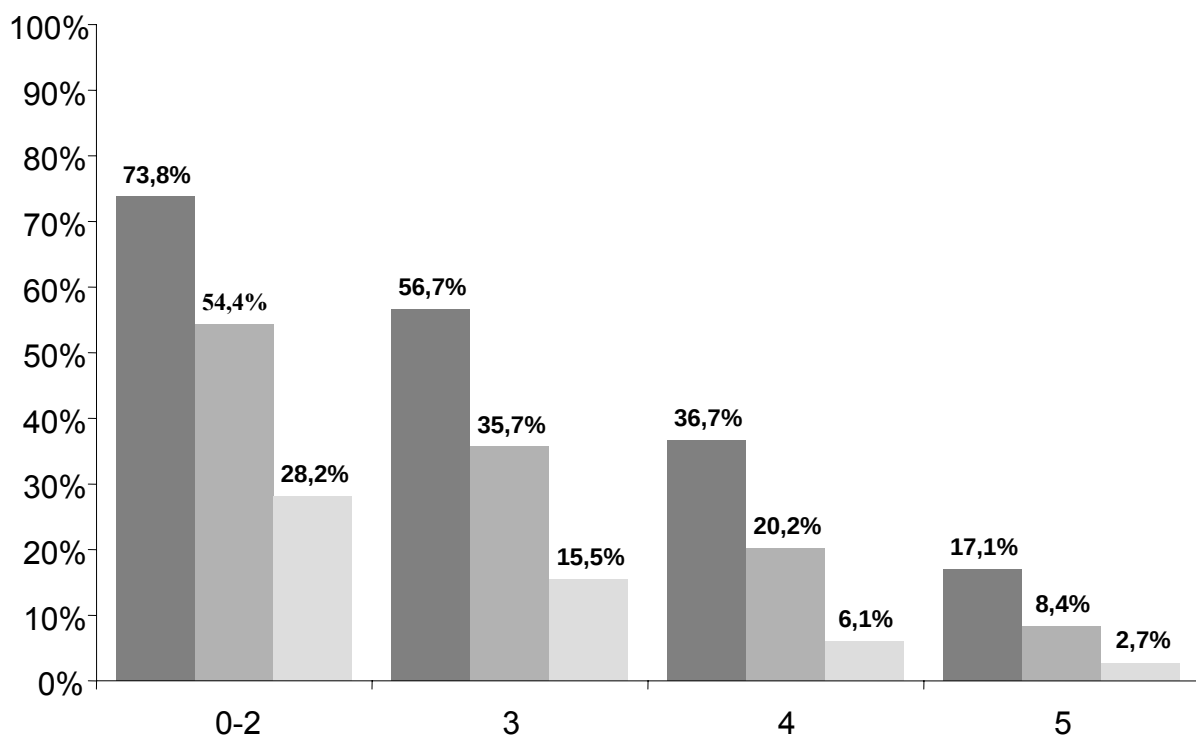


Figure 75 : Fréquence de précipitations > 0 mm (gris foncé), ≥ 1 mm (gris moyen) et ≥ 5 mm (gris clair) selon le niveau de REAP à Nantes.

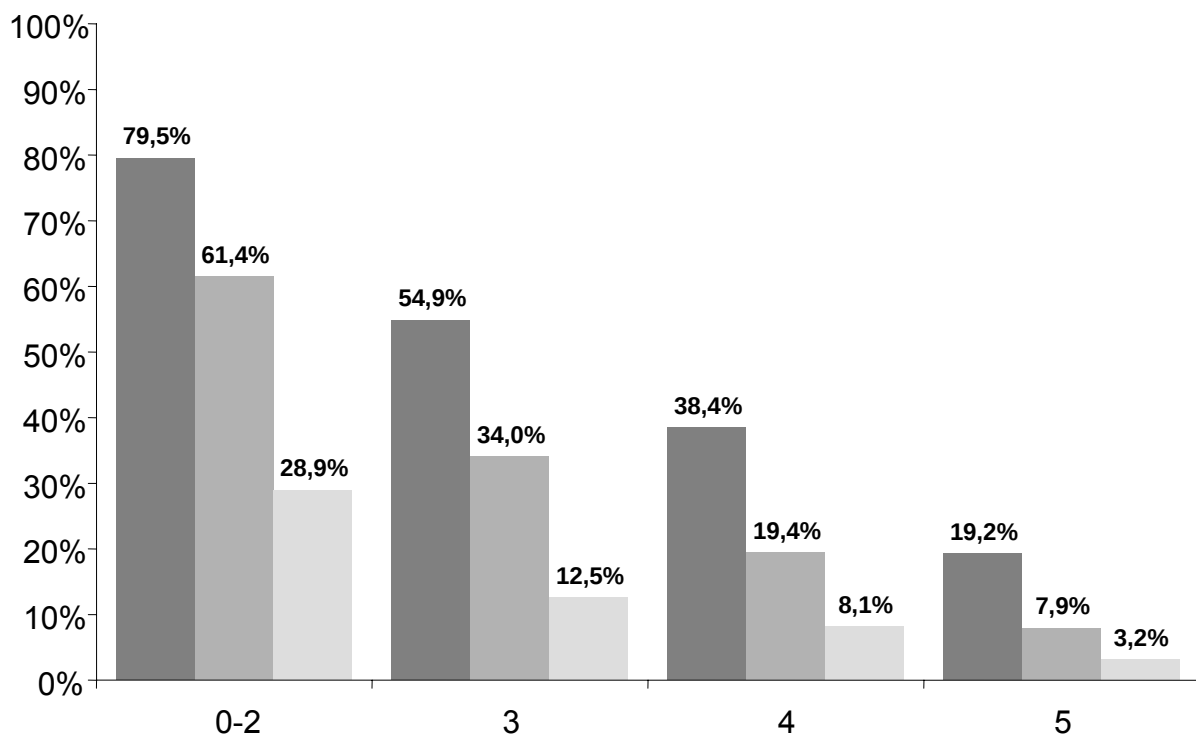


Figure 76 : Fréquence de précipitations > 0 mm (gris foncé), ≥ 1 mm (gris moyen) et ≥ 5 mm (gris clair) selon le niveau de REAP à La Roche-sur-Yon.

Les différences de fréquence sont significatives d'un niveau de risque à l'autre, pour un seuil de confiance de 95%, que ce soit pour l'ensemble des précipitations ou lorsque l'on considère seulement celles qui sont au moins égales à 1 et 5 mm. Les jours pluvieux, quel que soit le seuil retenu, se font de plus en plus rares quand le niveau de risque augmente. Les jours enregistrant des niveaux inférieurs à 3 sont pluvieux, pour 73,8% d'entre eux à Nantes et 79,5% à La Roche-sur-Yon. On enregistre des précipitations pour, seulement, 17,1% des jours à risque 5 de Nantes et 19,2% de ceux de la Roche-sur-Yon.

On peut donc constater, de manière globale, que les différents niveaux de risque compris dans la pleine saison pollinique sont liés à des conditions météorologiques différentes. Les jours présentant des risques 4 et 5 se caractérisent principalement par une température maximale significativement plus élevée que pour les autres niveaux de risque et, par conséquent, une humidité relative minimale significativement plus faible. De même la fréquence des précipitations y est significativement moindre. Les facteurs connus entravant ou favorisant la production et la dispersion des pollens sont bien déterminés dans le cas présent.

3.5.2.3 Relations entre le REAP et les cumuls de précipitations

Nous nous intéressons plus particulièrement, maintenant, aux jours où ont été enregistrés des précipitations. Celles-ci sont classées en trois groupes selon les seuils retenus par Météo France pour distinguer les jours avec des précipitations significatives (Marceaux et Taboulot, 1994), seuils déjà évoqués précédemment :

- de 0,1 à 0,9 mm inclus
- de 1 à 4,9 mm inclus
- 5 mm et plus

Un test du χ^2 , employé dans le cadre de la comparaison entre deux variables qualitatives, est utilisé dans le cas présent pour étudier le lien entre les niveaux de risque et les trois classes de précipitations. Le tableau 22 présente les écarts entre les effectifs observés et les effectifs théoriques pour chaque station.

NANTES	0,1-0,9 mm	1-4,9 mm	> 4,9 mm
risque 0 à 2	-9,77	-0,55	10,32
risque 3	-3,50	-0,92	4,42
risque 4	7,25	2,96	-10,21
risque 5	6,02	-1,49	-4,53

LA ROCHE/YON	0,1-0,9 mm	1-4,9 mm	> 4,9 mm
risque 0 à 2	-12,82	4,35	8,47
risque 3	-6,61	7,78	-1,18
risque 4	9,15	-6,21	-2,94
risque 5	10,28	-5,93	-4,35

Tableau 22 : Ecart entre les effectifs observés et prévus dans le cadre du test du χ^2 établi à partir des niveaux de REAP et de classes de précipitations. En rouge = observés > prévus, en bleu = observés < prévus.

Les valeurs du χ^2 pour Nantes et La Roche-sur-Yon sont, respectivement, de 17,81 et 23,12. Ces valeurs sont significatives pour un risque d'erreur de 5% ; il existe donc un lien entre précipitations et niveaux de REAP. A la lecture du tableau 22, on peut noter une surreprésentation des risques élevés pour les jours faiblement pluvieux (moins de 1 mm). La proportion de jours à risque allergique nul à faible est plus élevée que celles des autres niveaux, pour les jours présentant des précipitations supérieures ou égales à 5 mm.

En se focalisant particulièrement sur les jours pluvieux, on note une plus forte proportion de jours à risque élevé quand les précipitations sont faibles.

3.5.2.4 Utilisation de types de temps et des conditions de précipitations de la veille

Dans cette section, nous n'avons pas voulu raisonner en fonction de conditions météorologiques prises une à une, mais englobées dans un ensemble. Cette classification en types de temps a déjà été testée, notamment pour ce taxon, en distinguant une période humide et une période sèche (Spieksma et Tonkelaar, 1986), ou en se basant sur la classification Bénichou (1995) pour une étude se rapportant à la Bourgogne (Laaidi, 2000). Dans les deux cas, les plus fortes concentrations polliniques sont enregistrées quand la masse d'air est sèche. En ce qui concerne la présente étude, nous allons créer des types de temps à partir des précipitations, de l'humidité relative et des températures (minimales et maximales), propres à

chaque station, enregistrées sur les périodes définies dans la section 3.5.2.1. Travaillant sur quatre niveaux de REAP (les risques 0 à 2 étant groupés), quatre types vont être ainsi créés à partir d'une classification en nuées dynamiques. Il est nécessaire de rappeler que cette méthode permet de générer, à l'aide d'un algorithme, un nombre voulu de classes qui soient les plus spécifiques possibles. Les caractéristiques des types de temps propres à chaque ville sont présentées dans le tableau 23.

LA ROCHE/YON	Précipitations (mm)	JP (%)	Tn (°C)	Tx(°C)	Hn (%)	Hx (%)	nombre d'observations
type 1	5,9	87,2	12,7	17,9	74,8	97,2	109
type 2	2,0	58,8	12,4	20,8	56,7	96,1	337
type 3	0,4	26,9	11,2	23,4	42,6	95,5	398
type 4	0,1	4,9	12,8	27,7	30,1	87,3	164

NANTES	Précipitations (mm)	JP (%)	Tn (°C)	Tx (°C)	Hn (%)	Hx (%)	nombre d'observations
type 1	3,7	78,3	12,8	19,0	73,3	98,1	161
type 2	1,8	55,0	12,3	21,0	55,5	96,4	340
type 3	0,5	24,0	12,3	24,1	42,1	93,6	359
type 4	0,1	7,3	14,3	27,6	30,7	81,9	165

Tableau 23 : Caractéristiques météorologiques moyennes des types de temps issus de la classification en nuées dynamiques. JP = jours de précipitations ; Tn = température minimale ; Tx = température maximale ; Hn = humidité relative minimale ; Hx = humidité relative maximale.

Les valeurs du test F de Fisher associées à l'analyse de variance réalisée sur les variables précipitations, température et humidité relative sont très hautement significatives ($p < 0,0001$). Les classes ainsi créées sont donc bien discriminées.

Dans un second temps, nous allons caractériser chaque jour J de nos deux périodes d'étude par les conditions de précipitations à J-1 et J selon les combinaisons suivantes :

- 0-0 = pas de précipitations à J et J-1
- 0-1 = précipitations à J seulement
- 1-0 = précipitations à J-1 seulement
- 1-1 = précipitations à J et J-1

Le tableau 24 présente la distribution de chaque combinaison pour chaque station.

	NANTES	LA ROCHE-SUR-YON
1-1	24,59%	25,10%
0-1	15,51%	15,38%
1-0	15,61%	15,87%
0-0	44,29%	43,65%

Tableau 24 : Distribution des combinaisons de précipitations pour les périodes d'étude respectives de chaque ville.

Les relations existant entre les combinaisons de précipitations et les types de temps d'une part, et les niveaux de REAP d'autre part vont être étudiées grâce à une ACM (Analyse des Correspondances Multiple). Cette méthode peut s'appréhender comme une extension de l'analyse des correspondances simple à plus de deux variables. Cette dernière est une technique descriptive/exploratoire destinée à analyser des tables à double entrée, ou plus, contenant certaines mesures de correspondance entre les lignes et les colonnes. Les résultats donnent une information dont la nature est proche de celle fournie par les techniques d'analyse factorielle, et nous permettent d'explorer la structure des variables catégorielles de la table. En d'autres termes, l'analyse des correspondances est une manière de décomposer la statistique globale du Chi^2 en identifiant un nombre restreint de dimensions dans lesquelles les écarts aux valeurs théoriques peuvent être représentés.

Les calculs de l'analyse des correspondances multiple portent sur une matrice appelée table de Burt. A partir des tables de contingence, cette procédure consiste à croiser les catégories les unes avec les autres. Par exemple, dans le cas de deux variables catégorielles, ce tableau est constitué de quatre parties :

- le tableau croisé de la variable «REAP» avec elle-même
- celui de la variable «REAP» avec la variable «type de temps»
- celui de la variable «type de temps» avec la variable «REAP»
- enfin celui de la variable «type de temps» avec elle-même

Les résultats des deux premières dimensions des ACM sur les données de Nantes et La Roche-sur-Yon sont présentés dans les graphiques 77 et 78.

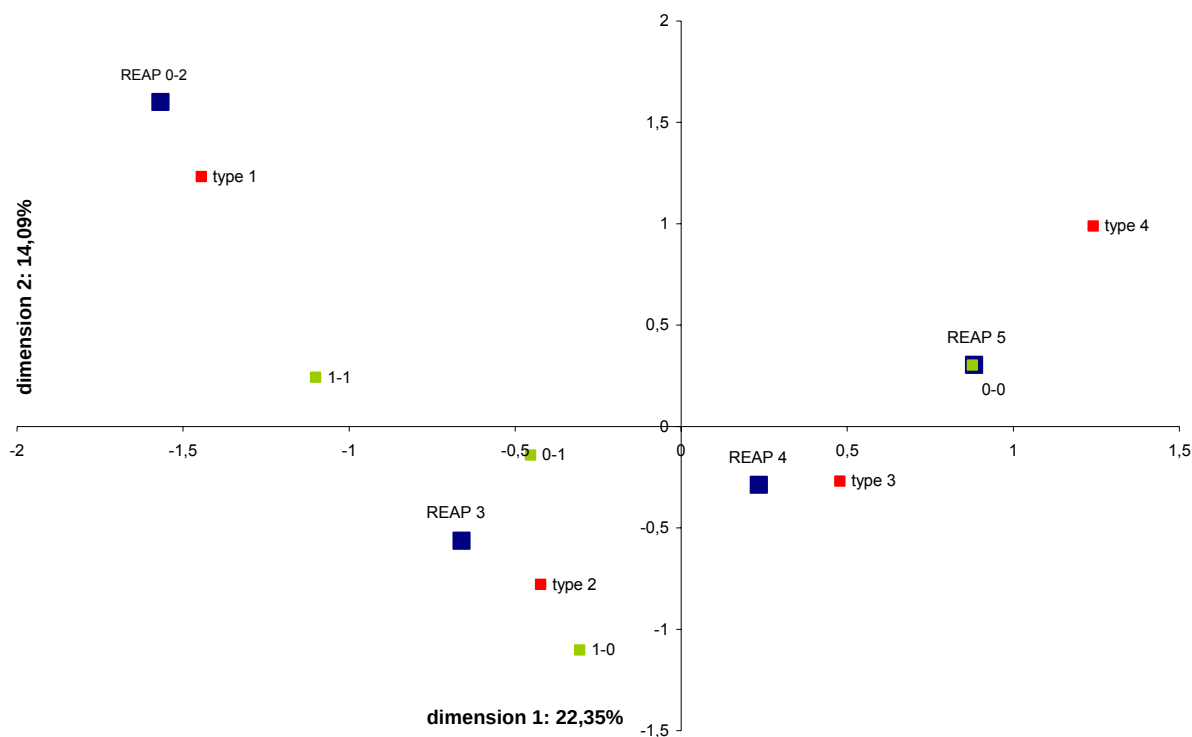


Figure 77 : Représentation des deux premières dimensions de l'ACM établie à partir du REAP, des types de temps et des combinaisons de précipitations à Nantes.

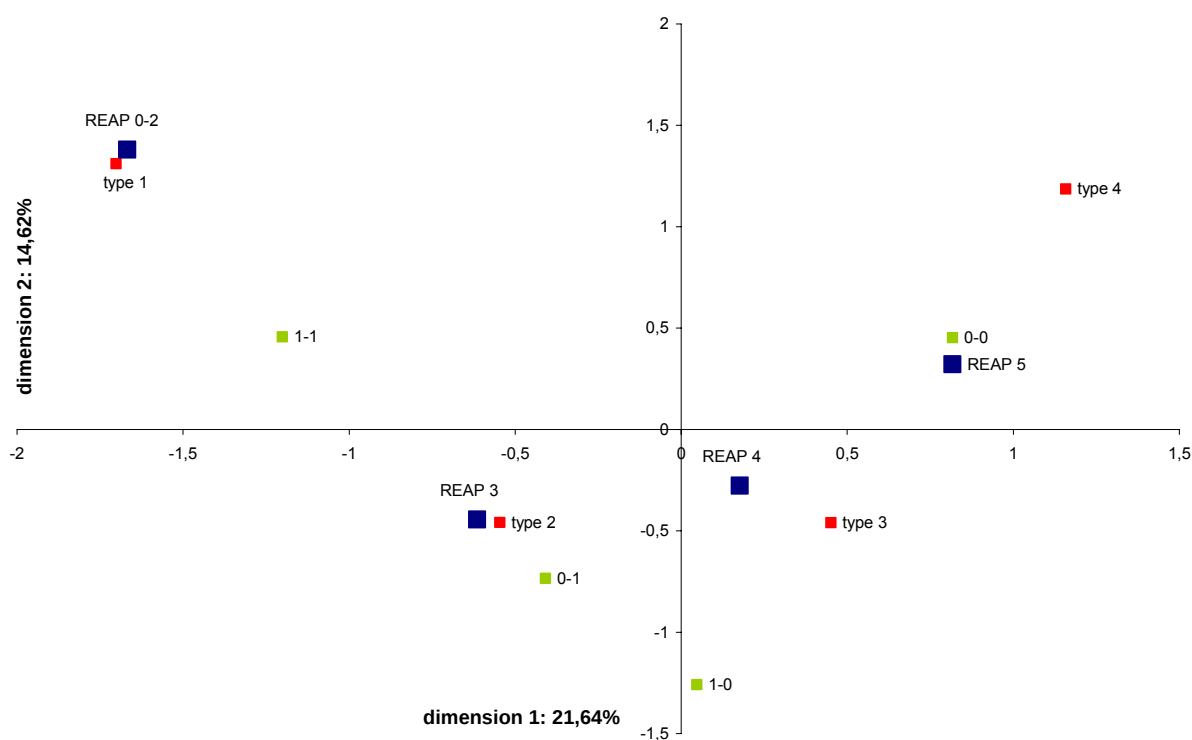


Figure 78 : Représentation des deux premières dimensions de l'ACM établie à partir du REAP, des types de temps et des combinaisons de précipitations à La Roche-sur-Yon.

Les deux premières dimensions n'expliquent que 36,44% (Nantes) et 36,26% (La Roche-sur-Yon) de l'inertie totale, cette dernière correspondant à la valeur totale du χ^2 de la table de Burt. La faiblesse des pourcentages semblerait s'expliquer par la méthode elle-même. En effet, la contribution de chaque variable avec elle-même représente un poids important dans la valeur totale du χ^2 de la table de Burt. Nous avons représenté les deux premières dimensions, mais nous allons surtout analyser la première qui est la plus aisément interprétable.

Cette première dimension oppose les jours à risques 4 et 5 (coordonnées positives) aux autres risques (coordonnées négatives). Les risques les plus élevés sont surreprésentés pour les types de temps 3 et 4, le risque 5 ayant des coordonnées proches de celles du type 4, qui correspond au type de temps le plus chaud et le plus sec. A l'opposé, les risques nul à faible ont des coordonnées proches du type de temps 1 qui peut être caractérisé comme frais, humide et souvent pluvieux. Chaque niveau de risque semble donc associé à un type de temps donné. Plus le temps est chaud, sec et peu pluvieux, plus les jours à risque allergique élevé sont surreprésentés et plus ceux à risque nul à moyen sont sous-représentés.

Le risque 5 est particulièrement surreprésenté quand il n'y a pas de précipitations, ni le jour du risque en question, ni la veille. Les risques 0 à 2 sont surreprésentés quand des précipitations sont enregistrées la veille et le jour même du risque. On peut également voir que, dans le cas d'une absence de précipitations le jour J, la probabilité d'avoir des risques élevés est plus faible s'il a plu à J-1. De même s'il pleut le jour du risque en question, on enregistre une plus faible proportion de jours avec des risques inférieurs ou égaux à 2 s'il n'a pas plu la veille, que dans le cas contraire. Cela montre que les conditions de précipitations de la veille ont une influence concomitante de celles du jour où sont enregistrées les concentrations polliniques analysées.

3.5.3 Le cas des variations interjournalières de REAP

Dans ce sous-chapitre nous allons étudier les variations des niveaux de risque en lien avec les évolutions interjournalières des différentes variables météorologiques. La terminologie suivante est utilisée:

- REAP – 2 = baisse du risque de deux crans ($5 \rightarrow 3$; $4 \rightarrow 0-2$)
- REAP – 1 = baisse du risque d'un cran ($5 \rightarrow 4$; $4 \rightarrow 3$; $3 \rightarrow 0-2$)

- REAP_0 = risque identique
- REAP + 1 et REAP + 2 = hausse du risque d'un et deux crans

Les variations interjournalières de trois crans (des risques 0-2 au risque 5 ou inversement) n'ont pas été retenues car elles concernent des effectifs trop faibles pour chaque ville: 0,69% du fichier de données à Nantes et 0,80% à La Roche-sur-Yon. Une fois retirées ces évolutions marginales, la proportion des différentes variations interjournalières est telle que présentée dans le tableau 25.

	NANTES	LA ROCHE-SUR-YON
REAP – 2	2,58%	4,13%
REAP – 1	22,62%	17,02%
REAP_0	49,50%	56,29%
REAP + 1	21,63%	19,44%
REAP + 2	3,67%	3,12%

Tableau 25 : Distribution des variations interjournalières de REAP pour les périodes d'étude respectives de chaque ville.

Dans 49,50% des cas à Nantes et 56,29% à La Roche-sur-Yon, le risque allergique est identique d'un jour à l'autre. La fréquence d'apparition décroît avec l'amplitude du risque. La figure 79 représente l'évolution des concentrations correspondant à chaque catégorie d'évolution de risque.

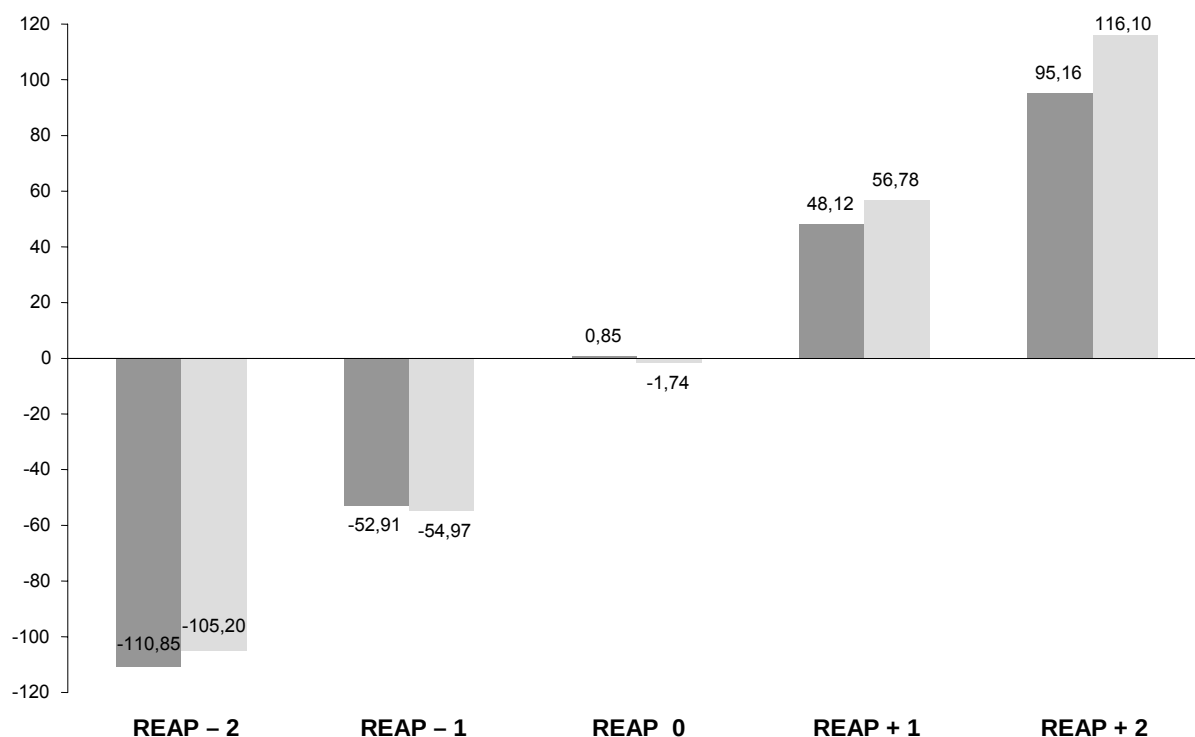


Figure 79 : Amplitude interjournalière moyenne des concentrations polliniques (gr/m³/j) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair).

L'amplitude moyenne des concentrations masque une grande variabilité des évolutions intragroupes. Les écarts-types oscillent ainsi entre 44,03 grains (à Nantes pour REAP + 1) et 95,6 grains (à La Roche-sur-Yon pour REAP + 2). Ainsi, il est possible de constater une augmentation interjournalière des concentrations de 200 grains tout en restant en risque 5 (de 120 à 320 grains par exemple) et c'est également vrai pour l'évolution contraire. A l'opposé, une évolution d'un grain peut faire passer d'un risque à un autre ; par exemple, une évolution interjournalière de 99 à 100 grains correspond à une hausse du risque d'un cran (risque 4 au risque 5). Le fait d'étudier les variations de niveaux de REAP revêt un but pratique pour les allergiques, mais la prudence s'impose dans l'analyse de ces évolutions, car elles masquent de fortes disparités quant à l'évolution des concentrations correspondantes.

Grâce au test de Mann-Whitney, il va être possible de tester si les amplitudes moyennes interjournalières des différentes variables météorologiques, liées aux cinq évolutions de REAP étudiées, sont significativement différentes. Nous allons présenter les résultats pour l'humidité relative minimale et les précipitations (figures 80 et 81). Ce sont les seules variables (avec l'humidité relative moyenne) dont l'amplitude interjournalière est significativement différente ($p < 0,05$), palier par palier de REAP - 2 à REAP + 2, pour les deux villes.

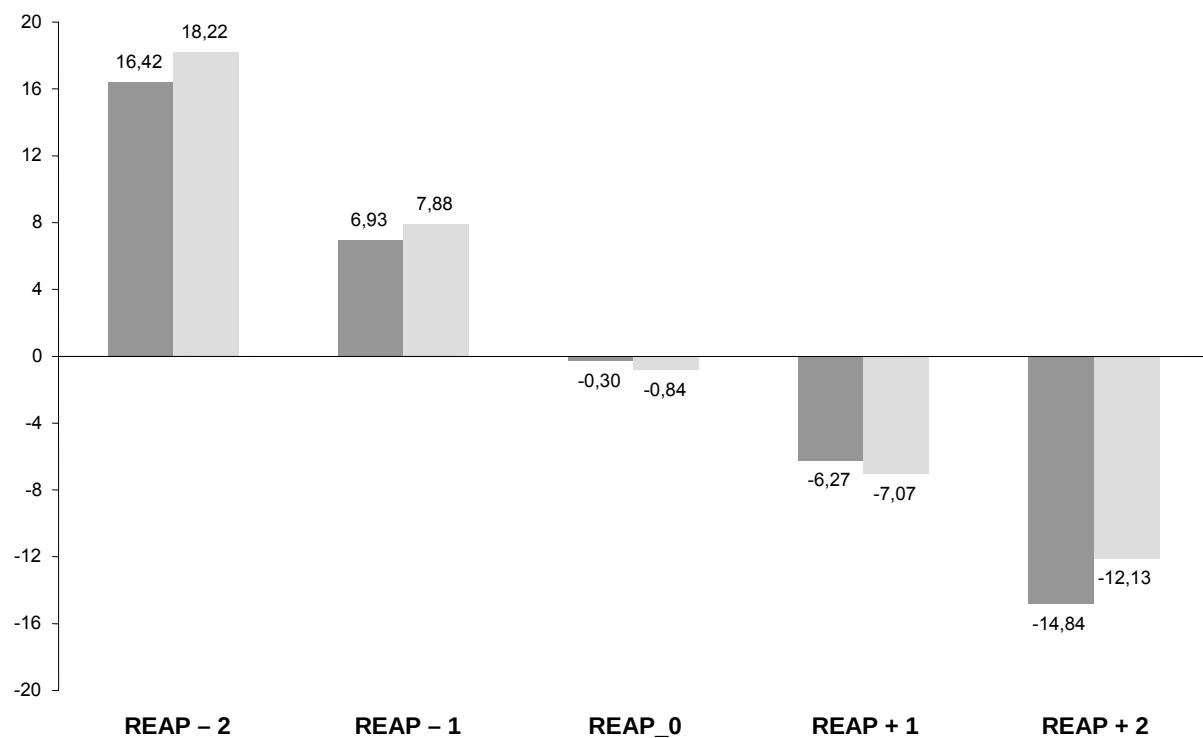


Figure 80 : Amplitude interjournalière moyenne de l'humidité relative minimale (en %) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair).

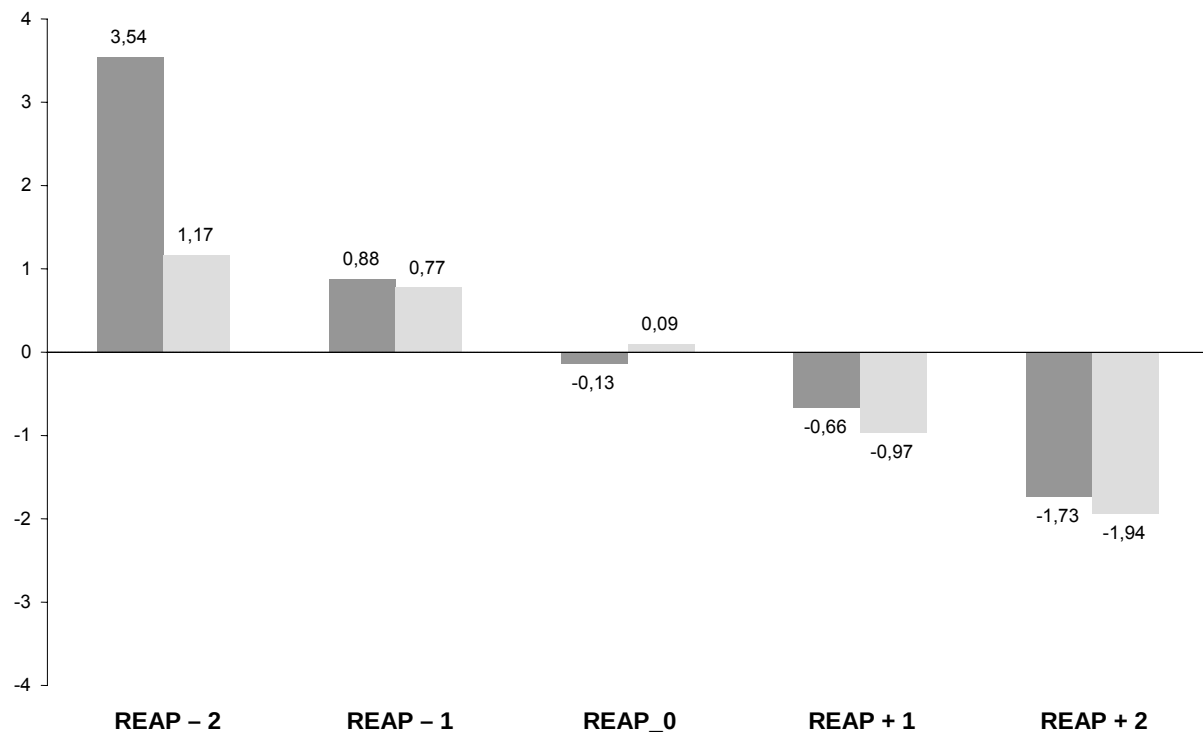


Figure 81 : Amplitude interjournalière moyenne des précipitations (en mm) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair).

L'importance de l'amplitude interjournalière de l'humidité relative minimale et des précipitations semble conditionner l'ampleur de la baisse ou de la hausse du REAP. Plus la diminution de l'humidité et des précipitations est importante, plus la hausse du niveau de risque le sera aussi, et inversement. Cependant, comme pour les concentrations, les écarts-types intragroupes sont importants : de 8,03% à 20,44% pour Nantes et de 11,2 à 17,8% pour La Roche-sur-Yon (humidité relative minimale). Pour les précipitations, ils s'étendent de 3 à 7,18 mm pour Nantes et de 3,88 à 11,83 mm pour La Roche-sur-Yon. Cela se traduit par des hausses de précipitations et d'humidité relative minimale accompagnant des hausses de risque. 23,17% et 20,09% des hausses de risque sont liées à une augmentation de l'humidité relative minimale pour Nantes et La Roche-sur-Yon respectivement ; les pourcentages sont de 18,92% (Nantes) et 18,78% (La Roche-sur-Yon) quand est enregistrée une hausse des précipitations. La prise en compte de toutes les évolutions de ces deux paramètres inclut, par conséquent, des évolutions très minimales qui ne sont pas susceptibles d'entraîner une diminution du risque. Les moyennes proches de 0 pour REAP_0 masquent des évolutions qui peuvent être, aussi bien, à la hausse qu'à la baisse pour les deux variables comme l'indiquent les écarts-types correspondants : 11,54 et 11,2% pour l'humidité relative minimale, 4,27 et 3,96 mm pour les précipitations à Nantes et La Roche-sur-Yon respectivement.

L'importance de l'amplitude interjournalière de la température maximale semble également conditionner l'ampleur de la baisse ou de la hausse du REAP. Elle est en effet très liée à l'humidité relative minimale. Par contre, on n'enregistre pas de différences significatives pour la température minimale et l'humidité relative maximale. L'amplitude interjournalière de ces variables ne semble pas jouer de rôle dans l'évolution du risque.

Malgré des évolutions moyennes significativement différentes, la très forte variabilité intragroupe dépendante du type de données rend impossible l'application pratique de ces résultats, qui n'ont, au bout du compte, qu'un but informatif. En effet, les hausses et les diminutions du niveau de risque sont liées à des évolutions contrastées des différents paramètres météorologiques car chaque classe d'évolution du REAP correspond à des évolutions de concentration très variables. Pour que les résultats soient exploitables, il serait nécessaire de créer des classes à partir des variations interjournalières de concentrations, mais les bornes des classes n'auraient plus aucun rapport avec les seuils du REAP des Poacées et ne seraient pas exploitables dans un but préventif.

3.5.4 Discussion et conclusion

Les risques élevés liés au pollen de Poacées dépendent de conditions météorologiques favorables à l'ouverture des anthères permettant la libération des pollens ainsi que leur dispersion. Ces conditions se caractérisent par un temps chaud, sec, avec des précipitations assez rares ; la température maximale, l'humidité relative minimale ainsi que le cumul et le nombre de jours de précipitations apparaissent comme étant les paramètres influençant le plus la survenue d'un risque donné, et ceux dont les évolutions peuvent entraîner l'évolution à la hausse ou la baisse du niveau de REAP. Cela se remarque aussi pour les risques nul à faible qui sont associés à des types de temps souvent pluvieux, humides avec des températures maximales relativement basses. Ces conditions retardent l'ouverture des anthères et tendent à précipiter vers le sol les pollens déjà présents dans l'air. Cette étude a permis, en tout cas, de confirmer le lien étroit existant entre ces diverses variables et les concentrations polliniques par l'intermédiaire du REAP développé pour ce taxon. Les seuils retenus pour son élaboration se sont avérés appropriés, car ils ont permis de particulariser chaque niveau de risque par des conditions de température, d'humidité relative et de précipitations significativement différentes et suivant une évolution progressive concomitante à l'évolution du REAP.

CONCLUSION DE LA PARTIE

A travers l'exemple de cinq villes de la moitié Nord de la France, nous avons pu démontrer que les six méthodes de cumul préalablement testées fournissaient des résultats statistiquement équivalents pour chacun des taxons étudiés, et de n'en retenir que cinq finalement. La méthode de Lejoly-Gabriel, la plus simple méthodologiquement, s'avérait ainsi aussi efficace que les autres. Les modèles issus de la régression linéaire multiple, dont le principal avantage était la possibilité d'inclusion de plusieurs types de variables, se sont révélés très sensibles aux années exceptionnelles, que ce soit au niveau des conditions climatiques ou de l'apparition du risque, sensibilité rencontrée également dans les modèles issus des méthodes de cumul. Nous avons démontré ainsi la difficulté d'élaborer des modèles globaux, même si plusieurs variables se retrouvaient dans plusieurs équations, et, par la même occasion, l'impossibilité de se priver des années exceptionnelles.

Alors que les prédictors des équations de régression se sont avérés très fluctuants avec le retrait ou l'ajout d'une année dans la série pollinique de base, les seuils d'action et les dates de démarrage de cumul se sont révélés, dans le même cas de figure, beaucoup plus stables. L'utilisation prospective des modèles pour l'année 2006, élaborés à partir des sites présentant au moins 15 ans de données polliniques, et ce pour le frêne, le bouleau et les Poacées, s'est avérée très intéressante. Dans le cadre des modèles issus des méthodes de cumul, nous avons pu constater, pour une bonne part d'entre eux, l'influence des dates de démarrage de cumul sur la qualité de la prévision, mais aussi, pour d'autres, l'influence du type de méthode employé et du seuil d'action associé. Ce constat débouche sur la nécessité de tester plusieurs modèles, pour l'ensemble des méthodes retenues, initiés à des dates différentes dans une optique préventive. Cela tout en veillant, pour chaque taxon, à ne jamais écarter les modèles établis à partir du mois précédant celui de la pollinisation – qui sont les plus fréquemment déterminés – et en préservant au moins deux semaines d'accumulation des unités de «chaleur».

En ce qui concerne l'étude sur les concentrations de Poacées, elle a essentiellement permis de confirmer l'influence de divers paramètres météorologiques sur les concentrations polliniques par une approche reprenant les seuils de risque du RNSA.

CONCLUSION GENERALE

Ce travail de thèse a essentiellement consisté en une présentation des différentes méthodes de prévision de la date d'apparition d'un risque d'exposition allergique aux pollens (REAP) et à l'application de plusieurs d'entre elles, tout ceci précédé d'une étude assez poussée sur les données polliniques et notamment la qualité de celles-ci.

La première partie s'est déclinée en une présentation globale de ces données qui a permis de constater une forte disparité inter-sites au niveau de la part occupée par les données manquantes. Si elles représentent globalement moins de 5% de l'effectif étudié, valeur tout à fait convenable, leur positionnement dans l'année a posé un certain nombre de problèmes notamment lorsqu'elles coïncident avec le démarrage de la saison d'une famille, d'un genre ou d'une espèce allergisante. Afin de répondre au mieux à l'objectif de notre travail, il a fallu effectuer plusieurs reconstitutions de valeurs manquantes qui ne devraient donc être considérées que comme des approximations. D'autres paramètres ont également influencé l'exploitation de ces données telles les dates de mise en route et d'arrêt des capteurs qui ne se sont pas toujours avérées adéquates, notamment dans le cadre d'une étude complète sur des espèces pollinisant dès le mois de février. Enfin, il a été établi que l'implantation en hauteur des capteurs et le mode de lecture des bandes sous microscope ont une influence certaine sur les valeurs des concentrations. Ainsi des divergences apparaissent au niveau des concentrations journalières selon le nombre de lignes de lecture. Une analyse complète des bandes étant techniquement et humainement impossible, l'homogénéisation du nombre de lignes lues est nécessaire, avec une préférence pour trois lignes dans le souci d'une meilleure représentativité. Ces concentrations polliniques, malgré les divers inconvénients rencontrés, constituent les seules données exploitables dans un registre aérobiologique, et retranscrivent finalement bien l'implantation des formations végétales de la zone d'action des capteurs, du moins pour les espèces et familles anémophiles. Ainsi, les quatre taxons de l'étude voient leur implantation relative suivre des logiques zonales ou régionales.

Nous nous sommes attachés à présenter, dans la deuxième partie, les différents critères de détermination de l'apparition d'un risque allergique et de la manière de le prévoir selon différentes méthodes. Dans le premier cas de figure, il s'est avéré que le choix d'un critère par rapport à un autre devait se faire en fonction de l'objectif recherché : un critère relatif dans le cadre d'une étude agronomique ou botanique, plus adapté par ailleurs à une comparaison

inter-sites, et un seuil brut dans le cadre d'une étude aérobiologique à visée sanitaire. Malgré les réserves qui peuvent être émises sur les valeurs brutes du fait de la méthode de comptage, il était nécessaire de s'appuyer sur ce type de seuil afin de répondre à l'objectif de notre travail en utilisant les niveaux de risque, propre à chaque taxon, élaborés par le RNSA. Quant aux méthodes de prévision de la date d'apparition du risque, nous avons retenu les plus couramment usitées dans la littérature, avec six méthodes de cumul de température et la régression linéaire multiple. Le calcul de la dormance, difficile à délimiter dans le temps et complexifiant les modèles, n'a pas été retenu, en partant notamment du principe qu'elle est systématiquement accomplie chaque hiver, à une date très probablement antérieure à celle du démarrage du cumul des unités de «chaleur» déterminée pour bon nombre de modèles.

Les résultats découlant de ces différentes méthodes, constituant la dernière partie de ce travail, ont mis en avant que les six méthodes de cumul étaient statistiquement équivalentes, quel que soit le taxon étudié. Ainsi la méthode de Lejoly-Gabriel, consistant à cumuler les températures maximales au-delà d'un seuil donné, s'est avérée aussi efficace que les autres méthodes plus complexes dont celle sinusoïdale des *Growing Degree Days* (GDD_2), préalablement retenue puis écartée de l'analyse. Cet état de fait montre donc que la complexité d'une méthode n'est pas un gage d'amélioration des performances prévisionnelles par rapport à une autre plus basique. L'application de différents modèles issus des méthodes de cumul testées, dans le cadre d'une prévision prospective pour l'année 2006, conduit à souligner, selon les modèles, le poids de la date de démarrage de ces cumuls dans la qualité des prévisions, ainsi que l'influence du type de méthode et du seuil d'action (température ou coefficient) associé. Quant aux équations issues de la régression linéaire multiple, par l'application de validations croisées, nous avons pu voir l'influence de l'ajout ou du retrait de certaines années sur la significativité des corrélations entre les dates d'apparition du risque et divers paramètres météorologiques. Malgré cela la température, critère le plus influent du développement, était assez souvent corrélée avec les dates d'apparition du risque, notamment durant les dernières périodes précédant la survenue de ce dernier. Ce résultat, et le fait qu'un nombre important de dates optimales de démarrage des modèles de cumul ont lieu relativement peu de temps avant l'apparition du risque, montrent l'importance des conditions météorologiques des dernières semaines sur le début de la saison pollinique d'une famille, d'une espèce ou d'un genre donné.

Au bout du compte nous n'avons pu que répondre partiellement au développement d'un modèle national. Le nombre de stations présentant des séries d'une longueur substantielle (15

ans) est trop faible et, surtout, celles-ci sont trop éloignées les unes des autres pour pouvoir effectuer des prévisions à l'échelle d'une zone regroupant plusieurs régions. A partir du réseau actuel, composé de 63 stations, dont les plus récentes auront fonctionné pendant au moins une quinzaine d'années, il sera probablement plus aisé d'élaborer des modèles adaptés à de grandes régions. Ainsi, il sera possible, peut-être, de faire ressortir des corrélations plus robustes et plus systématiques d'une ville à l'autre dans le cadre de la régression linéaire multiple. De même, en ce qui concerne les méthodes de cumul, nous pouvons supposer qu'il sera plus aisé de relier les valeurs des différents seuils d'action à des paramètres géographiques et météorologiques, comme cela a pu être constaté pour le bouleau à partir des cinq stations présentées dans le sous-chapitre 3.4.1. Il serait même possible, à priori, d'utiliser les seuils d'action et les dates de démarrages optimales déterminées pour une station donnée et de les appliquer à une autre ville, à condition qu'elles ne soient pas trop éloignées d'un point de vue climatique et d'y recalculer les cumuls d'unités de «chaleur» propres. Il convient de noter enfin que l'augmentation de la longueur des séries ne sera pas nécessairement accompagnée d'une amélioration de la précision de ces modèles, précision évaluée en fonction de l'erreur moyenne absolue (MAE pour *Mean Absolute Error*) entre autres.

Quant à l'emploi de seuils bruts de détermination, qui se basent sur les niveaux de risque développés par le RNSA, il pourrait constituer un frein à la mise en œuvre de modèles et d'équations d'échelle régionale. En effet, les concentrations polliniques étant très probablement influencées, dans une certaine proportion, par des paramètres techniques, l'extension des modèles peut s'avérer délicate. Une plus grande homogénéisation des hauteurs d'implantation, comme cela est constaté depuis environ dix ans, est un point essentiel dans le cadre d'une politique de prévision à plus grande échelle que la zone de couverture de chaque capteur.

La prévision des concentrations n'a, quant à elle, été que peu développée considérant que l'aspect le plus important de la prise d'un traitement anti-allergique résidait dans une prescription effectuée au moment adéquat. Par la suite, à la manière des bulletins hebdomadaires du RNSA, une simple utilisation des prévisions météorologiques et des observations phénologiques ponctuelles est suffisante pour assurer le bon déroulement de la prise des traitements. A travers les exemples nantais et yonnais sur les Poacées, famille qui a l'avantage de présenter une longue saison de pollinisation, nous avons pu confirmer les relations existant entre de fortes concentrations et des conditions météorologiques favorables à la production et à la dispersion des pollens.

Finalement, ce travail de thèse a eu pour but de déterminer les fondements d'une modélisation future d'échelle nationale et, point le plus important, a permis de mettre en avant la possibilité d'élaborer des modèles de bonne qualité, pour les quatre taxons concernés, à partir de seuils bruts de détermination de l'apparition d'un risque allergique. Une collaboration avec la division agrométéorologie de Météo France à Toulouse, sous la direction d'Emmanuel Cloppet, est progressivement mise en œuvre afin d'aboutir à des modèles de prévision opérationnels, élaborés de manière plus systématique, qui constitueront le prolongement des résultats de ce travail. Les moyens techniques et humains dont dispose Météo France permettront de répondre pleinement à l'objectif de prévention et de diffusion à l'attention d'un large public.

BIBLIOGRAPHIE

ADAMS-GROOM B., EMBERLIN J., CORDEN J., MILLINGTON W., MULLINS J., 2002: Predicting the start of the birch pollen season at London, Derby and Cardiff, United Kingdom, using a multiple regression model, based on data from 1987 to 1997. *Aerobiologia*, **18**, 117-23.

ALCALA A.R., BARRANCO D., 1992 : Prediction of flowering time in olive for the Córdoba olive collection. *Horticultural Science*, **27**, 1205-7.

ALCAZAR P., COMTOIS P., 2000 : The influence of sampler height and orientation on airborne *Ambrosia* pollen counts in Montreal. *Grana*, **39**, 303-7.

ALLEN J.C., 1976 : A modified sine wave method for calculating degree days. *Environmental Entomology*, **5**, 388-96.

ANDERSEN T.B., 1991 : A model to predict the beginning of the pollen season. *Grana*, **30**, 269-75.

ANTEPARA I., FERNANDEZ J.C., GAMBOA P., JAUREGUI I., MIGUEL F., 1995 : Pollen allergy in the Bilbao area (European Atlantic seaboard climate) : pollination forecasting methods. *Clinical Experimental Allergy*, **25**, 133-40.

ARON R.H., 1975 : A method for estimating the number of hours below a selected temperature threshold. *Journal of Applied Meteorology*, **14**, 1415-8.

ARON R.H., 1983 : Availability of chilling temperatures in California. *Agricultural Meteorology*, **28**, 351-63.

ARON R.H., GAT Z., 1991 : Estimating chilling duration from daily temperature extremes and elevation in Israel. *Climate Research*, **1**, 125-32.

ASERO R., MARZBAN G., MARTINELLI A., ZACCARINI M., LAIMER DA CAMARA MACHADO M., 2006 : Search for low allergenic apple cultivars for birch pollen-allergic patients : is there a correlation between *in vitro* assays and patient response. *European Annals of Allergy and Clinical Immunology*, **38**, 94-8.

ASHCROFT G.L., RICHARDSON E.A., SEELEY S.D., 1977 : A statistical method of determining chill unit and growing degree hour requirements for deciduous fruit trees. *Horticultural Science*, **12**, 347-8.

BAGAROZZI D.A., TRAVIES J., 1998 : Ragweed pollen proteolytic enzymes : possible roles in allergies and asthma. *Phytochemistry*, **47**, 593-8.

BENICHO P., 1995 : Classification automatique de configurations météorologiques sur l'Europe occidentale. Météo France, monographie n°8, 46 p.

BESANCENOT J.P., 2007 : Notre santé à l'épreuve du changement climatique. Editions Delachaux et Niestlé, 222 p.

BESANCENOT J.P., SULMONT G., 2002 : Réflexions sur le réseau français de surveillance aéropollinique : analyse de la situation présente et perspectives d'optimisation. Climat et Santé et Université de Picardie, 23 p.

BESANCENOT J.P., SULMONT G., LAAIDI M., THIBAUDON M., 2003 : The French network for aerobiology surveillance : present situation and perspectives for optimisation. *European Annals of Allergy and Clinical Immunology*, **35**, 158-63.

BIDABE B., 1967 : Action de la température sur l'évolution des bourgeons de pommier et comparaison des méthodes de contrôle de l'époque de floraison. *Annales de Physiologie Végétale*, **9**, 65-86.

BOULLARD B., 1999 : Dictionnaire de botanique. Editions Ellipses, 398 p.

BRICCHI E., FORNACIARI M., GIANNONI C., GRECO F., FASCINI D., FRENGUELLI G., MINCIGRUCCI G., ROMANO B., 1992 : Fluctuations of grass pollen content in the atmosphere of East Perugia and meteorological correlations (year 1989). *Aerobiologia*, **8**, 401-6.

BRYANT R.H., EMBERLIN J.C., NORRIS-HILL J., 1989 : Vertical variation in pollen abundance in North-Central London. *Aerobiologia*, **5**, 123-37.

CANNELL M.G.R., SMITH R.I., 1983 : Thermal time, chill days and prediction of budburst in *Picea sitchensis*. *Journal of Applied Ecology*, **20**, 951-63.

CESARACCIO C., SPANO D., DUCE P., SNYDER R.L., 2001 : An improved model for determining degree-day values from daily temperature data. *International Journal of Biometeorology*, **45**, 161-9.

CESARACCIO C., SPANO D., SNYDER R.L., DUCE P., 2004 : Chilling and forcing model to predict bud-burst of crop and forest species. *Agricultural and Forest Meteorology*, **126**, 1-13.

CESARACCIO C., SPANO D., SNYDER R.L., DUCE P., JONES H., 2006 : Improvement of chilling and forcing model to predict bud-burst. Communication for the 17th Conference on Biometeorology and Aerobiology.

CHADULE, 1997 : Initiation aux pratiques statistiques en géographie. Editions Armand Colin, 4^{ème} édition, 203 p.

CHAPPARD C., BONNEVIAL J., COLSON M., MATHERN G., EMONOT A., 2004 : Forecast of pollination dates and relation to onset of allergic pathology. *Aerobiologia*, **20**, 35-42.

CHRISTOPHE B., GILLET A., MEJECASE M., 2004 : Allergy to ragweed in the Rhône-Alpes and consumption of anti-allergics. *European Annals of Allergy and Clinical Immunology*, **36**, 333-6.

CHUINE I., COUR P., ROUSSEAU D.D., 1998 : Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. *Plant, Cell and Environment*, **21**, 455-66.

CHUINE I., COUR P., ROUSSEAU D.D., 1999 : Selecting models to predict the timing of flowering of temperate trees : implications for tree phenology modelling. *Plant, Cell and Environment*, **22**, 1-13.

CLOT B., 2001 : Airborne birch pollen in Neuchâtel (Switzerland): onset, peak and daily patterns. *Aerobiologia*, **17**, 25-9.

CLOT B., 2003 : Trends in airborne pollen : An overview of 21 years of data in Neuchâtel (Switzerland). *Aerobiologia*, **19**, 227-34.

COMTOIS P., GAGNON L., 1988 : Concentrations polliniques et fréquence des symptômes de pollinose : une méthode pour déterminer les seuils cliniques. *Revue Française d'Allergologie*, **28**, 279-86.

CORDEN J., MILLINGTON W., BAILEY J., BROOKES M., CAULTON E., EMBERLIN J., MULLINS J., SIMPSON C., WOOD A., 2000 : UK regional variations in *Betula* pollen (1993-1997). *Aerobiologia*, **16**, 227-32.

CORDEN J., STACH A., MILLINGTON W., 2002 : A comparison of *Betula* pollen seasons at two European sites; Derby, United Kingdom and Poznan, Poland (1995-1999). *Aerobiologia*, **18**, 45-53.

CORSICO R., 1993 : L'asthme allergique en Europe. In Spieksma F.T.M., Nolard N., Frenguelli G., Van Moerbeke D. *Pollens de l'air en Europe*, 19-29.

COTOS-YAÑEZ T.R., RODRIGUEZ-RAJO F.J., JATO M.V., 2004 : Short-term prediction of *Betula* airborne pollen concentration in Vigo (NW Spain) using logistic additive models and partially linear models. *International Journal of Biometeorology*, **48**, 179-85.

DAHL A., STRANDHEDE S.O., 1996 : Predicting the intensity of the birch pollen season. *Aerobiologia*, **12**, 97-106.

D'AMATO G., SPIEKSMAS F.T.M., 1992 : European allergenic pollen types. *Aerobiologia*, **8**, 447-50.

D'AMATO G., SPIEKSMAS F.T.M., LICCARDI G., JAEGER S., RUSSO M., KONTOUFILI K., NIKKELS H., WÜTHRICH B., BONINI S., 1998 : Pollen-related allergy in Europe. *Allergy*, **53**, 567-78.

DAVIES R.R., SMITH L.P., 1973 : Forecasting the start and severity of the hay fever season. *Clinical Allergy*, **3**, 263-7.

DECHAMP C., 1995 : L'ambrosie, un nouveau fléau. Editions Verso, 94 p.

DESSAINT F., CHAUVEL B., BRETAGNOLLE F., 2005 : L'ambroisie : chronique de l'extension d'un polluant «biologique» en France. *Médecine/Sciences*, **21**, 207-9.

DE WIT C.T., GOUDRIAAN J., VAN LAAR H.H., 1978 : Simulation of assimilation, respiration and transpiration of crops. Center for Agricultural Publishing and Documentation, Wageningen.

DREBORG S., FOUCARD T., 1983 : Allergy to apple, carrot and potato in children with birch pollen allergy. *Allergy*, **38**, 167-72.

DRIESSEN M.N.B.M., VAN HERPEN R.M.A., SMITHUIS L.O.M.J., 1990 : Prediction of the start of the grass pollen season for the southern part of the Netherlands. *Grana*, **29**, 79-86.

DURAND R., 1967 : Action de la température et du rayonnement sur la croissance. *Annales de Physiologie Végétale*, **9**, 5-27.

DURAND R., 1969 : Signification et portées des sommes de températures. *Bulletin technique d'information du Ministère de l'Agriculture et du Développement Rural*, **238**, 185-90.

DUTILLEUL P., LEGENDRE P., 1992 : Lack of robustness in two tests of normality against autocorrelation in sample data. *The Journal of Statistical Computation and Simulation*, **42**, 79-91.

EMBERLIN J., NORRIS-HILL J., 1991 : Annual, daily and diurnal variation of Urticaceae pollen in North-central London. *Aerobiologia*, **7**, 49-57.

EMBERLIN J., SAVAGE M., WOODMAN R., 1993 : Annual variations in the concentrations of *Betula* pollen in the London area, 1961-1990. *Grana*, **32**, 359-63.

EMBERLIN J., MULLINS J., CORDEN J., MILLINGTON W., BROOKES M., SAVAGE M., JONES S., 1997 : The trend to earlier Birch pollen seasons in the UK : A biotic response to changes in weather conditions ? *Grana*, **36**, 29-33.

EMBERLIN J., JAEGER S., DOMINGUEZ-VILCHES E., GALAN C., HODAL L., MANDRIOLI P., LEHTIMÄKI A.R., SAVAGE M., SPIEKSMÄ F.T., BARTLETT C., 2000: Temporal and geographical variations in grass pollens in areas of Western Europe : an analysis of season dates at sites of the European pollen information system. *Aerobiologia*, **16**, 373-9.

EMBERLIN J., DETANDT M., GEHRIG R., JAEGER S., NOLARD N., RANTIO-LEHTIMÄKI A., 2002 : Responses in the start of *Betula* (birch) pollen seasons to recent changes in spring temperatures across Europe. *International Journal of Biometeorology*, **46**, 159-70.

ENSA-M, 2005 : Métrologie des pollens dans l'air : étude intercomparative en région Languedoc-Roussillon. Rapport à destination de la DRASS Languedoc-Roussillon, 78 p.

ERIKSSON N., FORMGREN H., SVENONIUS E., 1982 : Food hypersensitivity in patients with pollen allergy. *Allergy*, **37**, 437-43.

- ESTRELLA N., MENZEL A., KRÄMER U., BEHRENDT H., 2006 : Integration of flowering dates in phenology and pollen counts in aerobiology : analysis of their spatial and temporal coherence in Germany (1992-1999). *International Journal of Biometeorology*, **51**, 49-59.
- FELBER F., CLOT B., 2003 : Dispersion du pollen et des graines, distribution des plantes. *Les cahiers du jardin de botanique*, **3**, 22-33.
- FREI T., TORRICELLI R., PEETERS A.G., WÜTHRICH B., 1995 : The relationship between airborne pollen distribution and the frequency of specific pollen sensitization at two climatically different locations in Switzerland. *Aerobiologia*, **11**, 269-73.
- FRENGUELLI G., BRICCHI E., ROMANO B., MINCIGRUCCI G., FERRANTI F., ANTOGNOZZI E., 1992 : The role of air temperature in determining dormancy release and flowering of *Corylus avellana* L. *Aerobiologia*, **8**, 415-8.
- FRENGUELLI G., TEDESCHINI E., VERONESI F., BRICCHI E., 2002 : Airborne pine (*Pinus* spp.) pollen in the atmosphere of Perugia (Central Italy) : Behaviour of pollination in the two last decades. *Aerobiologia*, **18**, 223-8.
- FRENZ D.A., MELCHER S.E., MURRAY L.W., SAND R.E., 1997 : A comparison of total pollen counts obtained 5.6 km apart. *Aerobiologia*, **13**, 205-8.
- FUERTES-ROFRIGUEZ C.R., GONZALEZ-PARRADO Z., VEGA-MARAY A.M., VALENCIA-BARRERA R.M., FERNANDEZ-GONZALEZ D., 2007 : Effect of air temperature on forecasting the start of Cupressaceae pollen type in Ponferrada (León, Spain). *Annals of Agricultural and Environmental Medicine*, **14**, 237-42.
- GALAN C., EMBERLIN J., DOMINGUEZ E., BRYANT R.H., VILLAMANDOS F., 1995 : A comparative analysis of daily variations in the Gramineae pollen counts at Córdoba, Spain, and London, UK. *Grana*, **34**, 189-98.
- GALAN C., FUILLERAT M.J., COMTOIS P., DOMINGUEZ-VILCHES E., 1998 : Bioclimatic factors affecting daily Cupressaceae flowering in southwest Spain. *International Journal of Biometeorology*, **41**, 95-100.
- GALAN C., GARCIA-MOZO H., CARIÑANOS P., ALCAZAR P., DOMINGUEZ-VILCHES E., 2001a : The role of temperature in the onset of the *Olea europaea* L. pollen season in southwestern Spain. *International Journal of Biometeorology*, **45**, 8-12.
- GALAN C., CARIÑANOS P., GARCIA-MOZO H., ALCAZAR P., DOMINGUEZ-VILCHES E., 2001b : Model for forecasting *Olea europaea* L. airborne pollen in South-West Andalusia, Spain. *International Journal of Biometeorology*, **45**, 59-63.
- GALAN C., GARCIA-MOZO H., VAZQUEZ L., RUIZ L., DIAZ DE LA GUARDIA C., TRIGO M.M., 2005 : Heat requirement for the onset of the *Olea europaea* L. pollen season in several sites in Andalusia and the effect of the expected future climate change. *International Journal of Biometeorology*, **49**, 184-8.

GARCIA-MOZO H., GALAN C., CARINANOS P., ALCAZAR P., MENDEZ J., VENDRELL M., ALBA F., SAENZ C., FERNANDEZ D., CABEZUDO B., DOMINGUEZ-VILCHES E., 1999 : Variations in the *Quercus* sp. Pollen season at selected sites in Spain. *Polen*, **10**, 59-69.

GARCIA-MOZO H., GALAN C., GOMEZ-CASERO M.T., DOMINGUEZ-VILCHES E., 2000 : A comparative study of different temperature accumulation methods for predicting the start of the *Quercus* pollen season in Córdoba (South West Spain). *Grana*, **39**, 194-9.

GARCIA-MOZO H., GALAN C., AIRA M.J., BELMONTE J., DIAZ DE LA GUARDIA C., FERNANDEZ D., GUTIERREZ A.M., RODRIGUEZ F.J., TRIGO M.M., DOMINGUEZ-VILCHES E., 2002 : Modelling start of oak pollen season in different climatic zones in Spain. *Agricultural and Forest Meteorology*, **110**, 247-57.

GONZALEZ-PARRADO Z., FUERTES-RODRIGUEZ C.R., VEGA-MARAY A.M., VALENCIA-BARRERA R.M., RODRIGUEZ-RAJO F.J., FERNANDEZ-GONZALEZ D., 2006 : Chilling and heat requirements for the prediction of the beginning of the pollen season of *Alnus glutinosa* (L.) Gaertner in Ponferrada (León, Spain). *Aerobiologia*, **22**, 47-53.

GOTTARDINI E., CRISTOFOLINI F., 1997 : Spring airborne pollen data in two sites in Trentino (Northern Italy): a comparison with meteorological data. *Aerobiologia*, **13**, 199-204.

GUERIN B., 1993 : Pollen et allergies. Editions Allerbio, 279 p.

HÄNNINEN H., 1987 : Effects of temperature on dormancy release in woody plants : implications of prevailing models. *Silva Fennica*, **21**, 279-99.

HÄNNINEN H., 1990 : Modelling bud dormancy release in trees from cool and temperate regions. *Acta Forestalia Fennica*, **213**, 1-47.

HIRST J.M., 1952 : An automatic volumetric spore trap. *Annals of Applied Biology*, **39**, 257-65.

HIRST J.M., 1953 : Changes in atmospheric spore content : diurnal periodicity and the effects of weather. *Transactions of the British Mycological Society*, **36**, 375-93.

HOUZE M.L., 2006 : Prévision et spatialisation des concentrations en ozone troposphérique en Bourgogne. Thèse de doctorat, Université de Bourgogne, 293 p.

ICKOVIC M.R., SUTRA J.P., THIBAUDON M., 1988 : From April 1st to July 30th in the Paris area, pollinosis symptoms compared to atmospheric pollen. *Annales de Sciences Naturelles et de Botanique*, **9**, 89-94.

JAEGER S., NILSSON S., BERGGREN B., PESSI A.M., HELANDER M., RAMFJORD H., 1996 : Trends of some airborne tree pollen in the Nordic countries and Austria, 1980-1993. *Grana*, **35**, 171-8.

JATO M.V., AIRA M.J., IGLESIAS M.I., ALCAZAR P., CERVIGON P., FERNANDEZ D., RECIO M., RUIZ L., SBAI L., 1999 : Aeropalynology of birch (*Betula* SP.) in Spain. *Polen*, **10**, 37-47.

JATO M.V., FRENGUELLI G., RODRIGUEZ F.J., AIRA M.J., 2000a: Temperature requirements of *Alnus* pollen in Spain and Italy (1994-1998). *Grana*, **39**, 240-5.

JATO M.V., RODRIGUEZ-RAJO F.J., SEJO M.C., 2000b : *Pinus* pollen in the atmosphere of Vigo and its relationship to meteorological factors. *International Journal of Biometeorology*, **43**, 147-53.

JATO M.V., RODRIGUEZ-RAJO F.J., MENDEZ J., AIRA M.J., 2002 : Phenological behaviour of *Quercus* in Ourense (NW Spain) and its relationship with the atmospheric pollen season. *International Journal of Biometeorology*, **46**, 176-84.

JONES S., 1995 : Allergenic pollen concentrations in the United Kingdom. PhD thesis, University of North London.

KADOCSA E., JUHASZ M., 2002 : Study of airborne pollen composition and allergen spectrum of hay fever patients in South Hungary (1990-1999). *Aerobiologia*, **18**, 203-9.

KRAMER K., 1994 : Selecting a model to predict the onset of growth of *Fagus sylvatica*. *Journal of Applied Ecology*, **31**, 172-81.

LAAIDI K., 1997 : Les pollinoses en Haute-Bourgogne : saison 1996. *Climat et Santé*, **17**, 27-40.

LAAIDI K., 1999 : Eco-épidémiologie des pollinoses en Haute-Bourgogne. Thèse de doctorat. Université de Bourgogne, 450 p.

LAAIDI K., 2000 : Utilisation des types de temps dans la prévision des jours à fort risque allergénique pendant la pollinisation des Poacées. *Climat et Santé*, **21**, 61-78.

LAAIDI K., 2001 : Pollen et pollinoses : seuils d'action clinique de quelques taxons allergisants. *Climat et Santé*, **22**, 95-113.

LAAIDI K., LAAIDI M., HOUNDENOU C., 1998 : Epidémiologie des pollinoses et facteurs météorologiques dans le Centre-Est de la France : le cas de la Haute-Bourgogne (1996-1998). *Climat et Santé*, **19**, 5-24.

LAAIDI K., LAAIDI M., 1999 : Airborne pollen of *Ambrosia* in Burgundy (France) 1996-1997. *Aerobiologia*, **15**, 65-9.

LAAIDI K., LAAIDI M., 2001 : Etude sur les prévisions de la date de début de pollinisation de quatre espèces végétales allergisantes en Bourgogne: contribution à la prévention des risques sanitaires liés à la pollution biologique de l'air. Ministère de l'Emploi et de la Solidarité, DRASS de Bourgogne, 57 p.

LAAIDI M., 1997 : Influence des facteurs météorologiques sur la concentration du pollen dans l'air. *Climat et Santé*, **17**, 7-25.

LAAIDI M., 1999 : Contribution à l'étude aéropollinique de la Haute-Bourgogne : introduction à une géographie des risques allergéniques. Thèse de doctorat. Université de Bourgogne, 509 p.

LAAIDI M., 2001a : Regional variations in the pollen season of *Betula* in Burgundy : two models for predicting the start of the pollination. *Aerobiologia*, **17**, 247-54.

LAAIDI M., 2001b : Forecasting the start of the pollen season of *Poaceae* : evaluation of some methods based on meteorological factors. *International Journal of Biometeorology*, **45**, 1-7.

LAAIDI M., THIBAUDON M., BESANCENOT J.P., 2003a : Two statistical approaches to forecasting the start and duration of the pollen season of Ambrosia in the area of Lyon (France). *International Journal of Biometeorology*, **48**, 65-73.

LAAIDI M., LAAIDI K., BESANCENOT J.P., THIBAUDON M., 2003b : Ragweed in France : an invasive plant and its allergenic pollen. *Annals of Allergy, Asthma and Immunology*, **91**, 195-201.

LAAIDI M., LAAIDI K., RIGOLLET S., 2003c : Pollen counts and allergies in Burgundy : assessment and perspectives. *European Annals of Allergy and Clinical Immunology*, **35**, 82-6.

LAAIDI M., THIBAUDON M., BESANCENOT J.P., 2004 : Comment prévoir la date de début de pollinisation d'une plante à partir de données météorologiques : l'exemple de l'ambrosie à Lyon. *European Annals of Allergy and Clinical Immunology*, **36**, 268-71.

LAHTI A., BJORKSTEN F., HANNUKSELA M., 1980 : Allergy to birch pollen and apple, and cross-reactivity of the allergens studied with the RAST. *Allergy*, **35**, 297-300.

LANDSBERG J.J., 1974 : Apple fruit bud development and growth ; analysis and an empirical model. *Annals of Botany*, **38**, 1013-23.

LARSSON K.A., 1993 : Prediction of the pollen season with a cumulated activity method. *Grana*, **32**, 111-4.

LATALOWA M., MIETUS M., URUSKA A., 2002 : Seasonal variations in the atmospheric *Betula* pollen count in Gdansk (southern Baltic coast) in relation to meteorological parameters. *Aerobiologia*, **18**, 33-43.

LEGENDRE P., LEGENDRE L., 1998 : Numerical Ecology. Elsevier Science BV, 853 p.

LEJOLY-GABRIEL M., 1978 : Recherches écologiques sur la pluie pollinique en Belgique. *Acta Geographica Lovaniensa*, **13**, 1-279.

LINVILL D.E., 1990 : Calculating chilling hours and chill units from daily maximum and minimum temperature observations. *Horticultural Science*, **25**, 14-6.

LYON F.L., KRAMER C.L., EVERSMEYER F.G., 1984 : Vertical variation of airspora concentrations in the atmosphere. *Grana*, **23**, 123-5.

- MAKRA L., JUHASZ M., BORSOS E., BECZI R., 2004 : Meteorological variables connected with airborne ragweed pollen in Southern Hungary. *International Journal of Biometeorology*, **49**, 37-47.
- MALANDAIN H., 2003 : Allergies associating pollens and foods. *European Annals of Allergy and Clinical Immunology*, **35**, 253-6.
- MALIK P., SINGH A.B., BABU C.R., GANGAL S.V., 1991 : Atmospheric concentrations of pollen grains at human height. *Grana*, **30**, 129-35.
- MARCEAUX S., TABOULOT S., 1994 : Atlas climatique de la Côte-d'Or. Météo France, Conseil Général de la Côte-d'Or, 127 p.
- MICHEL F.B., BOUSQUET J., 1997 : Les allergies : la fin d'une énigme. Editions Hachette, 383 p.
- MONTERO J.L., GONZALEZ J.L., 1983 : Diagramas bioclimáticos. Instituto Nacional para la conservación de la Naturaleza, Madrid.
- MOSEHOLM L., WEEKE E.R., PETERSEN B.N., 1987 : Forecast of pollen concentrations of Poaceae (grasses) in the air by time series analysis. *Pollen et Spores*, **29**, 305-22.
- MULLENDERS W., DIRICKX M., VAN DER HAEGEN D., BASTIN-SERVAIS Y., DESAIR C., 1972 : La pluie pollinique à Louvain-Heverlee en 1971. *Louvain Médecine*, **91**, 159-76.
- MUNUERA-GINER M., CARRION-GARCIA J.S., GARCIA-SELLES J., 1999 : Aerobiology of Artemisia pollen in Murcia (SE Spain) and its relationship with weather variables : annual and intradiurnal variations for three different species. Wind vectors as a tool in determining pollen origin. *International Journal of Biometeorology*, **43**, 51-63.
- MURRAY M.B., CANNELL M.G.R., SMITH R.I., 1989 : Date of budburst of fifteen tree species in Britain following climatic warming. *Journal of Applied Ecology*, **26**, 693-700.
- NILSSON S., PERSSON S., 1981 : Tree pollen spectra in the Stockholm region (Sweden), 1973-1980. *Grana*, **20**, 179-82.
- NORRIS-HILL J., 1997 : The influence of ambient temperature on the abundance of Poaceae pollen. *Aerobiologia*, **13**, 91-7.
- NORRIS-HILL J., EMBERLIN J., 1991 : Diurnal variation of pollen concentration in the air of north-central London. *Grana*, **30**, 229-34.
- OBTULOWICZ K., SZCZEPANEK K., RADWAN J., GRZYWACZ M., ADAMUS K., SZCZECLIK A., 1991 : Correlation between airborne pollen incidence, skin prick tests and serum immunoglobulins in allergic people in Cracow, Poland. *Grana*, **30**, 136-41.
- ONG E.K., SINGH M.B., KNOX R.B., 1994 : Grass pollen in the atmosphere of Melbourne : seasonal distribution over ten years. *Grana*, **34**, 58-63.

ONG E.K., TAYLOR P.E., KNOX R.B., 1997 : Forecasting the onset of the grass pollen season in Melbourne (Australia). *Aerobiologia*, **13**, 43-8.

ORLANDI F., VAZQUEZ L.M., RUGA L., BONOFILIO T., FORNACIARI M., GARCIA-MOZO H., DOMINGUEZ E., ROMANO B., GALAN C., 2005 : Bioclimatic requirements for olive flowering in two mediterranean regions located at the same latitude (Andalucia, Spain, and Sicily, Italy). *Annals of Agricultural and Environmental Medicine*, **12**, 47-52.

PARTON W.J., LOGAN J.A., 1981 : A model for diurnal variation in soil and air temperature. *Agricultural and Forest Meteorology*, **23**, 205-16.

PATHIRANE L., 1975 : Graphical determination of the main pollen season. *Pollen et Spores*, **17**, 609-10.

PAULI G., PAPANIKOLAOU I., NIEDERBERGER V., 2003 : Sensibilisation au frêne : allergènes spécifiques et/ou «croisants» ? *Revue française d'Allergologie et d'Immunologie clinique*, **43**, 120-4.

PEETERS A.G., 2000 : Frost periods and beginning of the ash (*Fraxinus excelsior* L.) pollen season in Basel (Switzerland). *Aerobiologia*, **16**, 353-9.

PRIETO-BAENAPABLO J.C., HIDALGO J., DOMINGUEZ E., GALAN C., 2003 : Pollen production in the Poaceae family. *Grana*, **42**, 1-7.

RADISIC P., SIKOPARIJA B., 2005 : *Betula* spp. Pollen in the atmosphere of Novi Sad (2000-2002). *Aerobiologia*, **21**, 63-7.

RANTIO-LEHTIMAKI A., KOIVIKKO A., KUPIAS R., MAKINEN Y., POHJOLA A., 1991 : Significance of sampling height of airborne particles for aerobiological information. *Allergy*, **46**, 68-76.

RASMUSSEN A., 2002 : The effects of climate change on the birch pollen season in Denmark. *Aerobiologia*, **18**, 253-65.

RICHARDSON E.A., SCHUYLER D.S., WALKER D.R., 1974 : A model for estimating the completion of rest for «Redheaven» and «Elberta» peach trees. *Horticultural Science*, **9**, 331-2.

RICKMAN R.W., KLEPPER B.L., PETERSON C.M., 1983 : Time distributions for describing appearance of specific culms of winter wheat. *Agronomy Journal*, **75**, 551-6.

RIEUX C., 2006 : Représentativité statistique du réseau national de surveillance aérobiologique (RNSA) en Rhône-Alpes. Faisabilité d'une modélisation. Rapport de stage de Master 2 professionnel, Université de Grenoble, 35 p.

RODRIGUEZ-RAJO F.J., FRENGUELLI G., JATO M.V., 2003 : Effect of air temperature on forecasting the start of the *Betula* pollen season at two contrasting sites in the south of Europe (1995-2001). *International Journal of Biometeorology*, **47**, 117-25.

RODRIGUEZ-RAJO F.J., DACOSTA N., JATO V., 2004 : Environmental factors affecting the start of pollen season and concentrations of airborne and concentrations of airborne *Alnus* pollen in two localities of Galicia (NW Spain). *Annals of Agricultural and Environmental Medicine*, **11**, 35-44.

RYBNICEK O., JAEGER S., 2001 : *Ambrosia* (ragweed) in Europe. *Allergy and Clinical Immunology International*, **13**, 60-6.

SANCHEZ-MESA J.A., SMITH M., EMBERLIN J., ALLITT U., CAULTON E., GALAN C., 2003 : Characteristics of grass pollen seasons in areas of southern Spain and the United Kingdom. *Aerobiologia*, **19**, 243-50.

SANCHEZ MESA J.A., GALAN C., HERVAS C., 2005: The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. *International Journal of Biometeorology*, **49**, 355-62.

SARVAS R., 1974 : Investigations on the annual cycle of development of forest trees. Autumn dormancy and winter dormancy. *Communicationes Instituti Forestalis Fenniae*, **84**, 101 p.

SMITH M., EMBERLIN J., 2005 : Constructing a 7-day ahead forecast model for grass pollen at north London, United Kingdom. *Clinical and Experimental Allergy*, **35**, 1400-6.

SNYDER R.L., 1985 : Hand calculating degree-days. *Agricultural and Forest Meteorology*, **35**, 353-8.

SNYDER R.L., SPANO D., CESARACCIO C., DUCE P., 1999 : Determining degree-day thresholds from field observations. *International Journal of Biometeorology*, **42**, 177-82.

SOFIEV M., SILJAMO P., RANTA H., RANTIO-LEHTIMÄKI A., 2006 : Towards numerical forecasting of long-range air transport of birch pollen : theoretical considerations and a feasibility study. *International Journal of Biometeorology*, **50**, 392-402.

SPIEKSMA F.T.M., den TONKELAAR J.F., 1986 : Four-hourly fluctuations in grass-pollen concentrations in relation to wet versus dry weather, and to short versus long over-land advection. *International Journal of Biometeorology*, **30**, 351-8.

SPIEKSMA F.T.M., NOLARD N., FRENGUELLI G., VAN MOERBOEKE D., 1993 : Pollens de l'air en Europe. Éditions UCB, 83 p.

SPIEKSMA F.T.M., , EMBERLIN J., HJELMROOS M., JAEGER S., 1995 : Atmospheric birch : *Betula* pollen in Europe : Trends and fluctuation in annual quantities and the starting dates of the seasons. *Grana*, **34**, 51-7.

SUTRA J.P., DESROZIER M., CERON J.P., FRAYSSINET P., MARTIN E., MERLIER C., PERARNAUD V., TRAULLE M.P., VEYSSIERE J.M., HOFF M., BURNICHON A., JULIEN P., RUFFALDI P., BOURGEOIS E., CHAMBAUD R., MARGUIER S., 1992 : La prévision statistique de la date initiale de pollinisation du bouleau en France. *Climat et Santé*, **7**, 53-84.

TABUENCA M.C., HERRERO J., 1965 : Influencia de la temperatura en la época de floración de frutales. *Anales de la Estación Experimental de Aula Dei*, **8**, 115-53.

TAUDORF E., MOSEHOLM L., 1988 : Pollen count, symptom, and medicine score in birch pollinosis. *International Archives of Allergy and Applied Immunology*, **86**, 225-33.

THIBAUDON M., 2003 : The allergy risk associated with pollens in France. *European Annals of Allergy and Clinical Immunology*, **35**, 170-2.

THIBAUDON M., LACHASSE C., FINET F., 2003 : Ragweed in France and in the Rhône-Alpes region (Lyon-Bourgoin-Grenoble-Roussillon). *European Annals of Allergy and Clinical Immunology*, **35**, 87-91.

THIBAUDON M., ELIAS K., BESANCENOT J.P., 2004 : Ambrosie et allergie le cas de la France. *Environnement, Risques et Santé*, **3**, 353-67.

THIBAUDON M., LACHASSE C., 2004 : Ragweed pollens under high surveillance. *European Annals of Allergy and Clinical Immunology*, **36**, 337-41.

TÖRNEVIK H., 1982 : An aerobiological model for operational forecasts of pollen concentration in the air. *SMHI Reports*, **38**, 1-38.

TOURAIN R., 1984 : La pollinose à *Ambrosia*. *Allergy and Immunology*, **16**, 275-6.

TOURAIN R., CORNILLON J., DE POUMEYROL B., 1966 : Pollinose et *Ambrosia*. La diffusion actuelle d'*Ambrosia* dans la région lyonnaise. Son rôle dans les maladies par allergie pollinique. *Bulletin mensuel de la Société Linnéenne de Lyon*, **6**, 279-85.

TRIGO M.M., TORO F.J., RECIO M., CABEZUDO B., 2000 : A statistical approach to comparing the results from different aerobiological stations. *Grana*, **39**, 252-8.

VALENCIA-BARRERA R.M., COMTOIS P., FERNANDEZ-GONZALEZ D., 2001 : Biogeography and bioclimatology in pollen forecasting . An example of grass in León (Spain) and Montréal (Canada). *Grana*, **40**, 223-9.

VAZQUEZ L.M., GALAN C., DOMINGUEZ-VILCHES E., 2003 : Influence of meteorological parameters on *Olea* pollen concentrations in Córdoba (South-western Spain). *International Journal of Biometeorology*, **48**, 83-90.

VIANDER M., KOIVIKKO A., 1978 : The seasonal symptoms of hyposensitized and untreated hay fever patients in relation to birch pollen counts : correlations with nasal sensitivity, prick tests and RAST. *Clinical Allergy*, **8**, 387-96.

WALLIN J.E., SEGERSTRÖM V., ROSENHALL L., BERGMANN E., HJELMROOS M., 1991 : Allergic symptoms caused by long distance transported birch pollen. *Grana*, **30**, 256-68.

WIHL J.A., IPSEN B., NÜCHEL P.B., MUNCH E.P., JANNICHE E.P., LÖVENSTEIN H., 1998 : Immunotherapy with partially purified and standardized tree pollen extracts. *Allergy*, **43**, 363-9.

WÜTHRICH B., 1995 : Pollenallergie. Aktuelles zur Epidemiologie, Pathogenese, Diagnostik und Therapie. *Deutsches Aerzteblatt*, **92**, 809-14.

WÜTHRICH B., SCHINDLER C., LEUENBERGER P., ACKERMANN-LIEBRIG U., 1995: Prevalence of atopy and pollinosis in the adult population of Switzerland. *International Archives of Allergy and Immunology*, **106**, 149-56.

YANG S., LOGAN J., COFFEY D.L., 1995 : Mathematical formulae for calculating the base temperature for growing degree-days. *Agricultural and Forest Meteorology*, **74**, 61-74.

ZALOM F.G., GOODDELL P.B., WILSON L.T., BARNET W.W., BENTLEY W.J., 1983 : Degree-days : the calculation and use of heat units in pest management. *DANR publications*, University of California, 2-10.

Liste des sites internets :

www.hist-geo.com

www.languedoc-roussillon.pref.gouv.fr

www.pollens.fr

www.univ-ubs.fr

www.web-provence.com

TABLE DES FIGURES

Figure 1 : Le capteur de pollen Hirst de Dijon.....	7
Figure 2 : Les composants du corps central du capteur pollinique.....	8
Figure 3 : Lames de lecture de Dijon (19-20/04/2005) et Castres (18-19/06/2005).....	9
Figure 4 : Nombre annuel de capteurs en fonctionnement sur la période 1987-2007 gérés par l'Institut Pasteur puis le RNSA.....	13
Figure 5 : Implantation des capteurs polliniques, gérés par le RNSA, par département en 1996 et 2007 (source des fonds de carte : www.hist-geo.com).....	14
Figure 6 : Lames de lecture de Dijon (11-12 mai, 17-18 mai, 19-20 mai et 26-27 mai 2004).....	23
Figure 7 : Classement par ordre croissant de la proportion de données manquantes sur la totalité des années disponibles ville par ville.....	27
Figure 8 : Classement par ordre croissant de la proportion de données manquantes, ville par ville, sur la période 1997-2005.....	28
Figure 9 : Proportion annuelle de données manquantes pour les 29 villes ayant des données sur la période 1997-2005.....	29
Figure 10 : Classement par ordre croissant de la durée moyenne des périodes lacunaires, ville par ville, sur la série 1997-2005.....	36
Figure 11 : Classement par ordre croissant de la date moyenne de démarrage des capteurs (avec écart-type), ville par ville et sur la période 1997-2005, en nombre de jours à partir du 1 ^{er} janvier.....	41
Figure 12 : Classement par ordre décroissant de la date moyenne d'arrêt des capteurs, ville par ville et sur la période 1997-2005, en nombre de jours à partir du 1 ^{er} janvier.	45
Figure 13 : Classement par ordre croissant de la hauteur (en mètres) du capteur par rapport au sol.....	50
Figure 14 : Une espèce de la famille des Poacées : le dactyle pelotonné (<i>Dactylis glomerata</i> L.) (Guérin, 1993).....	64
Figure 15 : Photographies du pollen de Poaceae (ø de 10 à 80 µm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Laaidi et Laaidi, 2001).....	65
Figure 16 : Un spécimen de bouleau : le bouleau pleureur (source : www.web-provence.com).....	67

Figure 17 : Photographies du pollen de bouleau (ϕ de 18 à 24 μm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Laaidi et Laaidi, 2001).....	68
Figure 18 : Un spécimen de frêne : le frêne commun (source : www.web-provence.com).	69
Figure 19 : Photographies du pollen de frêne (ϕ de 21 à 26 μm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Spieksma <i>et al.</i> , 1993).....	70
Figure 20 : Aperçu de quelques branches et feuilles d'un plant d'ambroisie (source : www.univ-ubs.fr).....	72
Figure 21 : Photographies du pollen d'ambroisie (ϕ de 19 μm) en microscopie optique (à gauche) et microscopie électronique (à droite) (Spieksma <i>et al.</i> , 1993).....	73
Figure 22 : Classement par ordre décroissant de la proportion de pollens de Poacées, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.....	78
Figure 23 : Classement par ordre décroissant de la proportion de pollens de bouleau, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.....	80
Figure 24 : Classement par ordre décroissant de la proportion de pollens de frêne, ville par ville, sur la période 1997-2005, et position par rapport aux autres taxons.....	82
Figure 25 : Classement par ordre décroissant de la proportion de pollens d'ambroisie, ville par ville, sur la période 1997-2005.....	84
Figure 26 : Poids factoriels des trois variables sur les deux premières composantes principales.....	86
Figure 27 : Résultats factoriels de la première composante principale où 1 cm = 1 unité (source du fond de carte : www.hist-geo.com).....	88
Figure 28 : Résultats factoriels de la deuxième composante principale où 1 cm = 1 unité (source du fond de carte : www.hist-geo.com).....	89
Figure 29 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au frêne à Amiens (1989-2005), critère par critère, en nombre de jours à partir du 1 ^{er} janvier.....	104
Figure 30 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au frêne à Montluçon (1990-2002 et 2004), critère par critère, en nombre de jours à partir du 1 ^{er} janvier.....	106
Figure 31 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au bouleau à Amiens (1987 et 1989-2005), critère par critère, en nombre de jours à partir du 1 ^{er} janvier.....	108

Figure 32 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé au bouleau à Montluçon (1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.....	110
Figure 33 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé aux Poacées à Amiens (1987-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.....	112
Figure 34 : Classement par ordre croissant de la date moyenne d'apparition du risque allergique associé aux Poacées à Montluçon (1989-2005), critère par critère, en nombre de jours à partir du 1^{er} janvier.....	114
Figure 35 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour le frêne.....	158
Figure 36 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour le frêne.....	159
Figure 37 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour le frêne.....	162
Figure 38 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour le frêne.....	164
Figure 39 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour le bouleau.....	179
Figure 40 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour le bouleau.....	181
Figure 41 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour le bouleau.....	184
Figure 42 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour le bouleau.....	186
Figure 43 : Valeurs d'erreur moyenne absolue (en jours) des cinq villes, par méthode, pour les Poacées.....	202
Figure 44 : Valeurs d'erreur moyenne absolue (en jours) des modèles issus des cinq méthodes de cumul, par mois et par ville, pour les Poacées.....	203
Figure 45 : Valeurs d'erreur moyenne absolue (en jours) des modèles des cinq villes, par mois et par méthode, pour les Poacées.....	205
Figure 46 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles et des villes, mois par mois, pour les Poacées.....	207
Figure 47 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Rouen, en nombre de jours à partir du 1^{er} janvier.....	223

Figure 48 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Amiens, en nombre de jours à partir du 1^{er} janvier.....	225
Figure 49 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Paris, en nombre de jours à partir du 1^{er} janvier.....	227
Figure 50 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Nancy, en nombre de jours à partir du 1^{er} janvier.....	229
Figure 51 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Lyon, en nombre de jours à partir du 1^{er} janvier.....	231
Figure 52 : Dates d'apparition du risque d'exposition allergique au pollen de frêne, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.....	232
Figure 53 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Rouen, en nombre de jours à partir du 1^{er} janvier.....	234
Figure 54 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Amiens, en nombre de jours à partir du 1^{er} janvier.....	236
Figure 55 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Paris, en nombre de jours à partir du 1^{er} janvier.....	239
Figure 56 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Nancy, en nombre de jours à partir du 1^{er} janvier.....	241
Figure 57 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Strasbourg, en nombre de jours à partir du 1^{er} janvier.....	243
Figure 58 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Lyon, en nombre de jours à partir du 1^{er} janvier.....	245
Figure 59 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Clermont-Ferrand, en nombre de jours à partir du 1^{er} janvier.....	246
Figure 60 : Dates d'apparition du risque d'exposition allergique au pollen de bouleau, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.....	248
Figure 61 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, en nombre de jours à partir du 1^{er} janvier, à Rouen (noir), Amiens (gris foncé), Nancy (gris moyen) et Strasbourg (gris clair).....	250
Figure 62 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Paris, en nombre de jours à partir du 1^{er} janvier.....	253
Figure 63 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Lyon, en nombre de jours à partir du 1^{er} janvier.....	255

Figure 64 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, en nombre de jours à partir du 1^{er} janvier, à Clermont-Ferrand (noir) et Saint-Etienne (gris).....	255
Figure 65 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Bordeaux, en nombre de jours à partir du 1^{er} janvier.....	259
Figure 66 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Toulouse, en nombre de jours à partir du 1^{er} janvier.....	259
Figure 67 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Nantes, en nombre de jours à partir du 1^{er} janvier.....	262
Figure 68 : Dates d'apparition du risque d'exposition allergique au pollen de Poacées, à Marseille, en nombre de jours à partir du 1^{er} janvier.....	264
Figure 69 : Dates d'apparition du risque d'exposition allergique au pollen d'ambroisie, à Lyon, en nombre de jours à partir du 1^{er} janvier et droite de tendance linéaire.....	266
Figure 70 : Valeurs d'erreur moyenne absolue mensuelle (en jours) de l'ensemble des modèles, mois par mois, pour l'ambroisie à Lyon.....	267
Figure 71 : Valeurs moyennes de température minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé) selon le niveau de REAP à Nantes.....	290
Figure 72 : Valeurs moyennes de température minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé) selon le niveau de REAP à La Roche-sur-Yon.....	290
Figure 73 : Valeurs moyennes d'humidité relative minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé), en %, selon le niveau de REAP à Nantes.....	291
Figure 74 : Valeurs moyennes d'humidité relative minimale (gris clair), moyenne (gris moyen) et maximale (gris foncé), en %, selon le niveau de REAP à La Roche-sur-Yon.....	292
Figure 75 : Fréquence de précipitations > 0 mm (gris foncé), ≥ 1 mm (gris moyen) et ≥ 5 mm (gris clair) selon le niveau de REAP à Nantes.....	293
Figure 76 : Fréquence de précipitations > 0 mm (gris foncé), ≥ 1 mm (gris moyen) et ≥ 5 mm (gris clair) selon le niveau de REAP à La Roche-sur-Yon.....	293
Figure 77 : Représentation des deux premières dimensions de l'ACM établie à partir du REAP, des types de temps et des combinaisons de précipitations à Nantes.....	298
Figure 78 : Représentation des deux premières dimensions de l'ACM établie à partir du REAP, des types de temps et des combinaisons de précipitations à La Roche-sur-Yon.....	298

Figure 79 : Amplitude interjournalière moyenne des concentrations polliniques ($\text{gr/m}^3/\text{j}$) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair)..... 301

Figure 80 : Amplitude interjournalière moyenne de l'humidité relative minimale (en %) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair)..... 302

Figure 81 : Amplitude interjournalière moyenne des précipitations (en mm) selon les variations du REAP à Nantes (gris moyen) et La Roche-sur-Yon (gris clair)..... 302

TABLE DES TABLEAUX

Tableau 1 : Tableau hebdomadaire des comptes polliniques de Chalon-sur-Saône pour la semaine du 28 mars au 3 avril 2006.....	10
Tableau 2 : Tableau journalier des comptes polliniques de Chalon-sur-Saône du 21 juin 2007.....	11
Tableau 3 : Récapitulatif des mises en routes, remises en route et arrêts de fonctionnement des sites de capture pollinique, de 1988 à 2007, pour un fonctionnement annuel.....	16
Tableau 4 : Durée moyenne annuelle, écart-type et coefficient de variation liés aux données manquantes sur la période 1997-2005.....	33
Tableau 5 : Nombre de grains de pollens comptés et total après conversion (coefficient de correction de 0,69) à Chalon-sur-Saône, du 4 mars 2007 à 8 heures au 5 mars 2007 à 8 heures : résultats pour deux lignes de lecture.....	54
Tableau 6 : Nombre de grains de pollens comptés et total après conversion (coefficient de correction de 0,46) à Chalon-sur-Saône, du 4 mars 2007 à 8 heures au 5 mars 2007 à 8 heures : résultats pour trois lignes de lecture.....	55
Tableau 7 : Classement des dix premiers taxons par ordre décroissant sur l'ensemble des 29 villes de l'étude.....	75
Tableau 8 : Classement des dix premiers taxons par ordre décroissant sur les villes situées au sud et au nord du 45^{ème} parallèle.....	76
Tableau 9 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique au pollen de frêne, sur les 11 années communes.....	168
Tableau 10 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique au pollen de bouleau, sur les 15 années communes.....	190
Tableau 11 : Présentation des variables corrélées significativement ($p = 0,05$) avec les dates d'apparition d'un risque d'exposition allergique aux pollen de Poacées, sur les 18 années communes.....	211
Tableau 12 : Caractéristiques géographiques des cinq villes de l'étude.....	276
Tableau 13 : Températures minimales (Tn), maximales (Tx) et moyennes (Tm) des mois de février (02) et avril (04) des années communes à chaque taxon.....	276

Tableau 14 : Dates moyennes d'apparition du risque allergique au pollen de frêne, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.....	281
Tableau 15 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour le frêne.....	281
Tableau 16 : Dates moyennes d'apparition du risque allergique au pollen de bouleau, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.....	282
Tableau 17 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour le bouleau.....	283
Tableau 18 : Dates moyennes d'apparition du risque allergique au pollen de Poacées, selon les critères du RNSA (REAP) et de la fréquence cumulée de 5%.....	284
Tableau 19 : Valeurs d'erreur moyenne absolue (en jours) de l'ensemble des meilleurs modèles issus des méthodes de cumul, selon le type de critère de détermination, pour les Poacées.....	285
Tableau 20 : Niveaux de REAP associés aux Poacées et concentrations correspondantes.	288
Tableau 21 : Distribution des niveaux de REAP pour les périodes d'étude respectives de chaque ville.....	289
Tableau 22 : Ecart entre les effectifs observés et prévus dans le cadre du test du χ^2 établi à partir des niveaux de REAP et de classes de précipitations.....	295
Tableau 23 : Caractéristiques météorologiques moyennes des types de temps issus de la classification en nuées dynamiques.....	296
Tableau 24 : Distribution des combinaisons de précipitations pour les périodes d'étude respectives de chaque ville.....	297
Tableau 25 : Distribution des variations interjournalières de REAP pour les périodes d'étude respectives de chaque ville.....	300

LISTE DES SIGLES ET ABREVIATIONS

Les organismes et congrès

ADEME : Agence de l'Environnement et de la Maîtrise de l'Energie
APPA : Association pour la Prévention de la Pollution Atmosphérique
CEMBREU : Centre Européen Médical Bioclimatique de Recherche et d'Enseignement Universitaire
DRASS : Direction Régionale des Affaires Sanitaires et Sociales
ENSA-M : Ecole Nationale Supérieure Agronomique de Montpellier
JIQA : Journées Interdisciplinaires de la Qualité de l'Air
RNSA : Réseau National de Surveillance Aérobiologique

Les critères et méthodes

AC : «Action Chaude» de la formule de Bidabe (1967)
ACM : Analyse des Correspondances Multiple
ACP : Analyse en Composantes Principales
Af : test de performance en prévision réelle
Chi² : test de comparaison de deux variables qualitatives
EEEE : sigle indiquant un problème technique caractérisé par une absence de données bi horaires dans les fichiers de données du RNSA
FT : Fraction-Time
GDD : Growing Degree Days
GDD_1 : Growing Degree Days par triangulation
GDD_2 : Growing Degree Days par une fonction sinusoïdale
GDH : Growing Degree Hours
Lej_Tx : méthode de Lejoly-Gabriel appliquée aux températures maximales
MAE : Mean Absolute Error
Q₁₀ : coefficient de développement
REAP : Risque d'Exposition Allergique aux Pollens
RMSE : Root-Mean-Square-Error
sum75 : somme cumulée de 75 grains
1%/j : fréquence journalière de 1%
1%/pen : fréquence journalière moyenne de 1% sur une pentade
2,5% : fréquence cumulée de 2,5%
2,5%/j : fréquence journalière de 2,5%
5% : fréquence cumulée de 5%
5%_1%/j : fréquence cumulée de 5% et fréquence journalière de 1%
30 g/j : concentration journalière de 30 grains

Les variables météorologiques et pédologiques

Gel : durée du gel (min)
Hn, Hx et Hm : humidité relative minimale, maximale et moyenne (%)

Insol : insolation (min)

JP, JP1 et JP5 : nombre de jours où les précipitations sont supérieures ou égales à 0,1 mm, 1 mm et 5mm

P : précipitations (mm)

Tn, Tx et Tm : température minimale, maximale et moyenne (°C)

Tn0, Tx0 et Tm0 : nombre de jours où la température minimale, maximale et moyenne est inférieure à 0°C

T-10 et T-20 : température mesurée dans le sol à – 10 cm et – 20 cm (°C)

TABLE DES MATIERES

REMERCIEMENTS

SOMMAIRE

INTRODUCTION GENERALE..... 1

PARTIE 1 : DONNEES POLLINIQUES ET TAXONS DE L'ETUDE.....5

1.1 PRESENTATION GENERALE DES DONNEES POLLINIQUES..... 7

1.1.1 Présentation du matériel..... 7

1.1.1.1 Fonctionnement du capteur pollinique.....7

1.1.1.2 Montage des lames et analyse des pollens..... 9

1.1.2 Présentation des sites du RNSA.....11

1.1.2.1 Le réseau de capteurs..... 11

1.1.2.2 Données de l'étude..... 18

1.1.3 Etude des données manquantes..... 19

1.1.3.1 Les causes..... 19

1.1.3.2 Détermination des causes : le cas de Chalon-sur-Saône et de Dijon..... 21

1.1.3.3 Répartition générale des valeurs manquantes..... 25

1.1.3.4 Etude de la variation interannuelle des valeurs manquantes par ville..... 30

1.1.3.5 Etude de la durée moyenne des périodes lacunaires..... 35

1.1.4 Etude sur les dates de mise en route et d'arrêt des capteurs..... 38

1.1.4.1 Mise en route et arrêt des capteurs : l'importance de leur choix..... 38

1.1.4.2 Comparaison des dates moyennes de démarrage..... 40

1.1.4.3 Comparaison des dates moyennes d'arrêt..... 44

1.1.5 Etude des concentrations polliniques quotidiennes : influence de différents paramètres..... 48

1.1.5.1 Influence de la position en hauteur du capteur..... 49

1.1.5.2 Influence d'autres paramètres techniques..... 53

1.1.5.3 Influence du nombre de lignes de lecture : le cas de quatre stations en 2007..... 56

1.1.5.4 Bilan..... 60

1.1.6 Conclusion.....61

1.2 PRESENTATION GENERALE DES TAXONS DE L'ETUDE..... 63

1.2.1 Les types de taxons retenus..... 63

1.2.1.1 <u>Les Poacées</u>	64
1.2.1.2 <u>Le bouleau</u>	66
1.2.1.3 <u>Le frêne</u>	68
1.2.1.4 <u>L'ambroisie</u>	71
1.2.2 Concentration et répartition des quatre taxons de l'étude sur l'ensemble du territoire français	74
1.2.2.1 <u>Présentation générale des résultats</u>	74
1.2.2.2 <u>Les Poacées</u>	77
1.2.2.3 <u>Le bouleau</u>	79
1.2.2.4 <u>Le frêne</u>	81
1.2.2.5 <u>L'ambroisie</u>	83
1.2.2.6 <u>Une autre approche de la présentation générale des résultats : application d'une Analyse en Composantes Principales (ACP)</u>	85
1.2.3 Conclusion	90
CONCLUSION DE LA PARTIE	91

PARTIE 2 : CRITERES ET METHODES LIES AU DEBUT DE LA SAISON POLLINIQUE..... 93

2.1 PRESENTATION ET ANALYSE DES DIFFERENTS CRITERES DE DETERMINATION DU DEBUT DE LA SAISON POLLINIQUE	95
2.1.1 Revue des différents critères s'appuyant sur des valeurs brutes	95
2.1.2 Revue des différents critères s'appuyant sur des valeurs relatives	97
2.1.3 Avantages et inconvénients des deux types de critères	100
2.1.4 Application de différents seuils pour le bouleau, le frêne et les Poacées : exemples d'Amiens et de Montluçon	102
2.1.4.1 <u>Le frêne à Amiens (1989-2005)</u>	102
2.1.4.2 <u>Le frêne à Montluçon (1990-2002 et 2004)</u>	102
2.1.4.3 <u>Le bouleau à Amiens (1987 et 1989-2005)</u>	107
2.1.4.4 <u>Le bouleau à Montluçon (1989-2005)</u>	109
2.1.4.5 <u>Les Poacées à Amiens (1987-2005)</u>	111
2.1.4.6 <u>Les Poacées à Montluçon (1989-2005)</u>	113
2.1.5 Quel critère choisir ?	115
2.1.6 Conclusion	119

2.2 PRESENTATION GENERALE DES METHODES DE PREVISION DE LA DATE D'APPARITION DU RISQUE..... 121

2.2.1 Détermination de la dormance.....122

2.2.1.1 Intérêt de la détermination..... 122

2.2.1.2 Présentation de différentes méthodes.....122

2.2.1.3 La détermination des unités de «froid» est-elle essentielle ?..... 127

2.2.2 La période de forçage des températures : présentation de plusieurs méthodes....127

2.2.2.1 Intérêt de la détermination de cette période..... 127

2.2.2.2 Les six méthodes présélectionnées dans ce travail..... 128

2.2.2.3 Présentation d'autres méthodes de cumul d'unités de «chaleur»..... 133

2.2.3 Utilisation d'une méthode plus globale : la régression linéaire multiple..... 136

2.2.4 Les critères de choix des meilleurs modèles.....137

2.2.4.1 Les critères destinés aux méthodes de cumul..... 138

2.2.4.2 Les critères destinés à la régression linéaire multiple..... 139

2.2.5 Conclusion.....140

CONCLUSION DE LA PARTIE..... 141

PARTIE 3 : PRESENTATION DES RESULTATS..... 143

3.1 OPERATIONS SUR LES DONNEES POLLINIQUES ET PRESENTATION DES DONNEES METEOROLOGIQUES..... 145

3.1.1 La reconstitution des données polliniques manquantes..... 145

3.1.2 Les données météorologiques utilisées..... 148

3.2 PREVISION DU FRENE, DU BOULEAU, DES POACEES : LE CAS DE CINQ VILLES DE LA MOITIE NORD DE LA FRANCE..... 151

3.2.1 Choix des villes..... 151

3.2.2 Prévision du frêne par les méthodes de cumul..... 152

3.2.2.1 Présentation des dates d'apparition et de l'application des méthodes..... 152

3.2.2.2 Présentation des meilleurs modèles..... 154

3.2.2.3 Comparaisons inter-méthodes et inter mensuelles..... 157

3.2.2.4 Bilan global..... 164

3.2.2.5 Etude des résidus.....165

3.2.3 Prévision du frêne par la régression linéaire multiple.....167

3.2.3.1 Etude des corrélations..... 167

3.2.3.2 Application de validations croisées.....169

3.2.3.3 <u>Equations de régression et résidus</u>	172
3.2.4 Prévision du bouleau par les méthodes de cumul	174
3.2.4.1 <u>Présentation des dates d'apparition et de l'application des méthodes</u>	174
3.2.4.2 <u>Présentation des meilleurs modèles</u>	176
3.2.4.3 <u>Comparaisons inter-méthodes et inter mensuelles</u>	179
3.2.4.4 <u>Bilan global</u>	186
3.2.4.5 <u>Etude des résidus</u>	187
3.2.5 Prévision du bouleau par la régression linéaire multiple	189
3.2.5.1 <u>Etude des corrélations</u>	189
3.2.5.2 <u>Application de validations croisées</u>	191
3.2.5.3 <u>Equations de régression et résidus</u>	195
3.2.6 Prévision des Poacées par les méthodes de cumul	197
3.2.6.1 <u>Présentation des dates d'apparition et de l'application des méthodes</u>	197
3.2.6.2 <u>Présentation des meilleurs modèles</u>	198
3.2.6.3 <u>Comparaisons inter-méthodes et inter mensuelles</u>	201
3.2.6.4 <u>Bilan global</u>	207
3.2.6.5 <u>Etude des résidus</u>	208
3.2.7 Prévision des Poacées par la régression linéaire multiple	210
3.2.7.1 <u>Etude des corrélations</u>	210
3.2.7.2 <u>Application de validations croisées</u>	212
3.2.7.3 <u>Equations de régression et résidus</u>	216
3.2.8 Conclusion	218
3.3 PRESENTATION DES RESULTATS POUR LES SERIES POLLINIQUES LES PLUS LONGUES	221
3.3.1 Résultats pour le frêne	222
3.3.1.1 <u>Rouen</u>	222
3.3.1.2 <u>Amiens</u>	224
3.3.1.3 <u>Paris</u>	226
3.3.1.4 <u>Nancy</u>	229
3.3.1.5 <u>Lyon et Bordeaux</u>	231
3.3.2 Résultats pour le bouleau	234
3.3.2.1 <u>Rouen</u>	234
3.3.2.2 <u>Amiens</u>	236
3.3.2.3 <u>Paris</u>	238
3.3.2.4 <u>Nancy</u>	240
3.3.2.5 <u>Strasbourg</u>	242
3.3.2.6 <u>Les sites du Centre-Est : Lyon et Clermont-Ferrand</u>	245
3.3.2.7 <u>Bordeaux</u>	248

3.3.3 Résultats pour les Poacées.....	250
3.3.3.1 Rouen, Amiens, Nancy et Strasbourg.....	250
3.3.3.2 Paris.....	252
3.3.3.3 Les sites du Centre-Est : Lyon, Clermont-Ferrand et Saint-Etienne.....	254
3.3.3.4 Les sites du Sud-Ouest : Bordeaux et Toulouse.....	258
3.3.3.5 Nantes.....	262
3.3.3.6 Marseille.....	263
3.3.4 Résultats pour l'ambrosie : le cas de Lyon.....	266
3.3.4.1 Résultats des méthodes de cumul.....	267
3.3.4.2 Résultats de la régression linéaire multiple et prévisions pour 2006.....	269
3.3.5 Discussion et conclusion.....	271
3.4 ETUDES COMPARATIVES COMPLEMENTAIRES.....	275
3.4.1 Détermination de relations entre seuils d'action et caractéristiques physiques : le cas d'Amiens, Rouen, Paris, Nancy et Strasbourg.....	275
3.4.2 Précision des modèles de prévision : comparaison en fonction de deux critères de détermination de l'apparition du risque.....	279
3.4.2.1 Le cas du frêne.....	280
3.4.2.2 Le cas du bouleau.....	282
3.4.2.3 Le cas des Poacées.....	283
3.5 ETUDE SUR LES CONCENTRATIONS POLLINIQUES : LES POACEES A NANTES ET LA ROCHE-SUR-YON.....	287
3.5.1 Matériel et méthodes.....	287
3.5.2 Relations entre REAP et conditions météorologiques.....	288
3.5.2.1 Description des données de Poacées.....	288
3.5.2.2 Conditions moyennes de température, d'humidité relative et de précipitations associées à chaque niveau.....	289
3.5.2.3 Relations entre le REAP et les cumuls de précipitations.....	294
3.5.2.4 Utilisation de types de temps et des conditions de précipitations de la veille.....	295
3.5.3 Le cas des variations interjournalières de REAP.....	299
3.5.4 Discussion et conclusion.....	304
CONCLUSION DE LA PARTIE.....	305
CONCLUSION GENERALE.....	307
BIBLIOGRAPHIE.....	311
TABLE DES FIGURES.....	325

TABLE DES TABLEAUX.....	331
LISTE DES SIGLES ET ABREVIATIONS.....	333
TABLE DES MATIERES.....	335

Résumé :

A l'heure actuelle, 16% des Français souffrent d'allergies à un ou plusieurs pollens. Les symptômes correspondants peuvent aussi bien se présenter sous une forme bénigne (rhinite, conjonctivite, toux) que sous une forme beaucoup plus grave : l'asthme. La prévision de la date d'apparition d'un risque d'exposition allergique aux pollens est donc nécessaire d'un point de vue sanitaire et préventif. Elle doit être la plus précise possible afin d'initier les traitements anti-allergiques au moment adéquat, dans un souci d'efficacité et de réduction des coûts liés à cette affection. La présente étude, qui a pour cadre géographique le territoire national français, concerne quatre taxons polliniques parmi les plus allergisants : le frêne, le bouleau, les Poacées et l'ambroisie. Ce travail comprend trois parties. La première partie consiste en une présentation générale des données polliniques composée d'une description des modes de recueil et d'analyse des pollens, d'une étude sur la qualité des séries polliniques et les divers problèmes rencontrés et, enfin, d'une description détaillée des quatre taxons de l'étude et de leur répartition géographique. La deuxième partie consiste en une description des différents critères de détermination du début de la saison pollinique et des méthodes inhérentes à sa prévision. Enfin, la troisième partie rassemble l'ensemble des résultats, pour les quatre taxons concernés, issus des méthodes de prévision sélectionnées et appliquées aux séries polliniques les plus longues. Cette dernière partie s'attache essentiellement à déterminer l'intérêt de l'utilisation des méthodes retenues, la qualité des modèles qui en sont issus et leur efficacité dans une optique préventive. Une étude sur les relations entre les concentrations polliniques et les variables météorologiques constitue le dernier volet de cette troisième partie.

Mots clés : Aérobiologie – Pollen – Allergie – Santé – Prévention – Risque d'exposition – Conditions météorologiques – Méthodes de prévision

Abstract :

At present, 16% of French people suffer from allergies to one or several pollens. The corresponding symptoms can be presented as under benign form (rhinitis, conjunctivitis, cough) as under much more serious form : asthma. Forecast of the starting date of an allergic exposure risk to pollens is necessary from a sanitary and preventive standpoint. Forecast has to be more precisely as possible in order to begin anti-allergic treatments at appropriate moment, with a view of effectiveness and reduction of the costs due to this disease. The present study, taking place in all the French territory, concerns four pollen taxa among the most allergenic : ash, birch, grasses and ragweed. This work comprises three parts. The first one consists of a global presentation of pollen data with a description of pollen collection and analysis methods, a study of the pollen series quality and the various problems found and, finally, it consists of a detailed description of the four taxa of the study including their geographical distribution. The second part describes various determining criteria of the beginning of the pollen season, and different forecast methods of this one. Finally, the third part collects all the results, for the four-concerned taxa, been due to the selected forecast methods applied to longest pollen series. This final part mainly focuses on the interest of each of the selected methods, the quality of models associated to each one and the efficiency of these models in a preventive perspective. A study of the links between pollen concentrations and meteorological data constitutes the last section of this third part.

Key words : Aerobiology – Pollen – Allergy – Health – Prevention – Exposure risk – Meteorological conditions – Forecast methods