HAL
open science

Analyse statistique temporelle de l'irradiation solaire globale quotidienne : modélisation d'une variable réduite à l'aide de modèles stochastiques A.R.M.A.

B. Guerrier, E. Boileau, C. Benard

To cite this version:

B. Guerrier, E. Boileau, C. Benard. Analyse statistique temporelle de l'irradiation solaire globale quotidienne : modélisation d'une variable réduite à l'aide de modèles stochastiques A.R.M.A.. Revue de Physique Appliquée, 1980, 15 (2), pp.93-102. 10.1051/rphysap:0198000150209300 . jpa-00244736

HAL Id: jpa-00244736

https://hal.science/jpa-00244736

Submitted on 4 Feb 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.

Revue Phys. Appl. **15** (1980) 93-102

Classification
Physics Abstracts
92.60

Analyse statistique temporelle de l'irradiation solaire globale quotidienne : modélisation d'une variable réduite à l'aide de modèles stochastiques A.R.M.A.

B. Guerrier, E. Boileau et C. Benard

Laboratoire des Signaux et Systèmes, C.N.R.S./E.S.E., Plateau du Moulon, 91190 Gif sur Yvette, France

(*Reçu le 20 juillet 1979, accepté le 5 octobre 1979*)

Résumé. — A partir de l'étude statistique des données météorologiques de trois sites climatiques différents, équatorial en altitude (Huallao), tempéré (Trappes) et méditerranéen (Carpentras), nous étudions les corrélations temporelles et la modélisation des fluctuations de la variable irradiation solaire globale quotidienne en utilisant une variable réduite et des modèles stochastiques linéaires.

Abstract. — From the statistical study of the meteorological data of three different types of climate, equatorial in altitude (Huallalo), temperate (Trappes) and mediterranean (Carpentras), we use a reduced variable and linear stochastic models to study the time correlations and the modelisation of the daily solar total irradiation fluctuations.

Notations :

$I(d, y)$: Irradiation solaire globale intégrée sur la journée pour le jour d de l'année y.

$\overline{I(d, y)_1}$: Moyenne locale décentrée.

$\overline{I(d, y)_2}$: Moyenne locale centrée.

$\overline{\overline{I(d)}}$: Moyenne sur les années de données de la moyenne locale centrée.

NA : Nombre total d'années de données ($NA' = NA - 2$).

α : Indice de la quinzaine ($1 \leqslant \alpha \leqslant 24$).

β : Indice du mois ($1 \leqslant \beta \leqslant 12$).

$\sigma_1(d, y)$: Ecart-type local décentré.

$\overline{\overline{\sigma}}(d)$: Ecart-type centré moyenné sur les années de données.

$z(d, y)$: Variable définie par : $\dfrac{I(d, y) - \overline{I(d, y)_1}}{\sigma_1(d, y)}$.

$v(d, y)$: Variable définie par : $\dfrac{I(d, y) - \overline{\overline{I(d)}}}{\overline{\overline{\sigma(d)}}}$.

$\rho_K(d, y)$: Coefficient d'autocorrélation d'ordre k.

$\overline{\rho_K}(\beta, y)$: Coefficient d'autocorrélation moyen d'ordre k pour le mois β.

$\overline{\overline{\rho}}_K(\beta)$: Moyenne sur NA' années de données de $\overline{\rho}_K(\beta, y)$.

$\varphi_{KK}(d, y)$

$\overline{\varphi}_{KK}(\beta, y)$, $\overline{\overline{\varphi}}_{KK}(\beta)$: Idem pour les coefficients d'autocorrélation partielle.

$P(\alpha, y)$, $q(\alpha, y)$: Paramètres caractéristiques de la quinzaine α pour l'année y.

$q(\alpha)$, $P(\alpha)$: Moyenne sur NA' années des paramètres $P(\alpha, y)$ ou $q(\alpha, y)$.

$I'(d, y)$: Prédiction de l'irradiation solaire globale quotidienne pour le jour d de l'année y.

Introduction. — Un appareil utilisant l'énergie solaire est un système dont les entrées sont des fonctions aléatoires du temps. Les effets de mémoire et corrélations qui peuvent être mis en évidence au niveau de ces entrées (irradiation solaire, température extérieure, vent ...) se répercutent sur le fonctionnement du système au cours du temps et sur l'allure des sorties. En particulier, les caractéristiques climatiques d'un lieu donné interviennent dans l'optimisation du dimensionnement et de la gestion d'un système.

Une première étude s'intéressant aux probabilités d'avoir des séquences plus ou moinx longues *d'irradiation stable* et utilisant un découpage discret des valeurs possibles prises par la fonction aléatoire étudiée a déjà été réalisée, réf. [1].

Article published online by EDP Sciences and available at http://dx.doi.org/10.1051/rphysap:0198000150209300

Dans cette étude, nous nous intéressons à la modélisation de l'irradiation solaire globale quotidienne en conservant le caractère continu de cette variable : à partir des données météorologiques de trois sites climatiques différents, équatorial en altitude (Huallao), tempéré (Trappes) et méditérranéen (Carpentras), et après étude des corrélations temporelles, nous utilisons une variable réduite et des modèles stochastiques linéaires pour représenter les fluctuations de l'irradiation solaire au cours du temps.

1. **Caractéristiques des modèles utilisés.** — Les modèles stochastiques que nous avons utilisés pour représenter les fluctuations de l'irradiation solaire globale quotidienne sont ceux présentés par Box et Jenkins, réf. [2].

Les modèles A.R.M.A., qui ont pour but de décrire le comportement statistique de processus stationnaires, sont linéaires : z_j, variable centrée modélisée, s'écrit comme somme d'une composante aléatoire a_j et de termes dépendant des valeurs prises par la variable dans le passé, sous la forme générale suivante :

$$\text{(a)} \quad z_j = a_j - q_1 a_{j-1} - q_2 a_{j-2} - \cdots + P_1 z_{j-1} + P_2 z_{j-2} + \cdots$$

Nous utiliserons dans la suite de l'exposé les modèles particuliers s'écrivant sous la forme suivante :

- Modèle à moyenne mobile d'ordre q (noté MAq)

$$z_j = a_j - q_1 a_{j-1} - q_2 a_{j-2} - \cdots - q_q a_{j-q} .$$

- Modèle autorégressif d'ordre p (noté ARp)

$$z_j = a_j + P_1 z_{j-1} + P_2 z_{j-2} + \cdots + P_p z_{j-p} .$$

- Modèle mixte d'ordre pq (noté Mixte pq)

$$z_j = a_j + \sum_{k=1}^{p} P_k z_{j-k} - \sum_{l=1}^{q} q_l a_{j-l} .$$

Remarque. — Les modèles stochastiques proposés par Box et Jenkins sont adaptés à la représentation de variables ayant une densité de probabilité paire : en effet, le bruit blanc a_j est supposé gaussien (notamment dans l'évaluation des paramètres P_k et q_l) et l'expression (a) peut s'écrire sous la forme :

$$z_t = a_t + \sum_{u=1}^{\infty} \psi_u a_{t-u}$$

ce qui implique que z_t soit également à densité de probabilité symétrique.

2. **Transformation préliminaire de la variable donnée : variable réduite.** — Nous disposons pour les sites de Huallao et de Trappes de 11 ans de données météorologiques de l'irradiation solaire globale quotidienne et de 9 ans pour le site de Carpentras.

Ces données se caractérisent par :

— un niveau moyen traduisant l'évolution saisonnière et variant lentement,

— des variations rapides d'un jour à l'autre dont l'amplitude varie avec l'époque de l'année (voir Fig. 1).

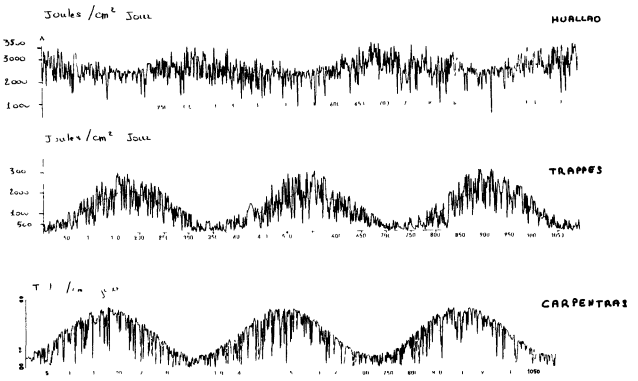


Fig. 1. — Irradiations solaires globales quotidiennes relevées sur les sites de Huallao, Trappes et Carpentras.

A partir de ces données, plusieurs transformations sont envisageables pour construire une variable adaptée aux modèles exposés ci-dessus : Un premier type de transformations utilisant entre autres des processus de type ARIMA (voir réf. [2], ch. 4), a été étudié : on se reportera aux articles réf. [3] et réf. [4].

Dans le travail présenté ici, pour stationnariser la variable en tenant compte des 2 phénomènes climatiques que nous venons d'énoncer, nous avons adopté la transformation suivante :

$$\frac{I(d, y) - \overline{I(d, y)}}{\sigma(d, y)}$$

où $\overline{I(d, y)}$ et $\sigma(d, y)$ représentent respectivement *le niveau moyen* autour duquel fluctue la variable et *l'écart-type* des fluctuations à court terme.

Dans la suite, deux définitions seront successivement utilisées pour ces grandeurs : soit par moyenne et écart-type local propres à l'année considérée, soit par estimation de la moyenne et de l'écart-type du processus. Cependant, ces transformations ne stationnarisent à priori pas complètement le processus. En effet, il peut subsister des différences d'effet de mémoire d'une période de l'année à l'autre, ce qui nous amène à envisager dans la suite une étude sur un laps de temps où le processus peut être considéré comme stationnaire.

3. **Etude statistique. Emploi d'expressions locales.** — 3.1 VARIABLE UTILISÉE. — 3.1.1 *Expression de la valeur moyenne.* — Dans un premier temps nous avons adopté pour centrer la variable une moyenne locale, ceci permettant d'envisager l'établissement du modèle avec un petit nombre de données ($\leqslant$ 1 an).

Soit :

$$\overline{I(d, y)_1} = \frac{1}{M} \sum_{j=d-(M-1)}^{d} I(j, y) \qquad (1)$$

(l'indice 1 se rapportera à une moyenne locale et décentrée).

Le décentrage de cette moyenne locale, lissage sur le jour d et les $(M - 1)$ jours précédant d, est nécessaire pour ne pas introduire de valeurs inconnues (postérieures au jour (d, y) considéré) au niveau de la prédiction.

a) Etude de M. — La série de moyennes ainsi formée doit conserver l'allure générale des variations saisonnières. L'optimisation de M jouera donc sur le compromis entre une valeur assez importante de M pour ne pas donner trop d'importance aux quelques jours qui précèdent et éviter des sauts rapides d'un jour à l'autre, et une valeur pas trop importante pour limiter l'influence du décentrage et ne pas aplatir les variations saisonnières par un trop long lissage.

Nous avons comparé pour différentes valeurs de M les expressions suivantes :

$$A_1(\alpha, M) = \left[\frac{1}{15}\frac{1}{NA'}\sum_y \sum_{d \in \alpha} (\overline{I(d, y)_1} - \overline{\overline{I(d)}})^2\right]^{1/2}$$

$$A_2(\alpha, M) = \left[\frac{1}{15}\frac{1}{NA'}\sum_y \sum_{d \in \alpha} (\overline{I(d, y)_2} - \overline{\overline{I(d)}})^2\right]^{1/2}$$

où :

- $NA' = NA - 2$ avec NA nombre total d'années de données. (L'élimination des années extrêmes dans la sommation est due à des simplifications de programmation.)
- α indice caractéristique de la quinzaine. Le découpage en quinzaines correspond à un découpage de l'année en 24 périodes de 15 jours.
- $\overline{I(d, y)_1}$ est définie par (1).
- $\overline{I(d, y)_2}$ représente la moyenne locale centrée sur M jours (M impair)

$$\overline{I(d, y)_2} = \frac{1}{M} \sum_{d-\frac{M-1}{2}}^{d+\frac{M-1}{2}} I(j, y) . \qquad (2)$$

- $\overline{\overline{I(d)}}$ représente la moyenne sur NA années de données de la moyenne locale centrée

$$\overline{\overline{I(d)}} = \frac{1}{NA} \sum_{y=1}^{NA} \overline{I(d, y)_2} . \qquad (3)$$

A la limite où $M = 1$, ces grandeurs sont des estimateurs de l'écart quadratique des fluctuations à court terme de I. A mesure que M augmente, les fluctuations rapides sont lissées et A_1 et A_2 décroissent, rapidement dans un premier temps, plus lentement ensuite : en effet, à partir d'un certain ordre de grandeur pour M, A_1 et A_2 ne sont plus que les estimateurs de l'écart quadratique des fluctuations lentes (moyenne sur un nombre croissant, M, de jours).

Mais pour ces grandes valeurs de M, A_1 est de plus en plus erroné en tant qu'estimateur de l'écart quadratique à cause du décentrage, c'est-à-dire que $\overline{I(d, y)_1}$ est de plus en plus erroné comme estimateur de la valeur moyenne locale de I. Nous avons retenu la valeur $M = 15$.

Il est à noter que la tendance à la décroissance de A_1 avec M s'inverse pour certaines quinzaines : voir pour exemples les figures 2 et 3 où les expressions A_1 et A_2 sont représentées pour le site de Trappes et pour différentes valeurs de M.

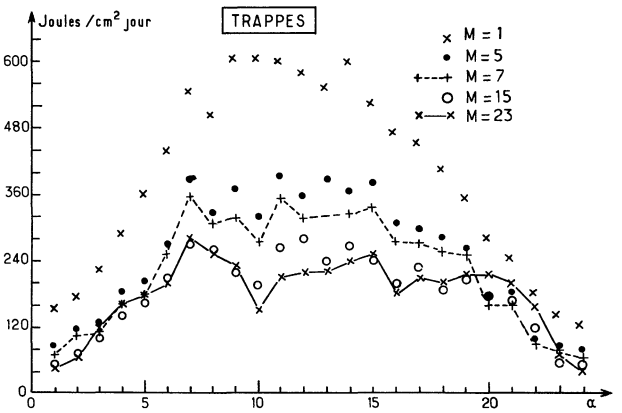


Fig. 2. — Site de Trappes : expression $A_1(\alpha, M)$ pour différentes valeurs de M.

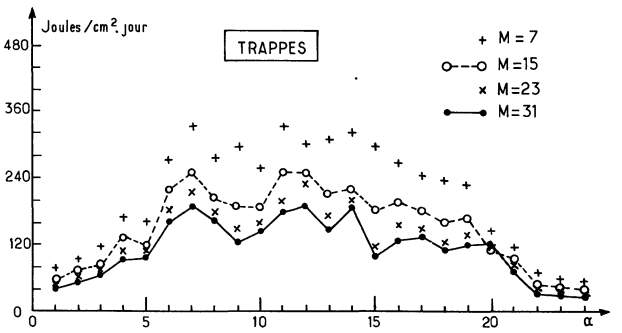


Fig. 3. — Site de Trappes : expression $A_2(\alpha, M)$ pour différentes valeurs de M.

b) Etude du décentrage. — Pour évaluer l'influence du décentrage, nous avons comparé l'écart aux moyennes locales centrées et décentrées à l'écart quadratique, sur des périodes de 15 jours ; soit respectivement les trois expressions suivantes :

$$B_1(\alpha) = \left[\frac{1}{15}\frac{1}{NA'}\sum_y \sum_{d \in \alpha} (I(d, y) - \overline{I(d, y)_2})^2\right]^{1/2}$$

$$B_2(\alpha) = \left[\frac{1}{15}\frac{1}{NA'}\sum_y \sum_{d \in \alpha} (I(d, y) - \overline{I(d, y)_1})^2\right]^{1/2}$$

$$B_3(\alpha) = \left[\frac{1}{15}\frac{1}{NA'}\sum_y \sum_{d \in \alpha} (I(d, y) - \overline{\overline{I(d)}})^2\right]^{1/2} .$$

L'écart quadratique a été calculé en estimant la moyenne théorique du processus par $\overline{\overline{I(d)}}$, moyenne sur les années de données de la moyenne locale (sur 15 jours) centrée : voir (3).

L'incertitude sur l'écart quadratique peut être appréciée à l'aide de l'écart quadratique de cet estimateur (B 3) considéré comme variable aléatoire. (voir réf. [5]). Les corrélations entre échantillons ont été négligées dans le calcul de l'incertitude : voir réf. [1].

Pour chaque quinzaine, l'incertitude a été calculée d'abord avec l'hypothèse d'une approximation gaussienne pour $I(d, y)$ puis dans le cas d'une évaluation empirique, et nous avons conservé la plus grande de ces deux estimations.

Les trois expressions définies ci-dessus ont été représentées sur les figures 4, 5 et 6 pour les sites (respectivement) de Huallao, Trappes et Carpentras : l'influence du décentrage est un peu moins sensible pour Huallao, l'irradiation moyenne variant moins d'une époque de l'année à l'autre pour ce type de climat.

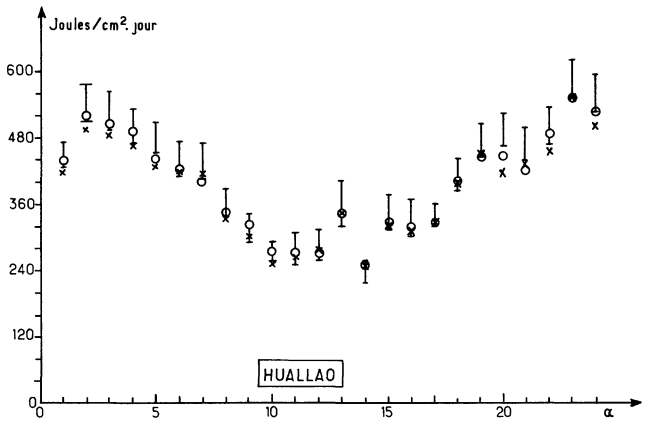


Fig. 4. — Site de Huallao : expressions $B_1(\alpha)$, $B_2(\alpha)$, $B_3(\alpha)$. I : incertitude sur l'écart quadratique ; X : écart/moyenne locale centrée ; 0 : écart/moyenne locale décentrée.

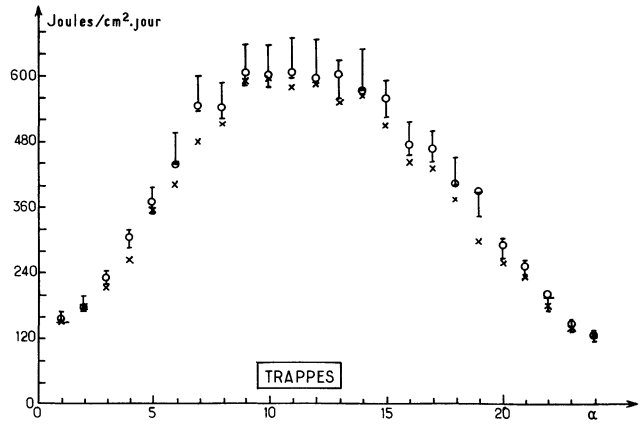
3.1.2 *Expression de l'écart quadratique local.* — Nous avons adopté, de même que pour la moyenne et pour les mêmes raisons, un *écart quadratique* local et décentré d'expression :

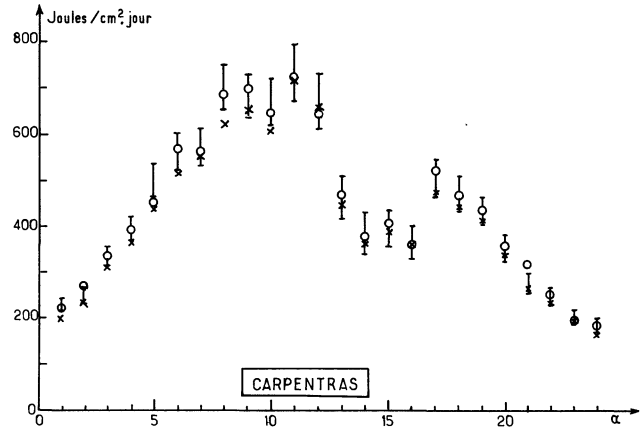
$$\sigma_1(d, y) = \left[\frac{1}{14} \sum_{j=d-14}^{d} (I(j, y) - \overline{I(j, y)_1})^2 \right]^{1/2} \quad (4)$$

(le coefficient est $\frac{1}{14}$ et non 1/15 de façon à avoir un estimateur sans biais).

La variable étudiée statistiquement dans la suite sera donc $z(d, y)$ donnée par :

$$z(d, y) = \frac{I(d, y) - \overline{I(d, y)_1}}{\sigma_1(d, y)} .$$





Figs. 5 et 6. — Idem pour Trappes et Carpentras.

3.2 Etudes des corrélations. — Les coefficients d'autocorrélation et d'autocorrélation partielle ont été estimés en utilisant les expressions proposées par Box et Jenkins (réf. [2]).

L'étude simultanée de ces deux séries de coefficients permet de choisir le modèle le mieux adapté, s'il existe, à la série temporelle étudiée et l'ordre de ce modèle (réf. [2]).

Pour chaque jour d, nous avons calculé le coefficient d'autocorrélation d'ordre k, $\rho_K(d, y)$, (avec $\rho_0(d, y) = 1$) sur N jours précédents, puis un coefficient moyen pour le mois :

$$\overline{\rho}_K(\beta, y) = \frac{1}{30} \sum_{d \in \beta} \rho_K(d, y)$$

où β est l'indice caractéristique du mois. (Le découpage en mois correspond à un découpage de l'année en douze périodes de 30 jours.)

Ce coefficient est ensuite moyenné sur les années :

$$\overline{\overline{\rho}}_K(\beta) = \frac{1}{NA'} \sum_{y} \overline{\rho}_K(\beta, y) .$$

De même $\overline{\varphi}_{KK}(\beta, y)$ représente le coefficient d'autocorrélation partielle moyen pour le mois β de l'année y,

et $\overline{\overline{\varphi}}_{KK}(\beta)$ le coefficient d'autocorrélation partielle moyenné sur les années pour le mois β.

L'étude mois par mois a pour but d'évaluer l'évolution des corrélations au cours des saisons. (Problème de l'éventuelle non stationnarité de la variable réduite étudiée.)

Les coefficients ont été calculés jusqu'à l'ordre 10 ($1 \leqslant K \leqslant 10$). Les calculs ont été faits sur des périodes de $N = 30, 60, 90$ ou 120 jours pour le site de Trappes : les résultats sont similaires, les variations saisonnières s'estompant un peu quand N augmente. N n'a pas été pris trop faible, l'erreur étant inversement proportionnelle à sa racine.

Sur la figure 7 sont représentés les coefficients du 1er ordre (moyennés sur les années et pour $N = 90$) pour les 3 sites : leurs valeurs peu élevées ($\leqslant 0{,}4$ pour Trappes et Carpentras et $\leqslant 0{,}3$ pour Huallao) indiquent des corrélations assez faibles et donc des effets de mémoire peu importants.

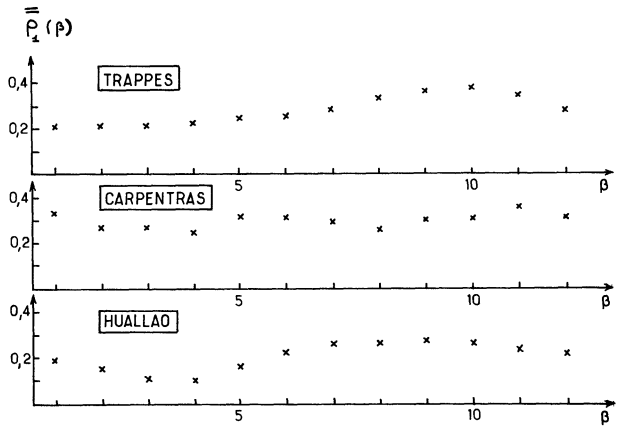


Fig. 7. — Coefficient d'autocorrélation moyen du premier ordre $\overline{\overline{\rho}}_1$ ($N = 90$).

Les coefficients d'ordre supérieur sont plus faibles :

Huallao

$|\overline{\overline{\rho}}_K| \leqslant 0{,}12$ et $|\overline{\overline{\varphi}}_{KK}| \leqslant 0{,}11$ $\quad 1 < K \leqslant 10$

Trappes

$|\overline{\overline{\rho}}_K| \leqslant 0{,}16$ et $|\overline{\overline{\varphi}}_{KK}| \leqslant 0{,}11$ $\quad 1 < K \leqslant 10$

Carpentras

$|\overline{\overline{\rho}}_K| \leqslant 0{,}21$ et $|\overline{\overline{\varphi}}_{KK}| \leqslant 0{,}14$ $\quad 1 < K \leqslant 10$.

Comme l'indique les différences entre $\overline{\rho}(\beta, y)$ et $\overline{\overline{\rho}}(\beta)$ ou entre $\overline{\varphi}(\beta, y)$ et $\overline{\overline{\varphi}}(\beta)$, certaines années ont, pendant certaines périodes, un comportement statistique différent de la moyenne des autres. (Voir par exemple le tableau I qui donne $\overline{\rho}_1(\beta, y)$ pour le site de Trappes.)

Ces remarques font apparaître que le modèle le mieux adapté à la série temporelle $z(d, y)$ ne s'impose pas de manière très tranchée ; de par leurs faibles valeurs les fonctions sont en effet souvent dans le domaine d'erreur. Leurs allures se rapprochent cependant plutôt de celle d'un modèle de type AR ou mixte suivant les époques de l'année. Pour comparaison, les trois types de modèle suivants seront envisagés :

AR 1

$$z(d, y) = Pz(d - 1, y) + a(d, y).$$

Mixte (1, 1)

$$z(d, y) = Pz(d - 1, y) - qa(d - 1, y) + a(d, y).$$

MA 1

$$z(d, y) = -qa(d - 1, y) + a(d, y).$$

3.3 Détermination des paramètres. — Pour tenir compte de l'évolution du comportement statistique de l'irradiation (non éliminée par la transformation $z(d, y)$), nous avons considéré des paramètres variables suivant l'époque de l'année et nous avons défini pour chaque quinzaine α un paramètre $P(\alpha, y)$ où $P(\alpha, y)$ est la moyenne des paramètres $P(d, y)$ calculés pour le second, sixième, dixième et

Tableau I. — *Trappes* $\overline{\rho}_1(\beta, y)$.

βy	2	3	4	5	6	7	8	9	10	$\overline{\overline{\rho}}(\beta)$
1	0,37	0,26	0,23	0,37	0,12	0,13	0,09	0,13	0,20	0,21
2	0,34	0,39	0,19	0,31	0,25	0,17	− 0,05	0,15	0,13	0,21
3	0,40	0,41	0,25	0,28	0,16	0,14	0,01	0,11	0,10	0,21
4	0,42	0,42	0,33	0,21	0,16	0,08	0,21	0,00	0,18	0,22
5	0,39	0,37	0,35	0,25	0,03	0,05	0,43	0,00	0,28	0,24
6	0,24	0,39	0,39	0,31	0,06	0,06	0,35	0,08	0,39	0,25
7	0,22	0,31	0,36	0,40	0,06	0,19	0,34	0,28	0,36	0,28
8	0,37	0,37	0,36	0,34	0,12	0,36	0,34	0,35	0,37	0,33
9	0,37	0,45	0,30	0,34	0,16	0,48	0,43	0,35	0,42	0,36
10	0,26	0,52	0,29	0,45	0,19	0,41	0,39	0,38	0,48	0,37
11	0,13	0,46	0,37	0,40	0,17	0,34	0,34	0,40	0,43	0,34
12	0,08	0,29	0,38	0,35	0,16	0,22	0,22	0,39	0,31	0,27

quatorzième jour de la quinzaine : La méthode retenue pour obtenir une bonne approximation des paramètres $P(d, y)$ consiste à minimiser par rapport aux $P(d, y)$ la somme des $[a(d', y)]^2$ où $[a(d', y)]$ est l'estimé de $a(d', y)$ conditionnellement aux valeurs des paramètres et de la série $z(d', y)$ connue sur une période de T jours.

Le choix de T est déterminé par la nécessité d'une précision suffisante (T assez grand pour que la moyenne n'oscille pas) d'une part et par celle de ne pas trop lisser les effets saisonniers (T suffisamment faible pour que $P(d, y)$ ne varie pas trop sur T jours).

3.3.1 *Processus* AR 1. — • La figure 8 représente les paramètres $P(\alpha)$ moyennés sur NA' années. Les paramètres $P(d, y)$ ont été calculés en minimisant la somme des $[a]^2$ sur les 90 jours précédant le jour d. (Seuls les jours antérieurs sont considérés de façon à déterminer le paramètre sur des données connues (en vue de la prédiction).)

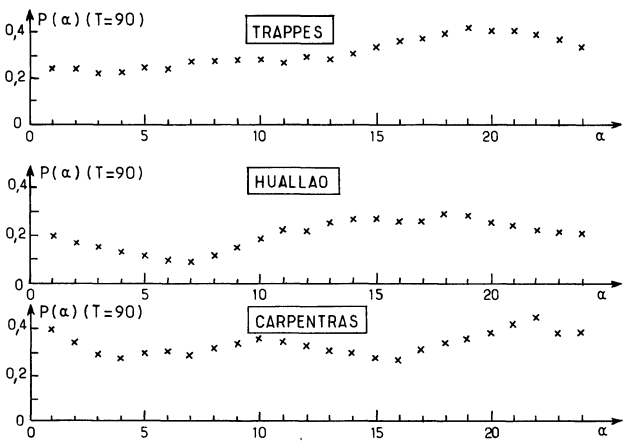


Fig. 8. — Paramètres moyennés sur les années dans le cas d'un AR 1 (T = 90 jours).

Le paramètre est positif et varie entre :

0,1 et 0,3 Huallao,
0,2 et 0,45 Trappes,
0,25 et 0,45 Carpentras.

L'effet de mémoire indiqué par ce type de modèle n'est donc pas très important, ce qui découle des résultats obtenus sur les corrélations.

• Pour certaines périodes et certaines années, le paramètre $P(\alpha, y)$ est assez différent du paramètre moyen $P(\alpha)$. Le comportement statistique n'est donc pas toujours similaire d'une année à l'autre. (Les paramètres très faibles correspondent aux périodes où les coefficients d'autocorrélation sont également faibles.)

• La variation entre deux périodes successives est généralement faible.

• Pour le site de Trappes, les paramètres ont été calculés en effectuant la minimisation pour T = 30 jours également (problème du choix de T). Les résultats moyennés sur les années ont été portés sur la figure 9 : les variations au cours de l'année (phénomènes saisonniers) sont plus importantes (lissage plus faible), mais les variations entre deux périodes successives le sont également. (Précision moins bonne.)

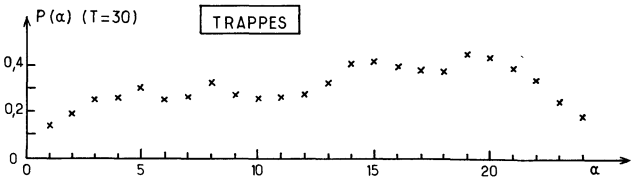


Fig. 9. — Idem pour le site de Trappes, T = 30 jours.

3.3.2 *Processus* MA 1. — Les paramètres associés au processus MA 1 sont légèrement plus faibles en valeur absolue que ceux associés à un AR 1 :

Huallao $0{,}1 \leqslant |q| \leqslant 0{,}3$.
Trappes $0{,}2 \leqslant |q| \leqslant 0{,}38$.
Carpentras $0{,}2 \leqslant |q| \leqslant 0{,}36$.

3.3.3 *Processus mixte* 1-1. — Sur les figures 10, 11 et 12 sont représentés les paramètres $P(\alpha)$ et $q(\alpha)$ moyennés sur NA' années (T = 90 jours). Le paramètre $P(\alpha)$ (caractère autorégressif du processus) est plus important que le paramètre $q(\alpha)$ (caractère moyenne mobile du processus).

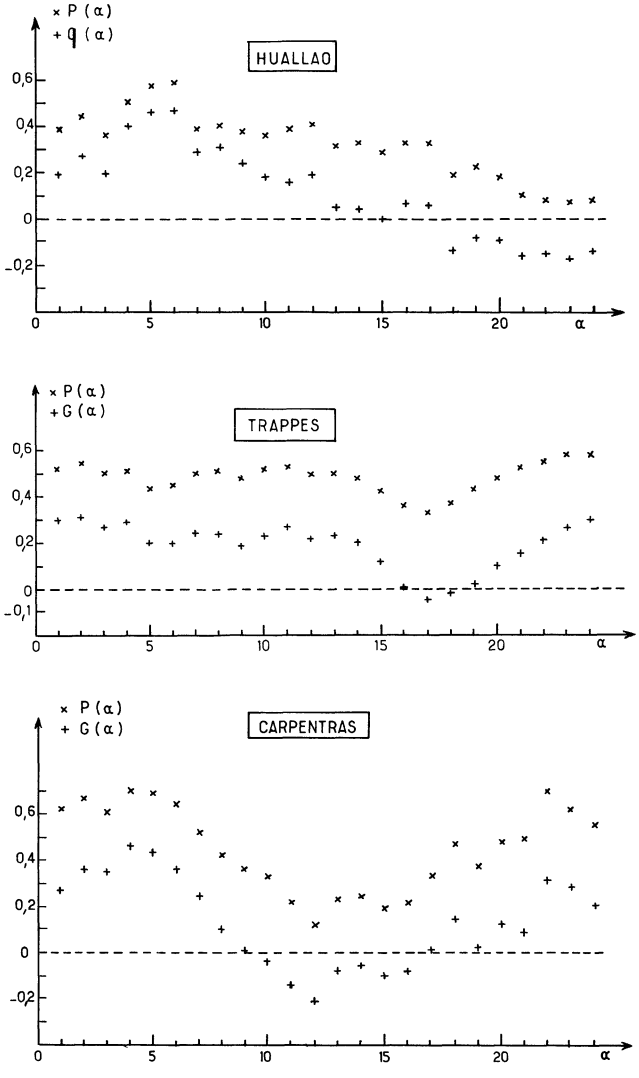
Leur différence, $P(\alpha) - q(\alpha)$, présente des fluctuations et des valeurs du même ordre que celles du paramètre des processus AR 1 ou MA 1 (en valeur absolue). Or cette différence est le coefficient du terme en $z(d-1, y)$ ou $a(d-1, y)$ suivant que l'on développe l'expression $z(d, y)$ par rapport aux z ou aux a, soit respectivement :

$$z(d, y) = a(d, y) + (P - q)\, z(d-1, y) + \\ + q(P - q)\, z(d-2, y) + \cdots$$

$$z(d, y) = a(d, y) + (P - q)\, a(d-1, y) + \\ + P(P - q)\, a(d-2, y) + \cdots$$

Les paramètres $P(\alpha)$ et $q(\alpha)$ du processus mixte, intervenant en eux-mêmes dans les coefficients des termes en $z(d-2, y)$ ou $a(d-2, y)$ fluctuent beaucoup plus suivant la quinzaine que leur différence ou les paramètres des processus AR 1 et MA 1. L'emploi de processus mixte à deux paramètres ne semble donc pas apporter un intérêt supplémentaire par rapport aux processus AR 1 et MA 1.

Les valeurs des paramètres confirment le relativement faible effet de mémoire et donc l'importance de la composante imprévisible $a(d, y)$ dans l'expression du processus. (Dans le cas du processus AR 1 par exemple, l'écart quadratique du bruit blanc est égal à $(1 - p^2)$ fois l'écart quadratique de la variable.)



Figs. 10, 11 et 12. — Paramètres moyennés sur les années dans le cas d'un Mixte 11 (T = 90 jours) pour les sites respectivement de Huallao, Trappes et Carpentras.

3.4 PRÉDICTION. — Dans ces conditions, l'emploi de ce type de modèles ne peut conduire à une prédiction précise. Cependant, pour tester la validité de ces différents modèles et en comparer le comportement, nous avons comparé les prédictions sur un jour (les modèles étant du premier ordre) obtenues à partir de ces modèles à des prédictions plus simples.

Les expressions (5) et (6) suivantes de l'erreur de prédiction à un jour ont été utilisées :

• Etude de prédiction pour un jour quelconque de la quinzaine α de l'année y :

$$\left[\frac{1}{15}\sum_{d\in\alpha}(I'(d,y)-I(d,y))^2\right]^{1/2}. \qquad (5)$$

• Etude de prédiction pour un jour quelconque de la quinzaine α d'une année quelconque :

$$\left[\frac{1}{15}\frac{1}{NA'}\sum_{y}\sum_{d\in\alpha}(I'(d,y)-I(d,y))^2\right]^{1/2} \qquad (6)$$

où $I'(d, y)$ représente la prédiction de l'irradiation globale quotidienne pour le jour d de l'année y et $I(d, y)$ la valeur réellement observée.

Trois possibilités ont été envisagées pour l'évaluation de $I'(d, y)$:

a) $I'(d, y) = I(d-1, y)$: prolongation de la valeur du jour précédent.

b) $I'(d, y) = \overline{I(d-1, y)}_1$: prolongation de la moyenne locale décentrée.

c) $I'(d, y)$ = prédiction obtenue à l'aide des modèles stochastiques.

Pour déterminer $I'(d, y)$ dans le cas *c*), on évalue dans un premier temps $z'(d, y)$ en posant $a(d, y)$, composante aléatoire dans l'expression de $z(d, y)$, nulle. (Ce qui correspond à sa valeur la plus probable compte tenu de l'hypothèse de distribution gaussienne de ces variables aléatoires.)

Nous avons alors :

AR 1 $\quad z'(d, y) = Pz(d-1, y)$

MA 1 $\quad z'(d, y) = -q[a(d-1, y)]$

Mixte $\quad z'(d, y) = Pz(d-1, y) - q[a(d-1, y)]$

où $[a(d-1, y)]$ est estimé en considérant la période des 90 jours précédents d. $I'(d, y)$ est ensuite calculé en posant :

$$I'(d, y) = z'(d, y)\,\sigma_1(d-1, y) + \overline{I(d-1, y)}_1 .$$

Paramètres locaux. — Dans le cas où l'on dispose de peu de données, variables et paramètres doivent être calculés localement : les paramètres utilisés pour la prédiction de la variable $z(d, y)$ pour chaque jour d'une quinzaine α sont ceux calculés pour la quinzaine $\alpha - 1$ et utilisent donc les 90 jours précédent cette quinzaine (voir calcul des paramètres). Pour les trois sites étudiés, des modèles de type AR 1, MA 1 et Mixte ont été testés : En moyenne (voir 3.4, expression (6)), les résultats obtenus avec l'un ou l'autre de ces modèles sont similaires : voir les tableaux II, III et IV où les erreurs de prédiction sont comparées pour les différents modèles.

Les figures 13, 14 et 15 représentent pour les différentes quinzaines les expressions (6) pour les trois prédictions (*a*), *b*) et *c*) avec un modèle AR 1) et la zone d'incertitude de l'estimé $B_3(\alpha)$ de l'écart quadratique : La prédiction obtenue à l'aide des modèles stochastiques est nettement meilleure que celle obtenue par prolongation de la valeur de la veille et légèrement meilleure que la prolongation de la moyenne sur les 15 jours précédents : L'étude statistique locale des corrélations apporte donc un supplément d'information.

L'erreur de prédiction (expression (6)) est de l'ordre de l'écart quadratique $B_3(\alpha)$.

Tableau II. — *Huallao* : *Erreur de prédiction* joules/cm^2-*jour*. *Variable* $z(d, y)$.

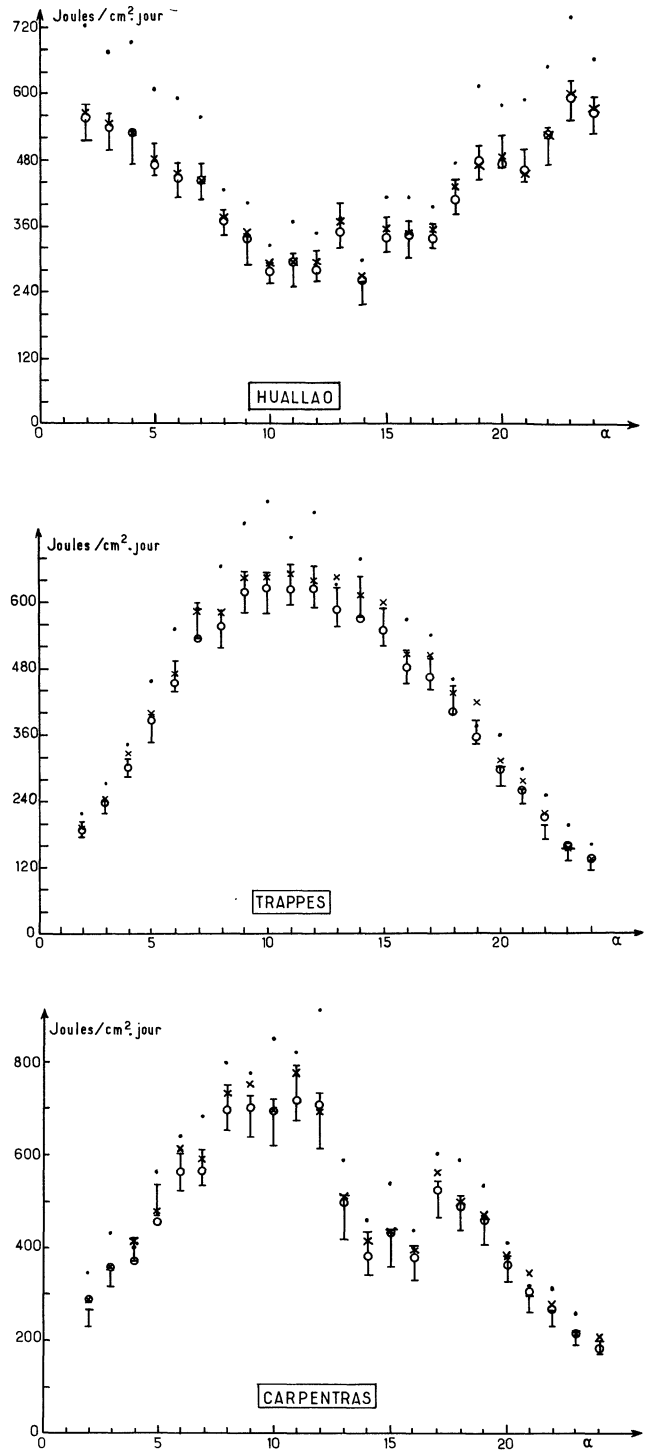
	Paramètres locaux			Para. moyen.	Para. uniq.
	MA 1	Mixte	AR 1	AR 1	AR 1
2	557	556	556	559	560
4	531	528	528	523	526
6	441	442	442	448	450
8	355	366	357	360	356
10	275	275	272	274	273
12	278	279	280	284	284
14	253	257	252	249	252
16	334	338	337	334	334
18	404	406	405	403	407
20	470	468	468	468	468
22	522	523	518	513	513
24	557	563	549	545	547

Tableau III. — *Trappes* : *Erreur de prédiction* joules/cm^2-*jour*. *Variable* $z(d, y)$.

	Paramètres locaux			Para. moyen.	Para. uniq.
	MA 1	Mixte	AR 1	AR 1	AR 1
2	184	183	184	183	183
4	305	297	300	305	301
6	454	459	453	452	452
8	561	560	557	556	555
10	632	631	627	630	630
12	619	633	624	617	617
14	579	576	571	574	574
16	483	484	480	478	478
18	411	404	402	397	400
20	301	293	295	295	294
22	215	208	208	206	206
24	132	134	133	131	131

Tableau IV. — *Carpentras* : *Erreur de prédiction* joules/cm^2-*jour*. *Variable* $z(d, y)$.

	Paramètres locaux			Para. moyen.	Para. uniq.
	MA 1	Mixte	AR 1	AR 1	AR 1
2	293	281	285	281	281
4	383	366	368	373	366
6	573	566	561	562	559
8	696	692	698	684	684
10	691	684	692	683	682
12	721	722	707	700	701
14	379	380	379	380	380
16	379	377	375	371	370
18	484	484	483	480	481
20	368	358	358	351	352
22	266	267	265	261	260
24	186	181	182	182	183



Figs. 13, 14 et 15. — Comparaison des 3 types de prédiction pour les sites respectivement de Huallao, Trappes et Carpentras. Variables $z(d, y)$. Expressions (6) avec :

- • $I'(d, y) = \underline{I(d-1, y)}$.
- X $I'(d, y) = \overline{I(d-1, y)}_1$.
- 0 $I'(d, y)$ = prédiction obtenu avec un modèle AR 1 avec paramètres locaux.
- I incertitude sur l'écart quadratique.

4. Etude statistique : emploi d'expressions moyennées sur plusieurs années de données. — 4.1 VARIABLE $z(d, y)$. PARAMÈTRES MOYENNÉS. — Dans le cas où l'on dispose d'un certain nombre d'années de données,

l'emploi de paramètres moyennés sur les années peut être envisagé : pour le processus AR 1, les expressions de type (6) ont été calculées pour les trois sites en utilisant soit des paramètres moyennés et saisonniers $P(\alpha)$, soit un paramètre unique P :

Huallao $P = 0{,}21$,
Trappes $P = 0{,}3$,
Carpentras $P = 0{,}34$.

Les résultats au niveau de la prédiction sont du même ordre, parfois légèrement meilleurs que dans le cas de paramètres locaux (voir Tabl. II, III et IV).

4.2 Cas d'une variable utilisant des expressions moyennées. — Dans le cas où l'on dispose de plusieurs années de données, nous pouvons envisager d'utiliser, outre des paramètres moyennés, une expression moyennée sur les années pour l'écart quadratique et la valeur moyenne dans la définition de la variable réduite soit :

$$v(d, y) = \frac{I(d, y) - \overline{\overline{I(d)}}}{\overline{\overline{\sigma}}(d)}$$

où $\overline{\overline{\sigma(d)}}$ est défini par :

$$\overline{\overline{\sigma(d)}} = \left[\frac{1}{14}\,\frac{1}{NA}\sum_{y}\sum_{1=d-7}^{d+7}(I(d, y) - \overline{I(d, y)}_2)^2\right]^{1/2}.$$

Cette variable nécessite pour sa construction un grand nombre de données mais a l'avantage d'utiliser des expressions centrées.

Sur les figures 16, 17 et 18 sont comparés les résultats (expressions du type (6)) obtenus pour un modèle AR 1 utilisant soit la variable $z(d, y)$ et des paramètres locaux, soit la variable $v(d, y)$ et des paramètres moyennés. (Dans les calculs l'année sur laquelle se fait la prédiction a été éliminée de l'estimation de la valeur moyenne et de l'écart quadratique du processus de façon à ne pas introduire dans l'expression de

$$I'(d, y) = v'(d, y).\overline{\overline{\sigma}}(d) + \overline{\overline{I}}(d)$$

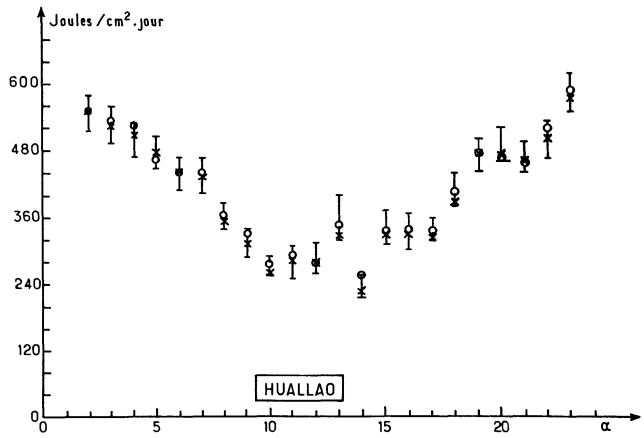
de valeurs supposées inconnues.)

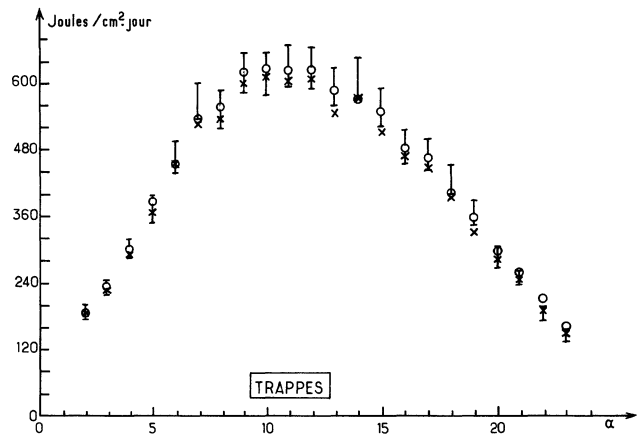
Les résultats sont meilleurs dans le cas de la variable $v(d, y)$, celle-ci évitant notamment l'inconvénient du décentrage.

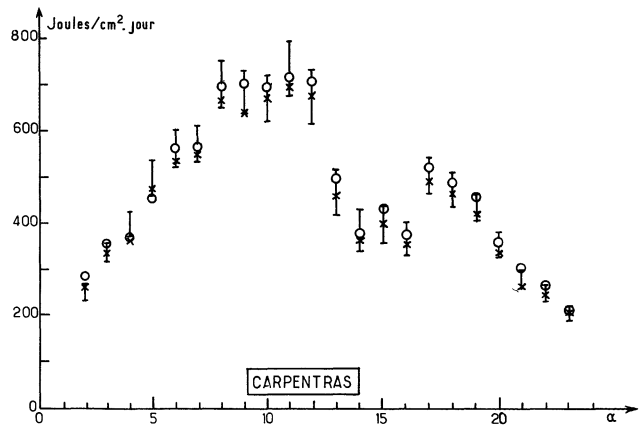
Les écarts obtenus avec $v(d, y)$ sont dans l'ensemble plus faibles que l'écart quadratique, les modèles utilisés apportent donc une information effective sur le comportement de la variable.

4.3 Simulation. — A partir de l'estimation de la moyenne et de l'écart quadratique du processus et des paramètres des modèles décrits précédemment, il est possible de simuler plusieurs années d'irradiation solaire globale quotidienne : pour plus de précisions voir réf. [3] et réf. [4].

Dans le cas de la variable $v(d, y)$ étudiée ici, l'écart quadratique du bruit blanc intervenant dans la simulation est constant, les différences d'amplitude saisonnière des fluctuations étant déjà prises en compte au niveau de la variable. Il n'y a donc pas de problème d'ajustement de l'écart quadratique du bruit blanc comme dans les réfs. [3] et [4]. D'autre part, par sa construction même, cette variable évite les







Figs. 16, 17 et 18. — Comparaison pour la prédiction des variables $z(d, y)$ et $v(d, y)$.

0 $I'(d, y)$ = prédiction obtenu avec la variable $z(d, y)$ et un modèle AR 1 avec paramètres locaux.

X $I'(d, y)$ = prédiction obtenu avec la variable $v(d, y)$ et un modèle AR 1 avec paramètres moyennés.

I incertitude sur l'écart quadratique.

effets de dérive qui peuvent intervenir avec des variables définies par accroissements comme c'est le cas dans les réfs. [3] et [4].

5. **Conclusion.** — A partir de l'étude statistique des données météorologiques de l'irradiation solaire globale quotidienne, nous avons testé des modèles simples (dépendance uniquement sous forme linéaire) à 1 ou 2 paramètres.

Vu le faible ordre de grandeur des paramètres, la comparaison des différents modèles au niveau de la prédiction n'est pas très fine. Les résultats obtenus indiquent cependant que ces modèles apportent un supplément d'information sur le comportement de la variable par rapport à des modèles simplistes.

A cause des faibles effets de mémoire observés, ces modèles peuvent surtout être employés pour la simulation de nouvelles années de données. Les résultats de l'ensemble de notre étude (réfs. [3] et [4]) mettent en évidence qu'une approche plus fine du comportement de l'irradiation nécessite l'intégration d'autres variables météorologiques telles que humidité, pression, vent, température, etc..., corrélées entre-elles : en effet, cet ensemble de variables propage une information plus grande sur leur passé qu'une seule composante telle que l'irradiation et l'augmentation des effets de mémoire correspondante devrait permettre l'introduction de modèles stochastiques multidimensionnels à paramètres plus élevés.

Bibliographie

[1] BENARD, C., BODY, Y., GOBIN, D., WIRGIN, A., *Caractérisation de la stabilité de l'intensité solaire, en France et au Pérou, par l'analyse temporelle de séries aléatoires* (La Météorologie, juin 1978).

[2] BOX, G. E. P. et JENKINS, G. M., *Time series analysis* (Holden day 1970).

[3] BOILEAU, E., *Discussion d'un modèle statistique en météorologie solaire, Revue Phys. Appl.* **14** (1979) 145.

[4] BOILEAU, E., GUERRIER, B., *Comparaison de modèles statistiques saisonniers et non saisonniers en météorologie solaire* (A paraître).

[5] FOURGEAUD, C. et FUCHS, A., *Statistique* (Dunod).