

Une démarche fondée sur les treillis de Galois pour l'aide à la qualification de l'état des milieux aquatiques

Agnès Braud, Corinne Grac, Silvia Pristavu, Emilie Dor, Florence Le Ber

► **To cite this version:**

Agnès Braud, Corinne Grac, Silvia Pristavu, Emilie Dor, Florence Le Ber. Une démarche fondée sur les treillis de Galois pour l'aide à la qualification de l'état des milieux aquatiques. SIDE 2009, 2009, Toulouse, France. pp.95-104. hal-00390697

HAL Id: hal-00390697

<https://hal.archives-ouvertes.fr/hal-00390697>

Submitted on 2 Jun 2009

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.

Une démarche fondée sur les treillis de Galois pour l'aide à la qualification de l'état des milieux aquatiques

A. Braud¹, C. Grac², S. Pristavu¹, E. Dor², F. Le Ber^{2,3}

¹LSIIT UMR 7005, F-67400 Illkirch
agnes.braud@unistra.fr

²LHyGeS UMR 7517 - ENGEES, F-67000 Strasbourg
{corinne.grac, florence.leber}@engees.u-strasbg.fr

³LORIA UMR 7503, F-54500 Vandœuvre-lès-Nancy

Résumé. L'évaluation de la qualité des cours d'eau est un problème majeur et complexe. Différents paramètres apportent des informations sur l'état de l'eau parmi lesquels des indices biologiques. Cinq indices ont été normalisés en France, mais leur interprétation est difficile, et plus particulièrement la corrélation de ces indicateurs. Le travail présenté dans cet article a pour objectif de déterminer des profils de stations (portions de cours d'eau) caractérisés par des valeurs pour ces indices. Nous proposons d'utiliser les treillis de Galois, comme outil d'extraction de connaissances permettant d'accompagner les biologistes dans l'explicitation de leur expertise et de les aider à mettre en évidence l'ensemble des interactions entre les compartiments biologiques. La finalité est d'intégrer ces profils dans un classifieur capable de qualifier une nouvelle station en fonction de ses valeurs pour les indices. La démarche complète est expliquée et les premiers résultats obtenus par ce processus sont détaillés. Ils montrent l'intérêt de cette approche pour l'identification de profils de stations pertinents par les biologistes.

Abstract. The evaluation of water quality is a major concern and a complex task. Several parameters bring information about this quality among which biological indices. Five indices have been normalized in France, but their interpretation is difficult, and more precisely their correlation. The work presented in this article aims at determining profiles for water stations (part of water bodies) characterized by values for those indices. We propose to use Galois lattices as a knowledge extraction tool to help biologists in making explicit their expertise and the interactions between biological compartments. The final goal is to integrate those profiles in a classifier allowing to qualify a new station according to its values for the indices. The whole process is explained and some results obtained on our data are detailed. They show the benefits for the biologists of

using this approach to identify relevant station profiles.

Mots-Clés. Treillis de Galois, concepts formels, règles d'implication, état biologique des masses d'eau, indices biologiques

Keywords. Galois lattices, formal concept, implication rules, biological water quality, biological indices

1 Introduction

L'évaluation de la qualité des cours d'eau est un problème majeur et complexe. La mise en place de mesures de contrôle et de sauvegarde de l'environnement, demandée notamment par la directive-cadre européenne sur l'eau (DCE), nécessite la formalisation de cette évaluation.

Différents paramètres apportent des informations sur l'état des cours d'eau : des paramètres biologiques (indiquant l'état des flore et faune aquatiques), physico-chimiques, physiques (par exemple l'état des berges). En France, la physico-chimie est mesurée très régulièrement sur ces réseaux. Elle peut permettre de détecter une pollution, mais reflète un état ponctuel et se révèle donc insuffisante comme outil d'évaluation. Au contraire, les populations présentes dans le milieu, information capturée par les paramètres biologiques, ne rendent pas compte immédiatement d'une pollution mais sont ensuite affectées durablement. Cinq indices biologiques ont été définis en France. Ils portent chacun sur un groupe taxonomique particulier et se basent sur les taxons de ce groupe présents dans l'eau. Ces groupes sont les macrophytes et les diatomées (flore), les invertébrés, les oligochètes et les poissons (faune). Jusqu'en 2006, sur les réseaux de surveillance nationaux (gestion par le Ministère de l'Ecologie et les Agences de l'Eau), c'est principalement l'indice invertébrés, puis dans une moindre mesure, les indices poissons et diatomées qui étaient utilisés en routine. Depuis 2006, sur les nouveaux réseaux issus de la DCE, l'utilisation de tous les indices biologiques hormis celui des oligochètes se généralise. Les mesures sont effectuées en des lieux précis, de dimensions normalisées et appelés *stations*.

Des données apportant des informations variées sont ainsi disponibles pour évaluer l'état de l'eau au niveau d'une station, mais le problème de leur interprétation demeure, et plus particulièrement la corrélation de l'ensemble des indicateurs. En ce qui concerne les indices biologiques, les experts sont spécialisés dans des groupes taxonomiques particuliers et il leur est donc difficile de proposer une interprétation globale de la qualité de l'eau, intégrant tous les indices. Ainsi croiser et chercher à interpréter l'ensemble des données biologiques, physico-chimiques, physiques, va devenir de plus en plus nécessaire. Cette problématique va être la même quels que soient les pays européens, et ce même si les indices utilisés diffèrent d'un pays à l'autre puisque basés sur des cortèges faunistiques et floristiques différents.

Pour mettre en évidence les influences relatives des indicateurs, les biologistes utilisent classiquement des techniques d'analyse de données telles que l'Analyse

des Correspondances Multiples (ACM) [4]. L'ACM vise à simplifier un tableau de données mettant en jeu plusieurs variables. Pour cela, de nouvelles dimensions sont calculées sur la base des variables initiales. Chaque dimension se voit associer un pourcentage indiquant la quantité d'informations qu'elle capture. L'intérêt de cette approche est de créer des dimensions capturant la plus grande quantité d'information possible. On étudie alors les données selon ces axes et notamment les plus porteurs d'informations. Pour cela on peut les représenter en considérant les dimensions deux deux, chaque couple de dimensions permettant de visualiser les données dans un plan. Ces méthodes reposent donc sur des projections où seulement une partie de l'information est représentée (la somme des quantités portées par les deux axes), ceci limitant les possibilités d'une interprétation globale.

Lafont [8] a proposé une autre approche, appelée système d'ambiance écologique. La qualité de l'eau y est étudiée selon quatre compartiments (qualité biologique générale, qualité biologique des sédiments, par exemple). Le système identifie les indices biologiques permettant d'évaluer la qualité de chacun de ces compartiments. Enfin, il distingue les différents habitats présents (par exemple les sédiments de surface fins), les associe aux compartiments et ajuste les résultats obtenus pour chaque compartiment en fonction de la surface des habitats et de leur fonctionnement (par exemple la capacité de purification). Ainsi l'habitat "sédiments de surface fins" permet d'évaluer le compartiment "qualité biologique des sédiments", ce compartiment étant associé à l'indice biologique oligochètes. Cette approche ne repose donc pas sur une méthode statistique, mais sur une démarche basée sur les connaissances expertes.

Notre objectif est de développer ce concept et de proposer un outil permettant d'accompagner les biologistes dans l'explicitation de leur expertise et de les aider à mettre en évidence l'ensemble des interactions entre indices biologiques. Nous proposons d'utiliser les treillis de Galois [2], comme outil d'extraction de connaissances sous une forme particulièrement bien adaptée à l'analyse par un expert.

Cet article présente une démarche pour la qualification de stations par les indices biologiques. Il s'agit de définir l'ensemble des profils de stations possibles en se basant sur les concepts mis en évidence par la construction d'un treillis de Galois et l'analyse de ces concepts par un hydrobiologiste. Un profil est la donnée de valeurs possibles pour les indices. L'étude réalisée porte sur un échantillon de stations de la plaine d'Alsace pour lesquelles nous disposons des valeurs concernant quatre indices biologiques (tous les indices sauf les poissons). Cette démarche alterne des étapes d'extraction automatique de connaissances et d'analyse des connaissances extraites.

Nous commencerons par décrire les données traitées. Nous introduirons ensuite les principes des treillis de Galois, puis nous présenterons la démarche proposée pour parvenir à un classifieur de stations. Les premiers résultats obtenus seront discutés avant de conclure.

2 Données

Dans le cadre d'un projet pour l'Agence de l'Eau Rhin-Meuse, le LHyGeS a conçu une base de données destinée à rassembler l'ensemble des informations biologiques, physico-chimiques, chimiques utiles à l'évaluation de la qualité des cours d'eau et à la compréhension des interactions entre les différents paramètres impliqués [6]. L'étude porte sur un échantillon de stations sélectionnées sur la plaine d'Alsace. La base est alimentée par des données issues de relevés effectués sur ces stations.

Les données utilisées dans cet article concernent 40 stations décrites par quatre attributs :

- l'Indice Biologique Global Normalisé (IBGN), indice basé sur les invertébrés (insectes, mollusques, ...)
- l'Indice Biologique Macrophytique en Rivière (IBMR) – les macrophytes sont des plantes aquatiques visibles à l'œil nu
- l'Indice Oligochètes de Bioindication des Sédiments (IOBS) – les oligochètes sont des vers vivant dans les sédiments
- l'Indice Biologique Diatomées (IBD) – les diatomées sont des algues microscopiques.

Ces indices font partie des cinq indices normalisés par l'AFNOR pour la France (voir par exemple [1]). Ils sont établis sur la base des taxons présents dans l'eau et les sédiments. Ces taxons ont des caractéristiques et des exigences particulières, chaque indice apporte des informations spécifiques concernant les quatre compartiments biologiques fondamentaux [8] : la qualité biologique générale, la qualité biologique des sédiments, la qualité biologique des eaux, la qualité biologique des (communautés de) poissons. Par exemple, l'IOBS apporte des informations sur l'état des sédiments.

Les valeurs pour trois stations sont présentées dans le tableau 1. Chaque valeur est un entier compris entre 1 (très bonne qualité) et 5 (très mauvaise qualité). Un IOBS incalculable indique une présence très faible ou l'absence d'oligochètes.

Code station	IBGN	IBMR	IOBS	IBD
BREI001	3	5	incalculable	3
BRUN002	2	4	3	3
HORG001	3	3	5	3

TAB. 1 – Classes de qualité de 3 stations pour 4 indices biologiques

3 Treillis de Galois

Les treillis de Galois ou treillis formels de concepts [2, 3, 5] permettent de traiter des données binaires, spécifiant des propriétés vérifiées ou non par un ensemble d'objets. On appelle *contexte* un triplet (O, T, I) où O est un ensemble

d'objets, T un ensemble d'attributs, I une relation de $O \times T$ dans $\{0, 1\}$, et $I(o, t) = 1$ exprime que l'objet $o \in O$ possède l'attribut $t \in T$. Ce contexte se représente naturellement sous la forme d'un tableau binaire.

Par exemple, considérons une version simplifiée de nos données sur les stations. $O = \{S1, S2, S3, S4\}$ est l'ensemble des stations et les attributs de $T = \{\text{Invertébrés}, \text{Macrophytes}, \text{Oligochètes}, \text{Diatomées}\}$ indiquent la présence ou non d'invertébrés, de macrophytes, d'oligochètes et de diatomées (tableau 2).

Station	Invertébrés	Macrophytes	Oligochètes	Diatomées
S1	1	0	0	1
S2	1	1	0	1
S3	1	0	1	1
S4	1	0	1	0

TAB. 2 – Exemple de contexte (binaire)

Soient $X \subseteq O$ et $Y \subseteq T$. On définit $f : 2^O \rightarrow 2^T$ tel que $f(X) = \{y \in T \mid \forall x \in X : xIy\}$ est l'ensemble des attributs partagés par tous les objets de X . Par dualité, on définit $g : 2^T \rightarrow 2^O$ tel que $g(Y) = \{x \in O \mid \forall y \in Y : xIy\}$ est l'ensemble des objets qui possèdent tous les attributs de Y . Le couple (f, g) est appelé *correspondance de Galois* entre les ensembles O et T . Les opérations $h = g \circ f$ et $h' = f \circ g$ sont des opérateurs de fermeture, qui sont appelés *fermetures de Galois*.

Ainsi pour notre exemple, si l'on prend $X = \{S1, S2\}$ alors $f(X) = \{\text{Invertébrés}, \text{Diatomées}\}$, ce qui signifie que les stations $S1$ et $S2$ ont en commun la présence d'invertébrés et de diatomées. Si maintenant on considère $Y = \{\text{Invertébrés}, \text{Diatomées}\}$, on a $g(Y) = \{S1, S2, S3\}$ car les trois stations $S1$, $S2$ et $S3$ possèdent à la fois les attributs Invertébrés et Diatomées.

Un *concept* du treillis de Galois est un couple (X, Y) , où l'*extension* X est telle que $X \subseteq O$ et $f(X) = Y$ et l'*intension* Y est telle que $Y \subseteq T$ et $g(Y) = X$. Dans notre exemple, $(\{S1, S2, S3\}, \{\text{Invertébrés}, \text{Diatomées}\})$ est un concept du treillis de Galois. Le treillis de Galois est construit à partir de l'ensemble des concepts issus du contexte et d'une relation d'ordre entre ces concepts (inclusion des extensions ou des intensions).

Le treillis de Galois obtenu sur nos données est présenté sur la figure 1. Les concepts sont décrits dans le tableau 3.

On peut calculer à partir du treillis la base des règles d'implication entre propriétés [7]. On peut également obtenir des règles d'association [9].

Dans les données sur lesquelles nous travaillons, les attributs peuvent prendre des valeurs dans le domaine $\{1, 2, 3, 4, 5, \text{incalculable}\}$. On parle alors d'attributs multi-valués, et nous nous situons dans le cadre d'un contexte multi-valué tel que défini dans [5]. Nous utilisons le logiciel Galicia¹ pour la construction de nos treillis, ce logiciel permettant le traitement de contextes multi-valués.

¹<http://www.iro.umontreal.ca/galicia/>

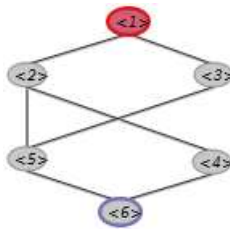


FIG. 1 – Treillis de Galois du contexte exemple

Numéro	Extension (stations)	Intension (indices)
1	{S1,S2,S3,S4}	{Invertébrés}
2	{S1,S2,S3}	{Invertébrés, Diatomées}
3	{S3,S4}	{Invertébrés, Oligochètes}
4	{S2}	{Invertébrés, Macrophytes, Diatomées}
5	{S3}	{Invertébrés, Diatomées, Oligochètes}
6	{}	{Invertébrés, Macrophytes, Diatomées, Oligochètes}

TAB. 3 – Concepts du treillis exemple

4 Démarche proposée pour la création du classifieur

Notre but est de pouvoir relier indices biologiques (d'une station) et caractéristiques du milieu telles que le niveau trophique, la qualité des sédiments, les pressions subies, . . . Pour cela, nous proposons une démarche conduisant à la création d'un classifieur qui permette de qualifier une station en terme de milieu à partir de ses valeurs pour les indices biologiques. Ce classifieur se basera sur un ensemble de profils de stations déterminés au cours du processus. Un profil est un ensemble cohérent de valeurs possibles pour les indices biologiques, qui correspond à un milieu particulier.

La démarche complète est décrite sur la figure 2. Elle part d'un ensemble de données composées des valeurs des différents indices biologiques pour un échantillon représentatif de stations tel que dans le tableau 1. Ce contexte multi-valué sera appelé C dans ce qui suit. La construction d'un treillis de Galois pour C fournit des concepts mettant en évidence des stations présentant des valeurs communes pour des indices biologiques. Ces concepts représentent des profils potentiels, et serviront de base pour la détermination des profils définitifs.

L'étape 1 met en œuvre un algorithme de construction de treillis à partir du contexte C . Les concepts obtenus, leur organisation hiérarchique dans le treillis et les règles d'implication servent de base à l'analyse de l'expert (étape 2). Le nombre de concepts construits est beaucoup trop important et tous ne représentent pas des profils cohérents. L'analyse de l'expert vise donc

à sélectionner les concepts pertinents, notamment en regroupant des valeurs proches pour certains indices. Le regroupement d'indices inclut deux cas : il permet de simplifier le problème et d'identifier des sous-treillis basés sur une sélection d'indices (le processus est itératif : étape 3a), ou il permet d'obtenir des profils à qualifier (étape 3b). Dans l'étape 3b, il est ensuite possible à l'expert de déterminer d'éventuels profils non présents dans l'ensemble. Afin de le compléter, l'expert crée des stations virtuelles. Une station virtuelle est décrite par un ensemble de valeurs d'indices correspondant, selon l'expert, aux caractéristiques d'un milieu particulier. L'ensemble des profils étant alors représenté parmi l'ensemble des stations (stations initiales et virtuelles), on construit un classifieur (étape 4) permettant de qualifier une nouvelle station (étape 5).

5 Premiers résultats obtenus sur les données

Le treillis construit à partir des 40 stations (étape 1) comprend 102 concepts, dont sont déduites 51 règles d'implication. Comme on peut le voir sur la figure 3, ce treillis est grand, rendant son interprétation difficile et une simplification serait intéressante.

Les règles mettent en valeur des implications entre des valeurs d'indices (pour le jeu de données considéré). Leur examen permet de détecter des corrélations entre indices et des règles similaires ayant des valeurs proches pour des indices. Chaque règle est issue d'un concept et le treillis permet d'identifier facilement les concepts liés (ayant des valeurs communes pour les indices). On peut donc envisager un regroupement des concepts sous-jacents si le concept obtenu est pertinent pour l'expert.

Ainsi dans l'étape 2, l'étude du treillis et des règles a permis d'identifier les règles suivantes, pour lesquelles les valeurs des indices sont identiques à l'exception de l'IBMR (la notation *indice : valeur* indique la valeur pour l'indice) :

Règle	Support
IBMR :3 IOBS :incalculable \Rightarrow IBGN :3 IBD :3	1
IBMR :5 IOBS :incalculable \Rightarrow IBGN :3 IBD :3	2

Une autre implication proche indique : IBMR :4 IOBS :incalculable \Rightarrow IBD :3 (support 6). A cette règle sont associées 6 stations qui ont des IBGN entre 2 et 4. La connaissance qu'a l'expert des stations permet de savoir qu'elles correspondent à des cours d'eau de tailles variées. L'IBD :3 indique une qualité d'eau de niveau assez faible, et l'absence d'oligochètes (IOBS :incalculable) complète cette analyse en indiquant que les sédiments sont en mauvais état.

Parmi les implications, une autre règle présente des caractéristiques similaires : IBGN :3 IBMR :4 IOBS :5 \Rightarrow IBD :3 (support 1). Ainsi, on peut également envisager de regrouper les valeurs d'IOBS 5 et incalculable qui sont contiguës dans l'échelle des valeurs (regroupement accepté par l'expert).

L'étude des règles nous a souvent conduits à former des concepts autour des deux indices IBD et IOBS. Il est donc apparu intéressant de construire un treillis portant uniquement sur ces indices (étape 3a), afin de confirmer les tendances ressorties de cette première analyse et de simplifier l'étude (voir figure 4). D'un point de vue biologique, l'IOBS apporte des indications sur l'état des sédiments et l'IBD sur l'état de l'eau. Le fait de cibler l'analyse sur ces deux compartiments est pertinent pour l'expert. Le treillis, plus petit, comporte 19 concepts, ce qui simplifie son étude.

Après analyse des concepts et implications de ce nouveau treillis, l'expert a jugé pertinents les regroupements du tableau 4. Les concepts correspondant au regroupement 2 sont indiqués sur le treillis de la figure 4, il s'agit des concepts 10 et 12, et ces deux concepts ont comme père le concept 8 dont l'intension est $\{IBD :4\}$.

Numéro	Intension	Nombre de stations concernées
1	IOBS : $\{\text{incalculable}, 5, 4\}$ + IBD :3	22
2	IOBS : $\{5, 4\}$ + IBD :4	4
3	IOBS :3 + IBD :3	7
4	IOBS :3 + IBD :4	2
5	IOBS : $\{\text{incalculable}, 5, 4\}$ + IBD :2	5

TAB. 4 – Regroupements pour le treillis IOBS+IBD

Pour chacun de ces regroupements, nous avons construit un sous-treillis basé sur les stations correspondantes pour ensuite procéder de la même façon (étape 3a). Pour le treillis issu du concept IOBS : $\{\text{incalculable}, 5, 4\}$ + IBD :3, cinq regroupements ont été identifiés (étape 2) (voir tableau 5). Ces concepts sont en cours de qualification par l'expert (étape 3b), et les sous-treillis issus des quatre autres concepts sont en cours d'analyse (étape 2).

IBMR : $\{5, 4\}$ + IBGN :2
IBMR : $\{5, 4\}$ + IBGN :3
IBMR : $\{5, 4\}$ + IBGN :4
IBMR :3 + IBGN :4
IBMR :2 + IBGN :4

TAB. 5 – Regroupements pour le treillis IOBS : $\{\text{incalculable}, 5, 4\}$ + IBD :3

6 Conclusions et perspectives

Nous avons proposé une démarche utilisant les treillis de Galois pour aider les experts dans la détermination de profils de milieux aquatiques décrits par des indices biologiques. Ce travail s'inscrit dans le cadre de la compréhension

globale des réponses apportées par les différents indicateurs disponibles pour évaluer la qualité des cours d'eau, l'influence de chacun sur la qualité globale de l'eau n'étant pas à l'heure actuelle connue. L'approche apporte une aide aux experts dans ce travail et offre un support pour la discussion entre experts des différents groupes taxonomiques.

L'analyse faite par l'expert nous a permis de confirmer que les treillis de Galois représentent un outil intéressant pour synthétiser l'information et faciliter sa compréhension. La structuration des concepts donnée par le treillis apporte un plus par rapport à un simple calcul de règles d'association. La démarche proposée a ainsi permis d'assister l'expert dans l'organisation de ses connaissances. L'itérativité du processus, permettant de construire des sous-treillis sur des indices ou valeurs d'indices particuliers, facilite l'analyse.

Les premiers résultats présentés ont été obtenus après analyse d'une hydrobiologiste experte en invertébrés. Nous devons les présenter à des experts des autres indices afin de valider les regroupements de valeurs proposés sur les indices.

L'étape suivante consiste à ajouter des profils manquants (étape 3b) et à construire le classifieur (étape 4), intégrant les profils manquants (stations virtuelles). Différentes approches peuvent être envisagées dont la construction d'un nouveau treillis.

Un autre travail consistera à intégrer aux profils les autres paramètres d'évaluation de l'eau : physico-chimiques et physiques. Pour cela, on pourra repartir des concepts identifiés et construire un treillis où les objets sont les stations et les attributs les concepts, les valeurs de la physico-chimie et les caractéristiques physiques.

Pour 5 stations parmi les 40 étudiées ici, des prélèvements continueront à être effectués régulièrement. La considération de l'évolution des stations au cours du temps est un aspect important dans le but de comprendre l'influence des pollutions et la réponse des différents indices. Nous comptons aborder cet aspect par la suite.

Ce travail se rapproche de celui proposé par Lafont [8]. Il en diffère car ce dernier propose une évaluation plus fine mais plus complexe, pondérant l'effet de chaque indice en fonction de l'étendue des habitats. Cependant, il ne considère a priori pas le recoupement des résultats obtenus sur plusieurs stations afin de les généraliser. La comparaison des résultats de ces deux approches représente une perspective intéressante.

En comparaison de l'ACM utilisée classiquement par les biologistes, notre approche facilite une interprétation globale, et facilite la création de profils qui sont suggérés par les concepts.

Notons enfin que nous avons considéré l'utilisation de techniques telles que les arbres de décision et de régression, afin de prédire la qualité globale en fonction des réponses des différents indices. Cependant ceci nécessite un étiquetage des différentes stations par la qualité globale. Cet étiquetage global étant justement le but du travail, les techniques supervisées se révèlent non adaptées.

Remerciements. Les auteurs remercient l'Agence de l'Eau Rhin-Meuse pour son soutien à ce projet, ainsi que les relecteurs de leurs conseils pour améliorer la présentation de ce travail.

Références

- [1] AFNOR. Qualité de l'eau : détermination de l'Indice Biologique Global Normalisé (IBGN). NF T90-350, 2004.
- [2] M. Barbut and B. Monjardet. *Ordre et classification – Algèbre et combinatoire*. Hachette, 1970.
- [3] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.
- [4] Brigitte Escofier and Jérôme Pagès. *Analyse factorielles simples et multiples*. Dunod, 1988.
- [5] B. Ganter and R. Wille. *Formal Concept Analysis : Mathematical Foundations*. Springer Verlag, 1997.
- [6] C. Grac, A. Herrmann, F. Le Ber, M. Trémolières, A. Braud, A. Handja, and N. Lachiche. Mining a database on alsatian rivers. In P. Gourbesville, J. Cunge, V. Guinot, and S.-Y. Liong, editors, *Proc. of the 7th Int. Conf. on Hydroinformatics*, volume III, pages 2263–2270, 2006.
- [7] J.-L. Guigues and V. Duquenne. Familles minimales d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences Humaines*, 95 :5–18, 1986.
- [8] M. Lafont. A conceptual approach to the biomonitoring of freshwater : the ecological ambience system. *Journal of Limnology*, 6 :17–24, 2001.
- [9] A. Napoli. A smooth introduction to symbolic methods in knowledge discovery. In H. Cohen and C. Lefebvre, editors, *Categorization in Cognitive Science*. 2006.

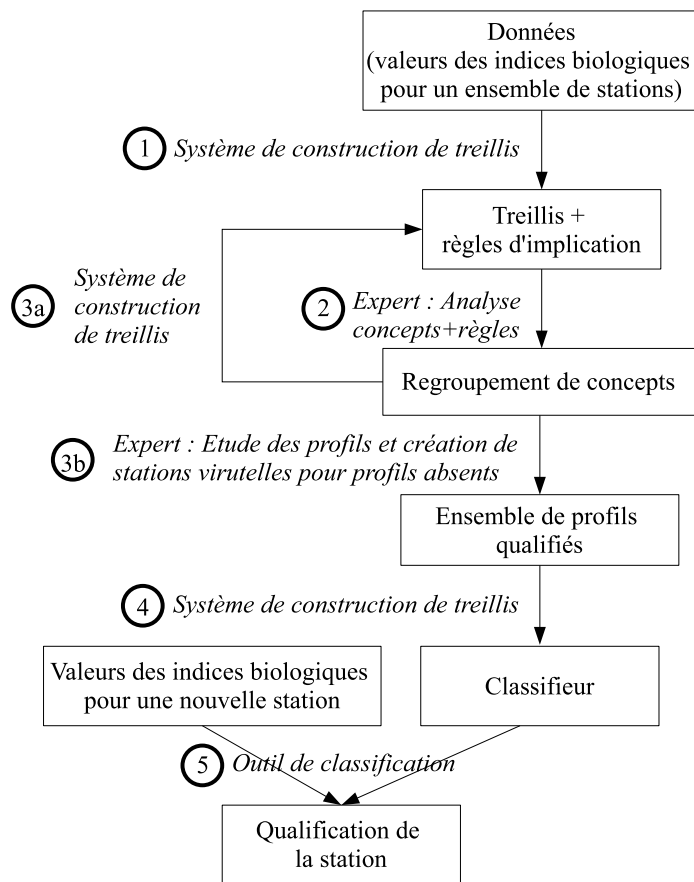


FIG. 2 – Démarche

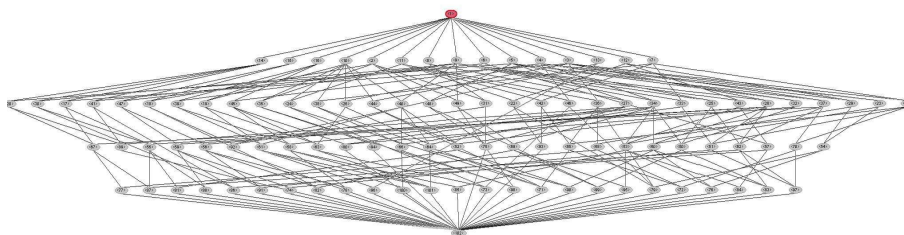


FIG. 3 – Treillis initial obtenu sur l'ensemble des données

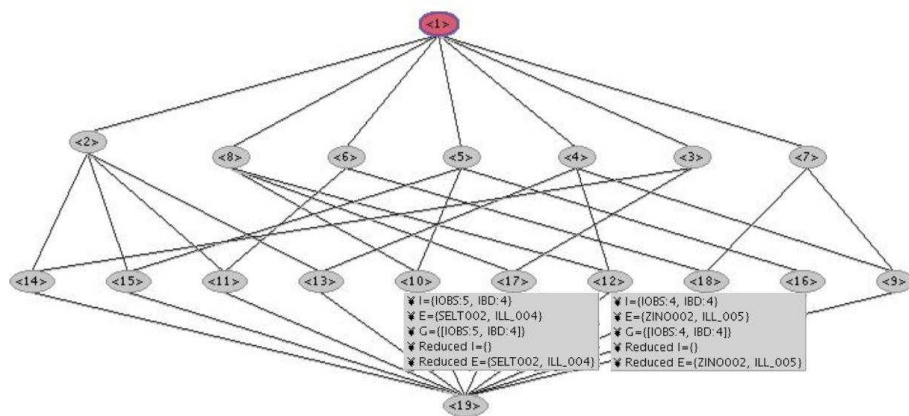


FIG. 4 – Treillis obtenu à partir des indices IBD et IOBS