

Analyse des tendances de la composition des eaux par stratification. Application pilote dans le cadre DPSIR de l'AEE

Assessment of trends in water composition by stratification. Pilot application within the EEA DPSIR framework

PHILIPPE CROUZET

Agence Européenne pour l'environnement
Kongens Nytorv 6, DK1050 Copenhague K Danemark
Tél : +45 3336 7191, Fax : +45 3336 7293, e-mail : philippe.crouzet@eea.europa.eu

BERNARD BEAUZAMY, CHARLINE CARLIER

Société de Calcul Mathématique
111 Faubourg Saint Honoré 75008 Paris
Tél : +33 (0)1 42 89 10 89, Fax : +33 (0)1 42 89 10 69, e-mail : bernard.beauzamy@scmsa.com, charline.carlier@scmsa.com

Assessment of trends in water composition by stratification. Pilot application within the EEA DPSIR framework. The request to assess the efficiency of measures in terms of environmental results requires the ultimate use of field observations that are often understood as useless expenses. In the area of surface water, elementary watershed stratification allows clarifying the relationships between causes (drivers) and observations and helps forecasting the time lag required to meet a target under constant effort. The constraints intrinsic to the classical statistical techniques lead to defining a probabilistic approach to the assessment of observations, after stratification. This complementary technique allows obtaining more information from the existing data.

Analyse des tendances de la composition des eaux par stratification. Application pilote dans le cadre DPSIR de l'AEE. Les demandes d'évaluation de l'efficacité des actions en termes de résultats environnementaux amènent à l'utilisation la plus approfondie des observations de terrain, souvent considérées comme des dépenses inutiles. S'agissant des eaux de surface, la stratification appliquée aux bassins versants élémentaires permet de clarifier les relations entre les causes (forces motrices) et les observations et de définir le délai nécessaire pour atteindre un objectif, à effort constant. Les contraintes spécifiques aux techniques statistiques classiques ont amené à définir une approche probabiliste de l'interprétation des observations après stratification, complémentaire à la précédente, de manière à tirer encore davantage d'information des données existantes.

I ■ INTRODUCTION

La prise en compte de l'environnement dans les politiques publiques, et tout d'abord européennes, a fait naître des exigences nouvelles en termes d'évaluation de l'efficacité de ces politiques (les buts affichés sont-ils atteints par l'application de ces politiques ?), d'efficience (les moyens nécessaires à cette atteinte sont-ils proportionnés ?) et enfin de pertinence (les buts sont-ils les bons, n'y a-t-il pas des effets pervers ?). L'Agence Européenne pour l'Environnement (AEE) est chargée de « soutenir le développement durable et de contribuer à une amélioration significative et quantifiable de l'environnement en Europe grâce à la fourniture d'information opportune, ciblée, appropriée et fiable aux décideurs et au public ».

L'AEE analyse l'évolution des composantes environnementales dans un but d'évaluation de l'état de l'environnement et

de mise en relation de ces évolutions avec les politiques environnementales et les secteurs d'activité. La série suivante de questions, dont la réponse demande des approches de complexité croissante, illustre les bases de cette analyse. En considérant par exemple la problématique simple de la teneur en nitrate de l'eau de rivière : « quelle est la tendance temporelle dans la rivière X ? », « quelle est la tendance dans les rivières des bassins versants soumis à la directive nitrate ? », « quel est le rôle des différents secteurs dans les tendances ? » et enfin « l'objectif d'une teneur T à l'année A sera-t-il respecté ? ».

La traduction de ces questions en méthodes de calcul, ciblées pour différents publics, requiert la mobilisation des observations de terrain pour la première question, de manière à établir un état (S de l'analyse DPSIR (££), et l'ajout de variables supplémentaires et d'hypothèses relatives aux forces motrices, aux pressions exercées, aux réponses et

à leurs interrelations ou impact (respectivement D, S, R et I de l'analyse DPSIR).

Les réponses attendues par les décideurs sont en outre produites dans un cadre fortement contraint : ce sont en effet les mêmes décideurs qui imposent par les choix budgétaires et réglementaires la capacité de fourniture des réseaux et ce sont également eux qui, au travers des obligations de rapportage, rendent telle ou telle information disponible au niveau européen.

L'utilisation des observations, qui sont la seule référence opposable, exige toutefois leur mise en perspective et leur contextualisation. Les méthodes développées à l'AEE, à la suite des travaux préparatoires engagés dès sa création en vue de la constitution de « réseaux représentatifs » destinés à la fourniture de données dans le cadre EIONET [9], ont été appliqués à l'échelle pilote sur le territoire français [8] à la suite de travaux conduits d'abord par l'Ifen en vue de remplir ses obligations de « Point focal national » de l'AEE. Cette application a permis de dégager des résultats prometteurs et des enseignements qui vont être appliqués pour une meilleure évaluation de l'état des eaux européennes, en complément des travaux plus ciblés conduits dans le cadre de la mise en œuvre de la directive cadre sur l'eau (Directive 2000/60/EC), notamment au sein de WISE.

II ■ MISE EN PERSPECTIVE ET CONTEXTUALISATION DES OBSERVATIONS

● II.1 ASYNCHRONISME ENTRE OBSERVATIONS ET ÉVALUATIONS

Les progrès de la métrologie en hydrologie quantitative et qualitative se traduisent par l'amélioration de la précision et de la pertinence de la mesure individuelle. En revanche, la continuité des mesures au cours du temps est sans cesse remise en cause par les modifications, réductions des programmes de mesure et par la nécessité de les ajuster aux nouvelles réglementations. Par ailleurs, l'ajustement progressif des programmes de surveillance à des finalités précises peut mettre en péril l'utilisation des données à d'autres fins ; par exemple calculer des débits massiques à partir de mesures orientées vers l'évaluation de la qualité en étiage est très acrobatique.

Assez paradoxalement, l'exactitude globale de l'information apportée par les réseaux se détériore alors que la précision de la mesure individuelle s'accroît. Comme la disponibilité d'informations de contexte spatialisée s'est en parallèle accrue de manière considérable au cours des 4 décennies passées, il est possible de remédier dans une certaine mesure à cette détérioration.

Le délai entre une cause et un effet est très variable. A l'échelle d'une masse d'eau, un rejet individuel épuré se traduit par une modification rapide de la composition en aval. A l'échelle d'un bassin versant, l'effet d'une politique d'action peut demander plusieurs décennies, notamment si des pollutions diffuses sont mises en jeu. En pratique, le nombre, le positionnement et la fréquence d'échantillonnage

des réseaux de mesure qualitatifs dans l'ensemble des pays d'Europe ont varié avec une dynamique plus rapide que les changements de composition résultant de l'application des politiques. L'instabilité de fait des réseaux d'observation implique donc soit de limiter l'analyse aux seuls cas heureux où leur permanence a été assurée (analyse de cohorte), soit de considérer les observations comme le résultat d'enquêtes par sondage. Il faut alors définir *a posteriori* les attributs de chaque observation en termes de population d'appartenance.

L'analyse de cohorte est mise à profit dans plusieurs analyses de l'AEE et elle confirme la relation causale entre des changements sur les bassins versants et les observations [5, p333]. Elle manque en revanche de la dimension spatiale et représentative exigée par les missions de l'AEE.

● II.2 LES APPORTS DE L'ANALYSE PAR STRATIFICATION

La variabilité des écoulements est en relation avec la réponse du bassin versant à la pluviométrie efficace. La variabilité de la composition dépend, toutes choses égales par ailleurs, de la densité des apports et de la dilution par les écoulements pour les substances conservatives au moins. La concentration est donc homogène avec une densité d'apports diluée par un volume ; elle est donnée par la formule :

$$C = M L^{-3} = (M L^{-2} T^{-1}) \div (L^3 L^{-2} T^{-1}) \quad (1)$$

où M représente une masse, L une longueur et T le temps.

Cette équation (1) aux dimensions élémentaire autorise à considérer comme très plausible que les valeurs des observations de concentration en provenance de bassins versants où la densité d'apport est proche et de dilution comparable appartiennent à une même population statistique. Une analyse statistique par stratification peut donc être réalisée en répartissant les observations en raison de leur strate d'appartenance. Cette approche a pour but d'une part de procurer une estimation la plus exacte possible des valeurs caractéristiques de la strate et d'en minorer l'incertitude. Elle est en outre de nature à estimer correctement les valeurs caractéristiques de la population (territoire ou classe de pression) et donc de permettre des comparaisons entre différents territoires et différentes périodes.

La méthodologie des analyses statistiques par stratification est bien codifiée. Elle s'applique toutefois aux enquêtes définies a priori ; dans le cas de l'utilisation des résultats de mesure, celles-ci ont été réalisées longtemps avant que la population ne soit définie et ses caractéristiques documentées. Dès le début de l'approche réalisée par l'Ifen pour définir les « réseaux représentatifs » demandés par l'AEE [10], la notion de « rétro-sondage » a été mise à profit. Elle consiste à estimer les conditions d'environnement de chaque point de mesure et à définir une classe de forces motrices auxquelles il est soumis. Un tirage en principe aléatoire amène alors à proposer un ensemble de points de densité spatiale homogène pour chaque strate [2].

Cette pratique a permis à la fois de répondre aux demandes de l'AEE et d'en mesurer les limites. En effet, une fois

éliminées les mesures non exploitables et les stations non sélectionnables, seul un ensemble d'effectif très pauvre pouvait satisfaire aux exigences du rétro-sondage : l'homogénéité de l'échantillon se paye alors en perte de précision et d'information et les tentatives pour pallier cela par des techniques géostatistiques n'ont pas abouti à des protocoles applicables pratiquement [3]. En revanche, le principe de l'analyse par stratification a toujours été validé.

Les travaux de l'Ifen ont été repris à l'AEE en considérant que les questions à résoudre imposaient de réduire au plus la variabilité des indicateurs statistiques. Partant du constat que les indicateurs les plus demandés sont des moyennes par strate, l'hypothèse de l'applicabilité du théorème central limite amène à envisager que la variance d'estimation dépend du quotient de la variance de l'échantillon par la racine carrée du nombre d'observations. Diminuer la variance par une stratification poussée et pertinente d'une part et tenir compte de tous les points de mesure existants d'autre part est la réponse la plus simple à ce problème.

● II.3 APPLICATION PRATIQUE

L'intérêt de la mise en œuvre des principes a été défini au cours de plusieurs travaux menés à l'Ifen et enrichis des remarques recueillies lors de débats avec les autres pays membres de l'AEE et avec le centre thématique. Le détail de la mise en œuvre est toutefois suffisamment délicat pour avoir rebuté la quasi-totalité des partenaires de l'AEE. Il a donc paru important à cette dernière de tester une analyse comparative de la performance des réponses apportées par l'analyse des mesures en provenance d'un sous échantillon des points (réseau « représentatif » fortement sélectionné selon les critères d'EIONET) d'une part et de l'application d'une méthodologie la plus rigoureuse possible portant sur la totalité des observations exploitables d'autre part. L'ensemble des protocoles et des résultats sont publiés [4] et donc seuls les principes importants sont discutés ici.

II.3.1 Population de Référence

Dans le cas d'un réseau EIONET, supposé représentatif, les stations constituent implicitement la population de référence. C'est donc en vertu de cette appréciation que les comparaisons de changement sont faites en termes de proportion de stations de telle ou telle qualité. Cette hypothèse implicite est également assumée quand, par exemple, la proportion de points où des pesticides sont détectés est assimilée ultérieurement à une proportion de rivières contaminées malgré les mises en garde de l'auteur [6].

Si en revanche une relation suffisamment documentée peut être établie entre un bassin versant (où s'expriment les causes) et les observations (qui reflètent les effets), alors l'ensemble des bassins versants élémentaires constitue la population de référence. Dans ce cas, les résultats sont « X % du territoire présente une eau de composition... » et non plus « on trouve dans Y % des points de mesure une eau... ». L'expression des résultats en termes de bassins versants offre des conditions de représentativité conformes aux attentes et permettant des comparaisons entre bassins versants ou entre agrégats de points de mesure soumis aux mêmes pressions.

La contrepartie du choix des bassins versants élémentaires comme population de référence oblige à les définir de manière homogène, à les connecter (en intégrant les passages de frontière) et à en identifier les drains principaux de manière à ce que chaque point de mesure sur un drain principal puisse être qualifié en raison de la combinaison et de l'intensité des pressions qui s'exercent à son amont. Dans le cas de l'étude pilote réalisée par l'AEE sur les données fournies par le France, les bassins élémentaires sont les « zones hydrologiques (ZH) » de la BD-Carthage, au nombre de 6190. Ce découpage n'est pas idéal, car la taille des bassins est hétérogène, de quelques km² à quelques centaines de km².

Enfin, attribuer à chaque station de mesure un positionnement sur son bassin versant et sur le drain principal est essentiel, faute de quoi une station pourrait être affectée improprement à un bassin. L'expérience a montré que, dans des cas particuliers, une seule station de mesure pouvait changer la valeur des statistiques de strate de manière considérable ; en contrepartie, l'analyse statistique préparatoire permet d'autant mieux de distinguer les valeurs hors échelle ('outliers') que la stratification est correcte et donc d'améliorer la qualité des données traitées. Le marquage de chaque station comme sélectionnable ou non est donc une étape délicate et indispensable.

II.3.2 Indicateur statistique

La sélection d'un indicateur statistique pertinent est affaire de compromis entre la demande d'évaluation, les données disponibles et les contraintes des méthodes statistiques utilisables. Dans le cas de l'analyse conduite par l'AEE, seuls la moyenne annuelle, le maximum et le minimum annuels des valeurs observées par point ont été demandés et collectés systématiquement dans le cadre EIONET. Les déciles haut et bas étaient également demandés, mais rarement obtenus.

S'agissant de l'évaluation d'une politique, les contraintes relevant de l'analyse de conformité ne s'appliquent pas nécessairement. Dans le dernier cas, une valeur seuil à ne pas dépasser s'interprète en général comme ne devant jamais être dépassée dans la ressource analysée, en pratique dans aucun des échantillons de surveillance. Dans une analyse de tendance au contraire, la référence à chaque mesure individuelle est rarement possible. On lui préfère des indicateurs plus agrégés se prêtant mieux à l'analyse et à l'estimation d'erreur. Il est alors important de bien définir la portée des résultats : il n'est pas équivalent de constater, au niveau d'une strate par exemple, que toute mesure, que chaque moyenne de point ou la moyenne de strate ou encore que la tendance de la moyenne de strate (par exemple la droite de régression en fonction du temps) sont inférieures à un seuil de référence.

La moyenne par point n'est pas le meilleur indicateur de l'évolution de la qualité des eaux, celle-ci étant le plus souvent définie par un dépassement temporaire de seuils. Elle est en revanche assez appropriée à l'évaluation du résultat de politiques de diminution des apports qui s'appliquent toute l'année et de plus c'est jusqu'à présent le seul indicateur calculable. L'impossibilité pratique de définir des valeurs de quantiles par strate et leur estimateur d'erreur a été démon-

trée par Armines, dans une étude pilotée par l'Ifen [3]. L'utilisation du krigeage spatial pour pallier les difficultés théoriques n'est pas envisageable en production de routine au niveau européen.

La moyenne de strate est facilement calculée à partir des moyennes par point. L'agrégation à n'importe quel niveau et l'estimation d'erreur sont également faciles à implémenter dans un système de calcul. Cet avantage est considérable en termes de production. En effet, les strates sont par définition à géométrie variable et l'appartenance d'un point à une strate est totalement contextuel puisqu'elle dépend des critères retenus pour la définir. La contextualisation de chaque point de mesure permet de tirer le meilleur parti d'un nombre très différent de mesures. En France, entre 1971 (premier Inventaire de la qualité des eaux) et 2005 le nombre de stations exploitables va de 400 à 1200 - 1900, selon le nombre de paramètres considérés.

En revanche, seule la moyenne annuelle pondérée par le temps (pour éliminer les biais résultant de fréquences de mesures différentes et irrégulières dans le temps) peut être agrégée. L'agrégation au niveau de la strate de moyennes pondérées par l'écoulement pose des problèmes théoriques de définition non résolus pour l'instant. De ce fait, les indicateurs produits jusqu'ici représentent la concentration d'un prélèvement effectué à débit constant à partir d'un point quelconque de mesure et la moyenne de strate la concentration d'un prélèvement effectué dans les mêmes conditions à partir du mélange de toutes les eaux des points de la strate.

● II.4 DÉFINITION ET CONSTITUTION DES STRATES

L'efficacité de la procédure requiert un nombre de strates aussi réduit que possible, sans masquer la réalité des différences entre les sources possibles d'impact. Les strates sont donc définies en fonction de classes de pressions potentielles respectant les principes de l'équation (1) et dont on sait, par les relations causales établies par ailleurs, qu'elles doivent se traduire par une signature particulière en termes de composition des eaux. Le travail de définition consiste donc à définir pour chaque paramètre pris en compte quels sont les types de pression qui en modifient la concentration et de définir les critères permettant de classer chaque bassin élémentaire dans une strate.

Une strate est définie par des classes de valeur de forces motrices, exprimant un potentiel de pression et non pas les pressions elles-mêmes. Cette distinction est essentielle dans le cadre DPSIR. Par exemple un fort pourcentage de terres labourées est une présomption de trouver du nitrate dans l'eau. Remplacer la force motrice « pourcentage de terre labourée » par une pression, par exemple « surplus en nitrate en $\text{kg an}^{-1} \text{ha}^{-1}$ » pourrait être envisagé. Dans ce cas toutefois, le surplus étant par nature variable au cours du temps, il ne serait plus possible de définir une tendance entre deux termes variant ensemble, et donc d'évaluer le résultat d'une politique. En résumant, il est très important que les critères de stratification soient les plus invariants possible, ce qui n'empêche pas d'envisager qu'un même point de mesure puisse successivement appartenir à différentes strates.

Tableau 1 : Critères de stratification utilisés

Nom de la strate	Objectif	Critères de stratification
1/ faibles pressions (code «F»), N = 14 %	Peu d'agriculture, présence marquée de territoires naturels, faible densité animale et humaine	AG1 < 25 %, FOR+AUT >= 40 %, pop < 40 h km ⁻² , Bétail < 50 eh km ⁻²
2/ impacts ordinaires (code «X»), N = 31 %	Valeurs moyennes des différentes activités humaines et agricoles	AG1 < 40 % OU (AG1+AG2) < 60 %, URB < 2.5 %, pop < 78 h km ⁻² , Bétail < 200 eh km ⁻²
3/ densément urbaine (code «V»), N = 4 %	Présence humaine en densités très élevées	URB > 10 %, pop >= 150 h km ⁻² ,
4/ mixte (critères «U» et «A» satisfaits simultanément) (code «M»), N = 9 %	Forte densités humaines et large aire de grandes cultures présentes ensemble	AG1 >= 40 % OU (AG1+AG2) >= 60 %, pop >= 78 h km ⁻²
5/ urbaine (code «U»), N = 6 %	Forte densité humaine, mais sans composante agricole additionnelle	AG1 < 40 %, (URB >= 2.5 % ET pop >= 78 h km ⁻²) OU (URB >= 1.5 % ET pop >= 90 h km ⁻²) (
6/ agricole intensive (cultures) (code «A»), N = 18 %	Large aire affectée aux grandes cultures et densité animale inférieure aux bornes, sans forte densité humaine	AG1 >= 40 % OU (AG1+AG2), pop < 78 h km ⁻² , Bétail < 200 eh km ⁻²
7/ agricole élevage (code «B»), N = 19 %	densité animale élevée, sans cultures majoritaires	pop < 78 h km ⁻² , Bétail >= 200 eh km ⁻²

AG1, AG2, FOR, URB et AUT : aires CORINE *land cover* 1990 respectivement d'agriculture intensive, non intensive, de forêts, artificialisées et autres (par ex. rochers, dunes, etc.), pop : nombre d'habitants, selon le RP 1999 et bétail : le cheptel, selon RA 2000, calculé en équivalent personne sur la base des émissions azotées [4]. Le pourcentage en première colonne est l'effectif final en nombre de ZH.

Dans la dernière étude réalisée par l'AEE, sept strates ont été définies et appliquées sur un découpage en 9 bassins versants, tels que définis antérieurement par l'Ifen [7] pour l'analyse des nutriments (nitrate, ammonium, phosphate et accessoirement DBO_5). Malgré ce découpage grossier, un effectif suffisant en stations n'est pas toujours trouvé dans chaque strate. Les définitions de strates sont reportées dans le *Tableau 1*.

Les critères et l'ordre de calcul (chiffre dans la colonne « nom de la strate » du *Tableau 1*) résultent d'une analyse statistique systématique des relations entre un critère et les observations par régression linéaire multiple incluant le temps. L'analyse porte à la fois sur un critère positif et un critère négatif. Par exemple, une forte corrélation linéaire positive entre la teneur en nitrate et le critère AG1 est confrontée à une corrélation linéaire négative avec le critère FOR. Ceci a mis en évidence des stations de bassins versants forestiers ayant de fortes concentrations en nitrate par suite de rejets industriels spécifiques. Dans ce cas, la station est marquée non sélectionnable dans ce cadre d'analyse.

La mise au point de la stratification est donc un processus avec rétroaction entre des jeux indépendants de données. Il est donc nécessaire que tous les bassins élémentaires soient munis d'une indication de strate pertinente. Pour cela, l'outil de stratification comporte une strate de contrôle alimentée par défaut ; à la fin du processus son effectif doit être nul, sinon les critères sont insuffisants ou ambigus. L'ensemble des processus de stratification, préparation des données, critères de sélection et ajustements est conduit avec le module « EuroWaternet » de l'application Nopolu® (Pöyry Environnement), installée et utilisée à l'AEE.

III ■ RÉSULTATS OBTENUS

L'analyse finale a produit des résultats pour les strates, au niveau national et par bassin. Un exemple de résultat est porté en *Figure 1*. Il montre pour le phosphore dissous réactif ('PO4') l'évolution des teneurs moyennes, France entière, pour différentes strates.

La strate 'F', non polluée et la strate 'M' qui cumule les apports urbains et agricoles sont clairement séparées par exemple. Le calcul impose une hypothèse sur la prise en compte des strates mal documentées, en l'espèce 70 % de l'effectif théorique doit être documenté, faute de quoi le calcul n'est pas fait. Par exemple, dans le cas de l'analyse sur un sous-bassin, il est possible qu'une strate soit présente en tant que ZH, mais sans un effectif suffisant de points de mesure.

Par ailleurs, ayant fixé une cible (ici $0,15 \text{ mg l}^{-1}$), la date à laquelle cette cible pourrait être atteinte est estimée en fonction de la qualité de l'ajustement linéaire. Dans le cas de la strate A, il y a par exemple 95 % de chances que ceci se passe en 2020 (2011 - 2028). La discussion détaillée est faite dans la publication de l'AEE [4].

Le découpage en strates et en sous-bassins rend intelligible l'information disponible et les tendances. En revanche, la nécessité d'utiliser un indicateur central masque l'évaluation des valeurs extrêmes et risque de brouiller l'analyse de l'efficacité. Une réflexion a été menée de manière à envisager un couplage stratification – analyse probabiliste plus pertinent car débarrassé des contraintes propres aux approches statistiques.

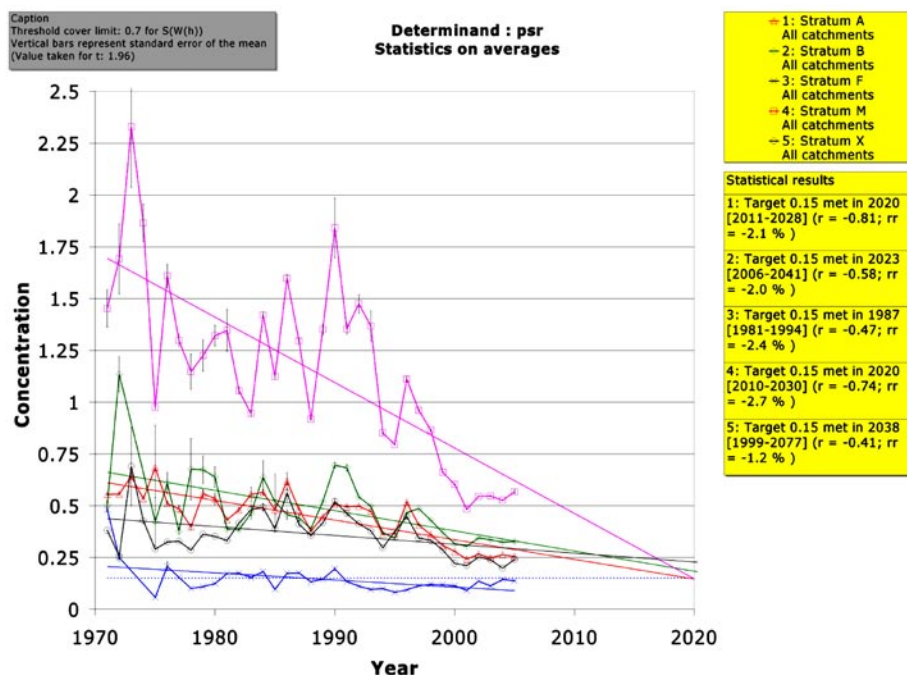


Figure 1 : Exemple de résultat (source: *Fig 24*, [4])

IV ■ APPROCHE PROBABILISTE DE VALORISATION DES INFORMATIONS STRATIFIÉES

● IV.1 JUSTIFICATIONS DE LA MÉTHODE PROBABILISTE

Les résultats obtenus par l'approche statistique répondent en grande partie aux questions posées, mais la mise en œuvre en est assez lourde, tout en amenant une perte de l'information contenue dans les données élémentaires, notamment des valeurs extrêmes dont l'évolution est cachée par la moyenne. Par ailleurs, les hypothèses sous-jacentes sont mal vérifiées et le calcul de l'intervalle de confiance se réfère implicitement à une symétrie de l'incertitude, ce qui est notoirement inexact.

L'AEE a donc travaillé de avec l'aide de la SCM à l'analyse critique de la méthode classique de manière à l'améliorer et pour élaborer une approche complémentaire, à base probabiliste[11], de manière à préparer ses futures évaluations, fixées à l'horizon 2011.

Une méthode probabiliste est en effet de nature à éclairer des aspects qui ne sont pas bien pris en compte par les techniques classiques et donc de compléter l'information produite. Sa mise en œuvre pose toutefois des difficultés dans l'organisation des données, mais élimine en contrepartie une grande partie des contraintes sous-jacentes à l'analyse statistique. Son application se décompose en trois étapes : *Création d'une table des occurrences, de la loi de probabilité et de la fonction de répartition.*

La *table des occurrences* recense, pour la zone étudiée et pour une année donnée, le nombre d'observations (stations si l'on ne dispose que de la moyenne, ou mesure individuelles) ayant présenté une concentration comprise dans un certain intervalle. Cette étape permet de construire un histogramme qui donne comme information le nombre d'observations comprises dans chaque intervalle de concentrations. Elle permet de donner une première approximation sur leur répartition dans la zone étudiée, sans préjudice de la stratification qui est un attribut supplémentaire des stations sous une certaine hypothèse. Toutefois, pour garantir l'absence de biais dans cette table il faut soit une distribution homogène des observations (assurée dans le cas de la moyenne), soit une reconstruction des observations manquantes [1].

La *loi de probabilité* se déduit très simplement de la table d'occurrence construite précédemment en divisant le nombre d'observation pour chaque intervalle par le nombre total des observations de la zone étudiée. Le résultat ainsi obtenu est la probabilité expérimentale qu'une observation de la zone étudiée présente une concentration comprise dans un certain intervalle.

L'inconvénient des résultats des deux étapes précédentes est qu'une partie de l'information est perdue puisque l'on représente uniquement des intervalles et donc seulement combien de valeurs se situent entre telle et telle borne. L'information de base est généralement déjà très pauvre (souvent des moyennes annuelles), on préférera donc utiliser la fonction de répartition. Au lieu de comptabiliser le nombre d'observations dans des intervalles, on comptabilise le nombre

d'observations ayant présenté une valeur C inférieure à un seuil α , ce qui constitue la *fonction de répartition*.

Il est donc possible de déterminer, pour une région donnée et pour une année donnée, la probabilité qu'une observation (individuelle ou agrégée) présente une concentration inférieure à un certain seuil

Avec une approche probabiliste, les informations apportées par chaque station sont conservées dans le résultat. Les probabilités sont un arrangement des données sous une forme qui permet une meilleure compréhension.

L'approche probabiliste complète l'approche statistique. Il est en effet assez simple de déterminer l'homogénéité d'une strate et ce faisant, de définir les stations ou les années hors norme.

● IV.2 LA MÉTHODE DE L'HYPERSURFACE PROBABILISTE EXPÉRIMENTALE (EPH)

La méthode la plus employée pour prévoir les valeurs dans un futur plus ou moins proche est la méthode de régression linéaire. Cette méthode est simple de compréhension et d'utilisation et permet de calculer un intervalle de confiance. Elle est applicable uniquement si les valeurs passées suivent une tendance claire et linéaire, ce qui n'est pas toujours vérifié. Lorsque cela n'est pas le cas une approche probabiliste permet de surmonter ce problème.

L'EPH est une méthode qui a été développée par la SCM [12]. Le résultat obtenu via cette méthode est, pour chaque point où l'information n'est pas connue, une loi de probabilité qui est une « propagation » de l'information à partir de points existants dans un espace multidimensionnel.

La *Figure 2* présente un exemple de résultat de la méthode de l'EPH appliquée dans le cas unidimensionnel de l'évolution de la probabilité observée qu'une station, dans la strate B du bassin de la Garonne, soit en dessous du seuil de 0.20 mg/l (jusqu'en 2005) et les prévisions de la méthode EPH (les trois points 2006 - 2008).

La méthode EPH permet également de reconstruire des données historiques manquantes. L'implémentation dans le système de calcul par stratification à l'AEE est en cours d'évaluation.

V ■ CONCLUSIONS

Les mesures en hydrologie sont aujourd'hui de plus en plus précises et de meilleure qualité. En revanche, le nombre de mesures et de stations tend à diminuer. Il convient alors de trouver une méthode qui prenne en compte dès le départ ce problème afin d'évaluer au mieux l'état environnemental des rivières.

Les méthodes statistiques utilisées actuellement par l'AEE sont simples d'utilisation et de compréhension mais elles entraînent une perte non négligeable dans l'information disponible et doivent s'appuyer sur des hypothèses statistiques invérifiables.

L'exigence en termes d'information fournie croît à l'inverse de la disponibilité en mesures. Il est donc nécessaire

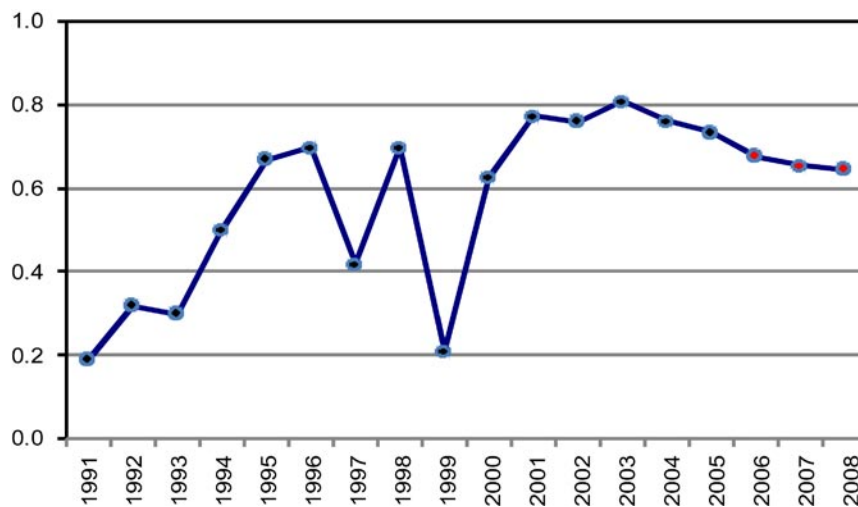


Figure 2 : Exemple de résultat obtenu avec l'EPH

d'utiliser en complément de ces méthodes une approche probabiliste qui permette de conserver toutes les informations disponibles pour chaque strate. Il n'y a pas de perte d'information mais seulement un arrangement des données de départ. La méthode probabiliste donne le cadre approprié pour conserver et utiliser toute l'information disponible, au prix toutefois de la nécessité de révision en profondeur des procédures de traitement.

Dans les deux cas, la régularité et la permanence des observations sont des préalables faute desquels l'interprétation des données dans un cadre de bassin versant et de forces motrices n'est pas possible.

VI ■ RÉFÉRENCES ET CITATIONS

- [1] BEAUZAMY B., ZEYDINA O. (2007) — *Méthodes probabilistes pour la reconstruction des données manquantes*. Paris : Société de Calcul Mathématique.
- [2] BETURE-CEREC & ARMINES (2000) — *Eurowaternet. Construction d'un réseau représentatif de qualité des cours d'eau. Phase I-Méthodologie : Analyse exploratoire des données. Rapport : Final, pour Ifen et Agences de l'eau, Orléans.*
- [3] BETURE-CEREC & ARMINES (2001) — *Eurowaternet. Construction d'un réseau représentatif de qualité des cours d'eau. Phase II-Rapport final. Rapport : Final, pour Ifen et Agences de l'eau, Orléans.*
- [4] CROUZET P., JENSEN M. T (2007) — *Assessing water quality in Europe using stratification techniques – Results of a prototype application using French data. reports / Technical 2007. Copenhagen : EEA. 10*
- [5] EUROPEAN ENVIRONMENT AGENCY (2005) — *The European environment – State and outlook 2005. Copenhagen : EEA.*
- [6] IFEN (2007) — *Les pesticides dans les eaux – Données 2005. Les dossiers de l'IFEN. 9*
- [7] JUIN D. (ED.) (2002) — *L'environnement en France. Paris et Orléans, La Découverte et Institut français de l'environnement.*
- [8] LEONARD J., CROUZET P. (1999) — *Construction d'un réseau représentatif. Contribution au réseau «EUROWATERNET» / Qualité des cours d'eau de l'Agence Européenne de l'Environnement. Jacques Theys. Orléans : Institut français de l'environnement. 13*
- [9] NIXON S. (1997) — *European freshwater monitoring network design. European Environment Agency ; Luxembourg : Office for Official publications of the European Communities. 10*
- [10] OIEAU & IFEN (1997) — *Réseau de suivi des eaux continentales de l'agence européenne de l'environnement. Projet pilote en France. (rapport : final (version française), pour Agences de l'eau, Institut français de l'environnement, Office international de l'eau. Orléans.*
- [11] SOCIETE DE CALCUL MATHEMATIQUE (2007) — *Improving the results and their attached uncertainty in stratified assessments of drivers – water composition relationships. Rapport final, pour EEA Paris.*
- [12] SOCIETE DE CALCUL MATHEMATIQUE (2007) — *Méthodes probabilistes pour l'analyse des incertitudes liées à la sûreté des réacteurs nucléaires. L'Hypersurface Probabiliste. Construction Générale et Applications. Rapport final, pour Institut de RadioProtection et de Sûreté Nucléaire. Paris.*