

# Hydrologie et Statistiques

J.M.Masson<sup>a</sup>

## Qui utilise les statistiques en hydrologie ?

Ce sont des ingénieurs hydrauliciens, agronomes, I.S.I.M, dont la formation de base en statistiques est assez légère, mais qui ont acquis une certaine culture dans ce domaine en lisant des ouvrages et des revues spécialisées, essentiellement Anglo-saxonnes. Rares sont en effet les ouvrages en langue française dans les bibliothèques des Ingénieurs (Ceresta, Dagnélie...), alors qu'en langue anglaise, on trouve aussi bien des ouvrages de base (Cramer, Kendall et Stuart, Draper et Smith...), que des ouvrages de statistiques appliquées à l'hydrologie (Haan, Yevjevich, Kottegoda...) ou des revues spécialisées (Water Ressources Research, Journal of Hydrology, Technometrics, Applied Statistics...).

Cette culture statistique, acquise par la bibliographie et la pratique, est en partie transmise par des interventions d'enseignement dans des formations d'ingénieurs ou de recherche. Elle permet de suivre l'évolution des méthodes, d'identifier, voir de formuler des problèmes statistiques, de critiquer des solutions apportées par d'autres, mais on ne peut pas dire qu'elle débouche sur une recherche en matière de statistiques appliquées à l'hydrologie.

Les recherches qui s'effectuent dans ce domaine sont surtout le fait d'instituts nationaux ou universitaires (Wallingford en Grande Bretagne, Wageningen aux Pays Bas, Fort Collins aux Etats Unis) où des équipes importantes comportant des statisticiens (Hosking, Buishand...) sont impliquées dans des problèmes hydrologiques.

La seule structure équivalente en France est celle d'E.D.F., avec des statisticiens (Bernier, Duband, Morlat ...) qui ont souvent des responsabilités opérationnelles comme la gestion d'un réseau de mesure destiné aux prévisions d'apports aux barrages. Un certain nombre de méthodes statistiques récentes et novatrices viennent de cette équipe (Loi des fuites, Gradex, DPFT).

En ce qui concerne les hydrologues de l'ORSTOM auxquels mon passé et mon travail me permettent de m'assimiler, je crois qu'il leur manque une structure d'appui qui pourrait leur venir en aide quand se posent des problèmes statistiques complexes. L'expérience montre que les universitaires, confrontés à leurs propres axes de recherche et à leurs comités d'évaluation, peuvent rarement jouer ce rôle qui pourrait être tenu par un statisticien travaillant au sein d'une équipe impliquée dans des problèmes hydrologiques. Ceci, en tenant compte que, même à l'ORSTOM, les problèmes hydrologiques auront de plus en plus tendance à associer qualité et quantité.

<sup>a</sup>Laboratoire d'hydrologie et modélisation U.S.T.L. 34095 Montpellier cedex 05

Pour qualifier la nature des problèmes rencontrés en hydrologie statistique, j'emprunterai à Morlat (1969) cette citation de Halphen : *"Il ne s'agit ni de problèmes d'hydrologie ni de problèmes de statistique, mais bien réellement de problèmes d'hydrologie statistique qui ne se réduisent pas seulement à l'application de la statistique à l'hydrologie."*

## Rôle des statistiques en Hydrologie

Les lignes qui suivent s'inspirent, pour certaines parties, d'une réflexion de T. Lebel sur l'utilisation des statistiques en hydrologie à l'ORSTOM, pour d'autres, de discussions avec C. Bocquillon, Directeur du Laboratoire d'Hydrologie et Modélisation - Université Montpellier 2.

En hydrologie, on peut considérer, comme T. Lebel, qu'il n'existe des lois physiques véritables que pour un nombre très limité de processus élémentaires qu'on peut rarement isoler comme tels dans la nature.

Avec C. Bocquillon, on peut aller plus loin et dire qu'en hydrologie, il n'y a pas de solutions déterministes. En effet, compte tenu des incertitudes inévitables sur les conditions aux limites au temps  $t$ , l'écriture des équations hydrodynamiques du transfert et de la diffusion d'un point  $A$  vers un point  $B$  voisin, conduit à obtenir en  $B$  au temps  $t + \Delta t$  une distribution statistique d'états possibles. Inversement, un état donné en  $B$  à l'instant  $t$  peut résulter de nombreux antécédents en  $A$  au temps  $t - \Delta t$ . A partir d'un nuage de pollution, on ne peut donc pas remonter, de manière déterministe, au point origine.

Comme pour l'étude de la turbulence en hydraulique, on devrait donc admettre qu'en hydrologie il n'y a pas d'autre approche réaliste qu'une approche statistique. Cette approche est rendue difficile parce que l'hydrologie est une science historique : les événements sont rarement reproductibles et l'observation de telle crue exceptionnelle, donc riche en information, doit s'accomoder des pannes de tel ou tel appareil, ou d'une courbe de tarage approximative.

Historiquement, la panoplie des outils statistiques utilisés en hydrologie s'est longtemps limitée à deux méthodes de base :

- Etude et ajustement de lois théoriques permettant de formuler analytiquement la distribution de variables aléatoires.
- La régression linéaire simple et multiple.

Pour traiter le premier problème, l'ORSTOM a développé le logiciel DIXLOI. Il vient d'être soumis à une évaluation (Lubes et Masson, 1991) qui montre qu'il y a encore des progrès à faire dans ce domaine.

La mise en oeuvre des régressions linéaires fait appel à des logiciels du commerce. Ces logiciels presse-bouton (la régression progressive, "Stepwise", ne choisit-elle pas elle-même les meilleures variables explicatives ?) ne contribuent guère à améliorer la formation statistique des utilisateurs. On pourrait envisager une formation à l'interprétation des résultats, une initiation aux méthodes qui favorisent la robustesse (Ridge Regression) et aussi une mise en garde contre les corrélations fallacieuses qui laissent penser qu'on peut faire des prévisions alors qu'il n'en est rien.

Les régressions servent en effet souvent à émettre des prévisions à courte échéance. Notons que certains collègues font des prédictions, ce qui, à mon avis, est une erreur. Pour le dictionnaire Larousse, une prédiction résulte d'une loi exacte (les astronomes prédisent les éclipses) ou d'une révélation (Nostradamus). Les méthodes hydrologiques étant entachées d'incertitudes, on ne peut parler, comme en météo, que de prévisions.

Parmi les outils statistiques de développement plus récent, citons :

- Les méthodes d'analyse multivariée autres que la régression (ACP, AFC, analyse hiérarchique). Elles sont pratiquées surtout sur de gros systèmes (Logiciel SAS) pour essayer de synthétiser les tableaux de données constitués par de nombreuses séries d'observations menées en parallèle sur des sites géographiques voisins. Les résultats sont souvent triviaux ou décevants.
- L'étude des champs aléatoires est devenu un domaine d'application important pour modéliser la distribution spatiale des variables climatiques ou des descripteurs hydrologiques du terrain. Les objectifs sont nombreux : Représentativité d'une mesure ponctuelle, calcul d'une moyenne spatiale et des erreurs associées, cartographie...

Les méthodes utilisées sont les méthodes géostatistiques et en particulier le Krigeage. Développé théoriquement par Matheron au début des années 70, la méthode a été vulgarisée dans ses applications par Delhomme en 1978. Depuis cette date, les publications consacrées aux applications du Krigeage n'ont cessé de se multiplier, au fur et à mesure de son utilisation par les grands organismes scientifiques souvent réticents à remplacer leurs méthodes "maison".

Si le Krigeage est un outil commode et objectif de calcul d'une lame d'eau, on peut s'étonner de le trouver encore au coeur de grands programmes de recherche où, à mon avis, il est mal utilisé. En effet, Matheron a développé sa méthode pour estimer la quantité globale (intégrale) de matière utile contenue dans un gisement minier, mais en aucun cas pour estimer une valeur ponctuelle ! (la méthode ne permet d'ailleurs pas de trouver une valeur supérieure aux valeurs mesurées). D'autre part, en analysant les estimations de pluie Krigée, on ne peut que retrouver la méthode d'interpolation : on étudie le Krigeage, pas la pluie, que l'on connaît toujours aussi mal.

Une étude sur le réseau pluviographique du Marché Gare de Rungis, avec des distances maximales de 800 m et des intervalles de temps de 5 minutes à la demi-heure (Bergaoui), a montré qu'à ces échelles, la pluie n'avait aucune structure. Pour quelle distance et quel pas de temps se structure-t-elle à Rungis ? Est-ce la même chose en Afrique ? N'est-il pas prétentieux d'étudier tout de suite le phénomène à 2 dimensions (Lois Surface-Hauteur-durée-fréquence). Pourquoi ne pas commencer, comme en écologie, par un transect et l'étude des lois distance-Hauteur-durée-fréquence.

A ces études des champs aléatoires, je pense qu'on pourrait rattacher le problème de la régionalisation. Deux vecteurs régionaux sont en concurrence à l'ORSTOM.

L'un est principalement destiné à la critique des données. J'ai pu vérifier sur un exemple que des méthodes plus simples donnaient les mêmes résultats. L'autre manque pour l'instant d'explications théoriques clairement écrites.

Un autre domaine de l'hydrologie statistique est celui des séries chronologiques. L'objectif est le plus souvent de créer des séries artificielles pour simuler le fonctionnement d'aménagements, plus rarement il est d'émettre des prévisions.

Si les modèles autorégressifs se montrent souvent capables de bien décrire les séries hydrologiques, on peut s'interroger sur l'intérêt de la composante "moyenne mobile" des modèles de Box et Jenkins (ARMA, ARIMA, SARIMA). Par stagiaire marocain interposé, j'ai pu constater les risques que prennent des enseignants non hydrologues recommandant ces modèles, ainsi que l'utilisation de logiciels presse-bouton (MINITAB). Les modèles autorégressifs, associés à un filtrage linéaire optimal (Kalman) restent des valeurs sûres pour les prévisions hydrologiques.

En matière de séries chronologiques, la tendance est à l'utilisation de méthodes de désagrégation. Le processus est étudié sur un grand pas de temps (année), puis désagrégé en saisons, puis en mois, puis en jours, ... Ceci permet de respecter les lois des variables considérées sur les différents pas de temps.

## Conclusion

On peut dire que les statistiques interviennent à toutes les étapes des études hydrologiques, depuis le recueil des données et leur critique, pendant l'élaboration des modèles et jusqu'à l'évaluation de l'impact d'aménagements.

Il est trivial de rappeler que pour être efficaces, les méthodes statistiques doivent être appliquées avec discernement, ce qui suppose, à mon avis, d'être impliqué dans les problèmes hydrologiques au sein d'une équipe d'hydrologues. Reste un problème d'enseignement et de formation, mais dans ce domaine, je me refuse d'être juge et partie.

## Références bibliographiques

- Bergaoui M. 1987. Etude de la distribution spatio-temporelle des précipitations à des échelles fines de temps et d'espace. Thèse de doctorat Mécanique UM2.
- Delhomme J.P. 1978. Application de la théorie des variables régionalisées dans les sciences de l'eau. Bulletin du BRGM (2) III, 4. 1978.
- Lubes H. et Masson J.M. 1991. Etude d'évaluation de logiciels d'ajustements de lois statistiques sur des variables hydrologiques. Laboratoire d'Hydrologie ORSTOM - L.H.M. UM2.
- Matheron G. 1970. La théorie des variables régionalisées et ses applications. Les cahiers du C.M.M., Fascicule 5, Ecole des mines de Paris.
- Morlat G. 1969. Problèmes statistiques en hydrologie. Séance organisée par l'Association Internationale pour la Statistique dans les Sciences Physiques, Londres, 9 septembre 1969.