POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

2500 chemin Polytechnique
Montréal (Québec)
H3T 1J4 CANADA

Rapport sur la dissertation de thèse de doctorat de Mathieu Lucas
INRAE Villeurbanne, 3 juin 2023

Résumé

La dissertation de M. Lucas, intitulée “Comment valoriser les données anciennes pour l’analyse fréquentielle des crues: application au Rhône à Beaucaire de 1500 à 2020”, porte sur l’exploitation de données indirectes antérieures à la mise en place de mesures systématiques des niveaux d’eau et des jaugeages des rivières, e.g., les repères de crues historiques, avec une prise en compte rigoureuse des incertitudes associées. Elle s’articule autour de trois contributions: (1) un bilan des données de crue disponibles et une analyse approfondie des changements morphologiques du site d’étude du Rhône à Beaucaire, (2) un article accepté avec révisions mineures dans Journal of Hydrology portant (a) sur la modélisation des courbes de tarage en incorporant la présence de plusieurs sources d’incertitude à l’aide d’une approche bayésienne ce qui permet de simuler des séries de débit à un pas de temps journalier sans valeur manquante et (b) l’analyse fréquentielle des séries de débit simulées avec une analyse de sensibilité de la longueur de la série sur l’incertitude des estimations (e.g., des niveaux de retour élevés) et (3) la comparaison de modèles emboîtés prenant en compte l’incertitude de différentes quantités, telles que le niveau du seuil de perception, issues de l’utilisation des données historiques afin d’identifier la prépondérance des différentes sources d’incertitude.

Commentaires

Mes commentaires sur la dissertation s’organisent selon les trois contributions énumérées ci-dessus.

<u>(1) Bilan des données de crue disponibles</u>

- J’ai trouvé le chapitre 1 un peu long (56 pages) et, il me semble, un peu déséquilibré en ce sens par rapport aux deux autres chapitres (d’une trentaine de pages chacun).
- On trouve à plusieurs reprises des renvois vers les chapitres ultérieurs ce qui ne rend pas, selon moi, la lecture très fluide. Par exemple, il est, à plusieurs reprises, question de la série de débit journalière simulée dans le chapitre 2.
- J’ai regretté de ne pas voir davantage développé la perspective de couplage entre un algorithme Monte Carlo et le modèle hydraulique afin de propager les incertitudes.
- Partout dans ce chapitre, et dans le reste de la dissertation, j’étais mal à l’aise avec l’utilisation de l’expression “série de débit continue”. En effet, au sens mathématique du

terme, “continu” désigne une variable ou un processus qui prend ses valeurs de façon continue, dans les réels pour une variable, dans un espace de dimension infini pour un processus. Dans la dissertation, on entend plutôt, il me semble, une série à pas de temps discret (journalier) sans valeurs manquantes.

(2) Modélisation des courbes de tarage et analyse fréquentielle des débits

- Concernant la procédure de segmentation basée sur les résidus d’une courbe de tarage de référence, je n’ai pas compris d’où proviennent les connaissances *a priori* sur:
  - la moyenne des résidus pour chaque sous-période;
  - le nombre de segments à considérer.
- Concernant l’équation (2.1) qui décrit les incertitudes des hauteurs d’eau mesurées, pourrait-on considérer des erreurs multiplicatives pour certaines d’entre elles plutôt que de les considérer toutes comme étant additive?

  Autrement dit, serait-il possible de motiver l’utilisation d’erreurs additives pour tous les types d’erreurs?
- J’aurais aimé voir plus de détails sur l’élaboration des distributions *a priori* des paramètres de la courbe de tarage.
- Concernant les jaugeages douteux à Pont de Beaucaire, quels critères ont-ils été utilisés pour les écarter?

  N’aurait-il pas été possible d’augmenter l’incertitude pour ces jaugeages plutôt que de les éliminer?
- Concernant l’identification des périodes de stabilité dans la relation hauteur-débit, y a-t-il une procédure pour garantir qu’il y a un mimimun d’observations dans chaque période?
- À partir de la p.63 et par la suite, il est question de contrôles hydrauliques ou contrôles affectant les relations hauteur-débit. J’ai trouvé que ce n’était pas très clair, ce que l’on entendait par ces contrôles. Est-ce que ce sont les paramètres de la courbe de tarage ou plutôt une quantité ayant une interprétation hydraulique qui peut regrouper plusieurs paramètres du modèle? Je pense que ce serait bien d’expliquer ce qu’est un contrôle hydraulique, quels sont les contrôles hydrauliques considérés et quels paramètres du modèle sont reliés à chacun des contrôles. Il y a des éléments de clarification par la suite mais il me semblerait avantageux, pour faciliter la lecture, de définir l’ensemble des éléments dès qu’ils sont abordés.
- Concernant la figure 2.8, j’ai trouvé qu’elle était très dense et qu’il gagnerait à y avoir plus d’explications.
- Concernant la figure 2.9, comment est calculée la différence entre les bornes d’incertitude à 95% et les observations originales?
- Concernant la figure 2.10, il me semble qu’il manque d’explications pour bien comprendre comment la partition en les différentes sources d’incertitude est obtenue.

- À propos des tests de Mann-Kendall pour déterminer la présence de tendances, le test est effectué 500 fois avec $\alpha = 0.05$, la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie. En principe, le rejet de l'hypothèse nulle devrait donc se produire $0.05 \times 500 = 25$ fois par simple malchance. Est-ce que cela est cohérent avec les résultats obtenus dans les analyses, i.e., environ 80% des fois, l'hypothèse nulle n'est pas rejetée?

- Au lieu d'envisager une approche régionale qui semble difficile à mettre en oeuvre, serait-il possible de développer une approche de régression, où certains des paramètres du modèle sont définis comme des fonctions de covariables tels que des covariables physiographiques? Ce pourrait être une piste pour rendre l'estimation du paramètre de forme plus robuste.

- Devant les difficultés d'analyse décrites, je me suis demandé s'il ne serait pas possible de créer des cas d'études synthétiques idéalisés à l'aide de modèles hydrauliques afin de tester certaines hypothèse de façon plus systématique. Par exemple, dans quels cas de figure des séries de hauteur d'eau plus longues sont utiles si (1) les hauteurs d'eau sont incertaines (de façon contrôlée, e.g., en ajoutant un bruit aléatoire) et (2) les jaugeages sont disponibles avec une certaine densité (e.g., à chaque pas de temps, un pas de temps sur deux, certaines périodes oui d'autres non, . . .).

- Je n'ai pas compris comment le modèle des courbes de tarage est utilisé pour générer une série complète de débit, en particulier, comment la simulation procède-t-elle pendant les années sans mesures de hauteurs d'eau correspondant à la construction de la station hydro-électrique?

- J'aurais bien aimé voir les équations pour le modèle bayésien des courbes de tarage multi-périodes et plus d'explication sur l'obtention de la distribution *a posteriori* ainsi que des garanties sur la convergence de la chaîne MCMC.

(3) Modélisation et comparaison des sources d'incertitude des données historiques

- À la page 93, r jeux de paramètres sont extraits de la distribution *a posteriori*. Il s'agit, je suppose, d'un tirage aléatoire qui est représentatif de la distribution *a posteriori*. J'aurais aimé avoir un peu d'information pour justifier le nombre de tirage aléatoire choisi, par exemple en raison de la forme ou du support de la distribution *a posteriori*.

- Dans l'équation (3.10), je n'ai pas compris si les bornes y_i^{inf} et y_i^{sup} sont supposées connues ou bien si elles font partie de l'inférence.

- Au niveau de l'équation (3.12), j'aurais bien aimé des explications sur la logique sous-jacente de cette définition des fréquences empiriques.

- À la sous-section 3.3.2, on s'intéresse à vérifier la stationnarité des débits max annuels sur la période régulière (appelée "continue" dans la dissertation). Du fait qu'il a été nécessaire de découper la période en plusieurs segments pour faire l'ajustement des courbes de tarage, cela suppose qu'il y a eu des changements importants dans la relation hauteur-jaugeage. On pourrait penser, il me semble, que d'emblée, la série des jaugeages et celle des maxima annuels serait non-stationnaire. Sinon, comment réconcilier une hypothèse de stationnarité des maxima annuels avec les courbes de tarage sur des segments multiples? J'aurais aimé voir une discussion à ce sujet.

- Tout d'un coup, le concept d'homogénéité est utilisé et il semble être employé à la place du concept de stationnarité. Je trouve que l'homogénéité est une notion assez floue et qu'il est préférable de ne parler que de stationnarité, un concept rigoureusement définie en statistiques.

- À la figure 3.3, on compare le nombre de fois où le débit a dépassé un seuil S_3 versus le nombre de fois où il a dépassé un seuil S_4 avec $S_3 < S_4$. On a donc forcément une imbrication des résultats, i.e., si le débit a dépassé S_4, il a forcément dépassé aussi S_3. Je ne suis pas sûre du gain à regarder ces deux séries de nombre d'occurrences qui ne sont clairement pas indépendantes.

- Concernant les comportements des queues de distributions dans le cadre de la théorie des valeurs extrêmes, la distribution de Weibull est décrite comme étant bornée supérieurement (ce qui est le cas de la loi Uniforme notamment) et non pas comme étant à queue légère. C'est plutôt la loi de Gumbel qui est la représentante des lois à queue légère (décroissante exponentielle comme la loi normale).

- À la page 107, j'aurais aimé voir des explications sur les liens entre la présence de corrélation entre les paramètres d'un modèle et la réduction en incertitude des estimateurs.

- Au vue de la difficulté à identifier t^* (la distribution *a posteriori* est très large), il me semble que ce doit être parce que la vraisemblance est peu sensible à ce paramètre. Cela ne me paraîtrait pas très étonnant car, la vraisemblance ne dispose pas d'information, par définition, avant la première occurrence de crue et n'a donc pas de moyen de favoriser une date plutôt qu'une autre, parmi les années antérieures à la première occurrence de crue. Si c'est bien le cas, est-ce que ce n'est pas alors perdu d'avance d'essayer d'estimer t^*? Par exemple, je crois que la vraisemblance de le loi binômiale ne doit pas être très sensible au nombre d'essai lorsque celui-ci devient grand, puisque dans ce cas, elle peut être approximée par la loi normale.

- Par ailleurs, la probabilité de succès de la loi binômiale est de plus en plus petite lorsque le niveau du seuil de perception augmente. On se trouve alors à vouloir estimer la probabilité d'événements rares. J'aurais aimé avoir des éléments pour justifier l'approche proposée dans ce cadre, autrement dit, est-ce que l'utilisation de la loi binômiale est adaptée en cas de faibles probabilités de succès?

- De manière plus générale et n'étant pas une modélisatrice avérée de l'approche bayésienne, j'aurais aimé avoir des informations sur la façon dont on doit ou peut s'y prendre pour déterminer des distributions *a priori* informatives? Est-ce que ces informations peuvent reposer sur les données ou doivent provenir plutôt de connaissances d'expert.e.s?

- Autre question générale, comment faire pour prendre une décision en tenant compte de l'incertitude? Par exemple, pour dimensionner une digue à l'aide d'une distribution *a posteriori*, comment s'y prendre? On se limite à la valeur postmax? Et que faire en présence de non-stationnarité, notamment celle qui est potentiellement induite pas le changement climatique?

Évaluation globale

Je trouve la question centrale de la dissertation de thèse de M. Lucas très intéressante: la connaissance de crues historiques antérieures à la mise en place d'instrumentation fournissant des mesures systématiques peut transformer l'estimation du risques d'inondations. En général, les séries de mesure, en raison de leur longueur limitée, contiennent peut d'événements extrêmes ce qui entraîne une sous-estimation de la probabilité de ces événements. De plus, l'incertitude associée à cette probabilité estimée est très élevée. Ceci est bien illustrée dans le chapitre 3 de la dissertion lorsque des estimations avec la série tronquée sont comparées à celles de la série complète. Étant donné l'énorme travail d'archivage, il est tout naturel de se poser la question de comment incorporer ces données d'archives afin d'affiner l'estimation des risques d'inondation.

Il faut d'abord faire le pont entre le travail d'archivage des historien.ne.s et les approches hydro-statistiques. C'est l'objet du premier chapitre qui, non seulement fait le point sur les données disponibles à Beaucaire, mais propose des analyses préliminaires pour comprendre l'influence de l'évolution du contexte hydrologique sur les hauteurs d'eau et les débits. Non seulement j'ai trouvé ce travail de recensement historique très intéressant, mais il me semble qu'il est aussi indispensable pour savoir comment exploiter les données historiques dans la modélisation hydro-statistique. J'ai trouvé que l'analyse fréquentielle intégrant les différentes sources d'incertitude dans la chaîne de modélisation de façon rigoureuse, contenue dans le chapitre 2, donne l'exemple et pourra servir de base de comparaison à d'autres propositions et variantes. À ma connaissance, les travaux existants ne considèrent qu'un nombre assez limité de sources d'incertitude. De plus, j'ai apprécié la façon dont la prépondérance de ces sources d'incertitudes est analysée en créant des scénarios plutôt que de recourir à des méthodes traditionnelles (e.g., l'analyse de variance). Enfin, il me semble que les modélisations proposées au chapitre 3 sont prometteuses en permettant, pour le site du Rhône à Beaucaire, l'incorporation des données historiques avec prise en compte de l'incertitude. Ces analyses mettent en évidence dans quelle mesure les données historiques sont utiles à l'analyse fréquentielle et quelles sont les faiblesses et les directions d'amélioration à considérer.

Recommendation

J'ai donc globalement une évaluation très positive et je recommende vivement la soutenance de la thèse de doctorat de Mathieu Lucas. Mes commentaires énoncés ci-dessus pourront servir de base à la discussion lors de la soutenance. Je fournis également le manuscrit annoté avec des suggestions de corrections mineures.

Julie Carreau, Ph.D. HDR
Professeure adjointe
Département de mathématiques et de génie industriel
julie.carreau@polymtl.ca