

SUJET DE STAGE 2019 – 6 mois

« Analyse et validation de la modélisation statistique des maxima annuels de pluies. »

Contexte :

Le laboratoire de Modélisation des Transferts dans l'Environnement (LMTE) du CEA de Cadarache est en charge de réaliser les études réglementaires d'évaluation du risque inondation externe par remontée de la nappe phréatique, au droit des installations du centre. En vue de réaliser ces études, le LMTE se base sur les recommandations du guide n°13 de l'Autorité de Sûreté nucléaire relatif à la protection des installations nucléaires de base contre les inondations. La démarche globale consiste à évaluer l'impact de précipitations extrêmes sur les niveaux de nappe, grâce à leur « injection » dans une chronique piézométrique représentative des fluctuations de la nappe, reconstituée à l'aide d'un simulateur numérique « pluies / niveaux ».

Préalablement à l'utilisation du simulateur « pluies / niveaux », cette démarche nécessite le calcul de pluies centennales¹, représentatives du site considéré. On dispose dans certains cas de plusieurs années de chroniques météorologiques relevées sur le site et à partir desquelles on souhaite estimer les valeurs des pluies centennales. Ces pluies étant par définition des événements rares, leur estimation à partir des chroniques nécessite l'utilisation de méthodes et outils statistiques avancés, fondés sur la théorie des valeurs extrêmes.

Pour cela, le LEMS (Laboratoire d'Etudes des Systèmes Innovants) en collaboration avec le LMTE a développé des scripts sous le logiciel R, permettant d'estimer les lames d'eau de période de retour centennale (quantiles à 99%) et des intervalles de confiance à 95% associés, à partir des chroniques pluviométriques enregistrées au pas de temps horaire sur le site de Cadarache (environ 60 années de chroniques disponibles). Plusieurs méthodes statistiques fondées sur la théorie des valeurs extrêmes ont ainsi été implémentées et testées :

- La méthode d'échantillonnage des maxima annuels (AMS pour « Annual Maximum Series ») qui consiste à retenir dans une chronique temporelle la plus forte valeur observée chaque année. La loi de l'échantillon est ensuite approchée par une loi généralisée des valeurs extrêmes (loi GEV).
- La méthode d'échantillonnage POT (« Peak Over Threshold ») qui consiste à ne retenir que les valeurs au-dessus d'un seuil donné. La loi de l'échantillon est alors approchée par une loi généralisée de Pareto (loi GP).

Dans les deux cas, les paramètres des lois GEV ou GP sont estimés à partir des données soit par maximisation de la vraisemblance, soit par la méthode des moments. Des intervalles de confiance pour ces paramètres sont aussi calculés. Une fois ajustés, les modèles statistiques GEV ou GP sont ensuite utilisés pour prédire les quantiles à 99% (pluies centennales), assortis d'intervalles de confiance (résultant des incertitudes sur les paramètres des lois).

Dans les études réalisées au LMTE, l'inertie de la nappe phréatique nécessite de fournir en entrée du simulateur « pluies / niveaux » les lames d'eau centennales correspondant à différentes durées de cumuls de pluies (pluie journalière, décadaire, cumulée sur vingt jours, ...). La méthode d'estimation précédemment détaillée pour une pluie journalière est donc étendue à une pluie cumulée et moyennée sur un ou plusieurs jours (quantile à 99% assorti de son intervalle de confiance à 95%). Pour constituer l'échantillon des pluies cumulées sur k jours (avec méthode AMS ou POT), les données journalières sont cumulées et moyennées sur k jours par fenêtre glissante avant d'extraire les maxima annuels (ou les valeurs dépassant un seuil donné pour la méthode POT). En procédant ainsi, la corrélation des données est accentuée (corrélation augmentant avec k) et le maximum annuel extrait s'apparente de moins en moins à l'observation de la valeur maximale d'une série de répliques identiquement et indépendamment distribuées d'une variable aléatoire (hypothèse nécessaire pour appliquer la théorie des valeurs extrêmes). La convergence vers une loi des extrêmes n'est donc plus assurée. Une récente analyse sur les données du site de Cadarache a d'ailleurs

¹ dont la probabilité annuelle d'occurrence ou de dépassement est de 1%.

révélé les problèmes d'adéquation des lois des extrêmes pour des échantillons correspondant à certaines durées de cumuls de pluies, au-delà de dix jours par exemple.

Il est donc nécessaire d'identifier la durée de cumul maximale pour laquelle l'approximation par une loi d'extremum généralisée reste valable et vérifier que la méthode de constitution de l'échantillon permet de conserver l'indépendance des maxima annuels. Une comparaison avec une méthode de constitution par blocs indépendants (*i.e.* sans recouvrement) pourra aussi être réalisée.

Objectifs du stage :

L'objectif de ce stage est donc de déterminer la durée de cumul de pluies maximale pouvant être considérée, telle que les propriétés statistiques des échantillons (à savoir indépendance, stationnarité et homogénéité) sont vérifiées et l'approximation par une loi des extrêmes demeure valide. Les scripts R actuels estimant les lois et implémentant quelques tests statistiques, ainsi que les jeux de données de chroniques pluviométriques seront fournis. Il conviendra alors de :

- prendre en main les scripts R développés par le LEMS et le LMTE ;
- réaliser la bibliographie relative à la théorie des valeurs extrêmes et aux différents tests statistiques ;
- reprendre les scripts afin de compléter certains tests statistiques (si besoin), d'implémenter la méthode d'échantillonnage par blocs indépendants. Les deux méthodes d'échantillonnage seront ensuite testées et comparées afin d'identifier les durées de cumul statistiquement acceptables dans les deux cas.

Discipline : Mathématiques appliquées, Statistiques et probabilités

Compétence software : Langage R

Formation souhaitée : BAC +5 (Stage pour dernière année d'Ecole d'ingénieur ou Master 2)

Localisation : Laboratoire d'Etudes et Modélisation des Systèmes (DER/SESI/LEMS), CEA de Cadarache

Encadrant CEA et contact : Amandine MARREL, Sabine BODEI – [amandine.marrel\[at\]cea.fr](mailto:amandine.marrel[at]cea.fr)