

Mini-symposium StatEnv

Statistique et environnement

Résumé

L'objectif de ce mini-symposium est de montrer des avancées récentes des statistiques appliquées à l'environnement. Les thématiques développées concernent notamment les descripteurs des données environnementales et les mesures de risques. Il s'agit d'outils et de concepts à même de rendre compte de la complexité des données et phénomènes environnementaux : prise en compte de la dimension spatiale et temporelle, modélisation de la dépendance entre les sites, mise en place d'outils de prévention des risques.

Organisateur(s)

1. **Céline Vial**, Polytech Lyon/ ICJ.
2. **Véronique Maume-Deschamps**, ISFA/ICJ.

Liste des orateurs

1. **Alexandre Brouste**, Laboratoire Manceau de Mathématiques
Titre : Estimation of wind turbine production.
2. **Jean-Baptiste Aubin**, INSA/ICJ
Titre : Le relascope : un nouvel outil de visualisation et de test—Un exemple d'application à des données environnementales.
3. **Edith Gabriel**, Laboratoire de Mathématiques d'Avignon
Titre : Prédire l'intensité locale d'un processus ponctuel partiellement observé – Application à l'estimation de la distribution d'espèces en écologie.
4. **Jonathan El Methi**, MAP5
Titre : Estimation non-paramétrique de mesures de risque pour des lois conditionnelles à queues lourdes avec application à des extrêmes pluviométriques.

Céline Vial, Polytech Lyon/Institut Camille Jordan, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France, cvial@math.univ-lyon1.fr

Véronique Maume-Deschamps, ISFA/Institut Camille Jordan, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France, veronique.maume@univ-lyon1.fr

Alexandre Brouste, Université du Maine, Département de Mathématiques Avenue Olivier Messiaen, 72000 LE MANS Cedex 9, Alexandre.Brouste@univ-lemans.fr

Jean-Baptiste Aubin, INSA-Lyon et ICJ, Bat. Léonard de Vinci, 20, av. J. Capelle, F 69621 Villeurbanne, FRANCE, jean-baptiste.aubin@insa-lyon.fr

Edith Gabriel, IUT, Département STatistique et Informatique Décisionnelle, Site Agroparc, BP 61207, 84 911 Avignon Cedex 9, France, edith.gabriel@univ-avignon.fr

Jonathan El Methni, Université Paris Descartes, Laboratoire MAP5, UMR CNRS 8145, 45 rue des Saints Pères, 75006 Paris, France, jonathan.el-methni@parisdescartes.fr

1 Estimation of wind turbine production

Alexandre Brouste

Renewable energies are clean but intermittent. Their structural uncertainties have to be handled for a successful energetic transition. Practitioners consider several uncertainties to estimate the energy produced by a wind farm. Two industrial problems can be studied : estimation of the annual wind farm production in the investment phase (for planning expansion) and the short-term forecasting in the operational phase (for dispatching in the grid, for hedging). We will show how « classical » methods (parameter estimation in diffusion processes, GLM) can be used in this context.

2 Le relascope : un nouvel outil de visualisation et de test – Un exemple d’application à des données environnementales

Jean-Baptiste Aubin et Samuela Leoni-Aubin

Mots clés : dependence measure, independence test, longest run, test on heteroscedasticity.

Un nouvel outil graphique, le *relascope*, est présenté. Le relascope possède plusieurs applications : par exemple, des test d’indépendance entre variables ou d’homoscédasticité pour les processus peuvent être issus de son utilisation. Le *relascope* est construit de la façon suivante. Considérons n réalisations de $(X, Y) : (x_1, y_1), \dots, (x_n, y_n)$ où X et Y sont des variables aléatoires univariées continues. Dans le nuage de points, pour chaque réel δ , la longueur de la plus longue suite d’observations consécutivement du même côté de la droite d’équation $y = \delta$, noté $L_n(\delta)$, est calculée. Le *relascope* est le graphe obtenu en plaç δ en abscisse et $L_n(\delta)$ en ordonnée.

Des tests formels peuvent être déduits du relascope. L’idée soutenant ces tests est que, si Y dépend de X , alors, il existe δ pour lequel le nombre d’observations consécutivement (pour X) supérieures/inférieures à δ , $L_n(\delta)$, est significativement différente de la quantité analogue sous l’hypothèse nulle d’indépendance \mathcal{H}_0 . Quand \mathcal{H}_0 est vraie, la distribution exacte de $L_n(\delta)$ et une p -valeur associée $p_n(\delta)$ sont déduites pour chaque δ [1]. Le minimum de ces p -valeurs : $p_{min} = \min_{\delta} p_n(\delta)$ est considéré. p_{min} est alors comparé à la distribution empirique de p_{min} ’s obtenus en considérant des échantillons de la forme $(x_1, y_{p(1)}), \dots, (x_n, y_{p(n)})$ où $(p(1), \dots, p(n))$ est une permutation $(1, \dots, n)$. Si la probabilité d’être plus extrême que p_{min} est inférieur au niveau nominal α , alors l’hypothèse d’indépendance est rejetée. Cette méthode sera développée sur un jeu de données réelles croisant des concentrations en divers polluants chimiques et des caractéristiques physico-chimiques d’un écoulement en entrée d’une station d’épuration.

L’homoscédasticité d’un processus stationnaire peut être testée par la même méthode en prenant comme variable X le temps et comme variable Y le carré des résidus.

Remerciements. Les données proviennent du projet SIPIBEL (www.sipibel.org).

3 Prédire l’intensité locale d’un processus ponctuel partiellement observé – Application à l’estimation de la distribution d’espèces en écologie

Edith Gabriel, Florent Bonneau, Pascal Monestiez et Joël Chadoeuf

Prédire l’intensité d’un processus ponctuel est une question encore très peu traitée dans la littérature bien qu’omniprésente en écologie. Lorsque l’on s’intéresse à la biodiversité, un enjeu d’intérêt est de connaître ses changements, ses variations spatiales. Cependant, la dépendance spatiale en écologie est complexe car elle peut s’exprimer à plusieurs échelles et doit être prise en compte lors de l’estimation de l’intensité locale, c’est par exemple le cas pour la distribution spatiale d’une espèce. Beaucoup de données ponctuelles en écologie sont clusterisées, car liées à l’hétérogénéité environnementale, et ne sont que partiellement observées à cause du compromis fait entre la taille de la zone d’étude et les temps et moment d’observation. Notre objectif est de répondre aux questions suivantes :

- Ayant l'ensemble des observations du processus ponctuel dans une fenêtre donnée, comment estimer localement l'intensité du processus? Y a-t-il une taille optimale pour la grille d'interpolation?
- Si le processus est échantillonné, i.e. ayant une observation partielle du processus, comment prédire l'intensité locale dans des zones non échantillonnées?

Afin de répondre à ces questions, nous considérons un semis de points observé dans une grande fenêtre. Nous supposons que le processus sous-jacent est stationnaire, isotrope et obtenu par un processus à faible dépendance dont un paramètre est dirigé par un champ aléatoire stationnaire à une échelle supérieure. Dans un objectif de prédire l'intensité locale du processus ponctuel dans des zones non-échantillonnées, notre approche consiste à définir les caractéristiques du champ aléatoire à partir de celles du processus ponctuel, puis à interpoler l'intensité locale par un krigeage ordinaire revisité. L'originalité et l'intérêt de cette démarche est que le variogramme est défini en fonction de l'intensité globale et de la fonction de corrélation de paire du processus, et donc que les poids de krigeage dépendent de la structure du processus ponctuel. Nous sommes alors en mesure d'utiliser l'ensemble des points pour prédire localement (par définition du krigeage) et d'utiliser l'information à fine échelle du processus ponctuel. Ainsi nous unifions deux approches jusque là utilisées de façon orthogonale en pratique.

4 Estimation non-paramétrique de mesures de risque pour des lois conditionnelles à queues lourdes avec application à des extrêmes pluviométriques.

Jonathan El Methni , Laurent Gardes & Stéphane Girard

L'étude et la maîtrise des risques extrêmes est d'un grand intérêt afin d'anticiper des catastrophes et ainsi permettre l'aménagement du territoire. Cette mesure de risque est appelée Value-at-Risk et représente le quantile d'ordre $p \in]0, 1[$ de la fonction de survie de la variable aléatoire d'intérêt. Bien que cette mesure de risque soit la plus utilisée elle ne fournit qu'une information ponctuelle et donc sous estime l'impact du sinistre. Dans le but de lui trouver une alternative, d'autres mesures de risques ont été proposées afin de prendre en compte les incertitudes sur les événements extrêmes.

Nos travaux qui ont fait l'objet de l'article Elmethni *et al.* (2014) [2] consistent à introduire et à estimer une nouvelle mesure de risque appelé Conditional Tail Moment. Elle est définie comme le moment d'ordre $a > 0$ de la loi de la variable aléatoire d'intérêt au-delà du quantile d'ordre p . Estimer le Conditional Tail Moment permet d'estimer toutes les mesures de risque basées sur les moments conditionnels telles que la Conditional Tail Expectation, la Conditional Value-at-Risk ou la Conditional Tail Variance. On s'intéresse à l'estimation de ces mesures de risque dans un cadre extrême c'est-à-dire lorsque p tend vers 0 lorsque la taille de l'échantillon augmente. On suppose également que la loi de la variable aléatoire d'intérêt est à queue lourde et qu'elle dépend d'une covariable. Les estimateurs proposés combinent des méthodes d'estimation non-paramétrique à noyau avec des méthodes issues de la statistique des valeurs extrêmes. On établira le comportement asymptotique de nos estimateurs.

On appliquera nos travaux à un jeu de données pluviométriques fourni par le Laboratoire d'étude des Transferts en Hydrologie et Environnement de Grenoble. On dispose des hauteurs de pluies journalières en millimètres entre les années 1958 et 2000 sur 523 stations situées dans la région des Cévennes-Vivarais. Dans ce contexte, les variables d'intérêt sont les précipitations journalières et les covariables sont les coordonnées géographiques des stations. En conclusion, on pourra estimer des mesures de risque extrêmes en des sites où l'on ne dispose pas de mesure.

Références

- [1] AUBIN, J.-B., LEONI-AUBIN, S., *A nonparametric lack-of-fit test for heteroscedastic regression models*, Comptes Rendus Mathématique, 349, 2011, 215 - 217..
- [2] J. EL METHNI, L. GARDES, AND S. GIRARD., *Nonparametric estimation of extreme risk measures from conditional heavy-tailed distributions.*, Scandinavian Journal of Statistics, 41(4), 2014, 988-1012.