

Éditorial du numéro spécial : Statistique pour les données spatiales et spatio-temporelles et réseau RESSTE

Title: Editorial of the special issue:
Statistics for spatial and spatio-temporal data and RESSTE Network

Delphine Blanke¹ et Denis Allard²

Comme son nom l'indique, la Statistique spatio-temporelle permet de traiter des données présentant une structure spatiale et dynamique. Typiquement, il peut s'agir de données environnementales ou climatiques, collectées dans différentes stations à des intervalles de temps répétés, mais le schéma d'échantillonnage peut parfois être moins régulier, par exemple le long d'une trajectoire. Les processus étudiés sont souvent complexes, en raison de l'intrication entre espace et temps et des interactions entre les différentes (et souvent nombreuses !) variables impliquées. La prise en compte des dépendances aussi bien spatiales que temporelles représente un des enjeux de ce domaine très actif de la Statistique. Les défis posés par l'analyse et la modélisation de ces données sont donc multiples : ils concernent à la fois leur visualisation, la définition de modèles pertinents, les méthodes d'estimation pour ceux-ci et la mise en œuvre de ces méthodes pour des ensembles de données de (très) grande taille. En lien avec l'essor très rapide de la recherche dans ces domaines, l'objectif de ce numéro spécial est de présenter des méthodes récentes, leurs applications à des jeux de données et leur implémentation sous *R*.

Plusieurs contributions à ce numéro sont issues d'équipes participant au réseau RESSTE (RE-Seau Statistiques pour données Spatio-TEmporelles), créé en 2014 et soutenu par le département Mathématiques et Informatique Appliquées de l'INRA. Ce réseau compte à ce jour autour de 60 membres provenant de 19 équipes de recherche appartenant à des universités et des instituts de recherche, nationaux et internationaux. Il propose une animation scientifique autour des modèles, méthodes et algorithmes pour les données spatio-temporelles, en fédérant des statisticiens de différents horizons et des modélisateurs des autres disciplines ayant à traiter ce type de données, et en confrontant les approches et points de vue provenant de différents domaines des statistiques. Il a pour objectif de contribuer à une plus large diffusion des méthodes en Statistique spatio-temporelle, comme par exemple, avec ce numéro du Journal de la Société Française de Statistique, et à participer à l'émergence d'une communauté scientifique. Il a également pour

¹ LMA EA2151, Université d'Avignon, Campus Jean-Henri Fabre, 301 rue Baruch de Spinoza, BP 21239, 84916 Avignon Cedex 9

E-mail : delphine.blanke@univ-avignon.fr

² BioSP, INRA, 84914 Avignon

E-mail : denis.allard@inra.fr

ambition d'initier de nouveaux fronts de recherche répondant aux enjeux rencontrés par les scientifiques, et de favoriser les collaborations entre équipes de recherche. Une présentation détaillée de ses activités, des ateliers et des exposés effectués dans le cadre du réseau est consultable à l'adresse <http://informatique-mia.inra.fr/resste/>.

Ce numéro spécial débute avec un article de M. Valera-López, A. Pineda et J.R. León "*Application of satellite image to the implementation of two stochastic models for modeling the transport of chlorophyll-a on Lake Valencia (Venezuela)*" qui propose deux modélisations possibles pour le transport de polluants dans le lac Valencia au Venezuela. Les auteurs utilisent un modèle hydrodynamique déterministe pour modéliser le courant de surface et des processus de diffusion pour le transport des polluants. L'article propose une contribution originale avec, en particulier, la mise en œuvre d'un algorithme générant de manière aléatoire la position initiale des polluants à partir d'informations provenant d'images satellitaires. Ainsi la dispersion des polluants peut être simulée depuis le moment où les images ont été capturées par le satellite en prenant à la fois en compte leur position initiale et leur densité initiale dans le lac. Cette méthode permet une bonne approximation de la densité de déchet organique dans le lac. La simulation du transport des polluants est ainsi déterminée par la position initiale des particules et par la densité de pollution autour de celles-ci.

L'article proposé par T. Opitz "*Latent Gaussian modeling and INLA : A review with focus on space-time applications*" porte sur les modèles bayésiens hiérarchiques structurés par un processus gaussien latent. De tels modèles se sont montrés suffisamment riches et flexibles pour caractériser des structures complexes dans les données en grande dimension, spatiales ou spatio-temporelles. Pour calculer les quantités liées aux lois a posteriori, la méthode INLA ("Integrated Nested Laplace Approximation") est une alternative aux méthodes MCMC qui se base sur des approximations de Laplace et des techniques efficaces d'intégration numériques. Cette approche, efficace et rapide, s'appuie sur des structures de dépendance markovienne pour permettre l'utilisation de matrices creuses et ainsi éviter le verrou constitué par des matrices pleines de grande dimension. Dans le cadre d'applications spatio-temporelles, l'article passe en revue les principales notions théoriques et les outils inférentiels liés à la méthode INLA. Un scénario, basé sur des données spatio-temporelles (non gaussiennes) permet d'illustrer l'utilisation pratique du package R-INLA et donne la syntaxe de ses commandes principales.

Dans leur article "*Detecting and modeling multi-scale space-time structures : the case of wildfire occurrences*", É. Gabriel, T. Opitz et F. Bonneau s'intéressent à l'éclosion de feux recensés quotidiennement depuis 1981 dans le sud de la France (Bouches-du-Rhône). Avec la connaissance de la surface brûlée et d'autres covariables (climatiques, occupation des sols) dont ils font une analyse exploratoire, ils adoptent une approche basée sur un modèle de Cox log-gaussien leur permettant d'inclure l'information des covariables et des effets spatio-temporels (via la méthode INLA). Ils modélisent ainsi, à la fois, les variations spatio-temporelles à grande échelle et les interactions locales. La structure d'interaction résiduelle est étudiée grâce à la fonction K spatio-temporelle non homogène. Leur travail permet en particulier de mettre en évidence les effets inhibiteurs, dans l'espace et le temps, des très grandes surfaces brûlées.

Dans l'article de F. Lavancier et P. Rochet "*A tutorial on estimator averaging in spatial point process models*", les auteurs présentent un tutoriel sur une méthode de combinaison d'estimateurs

dont ils étudient les performances sur des modèles standards de statistique spatiale. Les modèles envisagés sont ici : les processus de Poisson inhomogènes, les processus déterminantaux (avec répulsion entre les points), de Thomas (avec présence de *clusters*) ou encore des modèles booléens issus de Poisson homogènes. Pour chacun de ces processus, plusieurs méthodologies existent pour l'estimation des paramètres impliqués et il n'est pas toujours clair de savoir laquelle employer. L'intérêt de l'approche présentée ici est d'étudier une combinaison d'estimateurs, basée sur l'estimation de l'erreur quadratique moyenne des méthodes existantes. Les résultats sont illustrés par des simulations montrant l'apport positif de cette combinaison par rapport au meilleur estimateur initial. Pour chaque exemple, le code pour implémenter les estimateurs avec le logiciel *R* est fourni permettant ainsi de reproduire la méthode utilisée.

La dernière contribution est un article collégial écrit par 12 membres du réseau RESSTE. Il s'intitule "*Analyzing spatio-temporal data with R : Everything you always wanted to know – but were afraid to ask*". Il vient en conclusion de ce numéro spécial avec la perspective d'ouvrir justement de nouvelles pistes dans l'exploration et la prévision de données spatio-temporelles sous *R*, en privilégiant l'approche géostatistique basée sur les processus gaussiens. Cet article présente d'abord les modèles, les méthodes et les techniques utilisés pour l'analyse et la prévision de ces processus. En se basant sur des données réelles de concentration de particules fines (PM₁₀, concentrations relevées quotidiennement pendant un an dans 507 stations de qualité de l'air en France), les auteurs présentent et comparent les packages de *R* les plus complets pour l'analyse des données spatio-temporelles, avec leurs avantages et inconvénients. Toutes les étapes sont illustrées par leurs implémentations sous *R* dans l'article, mais également dans un fichier R Markdown accessible sous forme de matériel supplémentaire à télécharger.

Nous remercions l'ensemble des auteurs ayant contribué à ce numéro spécial ainsi que les rapporteurs pour leur lecture attentive des articles et leurs suggestions pertinentes. Un grand merci également à Gilles Celeux, éditeur en chef, pour la confiance immédiate qu'il nous a accordée et son soutien constant lors de la préparation de ce numéro. Enfin, pour toute question relative à RESSTE, n'hésitez pas à visiter la page web du réseau et à contacter un des membres du bureau !