

# ANALYSIS AND SIMULATION OF SPATIO-TEMPORAL RAINFALL : DOES A CLASSIFICATION OF LOCAL RAINFALL DATA INTO STATISTICALLY HOMOGENEOUS RAINFALL TYPES HELP ?

Etienne Leblois<sup>1</sup> & Jean-Dominique Creutin<sup>2</sup>

<sup>1</sup> *Irstea, Centre de Lyon-Villeurbanne, UR Hydrologie-Hydraulique, etienne.leblois@irstea.fr*

<sup>2</sup> *LTHE, Grenoble, jean-dominique.creutin@ujf-grenoble.fr*

**Résumé.** La pluie est un phénomène spatialement et temporellement variable, sans homogénéité apparente. Or, la majorité des techniques statistiques supposent une homogénéité des données étudiées. Une tentative a donc été faite pour considérer les données issues d'un réseau pluviométrique comme une succession de fragments ressortant d'une famille délimitée de types de pluie homogènes, différant par leur intermittence, leur intensité, leur structure spatiale et temporelle. Une simulation très utilisable en découle. Cependant une autocorrélation interne à chaque de type de pluie et une succession qualitative des types de pluie correctement simulées ne suffisent pas à créer une corrélation temporelle suffisante ; il faut réintroduire la corrélation empirique des données à travers les limites de type, que l'on avait initialement négligée.

**Mots-clés.** Pluie, Homogénéité, Géostatistique, Modèle de Markov Caché, Classification, ...

**Abstract.** Rainfall is a continuous and continuously changing phenomenon with a complex mixture of dry/wet and non-Gaussian rainfall amount/intensities, with no obvious process homogeneity. On the other hand, most statistical techniques rely on the assumption that the data under study is statistically homogeneous. Therefore, an attempt was made to consider data out of a rain-gauge data set as a succession of chunks out of a finite number of homogeneous rainfall types. This proves usable as a basis for simulation. However, the autocorrelation of rainfall within a rainfall type and a correct sequence of simulated rainfall types are not enough to convey a strong enough time structure, and reintroducing the observable correlation of rainfall across rainfall type boundaries seems necessary.

**Keywords.** Rainfall, Homogeneity, Geostatistics, Hidden Markov Model, Classification, ...

## 1 Motivation hydrologique

Le fonctionnement hydrologique d'un territoire implique une hiérarchie de bassins versants emboîtés structurée par le réseau hydrographique. Il est tributaire des entrées atmosphériques : rayonnements de courtes et longues longueurs d'onde, vent, précipitations. Nous présentons dans cette contribution une tentative de description fonctionnelle de l'entrée pluviométrique.

Notre perspective est celle d'hydrologues intéressés à ce qui peut se produire en général. La variabilité des événements passés, indice de ce que la nature est capable de générer, sera exploitée statistiquement (la météorologie à base physique ne sera ici pas convoquée).

Des propriétés comme l'espérance mathématique et la variance de la pluie n'ont rien d'abstrait, car elles représentent la ressource en eau et l'aléa pluviométrique, sous réserve qu'on soit bien à l'échelle des bassins étudiés, ce qui oblige à prêter une grande attention à l'extension des pluies dans l'espace et dans le temps ; la dynamique générale des événements hydrologiques, relativement plus nerveuse dans les petits bassins et apparemment plus lisse dans les grands, en dépend. De même l'évaluation du risque de crues concomitantes sur différents affluents paraît pouvoir s'appuyer sur

l'extension spatiale des pluies fortes – sachant qu'un événement pluvieux tout à fait considérable peut résulter d'intensités modérées mais présentant une grande extension temporelle et spatiale [2, 16, 17]. Notre approche est donc aussi de statistique spatiale, et fait appel aux outils de la géostatistique.

La source première de connaissance est constituée par les données d'observation, pluviométriques ou radar. Nous interprétons ces données à l'aide des outils de la géostatistique, distribution ponctuelle, variogrammes spatiaux et temporels, dans une perspective voyant la pluie comme une fonction aléatoire se déployant dans l'espace [6, 13, 14].

En résumé on cherche une caractérisation des entrées pluviométriques, qui puisse être fonctionnelle : descriptive, elle doit aussi permettre des simulations libres et conditionnées ; pour être hydrologiquement pertinente, on attend qu'elle rende compte, et respecte en simulation, la distribution observée des cumuls, au moins pour une gamme raisonnable de superficies et de durées ; la variabilité saisonnière, enfin, doit être respectée.

## **2 Modèle conceptuel : un mélange temporel de pluies spatialement homogènes**

Il est 5 caractéristiques de la pluie que nous considérerons comme admises par l'ensemble de nos lecteurs : la pluie se déploie dans l'espace et dans le temps ; elle se cumule au sol au cours du temps ; elle est souvent advectée (vent) ; elle présente parfois des intensités très fortes ; elle est intermittente.

Pour parvenir à un schéma suffisamment précis (d'aucun diront suffisamment appauvri) pour que les mathématiques puissent nous venir en aide, nous précisons maintenant notre modèle conceptuel. Il se distingue des caractéristiques précédentes en ce qu'il est un jeu d'hypothèses librement consenti, de fait simplificatrices, éventuellement discutables. Il est le suivant :

1) Homogénéité spatio-temporelle du processus. On entend par là qu'au-dessus d'un domaine d'étude préalablement fixé (suffisamment modeste pour cette hypothèse puisse être réaliste), à tout instant ne se présente qu'un type de pluie.

2) A ce type de pluie en succèdera au bout d'un certain temps un autre, l'alternance entre les différents types de pluie devant être décrite. Les paramètres de cette alternance porteront la variabilité saisonnière, les types de pluie, eux, ne changeant pas.

3) Intermittence. Au sein d'un même type de pluie se présentent des zones pluvieuses et non pluvieuses, qui se combinent pour former un composite. On admettra qu'il y a indépendance, au pas de temps et d'espace élémentaires, des champs d'indicatrice et de pluie non-nulle ; cette hypothèse de travail a été énoncée par Barancourt, et al. 1992 ; elle facilite largement l'évaluation théorique des propriétés attendues du signal simulé [11].

4) Advection. Les nuages générant les pluies sont soumis à une advection par le vent, facteur qui se superpose à leur variabilité propre. L'effet de cette advection est d'étendre les pluies générées sur de plus vastes territoires, tout en lissant les cumuls locaux. En terme statistiques : le signal de pluie temporellement agrégé présentera des variances diminuées et des longueurs de corrélation augmentées.

5) Distributions non gaussiennes. D'évidence empirique les pluies non-nulles (dont les zéros ont été écartés) présentent occasionnellement des valeurs très fortes. Ceci oriente vers des distributions non gaussiennes, d'asymétrie droite. Il ne s'agit pas cependant de lois d'extrêmes, car les petites pluies aussi doivent être décrites.

Ce jeu d'hypothèses, révisable, va permettre de progresser vers une description formalisée et opératoire.

### **3 Principe de la simulation des pluies homogènes**

La simulation d'une pluie homogène s'appuie sur la technique géostatistique classique des bandes tournantes, qui en essence génère des champs Gaussiens de structure spatiale choisie. Le champ de précipitation non nulle est dérivé d'un premier champ Gaussien simulé. On applique à ce champ une distribution calée sur l'ensemble des valeurs de précipitations non-nulles observées. On aura recours à des distributions de type loi de Pareto généralisée (GPD), loi gaussienne inverse (IG) – notre préférée, bien souvent adaptée –, loi gamma, etc... L'opération normalisant ou dénormalisant un champ de valeurs entre la loi normale et une loi utilisateur est classiquement dénommée anamorphose.

Le champ d'indicatrice est obtenu par seuillage dans un deuxième champ Gaussien, simulé indépendamment. Des précautions doivent permettre de tenir compte de l'altération des corrélations entre les champs Gaussiens issus de la technique des bandes tournantes et les champs dérivés (anamorphosée ou seuillée) du monde utilisateur.

Le champ de précipitation final est élaboré comme composite multiplicatif de ces deux champs de précipitation non-nulle et d'indicatrice de pluie.

Nous faisons une distinction explicite entre les séries chronologiques perçues en des emplacements fixes au sol (sous bassins, pluviomètres) et ce qui se passe au niveau de processus atmosphériques soumis à advection. Nous postulons que le processus de génération des précipitations est à décrire au niveau des nuages. Techniquement nous sommes amenés à simuler sur des rétro-trajectoires. L'advection est approchée comme un écoulement horizontal incompressible, très souvent uniforme

Sur la trace de l'advection dans les données pluviométriques, voir [13, 14]. La technique de simulation des pluies homogènes est décrite dans [11]. L'impact de l'advection est discuté dans [3].

### **4 Segmentation des pluies historiques en séquences homogènes**

Les paramètres d'un type de pluie doivent être établis pour une population réputée homogène. Ce sera d'emblée le cas si les données disponibles sont des lames d'eau radar issues d'un événement particulier considéré comme une classe en soi [1, 4, 5, 6]. Mais quand on part d'observations de long terme effectuées sur un réseau pluviométrique, il faut construire cette homogénéité en délimitant des sous-ensembles de travail au sein du continuum. En pratique nous segmentons la chronologie en classant chaque pas de temps, et obtenons un calendrier d'occurrence de types de pluie.

Il existe de nombreuses techniques de classification, toutes plus séduisantes les unes que les autres. Du point de vue de la critique méthodologique on notera

- que la caractérisation numérique de l'objet (choix des descripteurs en sémantique, support spatial et temporel) et de la métrique utilisée sont fondamentales pour la pertinence des classifications obtenues. Cette interaction entre le choix des descripteurs de classification et la thématique d'étude est la raison fondamentale du caractère très peu opérant des classifications préexistantes, exogènes, à traiter de questions nouvelles ;

- que dans notre cas la classification ne réalise pas une mise en contraste d'individus radicalement et qualitativement différents, mais plutôt un pavage, une segmentation au sein d'un espace continu ; la raffinement de la classification à la fois raisonnable et utile dépend surtout de la force travail dont on dispose pour le traitement des résultats, sous réserve de la masse de données dont on dispose (les classes obtenues doivent contenir suffisamment d'individus pour que leur propriétés statistiques soient identifiables) ;

- les deux points précédents étant bien repérés, l'algorithme de classification est assez secondaire quant à la qualité finale de la classification. Nous apprécions la classification non supervisée de Kohonen (Self Organizing Map, SOM) pour sa particularité qui est de ventiler les individus dans un

espace de classes ordonné sur un espace graphique prédéfini, en privilégiant la continuité entre classes voisines. Ce type de résultat est interprétable par le géographe comme une carte, par le dynamiqueur comme un plan de phase, par le statisticien comme une classification. Cet aspect en fait une technique aisément appréhendable dans les approches multi-disciplinaires, et son usage se développe en climatologie-météorologie.

Ainsi donc, nous procédons à une classification interne aux champs de pluie, orientée par les besoins des phases suivantes de la méthode. Nous ne recourons à la classification que pour les pas de temps proprement pluvieux, le temps sec étant mis à part.

Cette approche est décrite dans un travail universitaire [10] et dans un article en préparation (Leblois, Creutin et al., 2016 ?)

## **5 Inférence des paramètres pour une pluie homogène**

La méthode utilisée pour inférer les paramètres de simulation d'une pluie homogène exploite les propriétés agrégatives des champs aléatoires stationnaires. Les structures ponctuelles des champs d'indicatrice et de précipitation non-nulle étant fixées, les propriétés du composite sont fixées de fait, et celles du champs composite agrégé sur des durées choisies sont numériquement évaluables, à défaut d'être aisées à expliciter. L'idéal serait donc d'inférer les paramètres d'un processus ponctuel compatible avec les éléments variographiques observables à différentes durées, et c'est en principe possible [13, 14]. Il reste cependant plus simple de décrire le processus directement à la future maille de simulation, quand les données le permettent. On cale empiriquement des modèles de variogramme sur les données, en choisissant ceux qui s'ajustent au mieux. Dans le cadre de l'approche à deux champs latents, les variogrammes des pluies non-nulles et de l'indicatrice s'apprécient séparément.

## **6 Obtention de simulations ; cas des simulations conditionnelles**

L'obtention d'une simulation conditionnelle, en géostatistique des champs gaussiens, est une opération assez simple : à une simulation libre (non conditionnelle) est ajoutée une interpolation (krigeage) des écarts aux valeurs cibles, réalisée avec la même structure spatiale et temporelle.

Dans notre cas, toute valeur de précipitation non-nulle observée doit être transformée en son équivalent Gaussien pour que soit conditionné le champ Gaussien sous-jacent.

Il faut réaliser une opération équivalente pour l'indicatrice de précipitation. Le seuillage n'admettant pas de fonction réciproque, nous recourons donc préalablement à un échantillonneur de Gibbs ou à l'algorithme de Metropolis-Hastings, pour fournir des valeurs Gaussiennes compatibles avec les valeurs d'indicatrice observées [10].

Ceci amène parfois à rejeter explicitement une valeur observée de pluie non compatible avec la loi marginale adoptée attribuée au jour en question (exemple : pluie de 0.2 mm isolée au milieu d'un jour très pluvieux par exemple, ou des données de pluie nulle alors que le type adopté n'en prévoit pas). Ces rejets sont indice d'erreurs dans les données ou d'inadéquation du modèle descriptif, et doivent être remontés ; en règle ils doivent rester peu nombreux.

Les simulations conditionnées, multiples et équiprobables, expriment bien l'incertitude de connaissance des pluies conditionnellement aux observations et, bien sûr, au modèle de variabilité adopté, d'où leurs usages pratiques et scientifiques [1, 18].

## **7 Obtention de simulations ; cas des simulations libres**

On utilisera, comme il est d'usage en enchaînement de types qualitatifs, une chaîne de Markov cachée issue d'un modèle qui aura été entraîné sur la séquence historique. Il est à noter que la variation saisonnière s'exprime commodément comme variation saisonnière des éléments paramétriques de la chaîne de Markov cachée (probabilités de transition et d'expression), les types

de pluie restant identiques à eux-même.

Les états cachés ne doivent pas être interprétés trop vite comme des types de temps ; de fait, ils sont explicitement construits sur la base d'une optimisation numérique sur la seule base du calendrier d'occurrence des types de pluies. Mais bien sûr rien n'empêchera d'examiner la contingence et les relations physiques qui relient les types cachés à la situation atmosphérique [7, 8, 9], voire aux types de temps d'une classification indépendante de cette situation atmosphérique. Auquel cas le simulateur devient utilisable en désagrégation stochastique du climat.

On aboutit ainsi à un simulateur continu susceptible de fournir des entrées approximatives, mais formalisées, permettant d'objectiver divers travaux centrés sur la variabilité ordinaire de la ressource en eau ou des aléas hydrologiques [15].

## **8 Utilité et limite de la décomposition en types homogènes de pluie**

La décomposition du signal pluviométrique en un mélange temporel de types homogènes de pluie présente d'indéniables avantages. Elle permet de recourir à des techniques de simulation efficaces et éprouvées. L'analyse conjointe des données du réseau pluviométrique, qui contient beaucoup de matériau, autorise la construction d'un nombre de pluie relativement élevé qui se combinent harmonieusement selon les aspects considérés. Par exemple l'un ou l'autre type pourra contribuer aux valeurs fortes, selon la durée et la superficie considérées ; un type de pluie extrêmement intense et extrêmement couvrant existe dans la plupart des cas étudiés, dont la présence freine en pratique la déstructuration des valeurs fortes inévitable dans les simulations méta-gaussiennes.

Par contre, la formulation adoptée revient à admettre que les pluies ressortant de types de pluies différentes peuvent être considérées comme conditionnellement indépendantes. Cette simplification n'est pas dommageable en pratique dans les applications de simulation conditionnée [18], où la présence des données d'observation vient réinjecter des traits éventuellement absents du modèle. Elle est plus gênante dans le cas de simulations libres de longue durée, pour lesquelles les éléments d'autocorrélation présents dans le modèle (autocorrélation correcte dans un type de pluie, séquençement correct des types de pluie) peinent à suffire. Une précaution élémentaire, à cet égard, est d'étudier la pluie sur une durée correspondant à une fraction raisonnable des temps caractéristiques des bassins-versants cibles (1/3 ou 1/4), afin que les cumuls hydrologiquement pertinents ne soient pas trop affectés par les changements de types de pluie. Les pluies cumulées sur des durées longues (ressource) redeviennent correctes, la loi de tels cumuls ne dépendant guère des corrélations entre pas de temps successifs.

## **9 Remerciements**

Le travail ici évoqué a été soutenu par des années d'échanges et travaux avec Maria-Helena Ramos, Clément Perchat, Jean-Marie Lepioufle, Kolbjorn Engeland, Isabelle Emmanuel, Sandrine Anquetin, Hervé Andrieu, Benjamin Renard, Isabelle Braud, Angélica Caseri, Sheng Chen. Les développements ont été aidés par la Ville de Marseille, le Conseil Général des Hauts-de-Seine, l'Agence Nationale de la Recherche (à travers les projets RIVES, AVUPUR, FLOODSCALE), le ministère norvégien de la recherche à travers le Sintef (Norvège), la Caisse Centrale de Réassurance, EDF (E&R), les structures de coopération bilatérales FAST (France-Australie) et STAR (France-Corée du Sud).

La présente contribution se fonde, en l'actualisant, sur une présentation donnée aux Journées de climatologie de Lyon [12] qui ne semble plus disponible en ligne. Elle a bénéficié des encouragements essentiels de Julie Carreau.

## Bibliographie

- [1] Caseri, A., Javelle, P., Ramos, M.H., Leblois, E. - 2015. Generating precipitation ensembles for flood alert and risk management. *Journal of Flood Risk Management*, 14 p.
- [2] Ceresetti, D., Anquetin, S., Molinie, G., Leblois, E., Creutin, J.D. - 2012. Multiscale evaluation of extreme rainfall event predictions using severity diagrams. *Weather and Forecasting*, vol. 27, n° 1, p. 174-188
- [3] Creutin, J.D., Leblois, E., Lepioufle, J.M. - 2015. Unfreezing Taylor's Hypothesis for Precipitation. *Journal of Hydrometeorology*, vol. 16, n° 6, p. 2443-2462
- [4] Emmanuel, I., Andrieu, H., Leblois, E., Flahaut, B. - 2012. Temporal and spatial variability of rainfall at the urban hydrological scale. *Journal of Hydrology*, vol. 430-431, p. 167-172
- [5] Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., Payrastre, O. - 2015. Influence of rainfall spatial variability on rainfall-runoff modelling: Benefit of a simulation approach? . *Journal of Hydrology*, vol. 531, p. 337-348
- [6] Emmanuel, I., Leblois, E., Andrieu, H., Flahaut, B. - 2011. Variabilité spatio-temporelle des précipitations aux échelles de l'hydrologie urbaine. *Houille Blanche-Revue Internationale de l'Eau*, n° 4, p. 31-36
- [7] Godart, A., Anquetin, S., Leblois, E. - 2009. Rainfall regimes associated with banded convection in the Cevennes-Vivarais area. *Meteorology and Atmospheric Physics*, vol. 103, n° 1-4, p. 25-34
- [8] Godart, A., Anquetin, S., Leblois, E., Creutin, J.D. - 2011. The contribution of orographically driven banded precipitation to the rainfall climatology of a mediterranean region. *Journal of Applied Meteorology and Climatology*, vol. 50, n° 11, p. 2235-2246
- [9] Godart, A., Leblois, E., Anquetin, S., Freychet, N. - 2010. Analysis of the relationship between banded orographic convection and atmospheric properties using factorial discriminant analysis and neural networks. *Journal of Applied Meteorology and Climatology*, vol. 49, n° 4, p. 1558-8424
- [10] Leblois, E. - 2012. Le bassin versant, système spatialement structuré et soumis au climat. *Mémoire d'HDR*, INPG Université de Grenoble.
- [11] Leblois, E., Creutin, J.D. - 2013. Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method. *Water Resources Research*, vol. 39, p. 3375-3387
- [12] Leblois, E., Jeanpierre, A.L. - 2011. Description stochastique des pluies et applications à des pluies de la région lyonnaise. *Journées de Climatologie de la commission climat et société de la société géographique de France* 17/03/2011-18/03/2011, Lyon, FRA. 10 p.
- [13] Lepioufle, J.M., Leblois, E., Creutin, J.D. - 2012. Variography of rainfall accumulation in presence of advection. *Journal of Hydrology*, n° 464-465, p. 494-504
- [14] Lepioufle, J.M., Modélisation spatio-temporelle d'un champ de pluie ; application aux pluies journalières du bassin versant de la Loire, *Thèse INPG*, 2009
- [15] Moncoulon, D., Labat, D., Ardon, J., Leblois, E., Onfroy, T., Poulard, C., Aji, S., Remy, A., Quantin, A. - 2014. Analysis of the French insurance market exposure to floods: a stochastic model combining river overflow and surface runoff. *Natural Hazards and Earth System Sciences*, vol. 14, n° 9, p. 2469-2485
- [16] Ramos M.H., Creutin J.-D., Leblois E., (2005), Visualization of storm severity, *Journal of Hydrology*, Vol 315, pages 295-307, doi:10.1016/j.jhydrol.2005.04.007
- [17] Ramos M.H., Leblois E., Creutin J.-D., (2006), From point to areal rainfall: linking the different approaches for the frequency characterisation of rainfalls in urban areas, *Water Science & technology*, Vol 54 No 6-7 pp 33-40, IWA Publishing
- [18] Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., Franks, S.W. - 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, vol. 47, p. W11516