

# PROPOSITION SUJETS DE THESE CONTRATS DOCTORAUX 2021-2024

**Appel ciblé** (merci de cocher la case correspondante):

**Contrat doctoral fléché FR Agorantic**

**Les critères pour cet appel sont :**

1. La dimension interdisciplinaire du sujet ;
2. L'adéquation entre le projet et les axes scientifiques de la FR ;
3. Priorité sera donnée aux codirections impliquant deux laboratoires/unités membres de la FR Agorantic (ou en cours d'association) ;
4. L'un des codirecteurs(trices) doit être un(e) enseignant(e)-chercheur d'Avignon Université rattaché(é) à une des deux Écoles Doctorales

**Directrice de thèse** : LACAUX Céline, Professeure, LMA

Mail et Téléphone : [Celine.Lacaux@univ-avignon.fr](mailto:Celine.Lacaux@univ-avignon.fr), 0490144492

**Co-directeur (obligatoire pour la FR Agorantic)** : JOSSELIN Didier,  
Directeur de recherche, ESPACE

**Titre en français** : Généralisation des modèles robustes de projection sur des données massives multi-échelles.

**Titre en anglais** : Generalization of robust model projections for massive multi-scale data

**Résumé en 7 lignes et Mots-clés :**

L'objectif de cette thèse est de développer des méthodes (géo)statistiques pour évaluer la capacité des modèles de projection climatique, réalisés à différentes granularités spatiales (modèles globaux et régionaux du GIEC, modèles locaux basés sur des relevés de type krigeage), à estimer correctement le changement du climat (pluviométrie, températures, notamment), quelle que soit l'échelle considérée, de la région à la parcelle, dans un contexte de données massives et de préservation du patrimoine environnemental.

**Mots-clés** : modèles robustes, effet du support spatial, patrimoine environnemental, (géo)statistiques, changement climatique, échantillonnage, données massives, données hydro-climatologiques, (changement d')échelles.

## 1 - Présentation détaillée du sujet :

Le sujet de thèse s'inscrit dans les axes (1) *Méthodologies et Interdisciplinarité* et (5) *Structuration et exploitation de corpus* de la FR Agorantic.

Il s'agit de chercher à s'affranchir des difficultés à maîtriser la variabilité statistique des mesures inhérentes à la taille des échantillons, selon les échelles auxquelles elles sont recueillies. Cette question est aujourd'hui d'autant plus cruciale que, d'une part, on connaît avec relativement peu de précision l'impact local du changement climatique global, que, d'autre part, le réchauffement constitue une trajectoire de risque avéré pour notre planète particulièrement en zone méditerranéenne (Blanchet, 2018, Molinié et al., 2012), et qu'enfin, les mesures affluent en grande quantité et de façon continue.

Le lien entre la taille des échantillons, la variance et la robustesse d'un indicateur est observé dans toutes les disciplines. Il se traduit de différentes façons : le problème des unités surfaciques modifiables en géographie (MAUP, Openshaw, 1984), le paradoxe en Simpson en économie et statistique (1954), le paradoxe de Robinson en sociologie (1950), le problème d'inférence écologique (King, 1997) en sciences, et plus généralement, le problème du changement du support (COSP, King et al., 2004). Il n'est pas résolu en tant que tel et constitue potentiellement l'objectif ultime de cette thèse, puisque, comme nous le croyons, il réside foncièrement dans un problème d'échantillonnage (spatial) (Louvet et al., 2016, Josselin, 2019).

Les enjeux scientifiques de cette thèse se situent également à plusieurs niveaux :

1. Proposer un modèle pour un indicateur statistique (par exemple, l'entropie de Shannon, Josselin et al., 2008) permettant de séparer sa dépendance en la structure spatiale des données (ou encore le support) de celle induite par la nature stochastique des caractères observés, tout en prenant en compte la complexité et les associations d'informations à travers les échelles : *peut-on estimer le biais induit par le support de façon systématique quelle que soit l'échelle ?* Une étude empirique sur des données simulées (type modèle géostatistique ou modèle de parcelles) puis sur des données réelles sera menée en amont. La dépendance en le support sera dégagée par ré-échantillonnage des données et la variabilité en l'échelle sera évaluée en utilisant des regroupements d'informations par pavages de type cellule de Voronoï, des méthodes issues de l'intelligence artificielle (réseaux de neurones, apprentissage profond) ou encore des modèles multi-niveaux.

2. Prendre en compte explicitement le biais lié à l'échantillon dans la généralisation des modèles ; *la taille des échantillons peut-elle devenir un paramètre modulable pour mieux maîtriser les changements d'échelle ?*
3. Evaluer la capacité de généralisation des modèles en agrégation ou désagrégation spatiale. *Peut-on, et sous quelles conditions, appliquer un modèle global au niveau de la parcelle ? De quelle façon peut-on combiner les connaissances locales pour obtenir une information fiable à un niveau plus élevé ?*
4. Mesurer et maîtriser l'impact de l'usage de données massives sur les variations de taille des échantillons, et, en conséquence, sur les méthodes appropriées pour les traiter. *Les méthodes proposées résistent-elles et répondent-elles à la montée en charge numérique ? Les scientifiques restent-ils capables de générer une information fiable et pertinente, dépourvue des biais du support tout en maîtrisant les données massives ?*

Les mesures considérées sont essentiellement des données de température et de pluviométrie, suivies depuis plus de 40 ans à l'échelle journalière sur les stations cévenoles de l'UMR ESPACE (Boudevillain, 2011). Elles sont croisées avec des données issues des modèles globaux et régionaux du GIEC, ayant des précisions variables. L'ensemble des données traitées forme des échantillons de taille très différente, comme de qualité, précision et complétude fluctuantes. Il constitue un corpus idéal de données pour rechercher des méthodes génériques d'agrégation/désagrégation, applicables potentiellement à tout problème d'échantillonnage spatial.

Les résultats obtenus dans cette thèse peuvent potentiellement concerner d'autres thématiques de la FR Agorantic : découpages politico-administratif optimaux, évaluation de patrimoines naturels ou urbains, analyse robuste des données foncières, décodage d'informations numériques textuelles dans tout type de données décrivant des contenus sociaux, économiques, culturels et patrimoniaux, pour peu que celles-ci soient regroupées dans des échantillons statistiques et/ou des découpages spatiaux.

## **2 - Profil du candidat(e) :**

Le/la candidat(e) doit être titulaire d'un master 2 en mathématiques (spécialité statistique), ou d'un diplôme équivalent (école d'ingénieurs) et doit avoir une formation solide en statistique. Il doit aussi avoir une forte appétence pour les applications, compte-tenu du caractère pluri-disciplinaire du sujet, et la réalisation de codes. Il/elle doit maîtriser R et/ou Python afin de mettre en œuvre les diverses méthodes à explorer et développer.

### 3 - Références bibliographiques :

- Blanchet, J., Molinié, G. & Touati. 2018. Spatial analysis of trend in extreme daily rainfall in southern France. *Clim Dyn* 51, 799-812 (2018). DOI : [10.1007/s00382-016-3122-7](https://doi.org/10.1007/s00382-016-3122-7)
- Boudevillain, B., Delrieu, G., Galabertier, B., Bonnifait L., Bouilloud, L., Kirstetter, P.E., Mosini, M.L. 2011. The Cévennes-Vivarais Mediterranean Hydrometeorological Observatory Database. *Water Resour. Res.*, 47 (W07701).
- Josselin D. Mahfoud I., Fady B., 2008, Impact of a Change of Support on the Assessment of Biodiversity with Shannon Entropy, In Headway in Spatial Data Handling, 13th International Symposium on Spatial Data Handling, Montpellier, France, 23-25 July 2008, DOI: [10.1007/978-3-540-68566-1\\_7](https://doi.org/10.1007/978-3-540-68566-1_7)
- Didier Josselin. 2019, Impact of the Scale on Several Metrics Used in Geographical Object-Based Image Analysis: Does GEOBIA Mitigate the Modifiable Areal Unit Problem (MAUP)?, *ISPRS International Journal of Geo-Information*, MDPI, 2019, 8 (3), pp.156. [10.3390/ijgi8030156](https://doi.org/10.3390/ijgi8030156). [hal-02103566](https://hal.archives-ouvertes.fr/hal-02103566)
- King G. 1997. *A solution to the ecological inference problem. Reconstructing individual behaviour from aggregate data.* Princeton University Press.
- King G., Rosen O., Tanner A. M. (Eds.). 2004. *Ecological inference. New methodological strategies.* Cambridge University Press.
- Romain Louvet, Didier Josselin, Cyrille Genre-Grandpierre, Jagannath Aryal. Impact des niveaux d'échelle sur l'étude des feux de forêts du sud-est de la France . *Revue Internationale de Géomatique*, Lavoisier, 2016. [hal-01338077](https://hal.archives-ouvertes.fr/hal-01338077)
- Molinié G., Ceresetti D., Anquetin S., Creutin J.D., Boudevillain B.: Rainfall regime of a Mountainous Mediterranean Region: Statistical analysis at short time steps. *J. Appl. Meteor. Climatol.*, 51, 429-448, 2012
- Openshaw S. 1984. *The modifiable areal unit problem.* Norwich: Geo Books, CATMOG 38.
- Robinson W. 1950. Ecological correlations and the behaviour of individuals. *American Sociological Review*, vol. 15, p. 351-357.
- Simpson E. 1951. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society - Series B (Methodological)*, vol. 13-2, p. 238-241.

#### **4 - Opportunités de mobilité à l'international du doctorant(e) dans le cadre de sa thèse : oui - non**

Une mobilité à l'international du doctorant est à envisager : il nous faut étudier les possibilités de laboratoires d'accueil.

Ce projet s'inscrit par ailleurs dans le projet interdisciplinaire *CliMedScale* du Défi Changement Climatique2020 du CNRS pour lequel ESPACE a été lauréat, offrant des opportunités intéressantes de connexion avec d'autres laboratoires français et, à terme, d'insertion dans des projets internationaux de plus grande envergure.

Maximum 7 pages

**Les sujets devront être adressés à**

**agorantic@univ-avignon.fr**

**Avant 15 mars 2021 midi**