



Sujet de thèse

La qualité des bases de données spatiales pour la modélisation des changements d'occupation du sol : élaboration de critères et analyse de leur robustesse

Mots clés : Modélisation, occupation du sol, qualité, critères d'analyse, robustesse

Directeur : Martin PAEGELOW (paegelow@univ-tlse2.fr)

Co-directeur : María Teresa Camacho Olmedo (camacho@ugr.es)

Contexte du projet de thèse

Des données d'occupation / d'usage du sol sont mobilisées dans nombre de projets d'aménagement du territoire, d'études d'impact et d'anticipation des changements à venir (modélisations / simulations prospectives de type « land change »). Dans la plupart des projets à petite et moyenne échelle, la donnée utilisée provient d'un nombre de bases de données restreint (CLC, ESA Landcover CCI, GLC30, Spot Thema, Modis Landcover, LUCAS, ...). A grande échelle, les bases de données spatialisées sont encore peu nombreuses et ne couvrent que localement le territoire même si une couverture nationale est en train de déploiement (OCS GE actuellement disponible sur l'emprise de l'ancienne région Midi-Pyrénées).

Les études sur la fiabilité de ces bases de données (cf. Baudry, 1991, Pontius Jr. & Malizia, 2004 ; Runfola & Pontius Jr., 2013, thèse de Colas Robert soutenue le 15 décembre 2016) sont encore trop peu nombreuses et nombre de produits n'ont pas encore fait l'objet d'études critiques.

Connaissant l'impact des données d'entrée sur les résultats de ces modélisations, l'analyse systématique, le développement de critères d'estimation et l'automatisation d'un tel benchmarking sont devenus des enjeux cruciaux pour les études des changements environnementaux.

Le sujet de thèse proposé s'inscrit au cœur des problématiques des laboratoires d'accueil et s'appuie sur les acquis de 15 ans de projets scientifiques internationaux financés et d'une dizaine de thèses récemment soutenues ou en cours portant sur la modélisation spatio-temporelle et (co-)encadrées par les directeurs de thèse.

Description détaillée du projet de thèse

Ce projet vise trois objectifs :

- Un **objectif thématique** consistant à tester la robustesse d'un certain nombre de bases de données géoréférencées. Il s'agit de tester et de comparer les bases de données d'occupation du sol les plus connues (MODIS Land Cover, CLC, GLC30, ESA Landcover CCI, SPOT Thema, ...) ainsi que des produits élaborés à échelle nationale ou régionale (OCS GE, ...) en ce qui concerne leur qualité descriptive catégorielle et leur fiabilité temporelle : les changements affichés d'une date à l'autre sont-ils vrais ?
- Un **objectif méthodologique** cherchant à établir des descripteurs (géo-)statistiques solides, ayant par conséquent une vocation généraliste et transposable pour se prononcer sur la qualité des données en question. Les sources d'erreur les plus fréquentes sont liées aux confusions entre catégories d'occupation du sol proches (une contraction de la légende réduisant le taux d'erreur) et à la résolution spatiale souvent faible induisant des faux changements (Robert, 2016, Runfolla et Pontius Jr., 2013). Si le développement de métriques détectant de telles sources d'erreurs semble simple, il reste néanmoins de les convertir en descripteurs plus universels et robustes ainsi que de scruter les produits systématiquement et plus finement : y a-t-il des milieux plus affectés que d'autres ? des types et des vitesses de changements plus exposés ?
- Un **objectif opérationnel** visant une cartographie multi-résolution du degré d'incertitude des bases de données d'occupation et d'usage du sol déployées sur certains territoires (France, Espagne, ...) :
 - Différentes résolutions spatiales et leur intégration
 - Différentes résolutions (catégorielles) de légende
 - Différents types de territoires (urbains, ruraux, montagnards, ...)

Problématique et verrous scientifiques du projet de thèse

La problématique est simple : il s'agit de pouvoir d'assortir les données d'occupation du sol pour l'apprentissage des modèles d'une probabilité d'erreur (pouvant être répercuté sur le calcul probabiliste des quantités de changements pour la simulation par chaînes de Markov). Aujourd'hui, il est paradoxal que la moindre classification d'image satellite est validée, mais que nombre de modélisations ne posent même pas la question de la qualité du « carburant » introduit dans le modèle.

Les verrous scientifiques : un premier enjeu consiste à définir des critères de qualité suffisamment universels pour prendre en compte la grande diversité des bases de données d'occupation du sol (résolutions spatiale, catégorielle, temporelle, téléonomie de la BD). Une seconde difficulté consiste à tester la significativité des descripteurs à l'aune des singularités spatiotemporelles (types de milieux et de dynamiques). Enfin – et sur un plan opérationnel, peut-on aboutir à des descripteurs suffisamment simples pour couvrir les trois aspects « multi-resolutionnels » ébauchés ci-dessus ?

Profil de candidat recherché

Cette thèse, fondamentalement interdisciplinaire, s'adresse à un(e) étudiant(e) diplômé(e) d'un master 2 (ou équivalent) en géomatique, géographie, socio-anthropologie, sciences de l'environnement. Les compétences requises sont :

- Esprit scientifique
- Maîtrise des SIG, des outils de la modélisation « pattern based » (comme TerrSet, DINAMICA, SLEUTH, LUCSIM, CLUE...)
- Maîtrise des techniques statistiques multivariées ; la maîtrise de R serait un plus
- Maîtrise de l'anglais
- Francophone (dans le cadre d'une co-tutelle France-Espagne, la rédaction pourra se faire soit en français, soit en espagnol)

Contacts et constitution du dossier de candidature

Martin Paegelow, Professeur de Géographie Université Toulouse 2 Jean Jaurès, **Directeur de thèse** : paegelow@univ-tlse2.fr

María Teresa Camacho Olmedo, Professeur de Géographie Universidad de Granada (Espagne), **Co-Directeur de thèse** camacho@ugr.es

Le dossier de candidature est à envoyer aux directeur, co-directeur, la présidente de la commission doctorale du laboratoire GEODE Christine Vergnolle-Mainar : christine.vergnolle-mainar@univ-tlse2.fr et le directeur de GEOED Didier Galop : didier.galop@univ-tlse2.fr

L'envoi se fait uniquement par mail. *Attention, le serveur de l'université n'accepte pas les messages dont la taille des pièces jointes excède 6 Mo. Pensez à utiliser un service de transfert de fichiers.*

Le dossier comprend :

- CV
- Lettre de motivation manuscrite (scannée) de 1 page
- Cursus post-bac avec indication des mentions / moyennes de notes / rang dans la promotion (joindre obligatoirement la photocopie des diplômes et le relevé des notes pour les diplômes post-bac ; pour les diplômes étrangers : joindre obligatoirement une photocopie (traduite) certifiée conforme)
- Texte de positionnement par rapport au sujet de thèse et expérience dans le domaine

Références

Baudry, J., 1991. Phénomènes de dépendances d'échelle dans l'étude du changement d'utilisation des terres in C. Mullon (éd.), *Le transfert d'échelle*, Paris, ORSTOM, p. 425-438.

Pontius Jr, R.G., Malizia, N.R., 2004. Effect of category aggregation on map comparison, in: *Geographic Information Science*. Springer, pp. 251–268.

Robert C., 2016, « Comprendre les changements d'utilisation des terres en France pour mieux estimer leurs impacts sur les émissions de gaz à effet de serre. De l'observation à la modélisation », thèse de doctorat sous la direction de Marianne Cohen Ladyss, soutenue le 15.12.2016

Runfola, D.S.M., Pontius Jr, R.G., 2013. Measuring the temporal instability of land change using the Flow matrix. *International Journal of Geographical Information Science* 27, 1696–17