

# Quelle géostatistique pour des DPE à la localisation incertaine ?

Marc Grossouvre 1\*      Didier Rullière 2†      Jonathan Villot 2‡

## Résumé

Planifier la transition énergétique requiert des décideurs une connaissance approfondie de leur territoire. Ce dernier est étudié à travers des données collectées de sources multiples, à diverses échelles et sous des contraintes de confidentialité. Ces données renseignent des objets, tels que les bâtiments. Ainsi, étudier la performance énergétique des bâtiments demande d'apprendre des Diagnostic de Performance Énergétique (DPE) réalisés, lesquels sont publiés avec un géolocalisation à adresse postale dont la position est incertaine. Ce travail montre que l'apprentissage des DPE observés pour prédire les DPE manquants peut être envisagé comme un problème d'interpolation spatiale. Il propose une manière de traiter le DPE en tant qu'information géolocalisée et de prédire sa valeur à l'échelle du bâtiment. La méthodologie du krigeage est appliquée à des champs aléatoires observés à des emplacements aléatoires pour trouver le meilleur prédicteur linéaire non biaisé (BLUP). Ce nouveau modèle, appelé krigeage de mixtures, estime une moyenne et une variance du DPE pour un bâtiment non observé. Ces estimations fournissent aux décideurs une valeur attendue du DPE ainsi qu'une incertitude sur la prédiction. Le cas spécifique d'une ville française est pris comme exemple. Des applications sur des données simulées sont aussi présentées, suggérant que le krigeage de mixtures puisse être utilisé plus largement pour contrôler la propagation d'incertitude. L'ensemble du travail depuis la collecte des données jusqu'à la production cartographique en passant par la modélisation et l'optimisation a été codé en R.

**Mots-clefs** : Géostatistique, climat, rénovation, multi-échelle, MAUP.

## Contexte et motivation

Pour massifier la rénovation énergétique des bâtiments, les collectivités locales souhaitent pouvoir identifier les passoires énergétiques pour proposer à leur propriétaire un accompagnement pro-actif pour identifier une stratégie de rénovation et les financements associés. Les DPE sont observés au logement ou au bâtiment et identifiés par une adresse. Ils fournissent une consommation d'énergie donnée en  $kWh/m^2/an$  (kiloWatt-heure d'énergie primaire par mètre carré et par an), à laquelle est associée une étiquette allant de A à G. Les DPE sont appariés aux fichiers fonciers par l'adresse afin d'identifier la parcelle support du logement observé. Il reste une incertitude sur la position du logement observé sur la parcelle.

Comme indiqué par Gotway and Young (2002), les modèles d'interpolation spatiale traditionnels tels que le krigeage ne sont pas adaptés à cette incertitude. Nous proposons un nouveau modèle capable de gérer à la fois indifféremment des données localisées au point ou avec une incertitude sur une zone géographique en introduisant la notion de grain. L'incertitude de position est modélisée par une distribution de mélange qui traduit à la fois l'incertitude sur les données d'entrée et sur les données de sortie du modèle.

## Modélisation

On suppose qu'à chaque point  $x$  de l'espace  $\chi$  (une commune) est associée une variable aléatoire  $Y(x)$  (la consommation énergétique d'un mètre carré d'habitation localisé en  $x$ ). Un grain  $g$  est un ensemble de points

---

\*U.R.B.S. SAS, Bâtiment des Hautes Technologie, 20 Rue Professeur Benoît LAURAS, 42000 Saint-Etienne, France, marcgrossouvre@urbs.fr

†Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Département GMI, Espace Fauriel, 29 rue Ponchardier, F - 42023 Saint-Etienne, France, didier.rulliere@emse.fr

‡Mines Saint-Etienne, Univ Lyon, CNRS, Univ Jean Monnet, Univ Lumière Lyon 2, Univ Lyon 3 Jean Moulin, ENS Lyon, ENTPE, ENSA Lyon, UMR 5600 EVS, Institut Henri Fayol, F - 42023 Saint-Etienne France, jonathan.villot@emse.fr

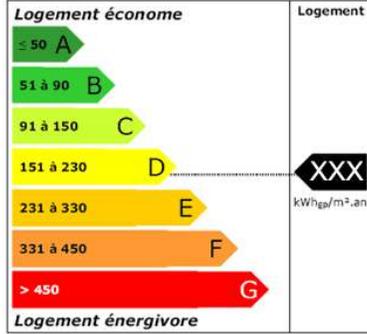


Figure 1: Vignette réglementaire du DPE jusqu'à 2021.

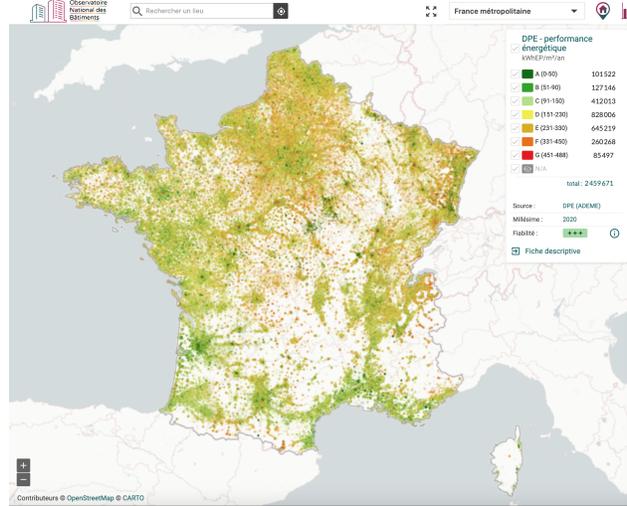


Figure 2: Carte des DPE inventoriés en France.

de l'espace associé à une variable de position aléatoire  $X_g$ . On note  $Y(g) = Y(X_g)$  la consommation énergétique d'un mètre carré localisé de façon aléatoire sur la parcelle  $g$ . Supposons qu'après une transformation adaptée,  $\forall x, E[Y(x)] = 0$ . Etant donné un vecteur d'observations  $\underline{Y}$ , pour tout  $g, g' \in \chi$ , et en particulier pour un singleton  $g = \{x\}$ , pour tout  $i \in 1, \dots, p$ , les poids  $\alpha(g)$  produisant le meilleur prédicteur linéaire non biaisé de  $Y(g)$  sont  $\alpha(g) = \mathbf{K}^{-1}\mathbf{h}(g)$ . Le modèle est codé sous forme de package R.

## Illustration

On étudie un quartier de la ville d'Angers sur lequel 362 observations sont fournies. Comme variables d'entrée, on utilise la longitude, l'altitude et l'année de construction. Sur la Figure 3 on visualise les DPE observés et sur la Figure 4 les DPE prédits.

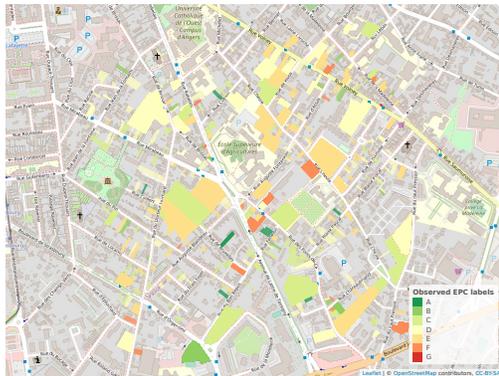


Figure 3: DPE observés.

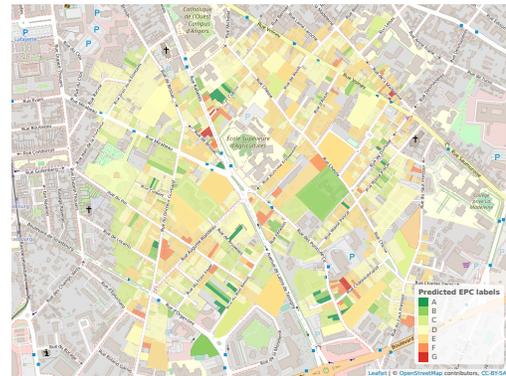


Figure 4: DPE prédits.

Les modèles usuels tendent à écraser les prédictions de DPE autour d'une tendance centrale, la prise en compte des incertitudes en entrée atténue cet effet.

## Références

Gotway, Carol, and Linda Young. 2002. "Combining Incompatible Spatial Data." *Journal of the American Statistical Association* 97 (February): 632–48. <https://doi.org/10.1198/016214502760047140>.