



vendredi 23 juin 2023,  
Rencontres R 2023

# **Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?**

Marc Grossouvre<sup>1</sup> sous la direction de Didier Rullière<sup>2</sup>  
et la supervision de Jonathan Villot<sup>3</sup>

<sup>1</sup>Doctorant CIFRE, U.R.B.S. SAS, marcgrossouvre@urbs.fr

<sup>2</sup>Mines Saint-Etienne - LIMOS - Univ Clermont Auvergne

<sup>3</sup>Mines Saint-Etienne - U.R.B.S. SAS

# Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?

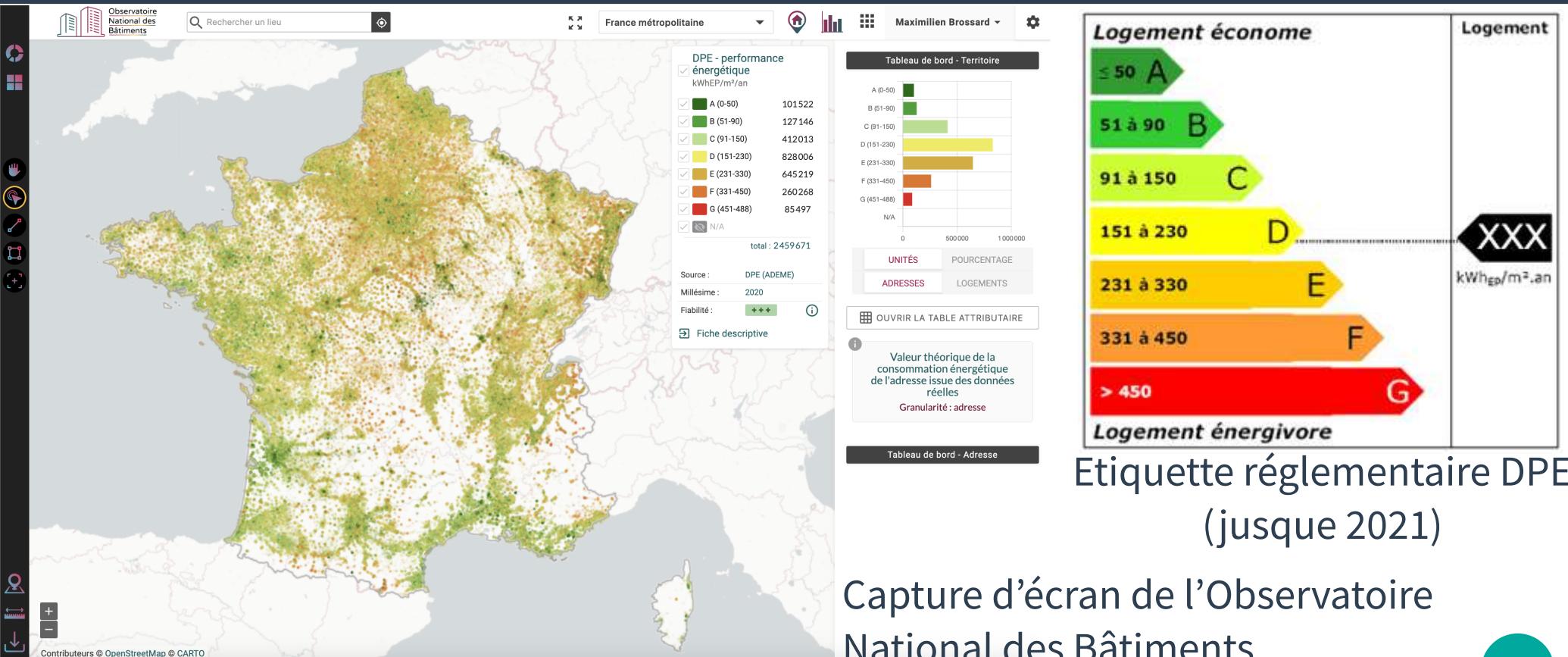
1. Que représentent les DPE fournis par l'ADEME ?
2. Une étiquette incertaine, une localisation incertaine
3. Intégrer ces aléas à un modèle géostatistique
4. Traitement des données et implémentation du modèle en R

# Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?



- 1. Que représentent les DPE fournis par l'ADEME ?**
- 2. Une étiquette incertaine , une localisation incertaine**
- 3. Intégrer ces aléas à un modèle géostatistique**
- 4. Traitement des données et implémentation du modèle en R**

# Environ 2 millions de DPE collectés par an



# 1. Que représentent les DPE fournis par l'ADEME ?

Code de la construction et de l'habitation,  
article L126-226



Légifrance

Le service public de la diffusion du droit

Loi du 22 août 2021

Le diagnostic de performance énergétique d'un bâtiment ou d'une partie de bâtiment est un document qui comporte la quantité d'énergie effectivement consommée ou estimée, exprimée en énergie primaire et finale, ainsi que les émissions de gaz à effet de serre induites, pour une utilisation standardisée [...].

# "un bâtiment ou une partie de bâtiment"

- La méthode n'est pas la même pour un logement ou pour un immeuble.
- Les étiquettes dans un même immeuble peuvent varier.



**1 rue d'Etrembières,  
Annemasse**

65 apparts en étiquette C,  
7 en étiquette D,  
5 en étiquette E,  
4 en étiquette F =  
passoires énergétiques.

Appart 705 :  
diagnostiqué C en 2018 ,  
diagnostiqué E en 2020...

# 1. Que représentent les DPE fournis par l'ADEME ?

Code de la construction et de l'habitation,  
article L126-226



Légifrance

Le service public de la diffusion du droit

Loi du 22 août 2021

Le diagnostic de performance énergétique d'un bâtiment ou d'une partie de bâtiment est un document qui comporte la quantité d'énergie effectivement consommée ou estimée, exprimée en énergie primaire et finale, ainsi que les émissions de gaz à effet de serre induites, pour une utilisation standardisée [...].

# "pour une utilisation standardisée"

- **Quels habitants respectent les standards ?**
  - Il y a ceux qui "surchauffent".
  - Il y a aussi ceux qui renoncent au chauffage.
  - Et ceux qui sont chauffés par les voisins...

Le DPE et la consommation réelle sont des données différentes. La consommation ne permet pas de prédire le DPE directement.



**1 rue d'Etrembières  
à Annemasse**

Consommation e.f. ENEDIS :  
360 MWh/an  
Surface habitable : 2 221m<sup>2</sup>

Consommation e.p. :  
418kWhep/m<sup>2</sup>/an  
⇒ étiquette F d'après la  
consommation

# 1. Que représentent les DPE fournis par l'ADEME ?

Une valeur réglementaire qui informe sur une partie du parc bâti



La cohérence entre les valeurs au logement et les valeurs au bâtiment n'est pas assurée.

Le DPE ne donne pas de consommation réelle.

Le DPE représente l'état d'un logement ou d'un bâtiment à un moment du passé.

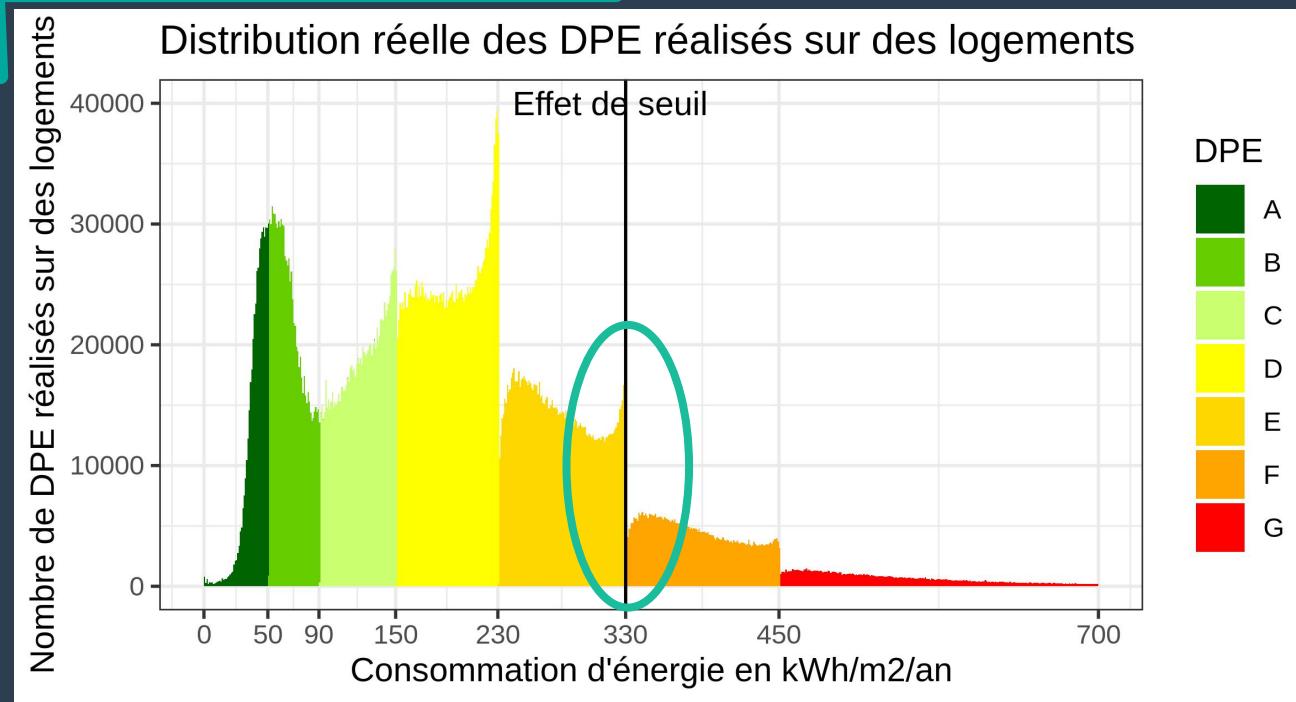
# **Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?**

- 1. Que représentent les DPE fournis par l'ADEME ?**
- 2. Une étiquette incertaine , une localisation incertaine**
- 3. Intégrer ces aléas à un modèle géostatistique**
- 4. Traitement des données et implémentation du modèle en R**

# Une étiquette incertaine : mesure humaine, effets de seuil, valeurs manquantes

Un seuil est une limite réglementaire du domaine associé à une étiquette DPE.

L'effet de seuil regroupe les phénomènes qui apparaissent à proximité de cette valeur.



# Une localisation incertaine : adresse, bâtiment et parcelle

nom\_rue

28 & 30 Rue Gabriel Vicaire \nRue Prosper Convert

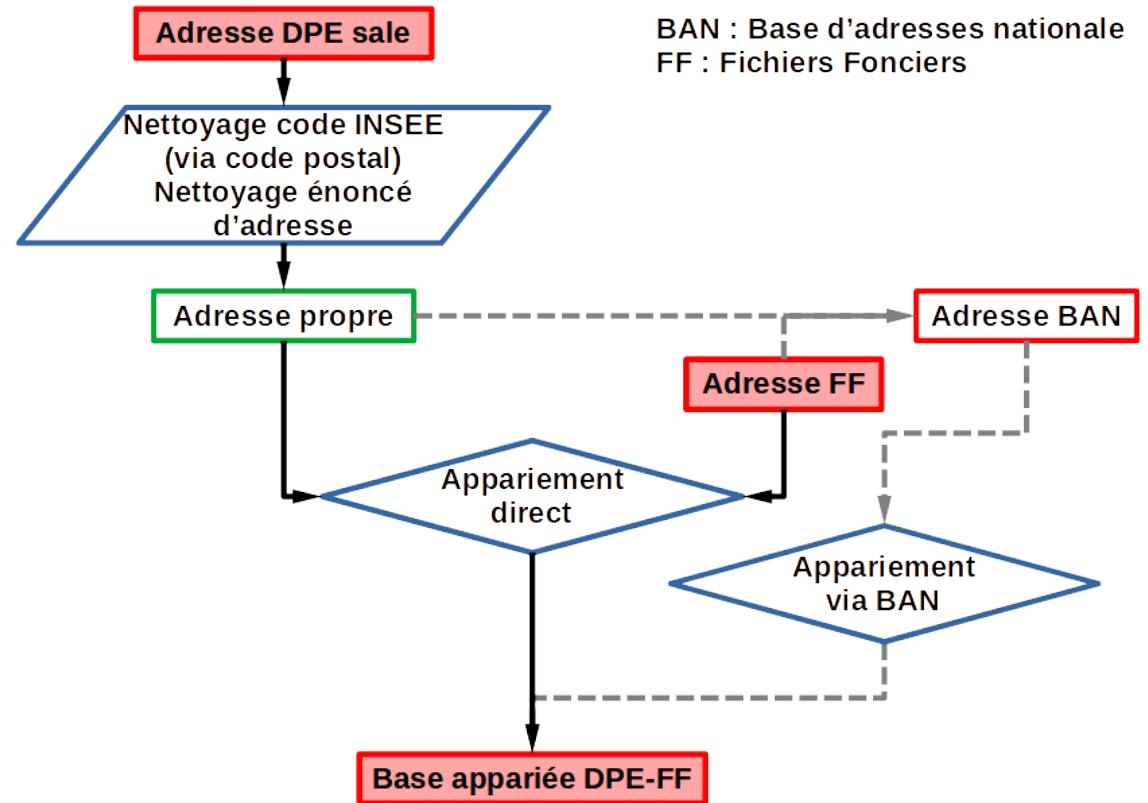
Comprendre : 28 et 30 rue Gabriel Vicaire,  
à l'angle de la rue Prosper Convert



# Il faut apparier les DPE aux fichiers fonciers pour les fiabiliser

Les fichiers fonciers (Ministère des finances) permettent de relier le DPE observé à une parcelle.

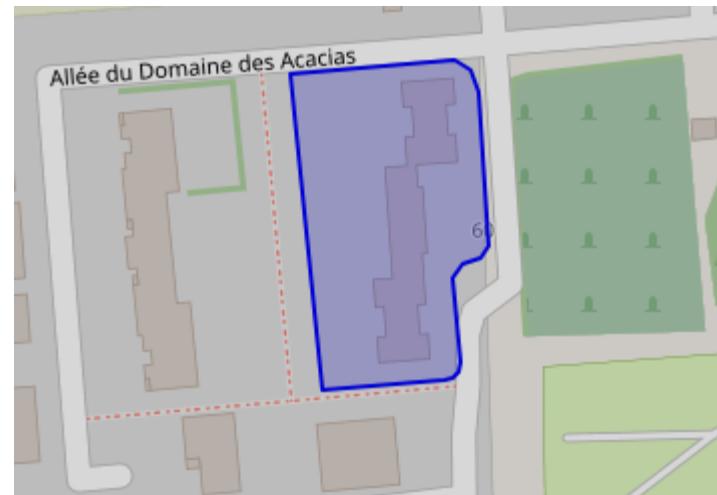
Le DPE renseigne sur l'un des logements de l'un des bâtiments de la parcelle.



# Une fois la parcelle associée à l'adresse identifiée, il est difficile de géolocaliser l'adresse sur la parcelle

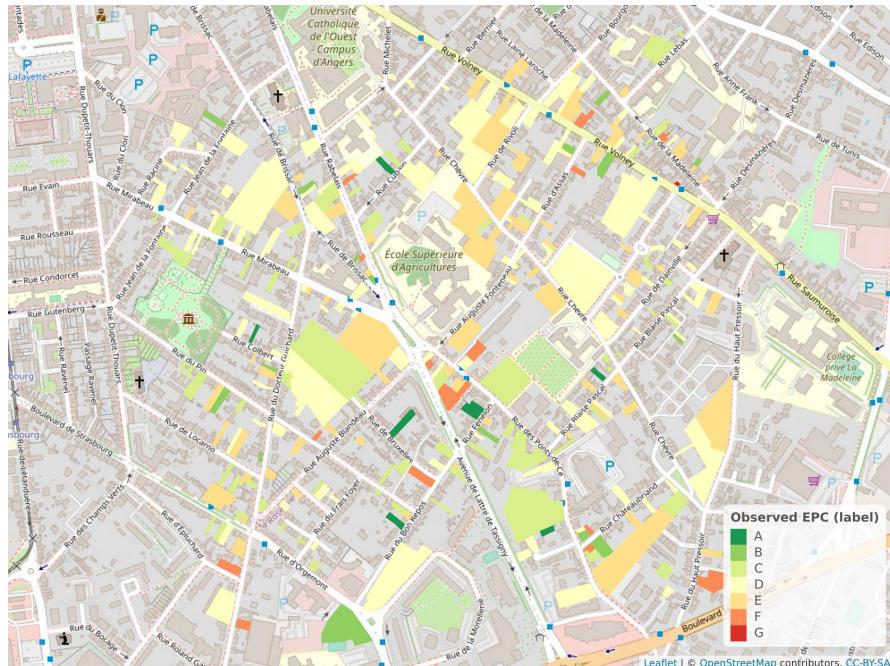
14, 22, 30, 40, 52 impasse des Acacias  
à Ars-sur-Formans (Ain).

- 5 maisons en bande,
- 1 seule parcelle,
- Observations : 3 étiquettes E, 2 étiquettes D.
- On ne sait pas où sont les logements observés.



# Beaucoup de valeurs manquantes.

## Un quartier d'Angers



- Environ 15% des adresses ont eu au moins un logement diagnostiquée.
- Que peut-on dire des adresses qui n'ont pas été observées ?
- Peut-on détecter les passoires thermiques ?

# Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?

1. Que représentent les DPE fournis par l'ADEME ?
2. Une étiquette incertaine, une localisation incertaine
3. Intégrer ces aléas à un modèle géostatistique
4. Traitement des données et implémentation du modèle en R

## 4. Intégrer ces aléas à un modèle géostatistique

**Problématique** : Comment prédire les DPE à l'adresse sans inspection physique ?

**Hypothèse** : Le DPE peut être modélisé et prédit comme une donnée géospatiale plutôt que d'ingénierie thermique.

**Verrou scientifique** : Nous devons gérer une incertitude à la fois sur la position des observations et leur valeurs.

# L'ADEME observe une distribution de mélange des DPE

Chaque logement a **un DPE aléatoire** : probablement C mais peut-être B ou D selon les travaux réalisés, le diagnostiqueur, le seuil...

**ET**

Chaque DPE de l'ADEME est associé à **un logement aléatoire** parmi les logements de la parcelle.

**DONC**

Chaque DPE de l'ADEME est donc  
**une valeur aléatoire associée à une position aléatoire.**

**Que peut-on faire de ça ?**

# Modélisation des DPE comme une distribution de mélange

Soient :

$x$  un point du territoire

$Y(x)$  la consommation d'énergie par mètre carré d'habitation

$g$  une parcelle

$X_g$  une position aléatoire sur la parcelle

Alors  $Y(g) = Y(X_g)$  est la consommation d'énergie par mètre carré d'habitation associée à une parcelle  $g$ .

Les DPE ADEME constituent un vecteur d'observations

$$\underline{\mathbf{Y}} = (Y(g_1), \dots, Y(g_n))^T.$$

# Covariance entre 2 distributions de mélange

En notant  $k(x, x') = \text{Cov}[Y(x), Y(x')]$  et  $\mu(x) = \mathbb{E}[Y(x)]$ .

Pour 2 parcelles  $g, g'$ , la covariance entre  $Y(g)$  et  $Y(g')$  peut s'exprimer :

$$\begin{aligned}\text{Cov}[Y(g), Y(g')] &:= \text{Cov}[Y(g), Y(g')] \\ &= \mathbb{E}[k(X_g, X_{g'})] + \text{Cov}[\mu(X_g), \mu(X_{g'})]\end{aligned}$$

En particulier

$$\begin{aligned}\text{Cov}[Y(g), Y(g)] &= \text{Cov}[Y(g), Y(g)] = \text{Var}[Y(g)] \\ &= \mathbb{E}[k(X_g, X_g)] + \text{Var}[\mu(X_g)]\end{aligned}$$

# Covariance entre 2 distributions de mélange

En notant  $k(x, x') = \text{Cov}[Y(x), Y(x')]$  et  $\mu(x) = \mathbb{E}[Y(x)]$ .

Pour 2 parcelles  $g, g'$ , la covariance entre  $Y(g)$  et  $Y(g')$  peut s'exprimer :

$$\begin{aligned}\text{Cov}[Y(g), Y(g')] &:= \text{Cov}[Y(g), Y(g')] \\ &= \boxed{\mathbb{E}[k(X_g, X_{g'})]} + \text{Cov}[\mu(X_g), \mu(X_{g'})]\end{aligned}$$

En particulier

$$\begin{aligned}\text{Cov}[Y(g), Y(g)] &= \text{Cov}[Y(g), Y(g)] = \text{Var}[Y(g)] \\ &= \boxed{\mathbb{E}[k(X_g, X_g)]} + \text{Var}[\mu(X_g)]\end{aligned}$$

# Le meilleur prédicteur linéaire non biaisé : Krigeage de distributions de mélanges

Pour une parcelle  $g$  qui n'a pas été observée, on cherche un prédicteur de  $Y(g)$  de la forme :  $\hat{Y}(g) = \sum_{i=1}^n \alpha_i Y(g_i) = \boldsymbol{\alpha}^T \underline{\mathbf{Y}}$ . Soit  $\mathbf{K}$  la matrice de covariance de  $\underline{\mathbf{Y}}$  et  $\mathbf{h}_g$  le vecteur de covariance entre  $\underline{\mathbf{Y}}$  et  $Y(g)$ . Alors si  $\underline{\mathbf{Y}}$  est d'espérance nulle, le meilleur prédicteur linéaire non biaisé (au sens de l'erreur quadratique) est :

$$\boldsymbol{\alpha} = \mathbf{K}^{-1} \mathbf{h}_g$$

On peut aussi le déterminer si  $\underline{\mathbf{Y}}$  n'est pas d'espérance nulle et estimer l'erreur de prédiction.

Grossouvre M., Rullière D., Villot J., *Spatial interpolation using mixture distributions: A Best Linear Unbiased Predictor*, 2023,  
preprint <https://hal.science/hal-03276127/>

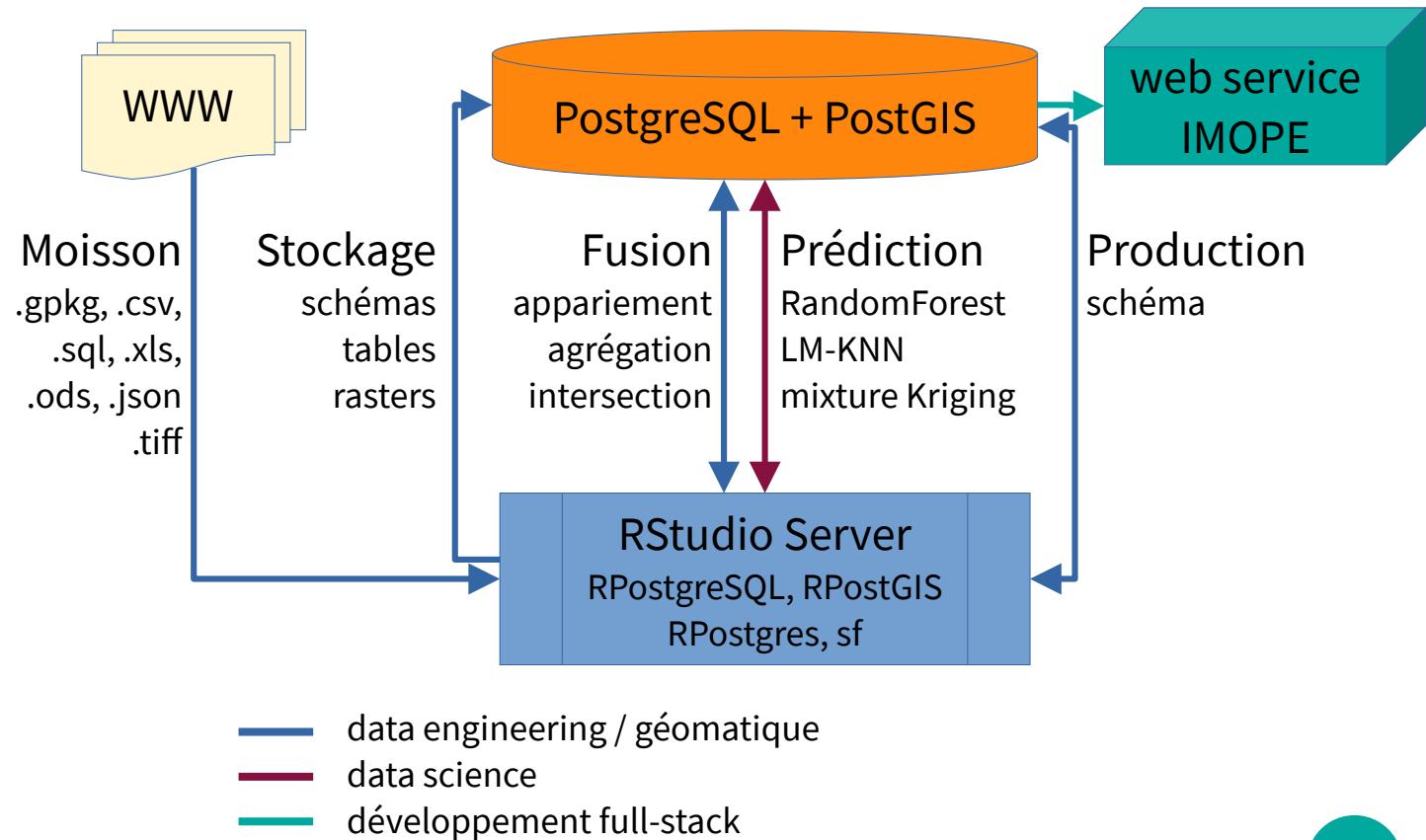
# **Quelle géostatistique pour des Diagnostics de Performance Energétique (DPE) à la localisation incertaine ?**

- 1. Que représentent les DPE fournis par l'ADEME ?**
- 2. Une étiquette incertaine, une localisation incertaine**
- 3. Intégrer ces aléas à un modèle géostatistique**
- 4. Traitement des données et implémentation du modèle en R**

# Traitement des données et implémentation du modèle en R

R pour tout !

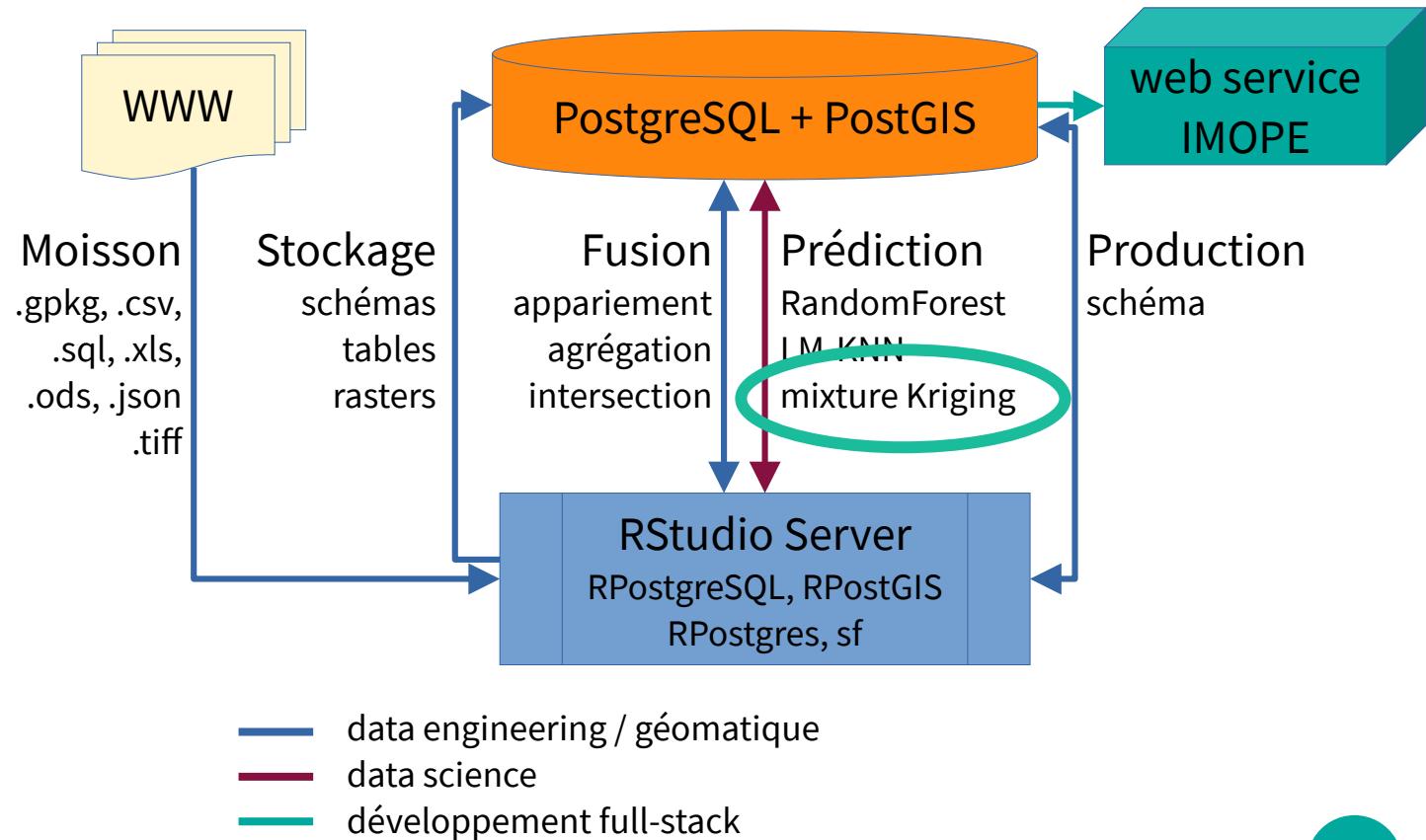
- Développer un package pour chaque type d'activité.
- Intégrer du C++ dans les packages.
- Lancer des requêtes SQL depuis RStudio.
- Intégrer au process le reporting (Rmd, Rnw)



# Traitement des données et implémentation du modèle en R

R pour tout !

- Développer un package pour chaque type d'activité.
- Intégrer du C++ dans les packages.
- Lancer des requêtes SQL depuis RStudio.
- Intégrer au process le reporting (Rmd, Rnw)



# Les questions techniques du package mixture Kriging (1/2)

$$\text{Cov} [Y(g), Y(g')] = \mathbb{E} [k(X_g, X_{g'})] + \text{Cov} [\mu(X_g), \mu(X_{g'})]$$

Pour 2 parcelles  $g, g'$ , si  $X_g, X_{g'}$  sont uniformes, on a :

$$\text{Cov} [Y(g), Y(g')] = \frac{1}{[g][g']} \sum_{(x,x') \in g \times g'} \text{Cov} [Y(x), Y(x')]$$

Donc le nombre de covariances à calculer augmente comme le carré de la densité de points.

Le noyau de covariance contient une exponentielle qui allonge le temps de calcul.

On choisit d'implémenter les noyaux de covariance en C++, RcppArmadillo.

## Les questions techniques du package mixture Kriging (2/2)

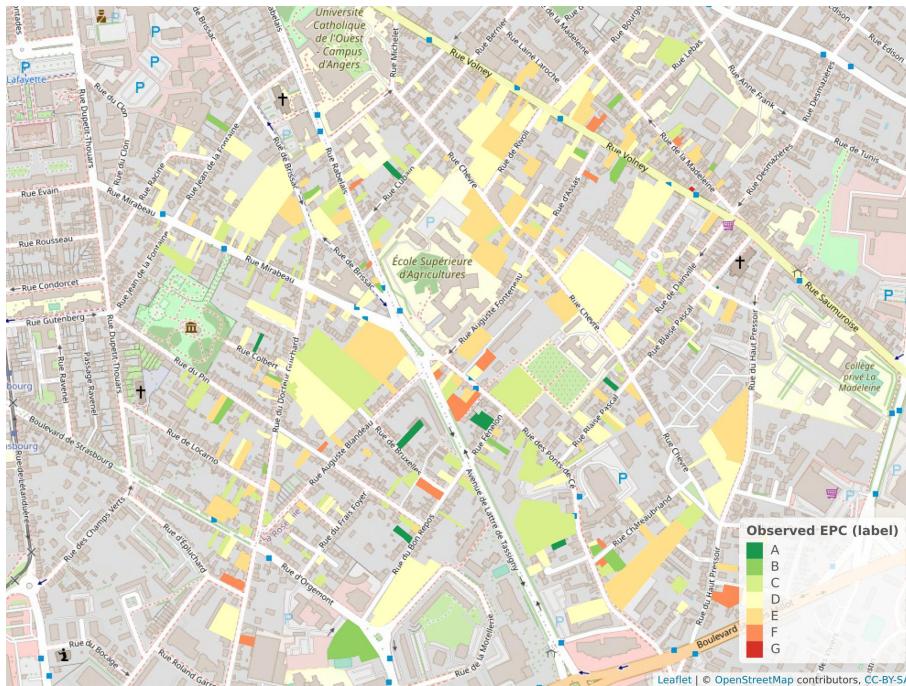
Il faut inverser la matrice de covariance des observations.

⇒ Difficile d'apprendre un grand nombre d'observations.

On cherche à combiner une famille de sous-modèles ayant le même effet qu'un grand modèle en utilisant des noyaux à support compact (en cours).

# Résultats d'un premier test sur un quartier d'Angers

## Les observations

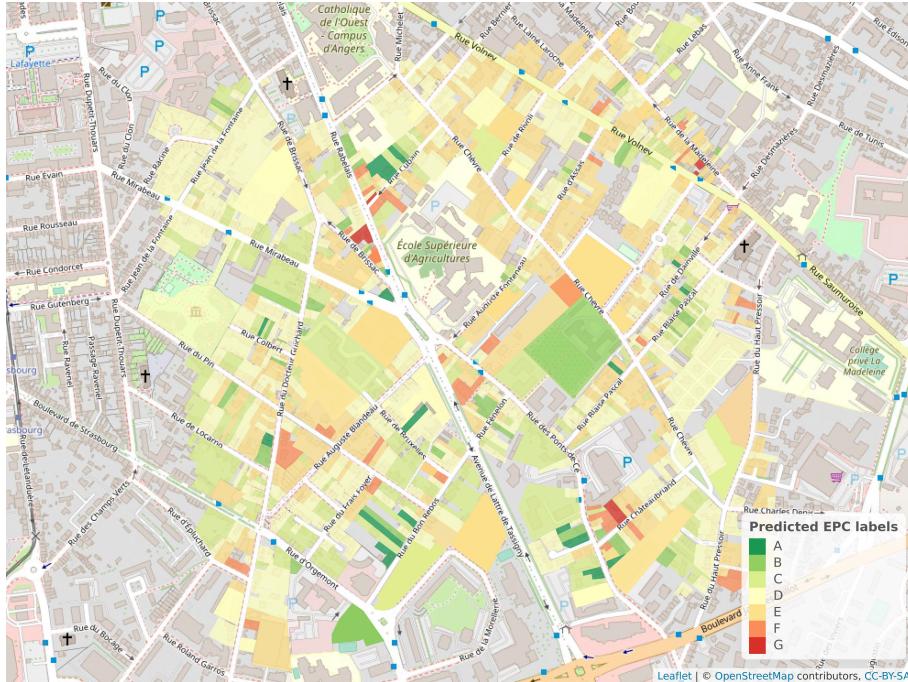


## Résultat des prédictions

True values	Predicted values						
	A	B	C	D	E	F	G
A	2	1	1	2	2	0	0
B	1	3	3	9	2	2	0
C	1	3	3	26	15	4	0
D	3	5	5	80	33	5	1
E	4	2	2	36	36	5	1
F	0	3	3	4	5	3	0
G	0	0	0	1	1	0	0

# Résultats d'un premier test sur un quartier d'Angers

## Les prédictions



## Résultat des prédictions

True values	Predicted values						
	A	B	C	D	E	F	G
A	2	1	1	2	2	0	0
B	1	3	3	9	2	2	0
C	1	3	3	26	15	4	0
D	3	5	5	80	33	5	1
E	4	2	2	36	36	5	1
F	0	3	3	4	5	3	0
G	0	0	0	1	1	0	0

# Conclusion

## Le DPE peut-être regardé comme une donnée géospatiale

- Il existe des observations de DPE abondantes.
- La principale limitation des modèles physiques thermiques, même simplifiés, est la nécessité d'une inspection physique.
- Plutôt que de tenter de reconstruire les variables physiques, il est possible d'envisager un modèle géospatial.
- Le modèle Mixture Kriging rend compte des incertitudes de position et de valeur des DPE observés.
- Avec 3 variables, on fait aussi bien que KNN avec 40 variables.

# Merci de votre attention. Avez vous des questions ?

L'Observatoire National  
des Bâtiments  
met à votre disposition  
les données ouvertes  
intégrées au long de ce  
process

[www.imope.fr/onb.html](http://www.imope.fr/onb.html)



## Annexes

Peut-on “lisser” les effets seuil ? Légitimité et validation ?

Géolocalisation, qu’en dit la BAN ?

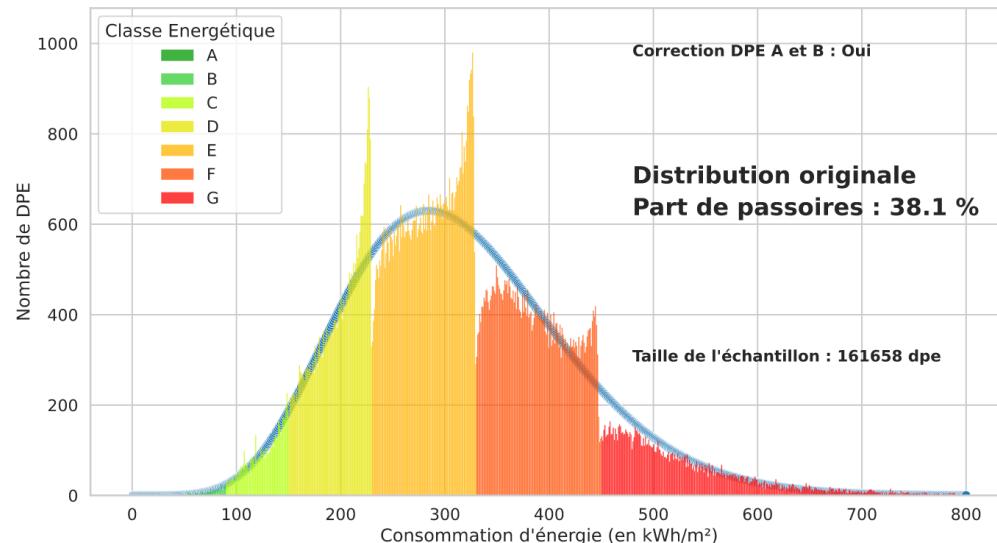
# Peut-on “lisser” les effets seuil ? Légitimité et validation ?

Source : Alternatives énergétiques, Yassine Abdelouadoud

Que mesure-t-on ?

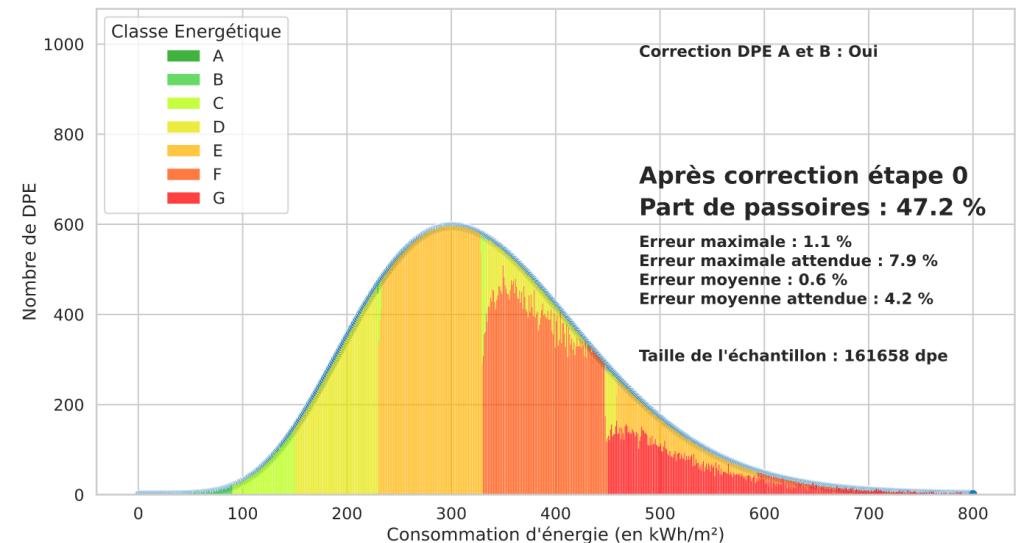
Année de construction : 1946 à 1974  
Méthode : 3CL

Type logement : Maison  
Combustible chauffage : Fioul



Année de construction : 1946 à 1974  
Méthode : 3CL

Type logement : Maison  
Combustible chauffage : Fioul



# Géolocalisation, qu'en dit la Base d'Adresses Nationale (BAN) ?

