HAL open science

# Impact de différents schémas d'échantillonnage sur la qualité prédictive de fonctions de pédotransfert entre deux réseaux de mesures de la qualité des sols

Benjamin Louis, Nicolas N. Saby, Dominique D. Arrouays, Claudy C. Jolivet, Line Boulonne, Eva Lacarce, Thomas T. Orton, Céline Ratié

## To cite this version:

Benjamin Louis, Nicolas N. Saby, Dominique D. Arrouays, Claudy C. Jolivet, Line Boulonne, et al.. Impact de différents schémas d'échantillonnage sur la qualité prédictive de fonctions de pédotransfert entre deux réseaux de mesures de la qualité des sols. 11. Journées d'Etude des Sols (JES), Mar 2012, Versailles, France. hal-02746023

**HAL Id: hal-02746023**

**https://hal.inrae.fr/hal-02746023**

Submitted on 3 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Impact de différents schémas d'échantillonnage sur la qualité prédictive de fonctions de pédotransfert entre deux réseaux de mesures de la qualité des sols

**LOUIS Benjamin, SABY Nicolas, ARROUAYS Dominique, JOLIVET Claudy, BOULONNE Line, LACARCE Eva, ORTON Thomas, RATIE Céline**

Journées d'étude des sols
20 mars 2012

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

# INTRODUCTION (1/5)

Contexte :

Suivi de la qualité des sols → **réglementations internationales**

versus

Réseaux de mesures → **échelles nationales** voire locales

Problème :

- Peu de standardisation *a priori* des protocoles d 'échantillonnages et d'analyses

# INTRODUCTION (2/5)

Demande :

Nécessité de construire des fonctions de pédotransfert (PTFs) entre les variables mesurées = **harmonisation *a posteriori***

$$VarA = PTF(VarB, \theta)$$

Besoin :

Avoir des **dispositifs d'échantillonnage différents** sur plusieurs sites de mesures **en parallèle** pour calibrer les PTFs

# INTRODUCTION (3/5)

QUESTION CLÉ :

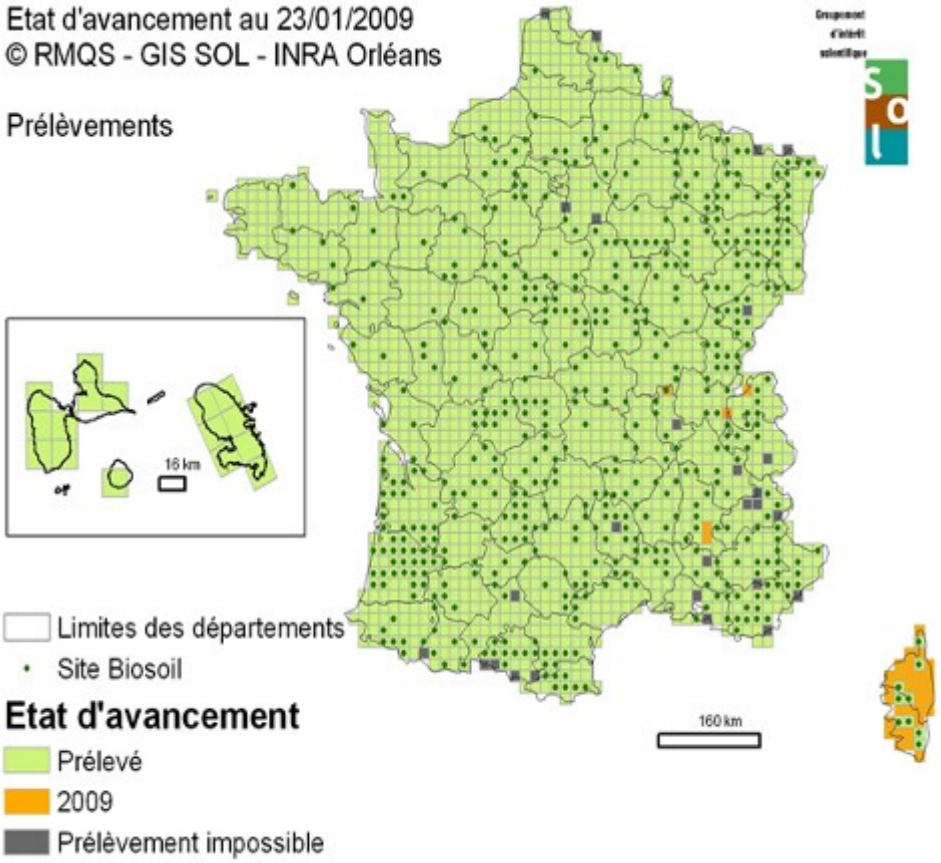
**Quels doivent être le nombre et la localisation des sites de calibrations pour assurer une bonne construction des PTFs ?**

# INTRODUCTION (4/5)

Sols forestiers français : deux réseaux calés sur une même grille de maille 16*16 km (524 sites)

-échelle nationale : **RMQS**

- échelle européenne : **BioSoil**






ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

# INTRODUCTION (5/5)

Comparaison des variables mesurées RMQS versus BioSoil :

→ **construction des PTFs** selon différents schémas d'échantillonnage

→ étude de la **qualité prédictive des PTFs** en fonction du nombre et de la localisation des sites échantillonnés

# MATERIELS ET METHODES (1/10)

## Comparaisons des dispositifs et variables utilisées

| | BioSoil | RMQS | Ressemblance |
|---|---|---|---|
| Réseau | Maille 16*16 km² | Maille 16*16 km² | = |
| Epaisseurs étudiées | 0-10 cm, 10-20 cm et 20-40 cm | 0-30 cm | ≠ |
| Carbone organique | Combustion sèche ISO 10694:1995 | Combustion sèche NF ISO 10694 | = |
| Potassium | HF & $HclO_4$ (total) ISO 14869-1:2001 | HF & $HclO_4$ (total) NF X 31-147 | = |
| Plomb | Eau régale (extractible) ISO 11466:1995 mod. | HF & $HclO_4$ (total) NF X 31-147 | ≠ |
| pH | Suspension 1/5 d'eau ISO 10390:1994 | Suspension 1/5 d'eau | = |

# MATERIELS ET METHODES (2/10)

**Construction des PTFs :**

Intuitivement : $$Var_{RMQS} = \alpha + \sum_i \beta_i * Var_{i,\,BioSoil}$$

Besoin de toutes les épaisseurs BioSoil ?

→ Sélection de modèle par méthode *« pas-à-pas »* à l'aide de tous les sites de mesures :

**Postulat : on obtient le meilleur modèle**

# MATERIELS ET METHODES (3/10)

**Schémas d'échantillonnage** :

- **Pour le nombre** : 50%, 40%, 30% , 20%, 10% et 5%
- **Pour la localisation** :
  - → échantillonnage aléatoire simple (**RS**) (x 1000)
  - → échantillonnage par hypercube Latin conditionné (Minasny et McBratney, 2006 **cLHS** et 2010 **DLHS**)

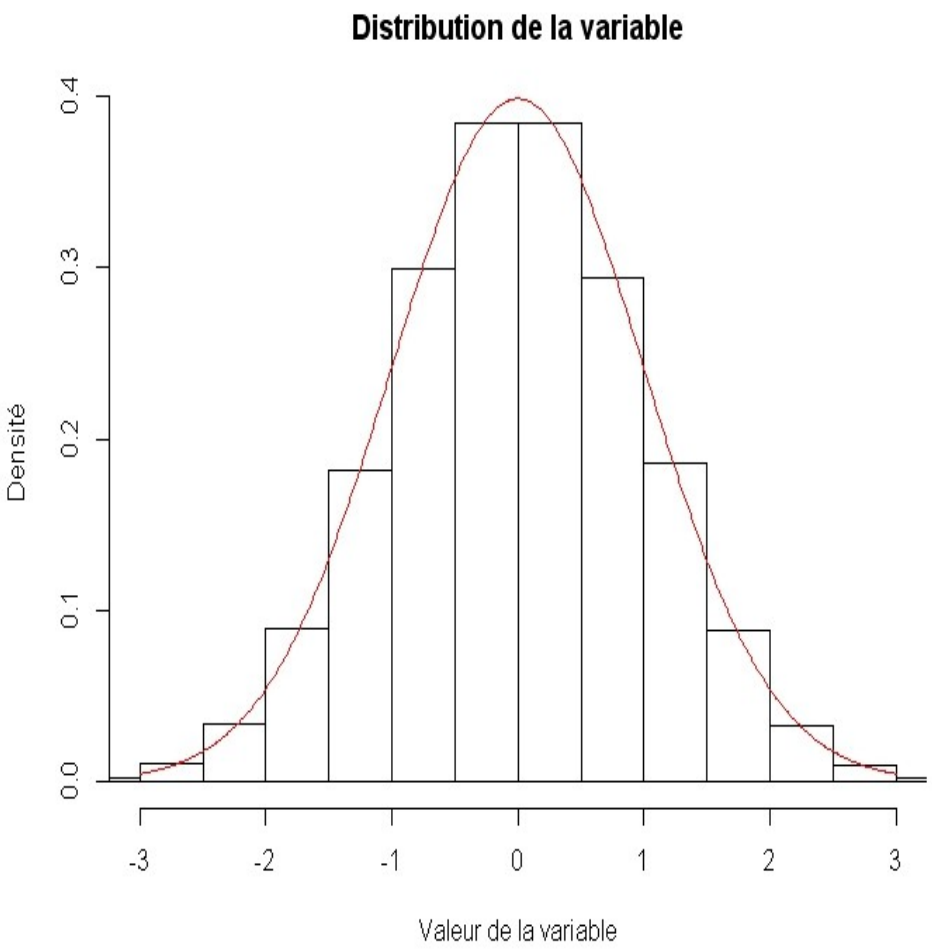
# MATERIELS ET METHODES (4/10)

**Échantillonnage par Hypercube Latin ?**

→ Permet d'obtenir un échantillon dont la distribution de toutes les variables est maximisée

→ Les variables ne peuvent pas être prises indépendamment

**→ cLHS**



# MATERIELS ET METHODES (5/10)

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

1 **Calcul de Obj**

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| **2** | $x_{1,2}$ | $x_{2,2}$ | $x_{3,2}$ |
| **4** | … | … | … |
| **6** | … | … | … |
| **7** | … | … | … |
| **9** | … | … | … |
| **10** | … | … | … |
| **11** | … | … | … |
| **12** | $x_{1,12}$ | $x_{2,12}$ | $x_{3,12}$ |

# MATERIELS ET METHODES (6/10)

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

1 **Calcul de Obj**

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 2 | $x_{1,2}$ | $x_{2,2}$ | $x_{3,2}$ |
| 4 | … | … | … |
| 6 | … | … | … |
| 7 | … | … | … |
| 9 | … | … | … |
| 10 | … | … | … |
| 11 | … | … | … |
| 12 | $x_{1,12}$ | $x_{2,12}$ | $x_{3,12}$ |

2

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 9 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

**Calcul de Obj**

# MATERIELS ET METHODES (7/10)

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

1 **Calcul de Obj**

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

< p, on accepte pas

Metro=exp(-ΔObj/T)

3

> p, on accepte

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 2 | $x_{1,2}$ | $x_{2,2}$ | $x_{3,2}$ |
| 4 | … | … | … |
| 6 | … | … | … |
| 7 | … | … | … |
| 9 | … | … | … |
| 10 | … | … | … |
| 11 | … | … | … |
| 12 | $x_{1,12}$ | $x_{2,12}$ | $x_{3,12}$ |

2

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 9 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

**Calcul de Obj**

# MATERIELS ET METHODES (8/10)

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

1 **Calcul de Obj**

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 3 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

< p, on accepte pas

Metro=exp(-ΔObj/T)

3

> p, on accepte

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 2 | $x_{1,2}$ | $x_{2,2}$ | $x_{3,2}$ |
| 4 | … | … | … |
| 6 | … | … | … |
| 7 | … | … | … |
| 9 | … | … | … |
| 10 | … | … | … |
| 11 | … | … | … |
| 12 | $x_{1,12}$ | $x_{2,12}$ | $x_{3,12}$ |

2

| Site | Var1 | Var2 | Var3 |
|---|---|---|---|
| 1 | $x_{1,1}$ | $x_{2,1}$ | $x_{3,1}$ |
| 8 | … | … | … |
| 9 | … | … | … |
| 5 | $x_{1,5}$ | $x_{2,5}$ | $x_{3,5}$ |

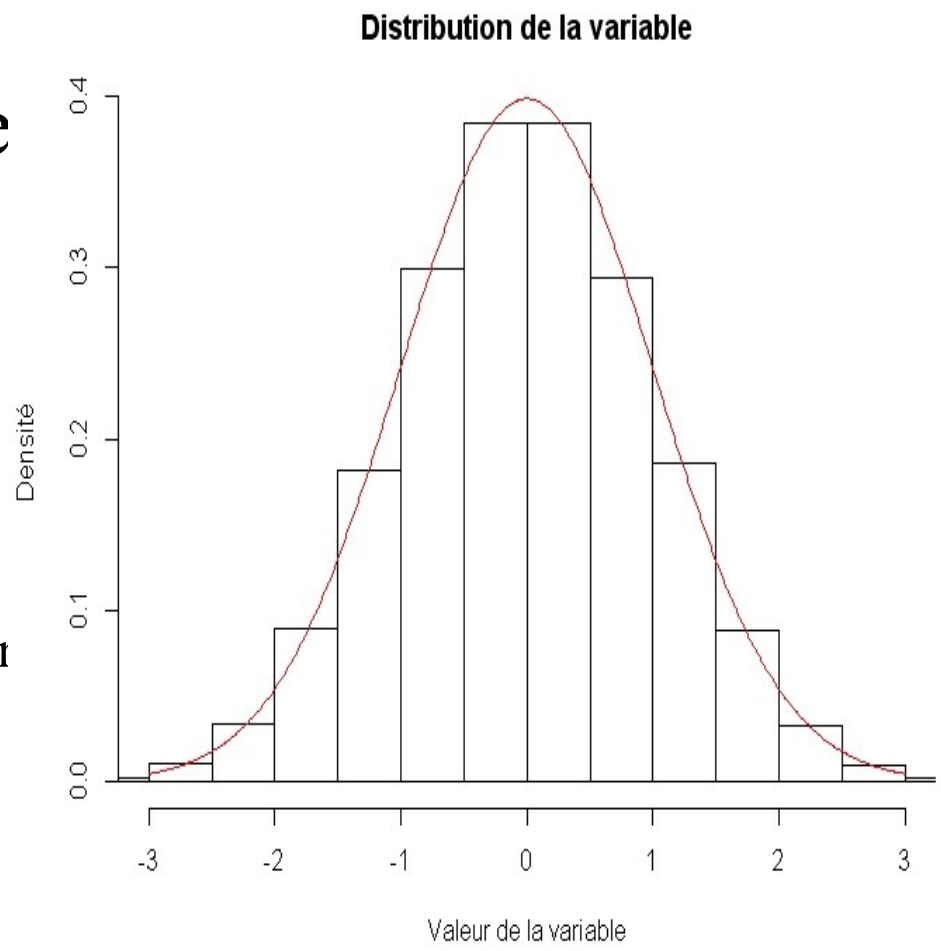
**Calcul de Obj**

4 100 x 50000

# MATERIELS ET METHODES (9/10)

- **DLHS** : même principe mais en donnant plus de poids aux valeurs extrêmes (100 x 50000)

→ inspiré par la D-optimalité (St. John and Draper 1975)






ALIMENTATION
AGRICULTURE
ENVIRONNEMENT
INRA

# MATERIELS ET METHODES (10/10)

Procédure d'évaluation de la qualité prédictive des PTFs :

i. Appliquer la stratégie d'échantillonnage pour obtenir un échantillon de taille n

ii. Calibrer les PTFs à l'aide des sites échantillonnés

iii. Prédire les valeurs des sites non échantillonnés

iv. Calculer la racine carrée de l'erreur quadratique moyenne de prédiction avec la formule:

$$RMSEP = \sqrt{\frac{1}{N-n}\sum(\hat{y} - y_{obs})^2}$$

# RESULTATS (1/2)

**PTFs sélectionnées :**

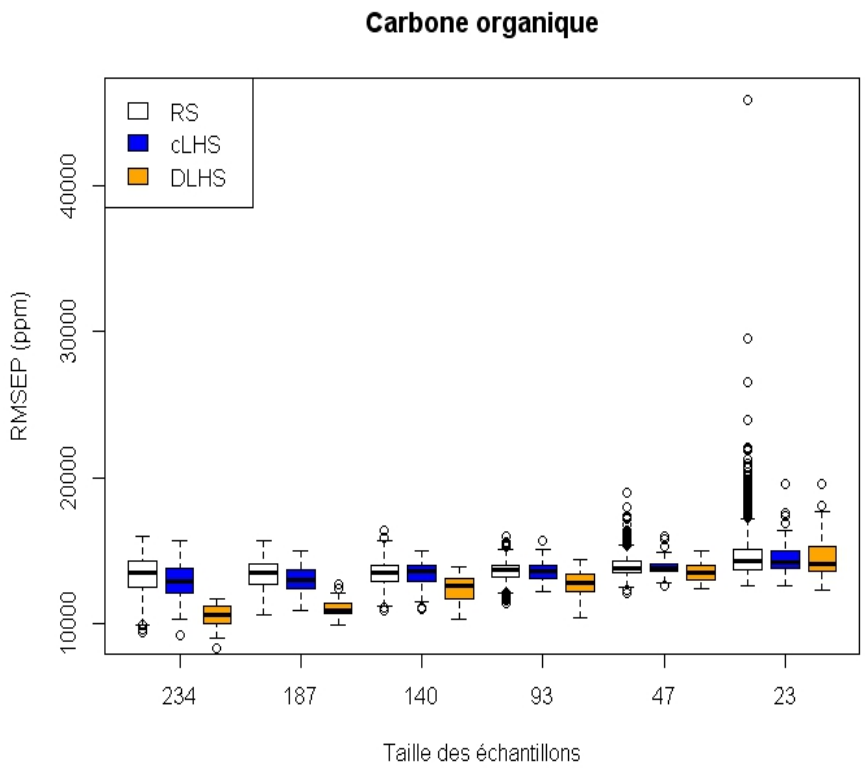
$$\log(C_{RMQS}) = \alpha + \sum_i \beta_i * \log(C_{BioSoil,i})$$

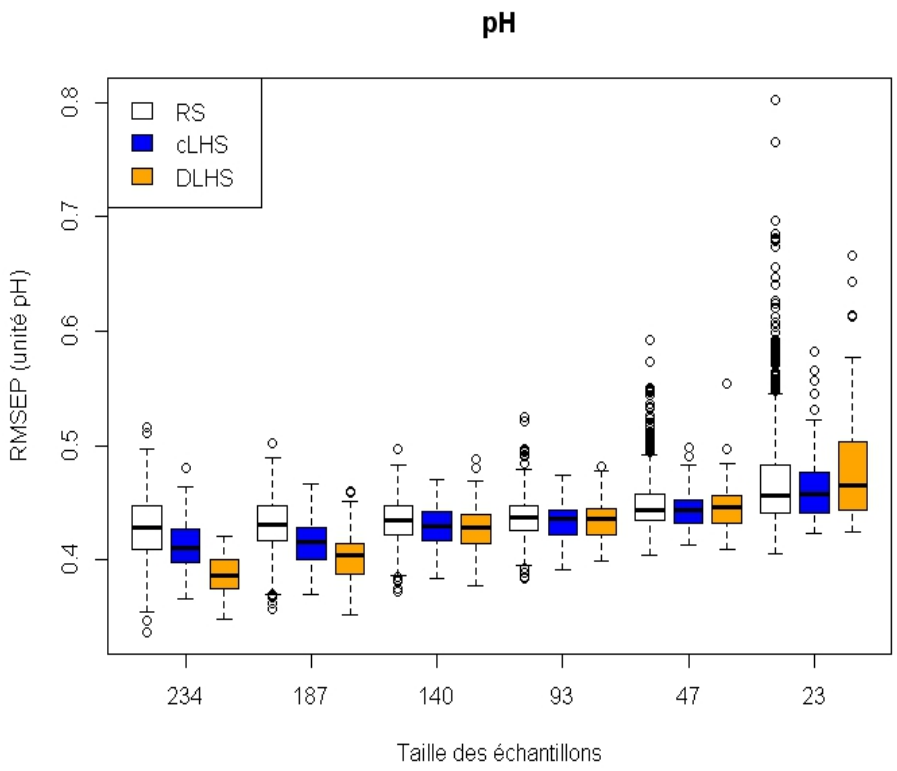
$$K_{RMQS} = \alpha + \beta_{01} * K_{BioSoil,01} + \beta_{24} * K_{BioSoil,24}$$

$$pH_{RMQS} = \alpha + \beta_{01} * pH_{BioSoil,01} + \beta_{24} * pH_{BioSoil,24}$$

$$\log(Pb_{RMQS}) = \alpha + \beta_{01} * \log(Pb_{BioSoil,01})$$

# RESULTATS (2/2)



# CONCLUSION

→ cLHS et DLHS permettent de réduire le risque d'erreurs extrêmes de prédiction

- En particulier :

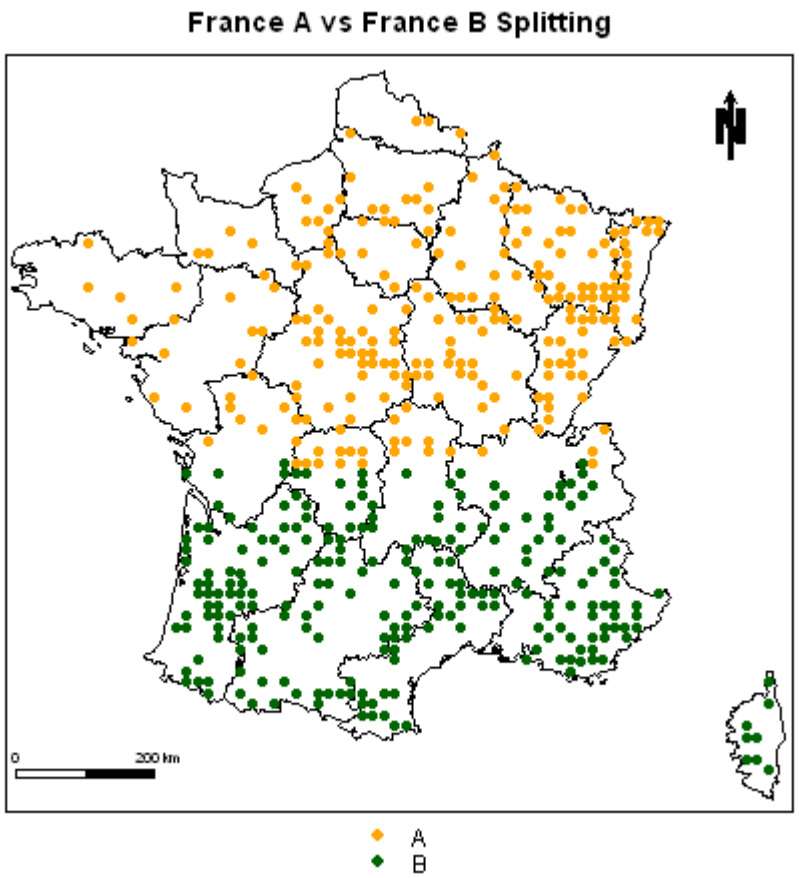
- Grande taille d'échantillon : DLHS

- Petite taille d'échantillon : cLHS

- Taille intermédiaire : dépend de la variable et de la relation attendue (linéaire ou pas?)

# Pour aller plus loin…

- Simulation d'un cas réel : France A vs France B
- Intervention de variables pédologiques
- Utilisation de méthode efficace pour la prédiction (Boosting regression trees,...)



# MERCI DE VOTRE ATTENTION!