



Traitements des données sols

Nicolas N. Saby, Manuel Pascal Martin, Christine Le Bas

► To cite this version:

Nicolas N. Saby, Manuel Pascal Martin, Christine Le Bas. Traitements des données sols. Séminaire du Département Environnement et Agronomie "Les Bases de données SOL", Sep 2014, Orléans, France. 31 p. hal-02792189

HAL Id: hal-02792189

<https://hal.inrae.fr/hal-02792189>

Submitted on 5 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.



Traitements des données sols

N Saby, M Martin, C Le Bas

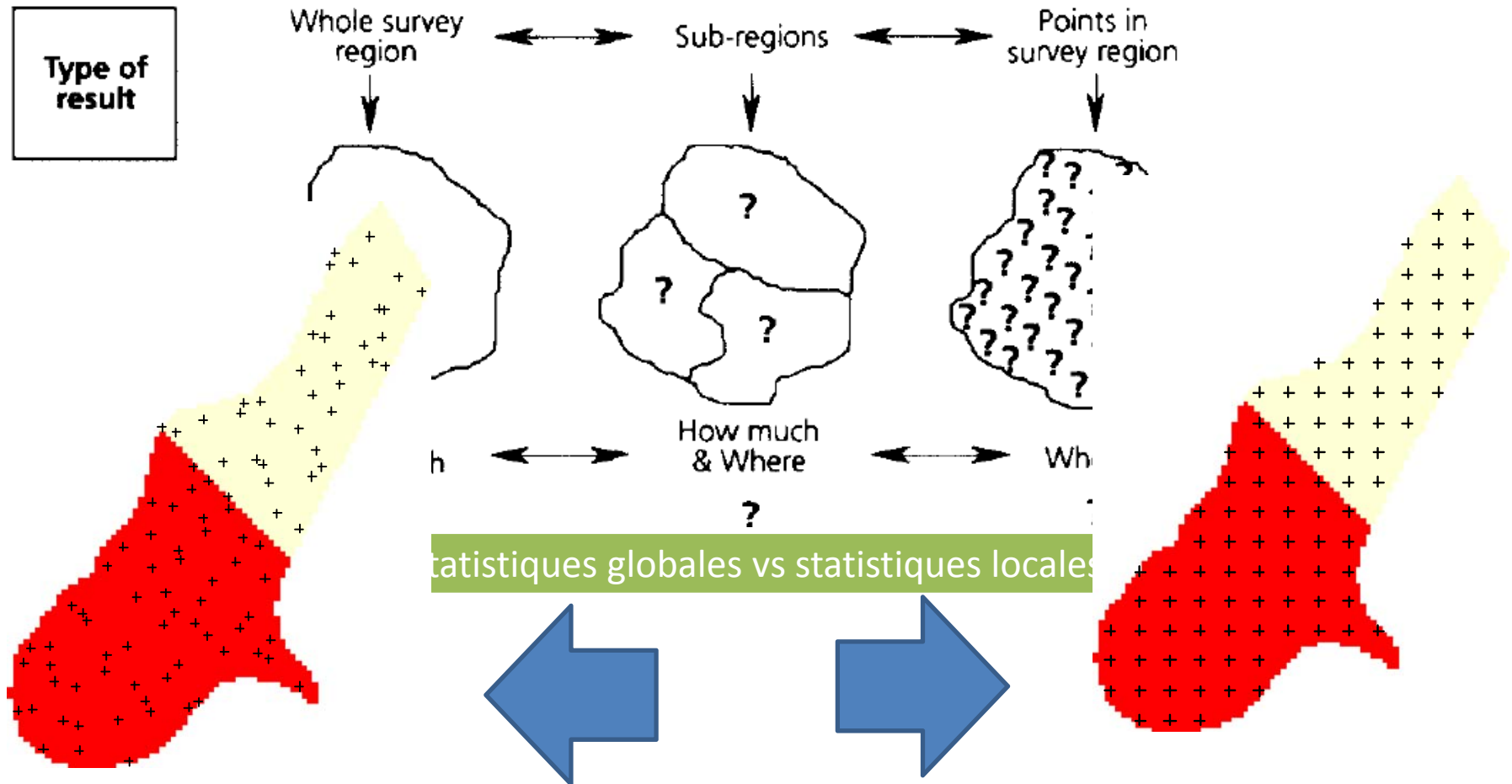




Introduction

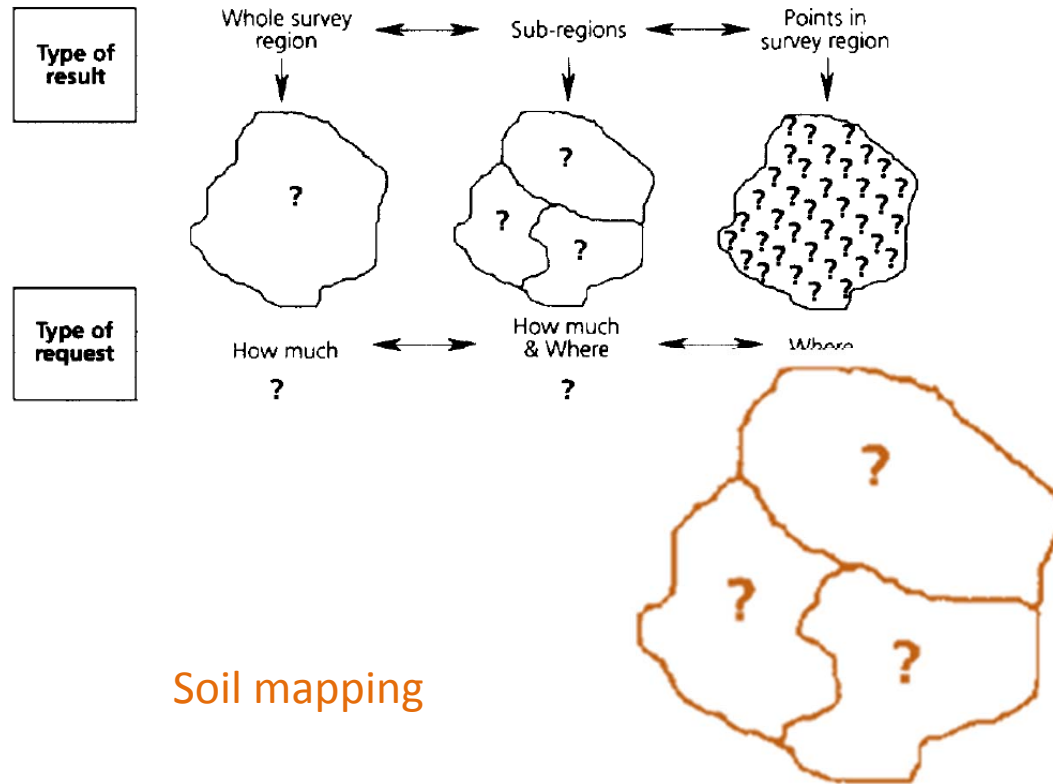
- Définition : Propriétés du sol au sens large (de la teneur en carbone jusqu'aux contaminants)
- Objectif triple des traitements
 - pour la compréhension de la distribution (spatiale et temporelle) des propriétés des sols => faire des cartes, détecter des évolutions et hiérarchiser les déterminants.
 - Pour renseigner les modèles mécanistes
 - pour traiter des questions plus larges des domaines de l'agronomie et de l'environnement (comptabilité carbone, cartographie de l'inégalité environnementale, etc.)

Échantillonnage statistique



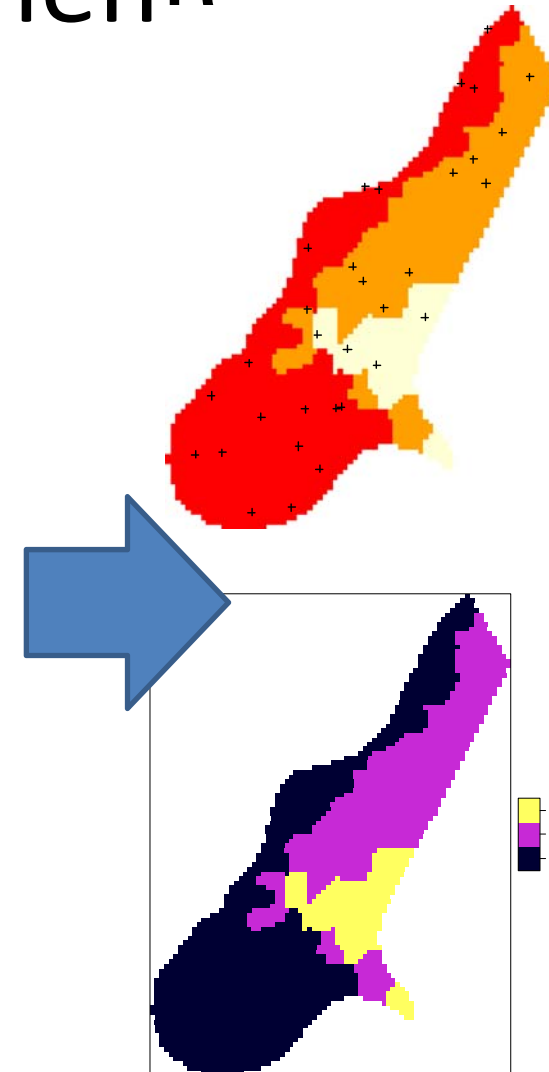
Domburg et al. *Geoderma* 1994

Échantillonnage orienté



Soil mapping

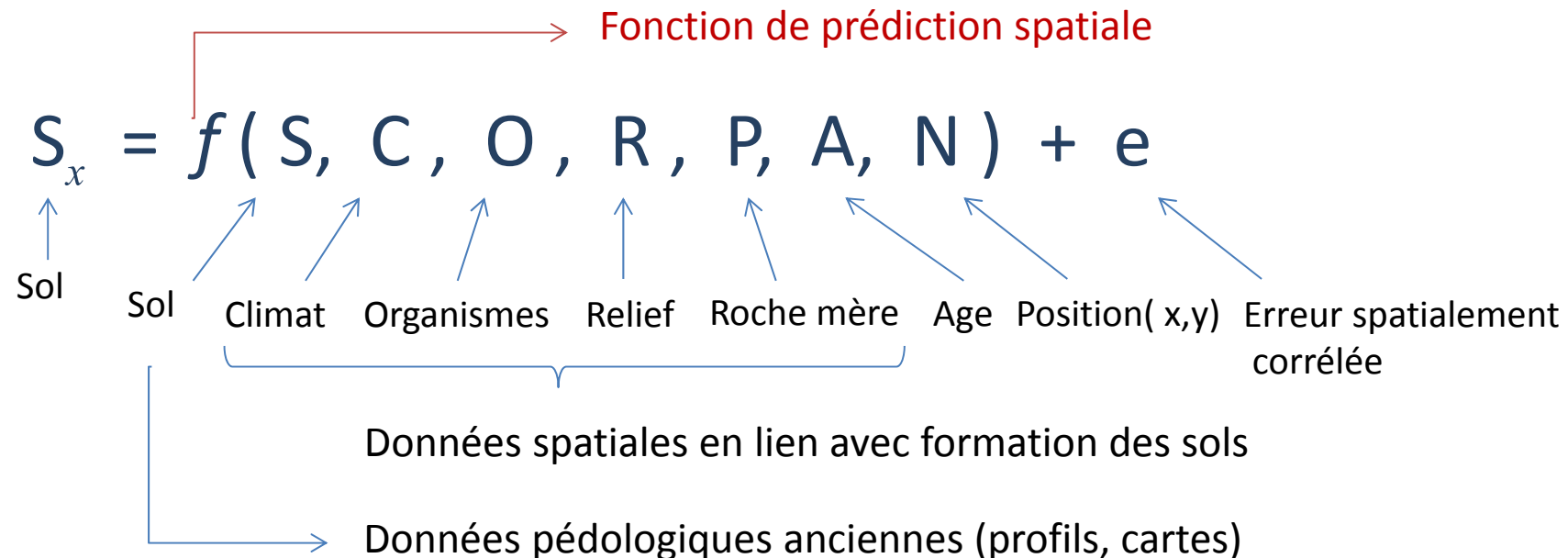
How much &
Where ?



Définition de la cartographie numérique

« Création et enrichissement de systèmes d'information pédologiques (...) par des modèles numériques inférant les variations spatiales et temporelles des sols et leurs propriétés à partir d'observations de sol et de données spatiales d'environnement » (P. Lagacherie)

Principe général (Jenny et al, 1941 puis McBratney et al, 2003)



Les traitements

Principe général *(Jenny et al, 1941 puis McBratney et al, 2003)*

$$S_x = f(S, C, O, R, P, A, N) + e$$

→ Fonction de prédiction spatiale

→ Analyse supervisée (Régression)

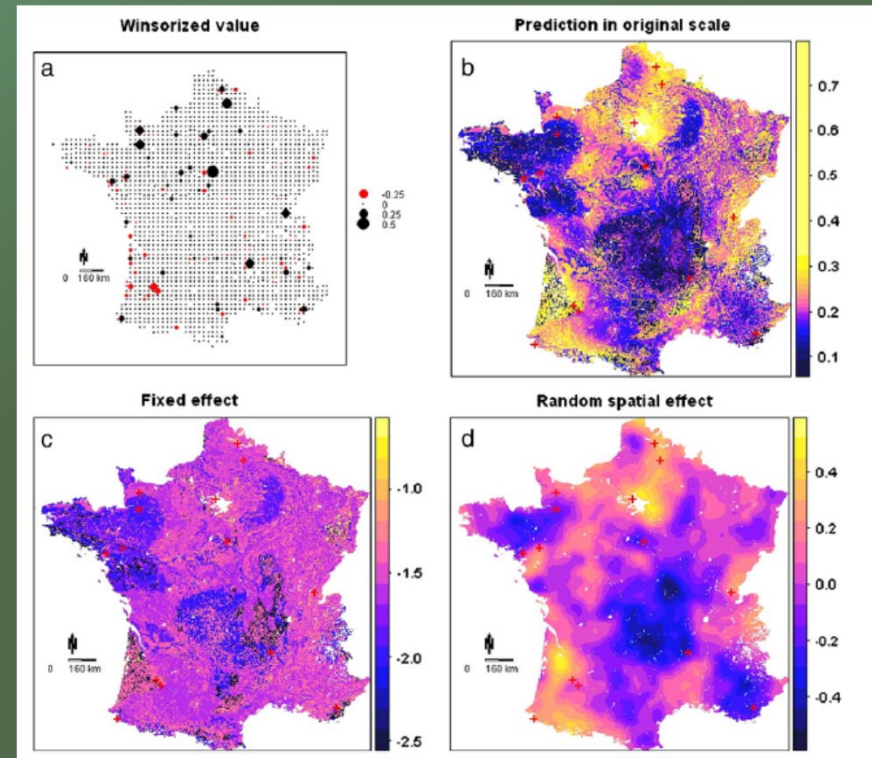
→ Analyse non supervisée (Clustering, ACP, etc.)

Plan

1. lister et hiérarchiser les déterminants de la variabilité spatio-temporelle des propriétés des sols
2. produire des statistiques globales (non locales) sur les propriétés des sols
3. Cartographier (statistiques locales) des propriétés des sols
4. Modéliser à l'aide de propriétés de sol disponibles dans les BDD

01

Lister et hiérarchiser les déterminants de la variabilité spatio-temporelle des propriétés du sol



Exemple 1 = déterminants des stocks de carbone

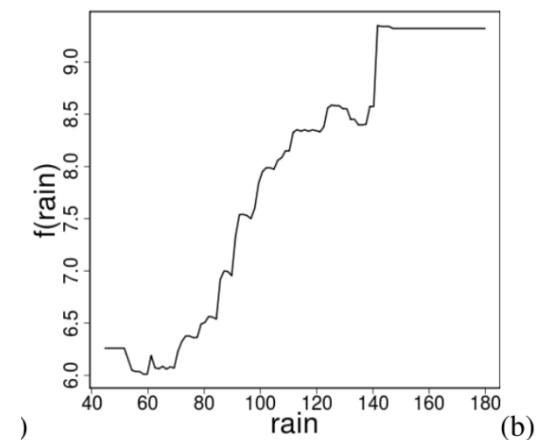
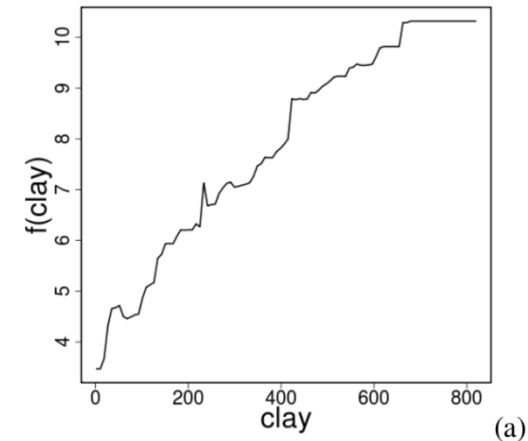
- Analyse supervisée
- Données = RMQS, composites de surface
- f = Régression multiple (GBM)
- Covariables SCORPAN = occupation du sol*, climat, texture*, comportement hydrique*, pH*, production primaire
- * : observé sur les profils constituant le jeu de données d'ajustement du modèle
- validation croisée

$$S_x = f(S, C, O, R, P, A, N) + e$$

Déterminants des stocks de C

Predictor	Cult model		Forest model	
	VIM	rank	VIM	rank
<i>lu3</i>	33.66	1	0.77	11
<i>lu2</i>	1.26	13	0.00	14
<i>lu1</i>	0.11	15	—	—
<i>lu.ipcc</i>	0	16	—	—
<i>a</i>	7.1	3	1.47	10
<i>b</i>	3.72	7	4.83	7
<i>rain</i>	6.6	4	13.27	3
<i>pet</i>	3.3	8	4.4	8
<i>temp</i>	3.03	9	1.83	9
<i>npp</i>	2.89	10	6.54	5
<i>wlogging</i>	1.34	12	0.06	12
<i>wregime</i>	1.14	14	0.03	13
<i>rf</i>	6.08	5	8	4
<i>clay</i>	22.55	2	29.55	1
<i>silt</i>	1.96	11	5.91	6
<i>ph</i>	5.26	6	23.35	2

Martin et al. / Biogeosciences, 8, 1053–1065, 2011



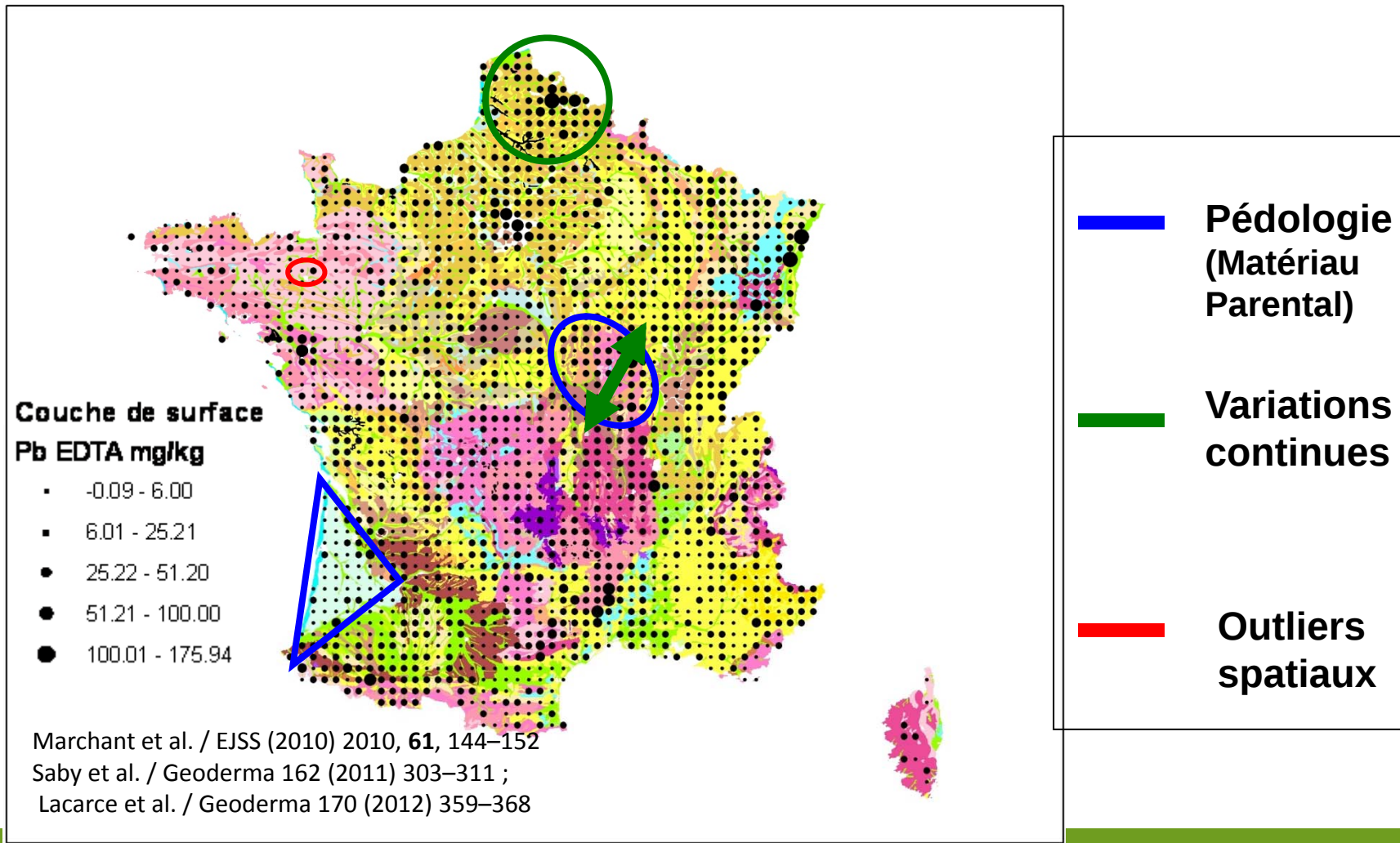


Exemple 2 = ETMs

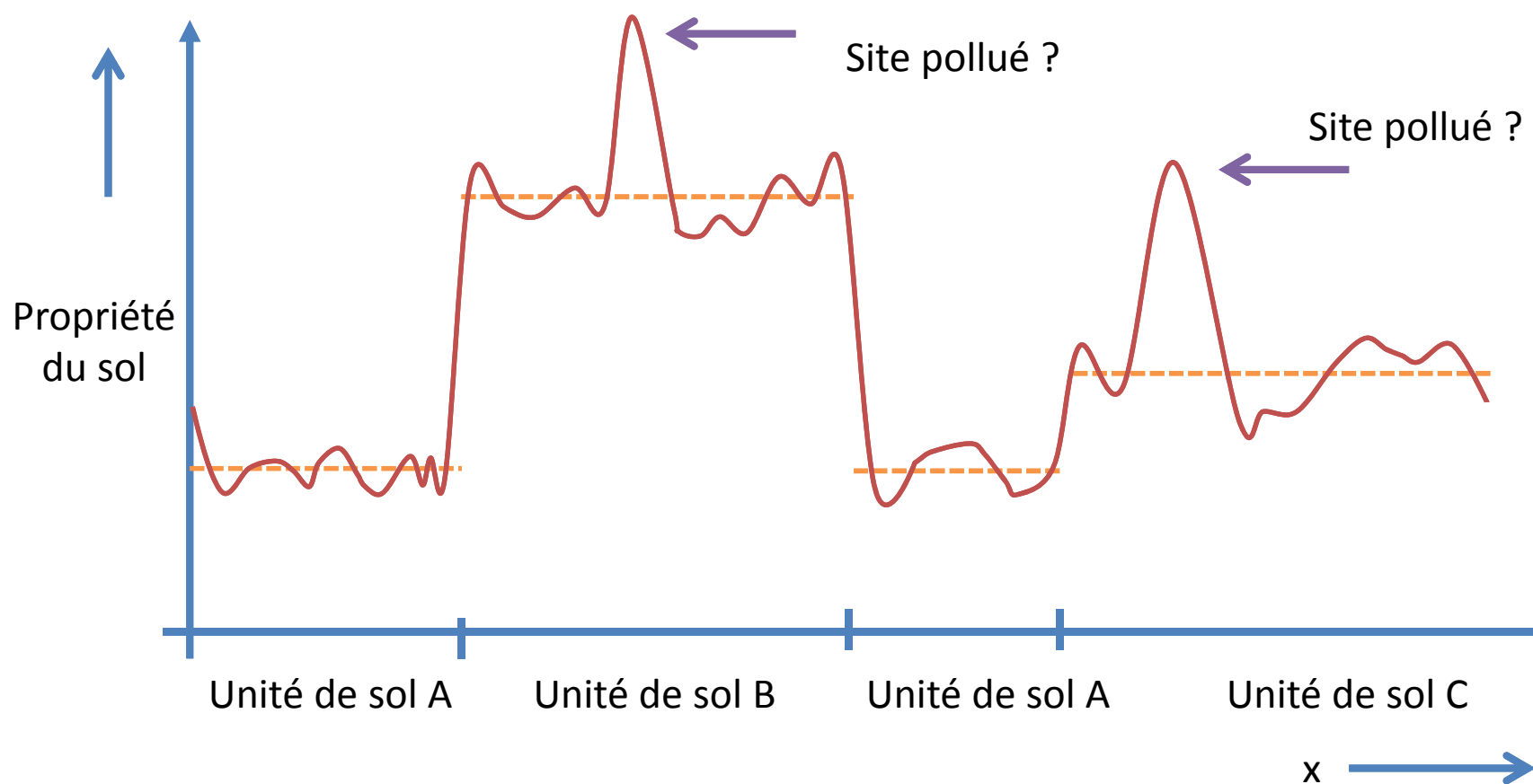
- Analyse supervisée
- Données = RMQS, composites
- f = Modèle linéaire mixte robuste
- Covariable SCORPAN = r matériau parental issue de la base au millionième (BDGSF)
- validation croisée

$$S_x = f(S, C, O, R, P, A, N) + e$$

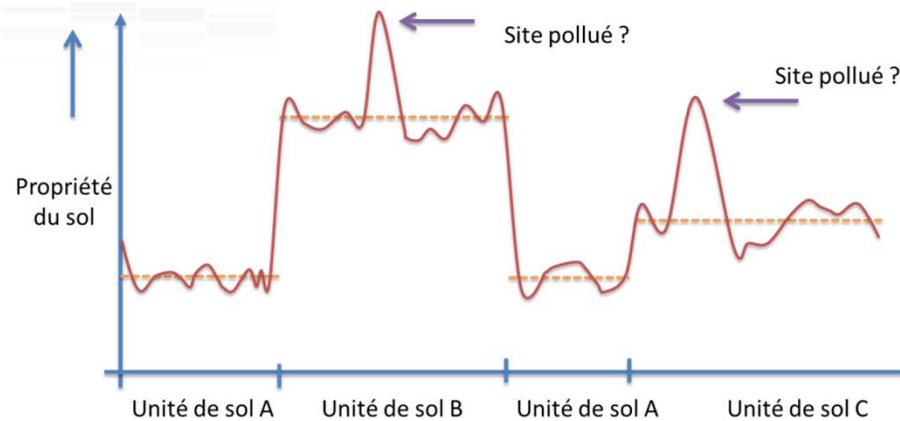
3 sources de variation spatiale



Exemple de dérivation en marche d'escalier (carte de sols)



Modèle linéaire mixte robuste



•
$$S(x, y) = m(x, y) + \varepsilon(x, y)$$

La propriété du sol

La « dérive », partie
déterministe ou expliquée
du modèle

Résidu stochastique, partie non expliquée,
spatialement corrélé

Modèle linéaire mixte robuste

$$S_x = f(S, C, O, R, P, A, N) + e$$

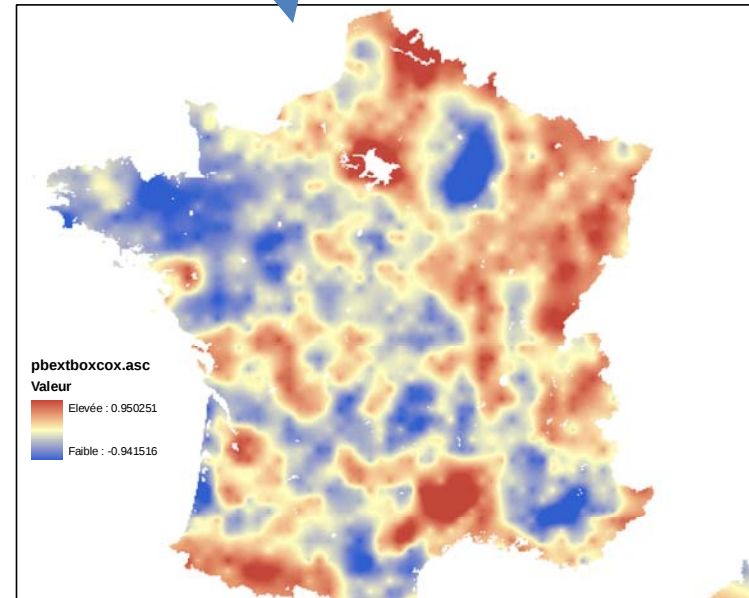
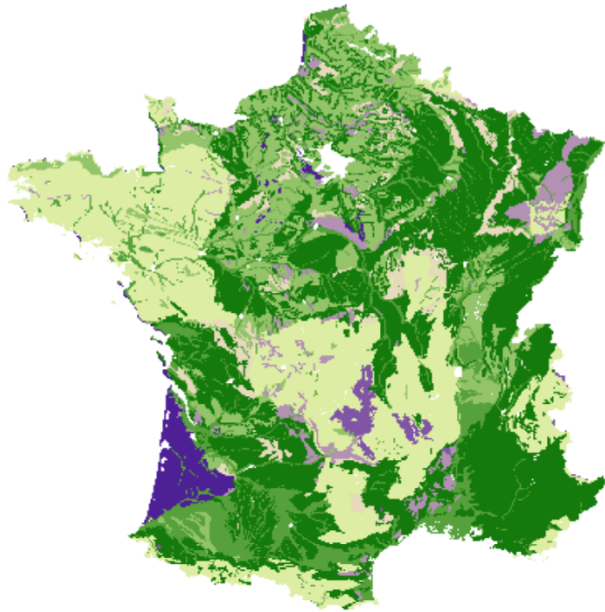
- $S(x, y) = m(x, y) + \varepsilon(x, y)$
 - La propriété du sol
 - La « dérive », partie déterministe ou expliquée du modèle
 - Résidu stochastique, partie non expliqué, spatialement corrélé

- Estimateurs et prédiction robustes

Séparer les effets

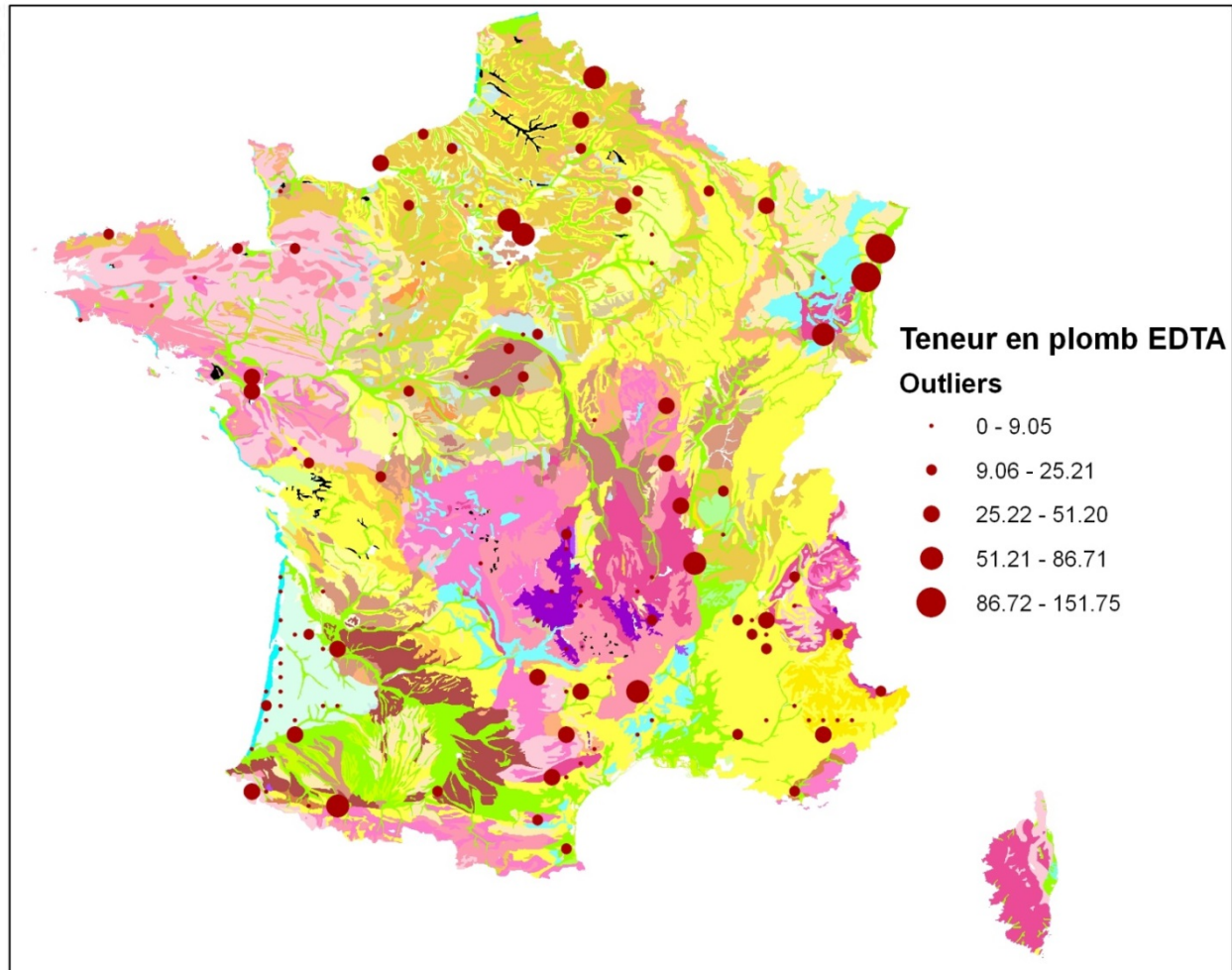
$$S(x, y) = m(x, y) + \varepsilon(x, y)$$

Teneurs moyennes en Pb EDTA
par matériau



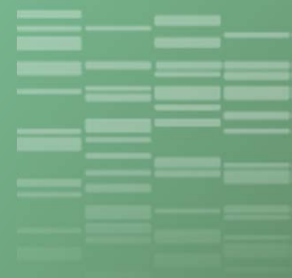
Effet des variations continues

Les outliers



Remarques

- Généricité des fonctions f
- Interactions entre variables non linéaires
- Validation (croisée)
- Expertise pédologique
- Expertise en géostatistique



02

**produire des statistiques globales
(non locales) sur les propriétés
des sols**

Exemple 1 Proposer des valeurs de référence

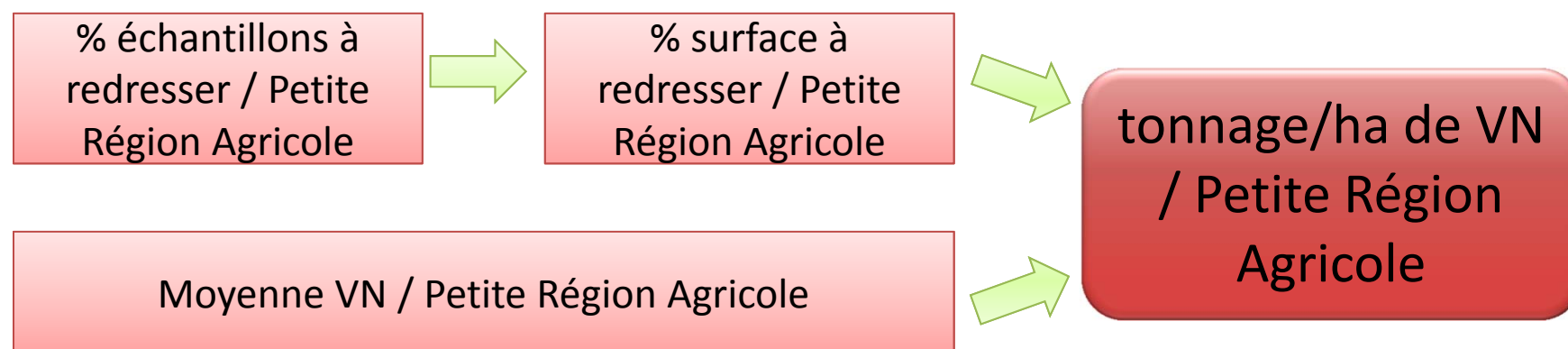
- => exemple du carbone
- Statistiques régionales sur le RMQS
 - Estimation des paramètres de distribution des stocks par grand type d'usage et incertitudes associées.
 - Estimations régionales utilisées dans le cadre des inventaires
 - Rapportage dans le cadre du Programme de développement rural 2014-2020 (Commission européenne)
 - ! Attention au calcul de la variance d'échantillonnage (mode d'échantillonnage du RMQS)

Exemple 2 = Estimer les besoins en Valeurs Neutralisantes à partir des données de la BDAT

- ❖ 2 objectifs de correction de pH : 6,3 et 6,8
- ❖ Période = 2005-2010 (> 650 000 pH et CEC)

$$VN = f(\Delta S/T, CEC)$$

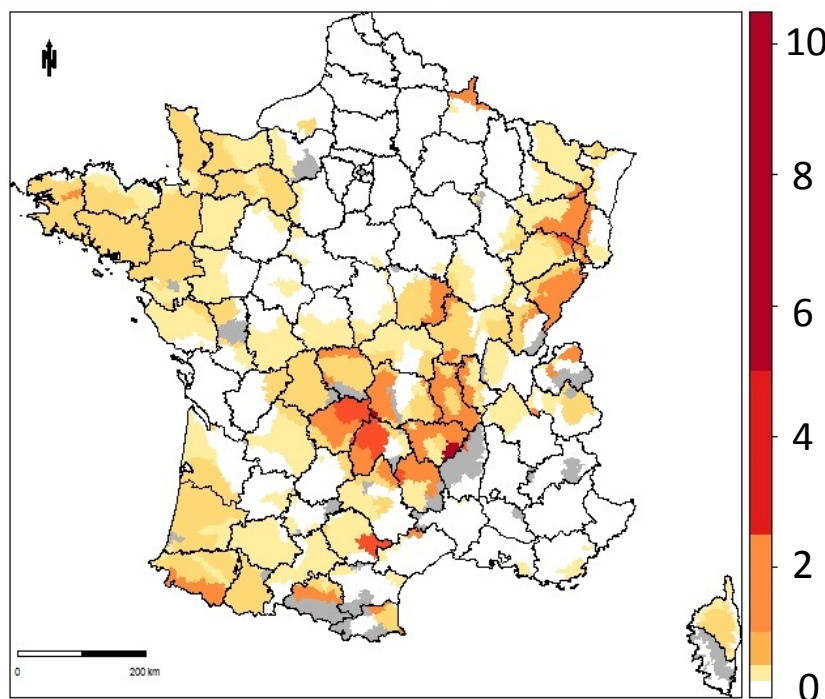
S/T	50	70	90	110	130
pH_{eau}	5,3	5,9	6,5	7,1	7,7



Estimer les besoins en VN

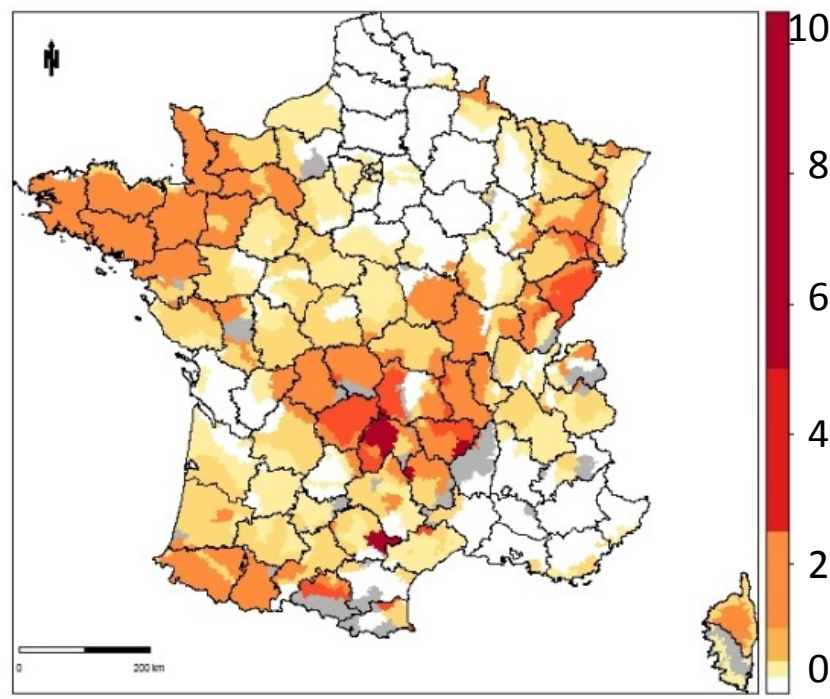
VN moyenne à apporter par PRA en t/ha

Correction de pH à 6,3



20 % des échantillons de la BDAT
Total : 4 millions de tonnes

Correction de pH à 6,8



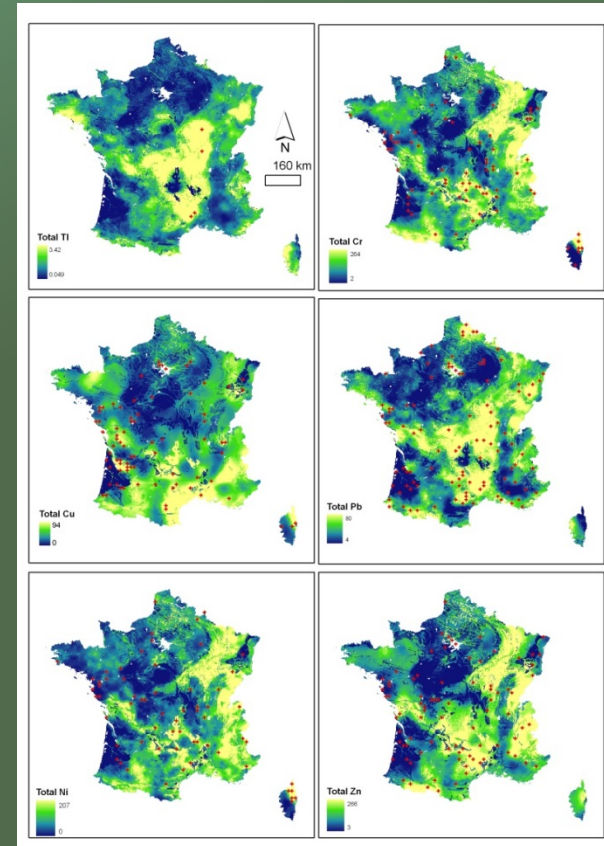
40 % des échantillons de la BDAT
Total : 10 millions de tonnes

Remarques

- Mode d'échantillonnage : il peut n'avoir pas été raisonné en adéquation avec la question posée
- Nombre d'individus dans chaque catégorie pour faire des analyses statistiques
- Ne pas oublier de calculer les incertitudes sur les valeurs

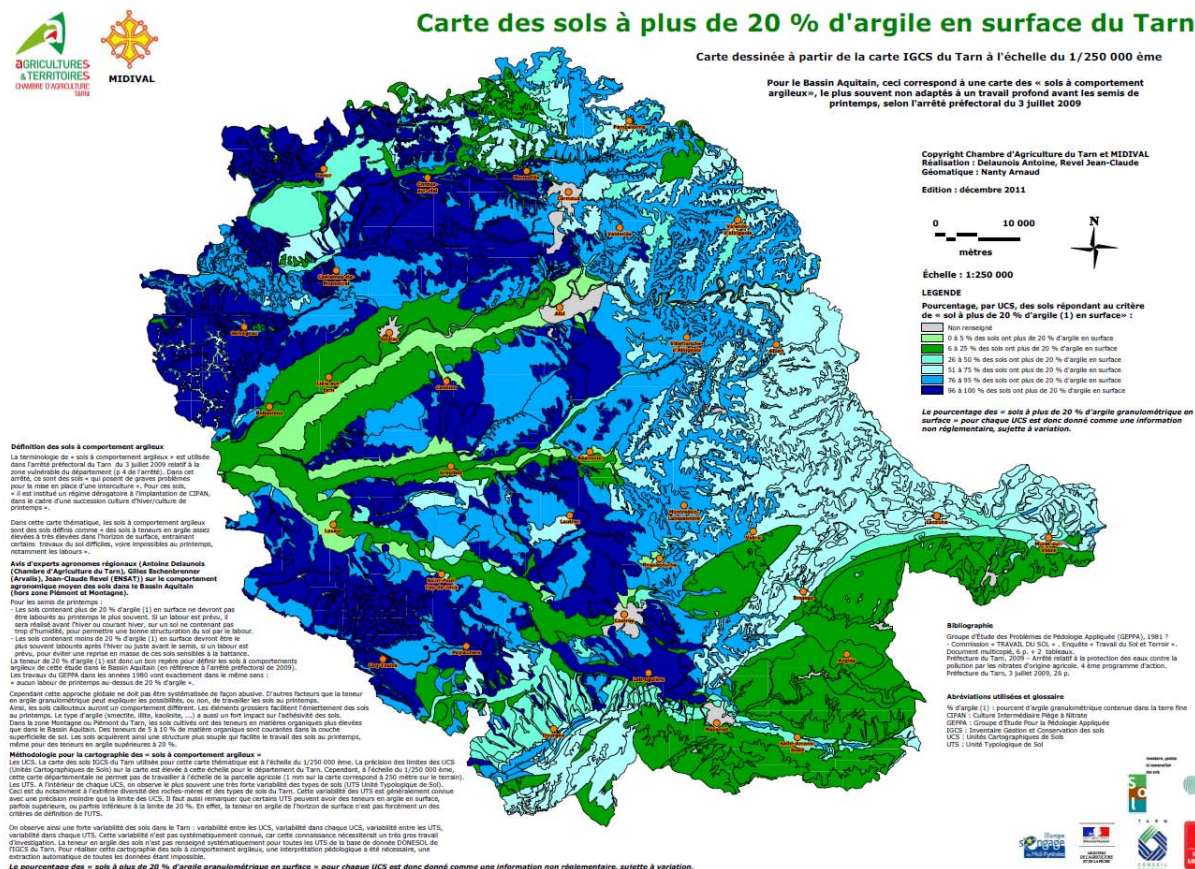
03

Cartographier (statistiques locales) des propriétés des sols



Exemple 1 = Cartographies thématiques

- Analyses non supervisée
- Expertise
- Données = IGCS UCS/UTS
- Covariable SCORPAN = sol
- Pas de Validation statistique





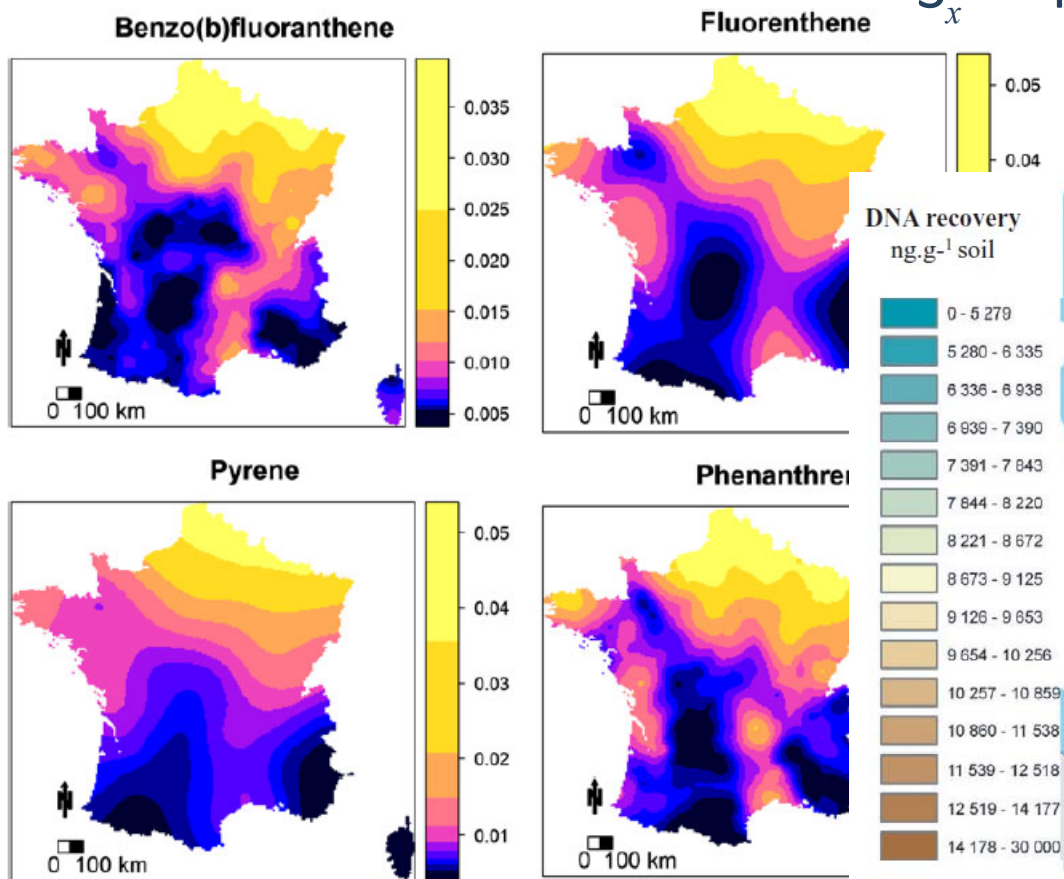
Exemple 2 = Cartographies à partir du RMQS

- Analyses supervisée
- Données = RMQS
- f = Modèle linéaire mixte
- Covariable SCORPAN = n
- Validation croisée
- Carte d'incertitude

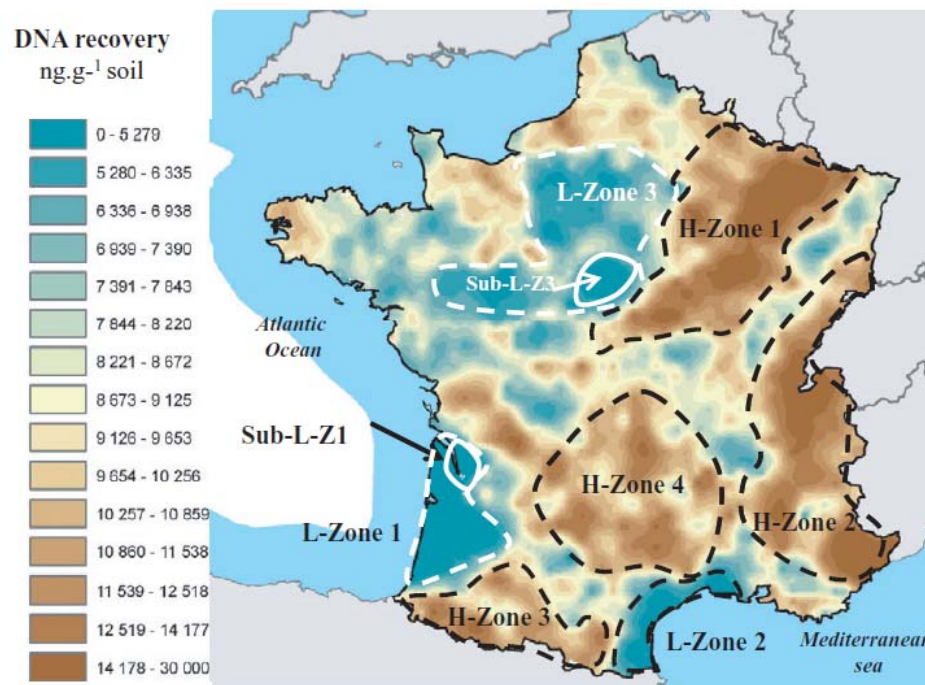
$$S_x = f(S, C, O, R, P, A, N) + e$$

Exemples purement spatiaux

$$S_x = f(S, C, O, R, P, A, N) + e$$



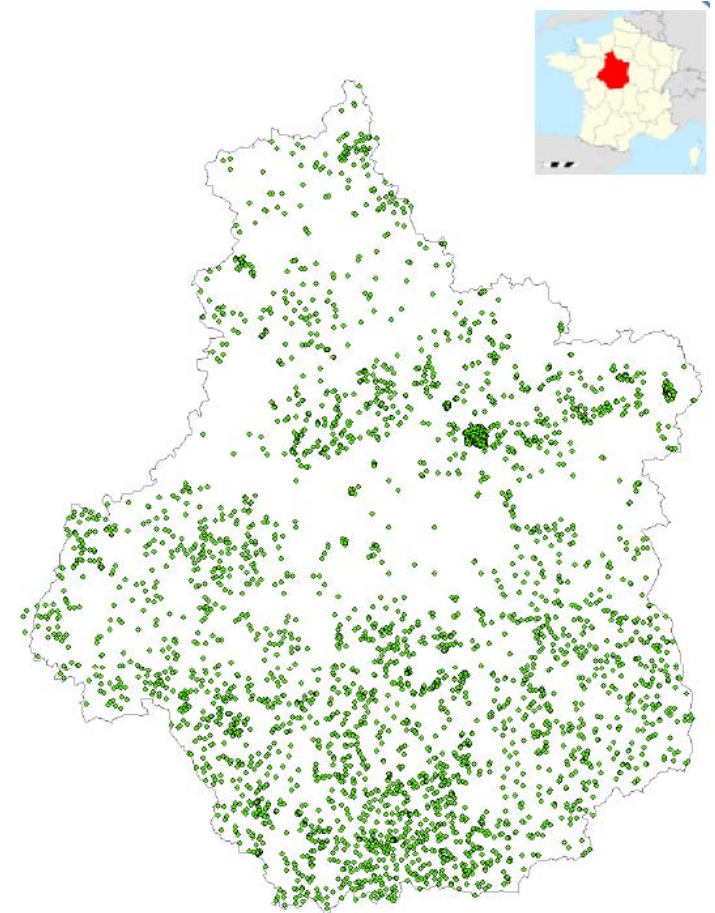
Villaneau et al., Environ Chem Lett (2013) 11:99-104



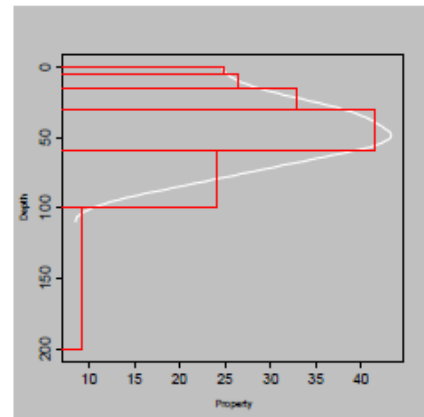
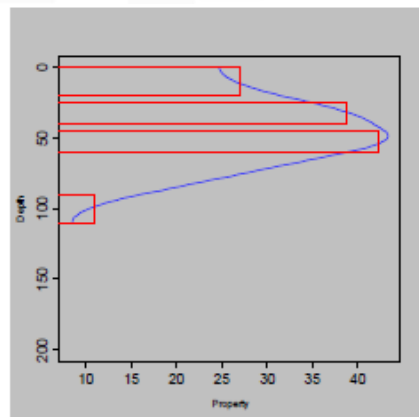
Dequiedt et al., Global Ecol. Biogeogr. (2011) 20, 641-652

Exemple 3 = Cartographie « 3D » de la granulométrie en Région Centre

- Analyses supervisée
- Données = IGCS profils
- Covariable SCORPAN =
$$S = f(S, C, O, R, P, A, N) + e$$
- f = couplage de régression multiple (BRT), d'un modèle géostatistique (LMCR) et d'une transformation permettant de travailler sur données compositionnelles
- Validation croisée, validation par le RMQS
- Carte d'incertitude

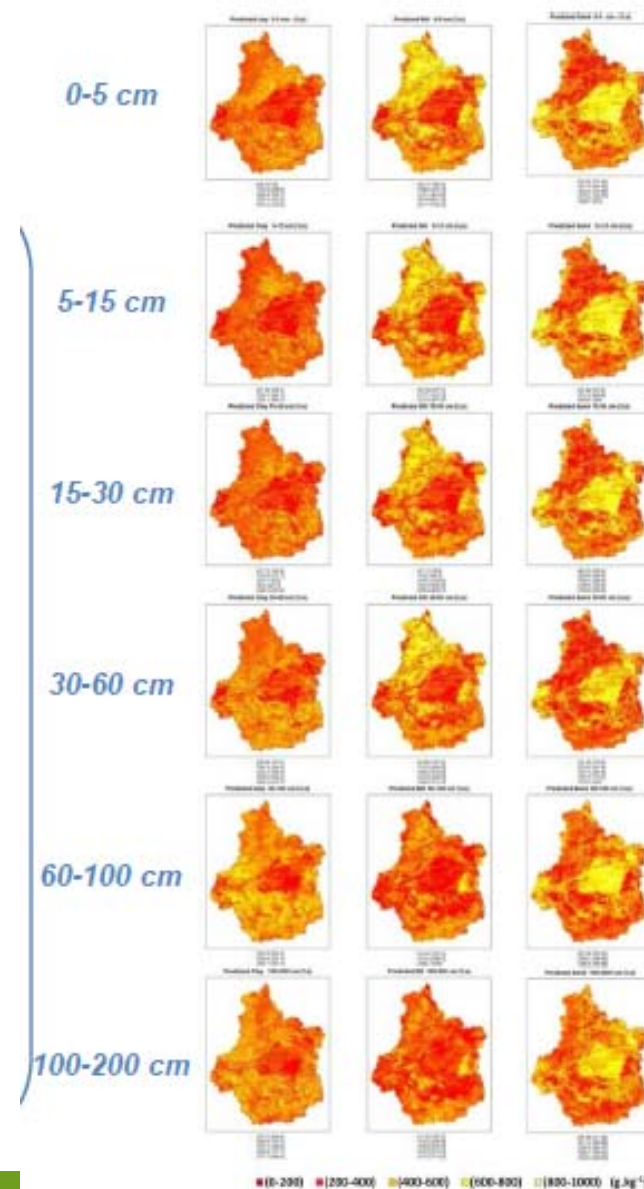


Cartographie « 3D » (2,5 D)

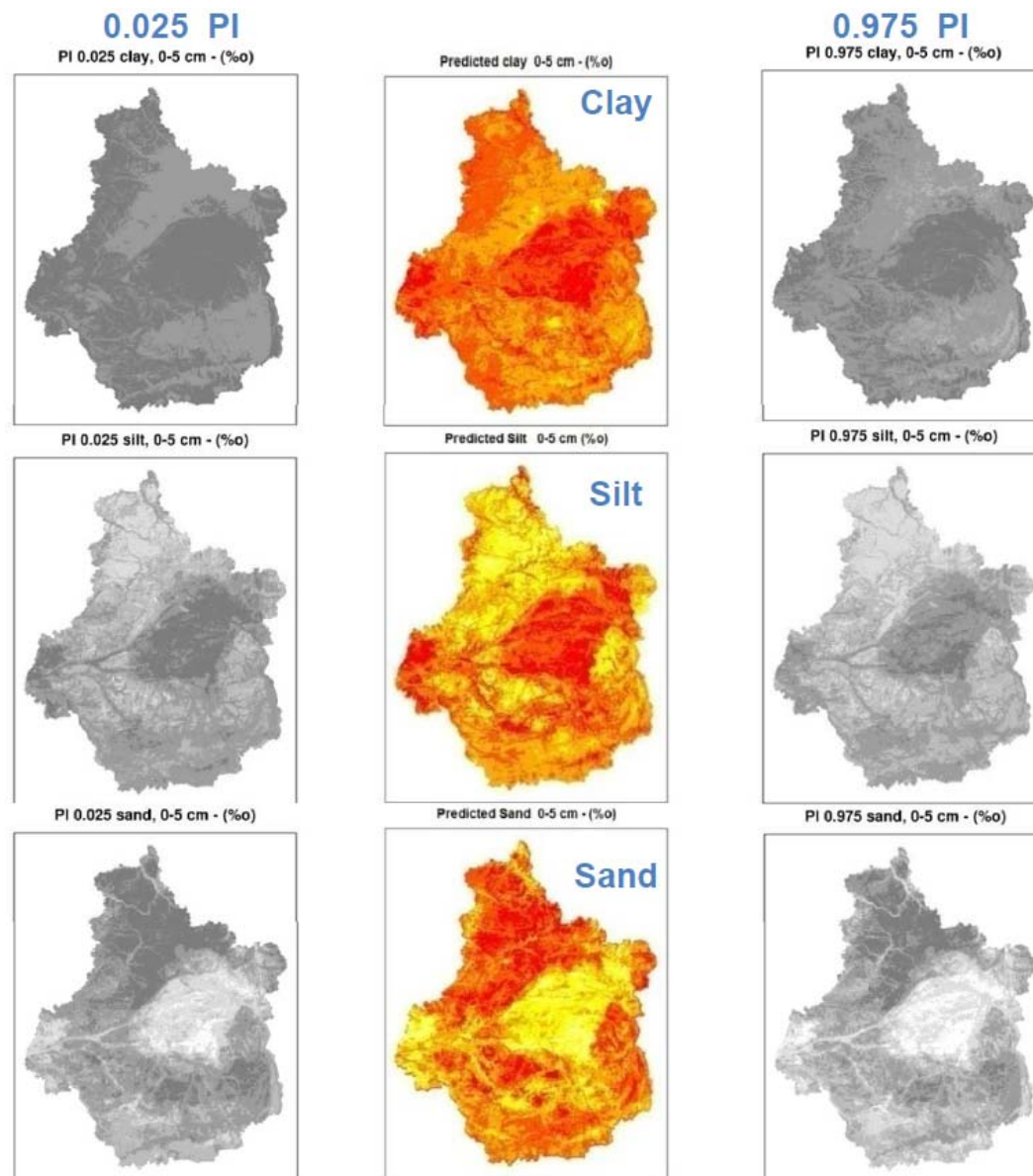
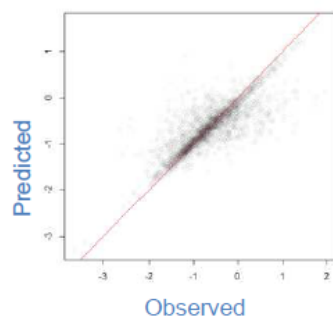
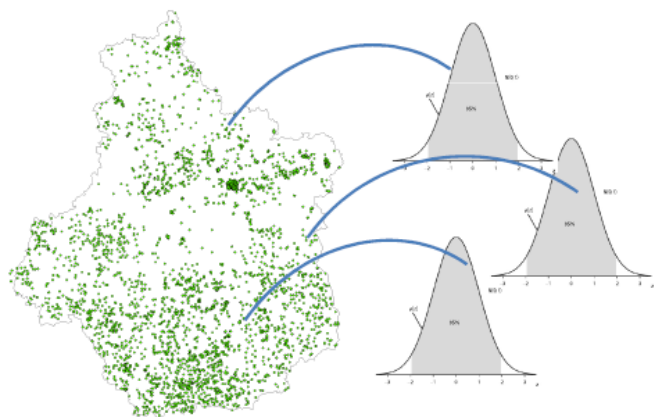


Spline type:
Equal-area
linear quadratic spline
(Bishop et al., 1999; Malone et al., 2009)

Modélisation des profils de granulométrie reconstitués sur 6 couches de référence



Cartes d'incertitude



■ (0-200) ■ (200-400) ■ (400-600) ■ (600-800) ■ (800-1000) (g.kg⁻¹)

Remarques

- réaliser (si possible) une carte des incertitudes associées
- Observations conjointes vs issues d'une carte
- attention aux biais liés
 - Aux covariables spatialement exhaustive
 - Aux variables sols et donc à la transformation
- incertitude peut être élevée pour production de carte à haute résolution
- spatialisation : sorties de modèles ou données d'entrée du modèle ?