



HAL
open science

Comparaison de trois méthodes d'estimation de la variance d'échantillonnage des estimations design-based de moyennes spatiales à partir d'un réseau systématique : application aux données du Réseau de Mesures de la Qualité des Sols

Nicolas N. Saby, D. J. Brus, Hakima Boukir, Dominique D. Arrouays

► **To cite this version:**

Nicolas N. Saby, D. J. Brus, Hakima Boukir, Dominique D. Arrouays. Comparaison de trois méthodes d'estimation de la variance d'échantillonnage des estimations design-based de moyennes spatiales à partir d'un réseau systématique : application aux données du Réseau de Mesures de la Qualité des Sols. 12. Journées d'Etude des Sols (JES), Jun 2014, Le Bourget du Lac, France. 350 p., 2014, 12e Journées d'Etude des Sols : Le sol en héritage. hal-02742923

HAL Id: hal-02742923

<https://hal.inrae.fr/hal-02742923>

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.

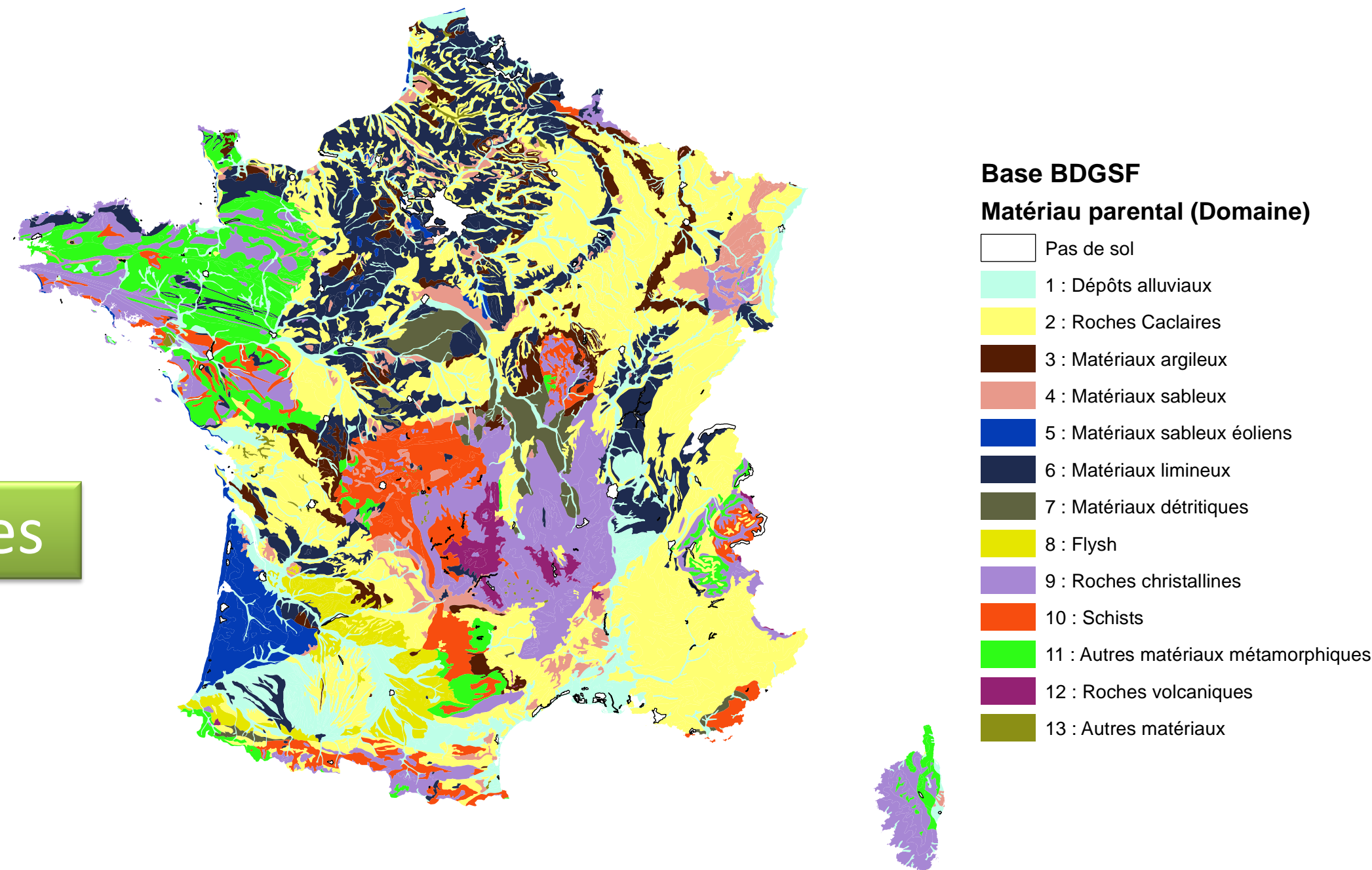
INTRODUCTION

Les 2200 sites du Réseau de Mesures de la Qualité des Sols (RMQS) ont été sélectionnés selon une grille systématique aléatoire (SY) de 16 km de côté. Cet échantillonnage SY permet d'avoir une bonne couverture spatiale. Les sites sont répartis de manière uniforme sur le territoire français conduisant à une bonne précision des estimations design-based (DB) des moyennes spatiales. Qui plus est, l'échantillonnage SY est très bien adapté à la cartographie comme le krigeage (méthode model-based). La méthode SY est donc un échantillonnage flexible.

L'estimateur DB de la moyenne spatiale est simple pour un SY mais, ce n'est pas le cas de la variance d'échantillonnage (ou erreur type) pour laquelle il n'existe pas d'estimateur non biaisé. Cette quantité est importante car elle renseigne sur le degré de fiabilité de la moyenne de l'échantillon par rapport à la moyenne de l'ensemble d'une population.

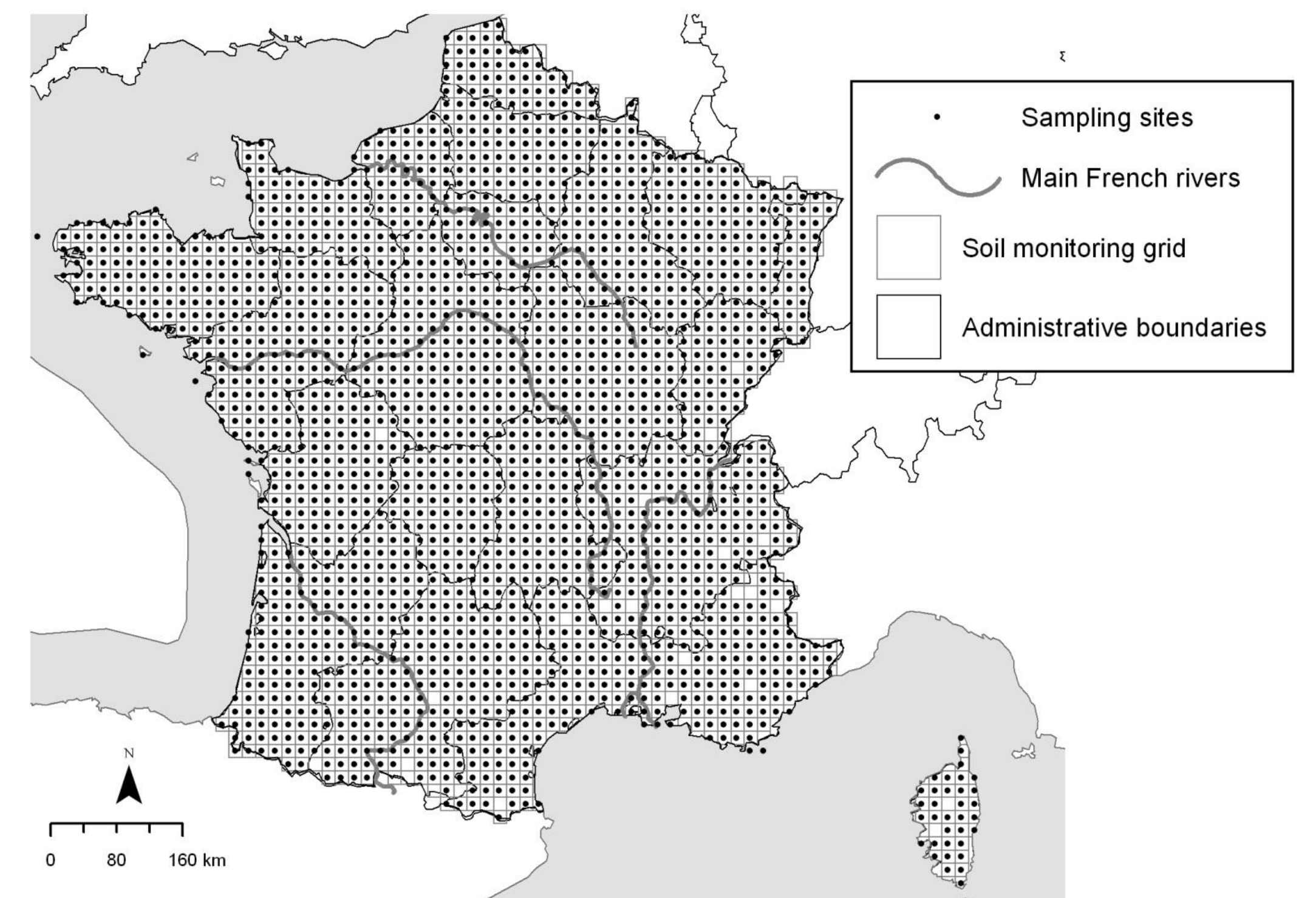
Ce travail présente et étudie trois approximations de la variance d'échantillonnage. Nous nous focalisons pour cela sur les teneurs en éléments traces métalliques (ETM = concentration total en Cd, Pb et Zn) mesurées dans le cadre du RMQS. L'objectif final est de fournir des statistiques descriptives des teneurs en ETM au sein de grand type de matériaux parentaux, appelé souvent des domaines.

MATÉRIEL



Carte des domaines

Zone d'étude et sites de surveillance



	ParMat 1	ParMat 2	ParMat 3	ParMat 4	ParMat 5	ParMat 6	ParMat 7	ParMat 8	ParMat 9	ParMat 10	ParMat 11	ParMat 12	ParMat 13
n	269	642	87	85	31	306	50	55	275	129	144	21	5
Cd	0,20	0,34	0,23	0,14	0,01	0,20	0,10	0,19	0,16	0,15	0,14	0,24	0,21
Pb	3,47	3,51	3,51	3,47	2,63	3,38	3,62	3,46	3,71	3,69	3,40	3,24	3,40
Zn	3,98	4,25	4,10	3,53	1,91	3,73	3,33	4,13	4,20	4,21	4,17	4,77	3,82

Tableau des estimations DB des moyennes spatiales par domaine. N est le nombre de sites,

MÉTHODE

Les 3 approximations retenues sont

SI = c'est le calcul classique (σ^2/n) qui nécessite de considérer la grille comme un échantillonnage aléatoire simple. Cela conduit généralement à une surestimation de la variance d'échantillonnage.

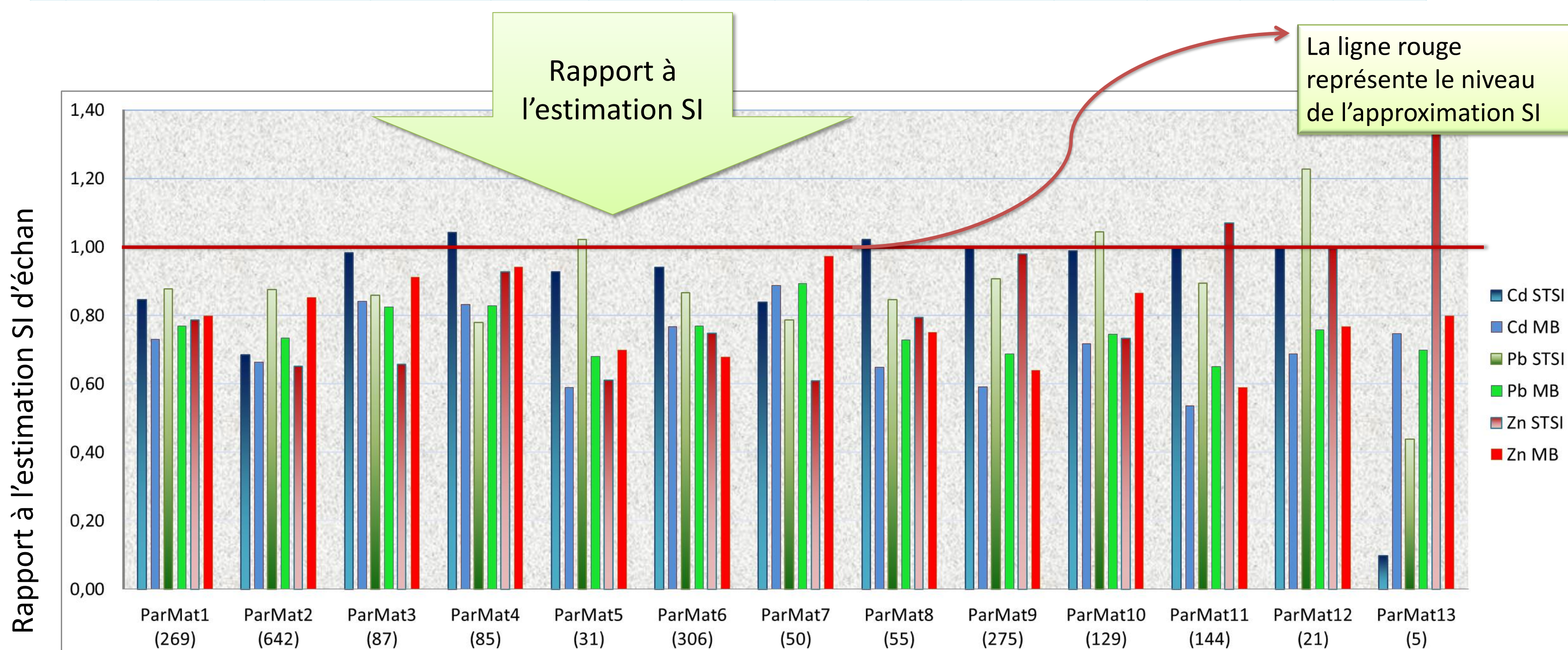
STSI = Dans ce cas, on considère l'échantillon SY comme un échantillonnage aléatoire stratifié. Pour cela, on regroupe par paire les sites sur la base de leurs coordonnées géographiques. On utilise la librairie R *spCosa*. Ces couples de sites sont alors considérés comme un échantillonnage aléatoire simple de taille 2 d'une strate non définie.

MB = Dans ce cas, la variance d'échantillonnage est prédite à partir d'un modèle stochastique de la variabilité spatiale. L'échantillonnage SY est ainsi utilisé pour calibrer un modèle (variogramme). Ce modèle est ensuite utilisé pour simuler R champs aléatoires de la variable d'intérêt. Chaque champ simulé est alors échantillonné S fois selon un SY produisant S estimations de la moyenne spatiale pour laquelle la variance d'échantillonnage peut être calculée. On obtient ainsi R variances que l'on moyenne pour produire une estimation MB de la variance d'échantillonnage. Cependant, en faisant l'hypothèse de géostationnarité de second ordre, cette procédure peut être simplifiée en évitant les simulations géostatistiques. On peut calculer la variance d'échantillonnage à partir de la moyenne des semivariances au sein de la grille et celle au sein du domaine.

RÉSULTATS

Tableau des 3 estimations de l'erreur type ($mg.kg^{-1}$) ($et =$ racine carré de la variance échantillonnage)

	ParMat1	ParMat2	ParMat3	ParMat4	ParMat5	ParMat6	ParMat7	ParMat8	ParMat9	ParMat10	ParMat11	ParMat12	ParMat13	
Cd	SI et	1,73E-04	2,88E-04	2,44E-03	9,48E-04	1,29E-04	1,54E-04	7,25E-04	1,61E-03	7,21E-04	9,24E-04	2,65E-04	5,42E-04	1,69E-03
	SY et	1,46E-04	1,97E-04	2,40E-03	9,89E-04	1,20E-04	1,45E-04	6,08E-04	1,64E-03	7,20E-04	9,13E-04	2,65E-04	5,40E-04	1,65E-04
	MB et	1,27E-04	1,91E-04	2,05E-03	7,88E-04	7,62E-05	1,18E-04	6,43E-04	1,04E-03	4,25E-04	6,61E-04	1,42E-04	3,73E-04	1,27E-03
Pb	SI et	1,044	0,706	4,230	8,988	31,096	0,603	5,054	3,310	8,673	2,719	2,184	3,833	24,396
	SY et	0,915	0,618	3,629	7,009	31,758	0,522	3,973	2,802	7,863	2,839	1,952	4,701	10,674
	MB et	0,802	0,518	3,485	7,447	21,163	0,463	4,510	2,410	5,953	2,022	1,420	2,905	17,042
Zn	SI et	5,585	7,440	33,458	92,614	3,777	2,991	51,111	31,410	24,668	17,986	29,601	48,186	50,723
	SY et	4,387	4,843	21,977	85,874	2,310	2,238	31,158	24,927	24,148	13,187	31,682	47,992	67,575
	MB et	4,454	6,345	30,474	87,078	2,639	2,026	49,714	23,544	15,746	15,551	17,448	36,913	40,516



Comment estimer le variogramme pour chaque domaine ?

L'approximation MB nécessite une estimation du variogramme pour chaque domaine afin de mettre en œuvre les simulations. Pour cela, les étapes suivantes sont nécessaires

1. Standardisation des mesures en retirant la moyenne du domaine et en divisant par l'écart type du domaine de telle sorte que les mesures ont une moyenne nulle et d'un écart-type de 1.
2. Estimer un seul variogramme V pour tous les domaines sur les mesures standardisées
3. Calculer le variogramme d'un domaine D en multipliant l'effet de pépite et le plateau du variogramme V par la variance du domaine D.

DISCUSSION & CONCLUSIONS

- ✓ Les approximations STSI ET MB sont généralement moins élevées que l'approximation SI et sont donc à privilégier.
- ✓ C'est cependant moins évident lorsque le nombre de sites est faible.
- ✓ L'approximation STSI est plus simple à calculer et moins gourmande en ressource de calcul.
- ✓ L'approximation MB est bien moins sensible au nombre de sites
- ✓ Ce travail permet de tirer des conclusions générales sur l'application des approximations de la variance d'échantillonnage dans le cas d'un réseau systématique nationale. Une étude basée sur des simulations pourrait préciser ces résultats.