



HAL
open science

Homosoil, une méthode pour l'extrapolation d'informations quantitatives du sol sur des zones inconnues

Christophe Rabillon

► **To cite this version:**

Christophe Rabillon. Homosoil, une méthode pour l'extrapolation d'informations quantitatives du sol sur des zones inconnues. [Stage] Autres régions du monde. Université d'Orléans (UO), FRA. 2014, 108 p. hal-02795942

HAL Id: hal-02795942

<https://hal.inrae.fr/hal-02795942v1>

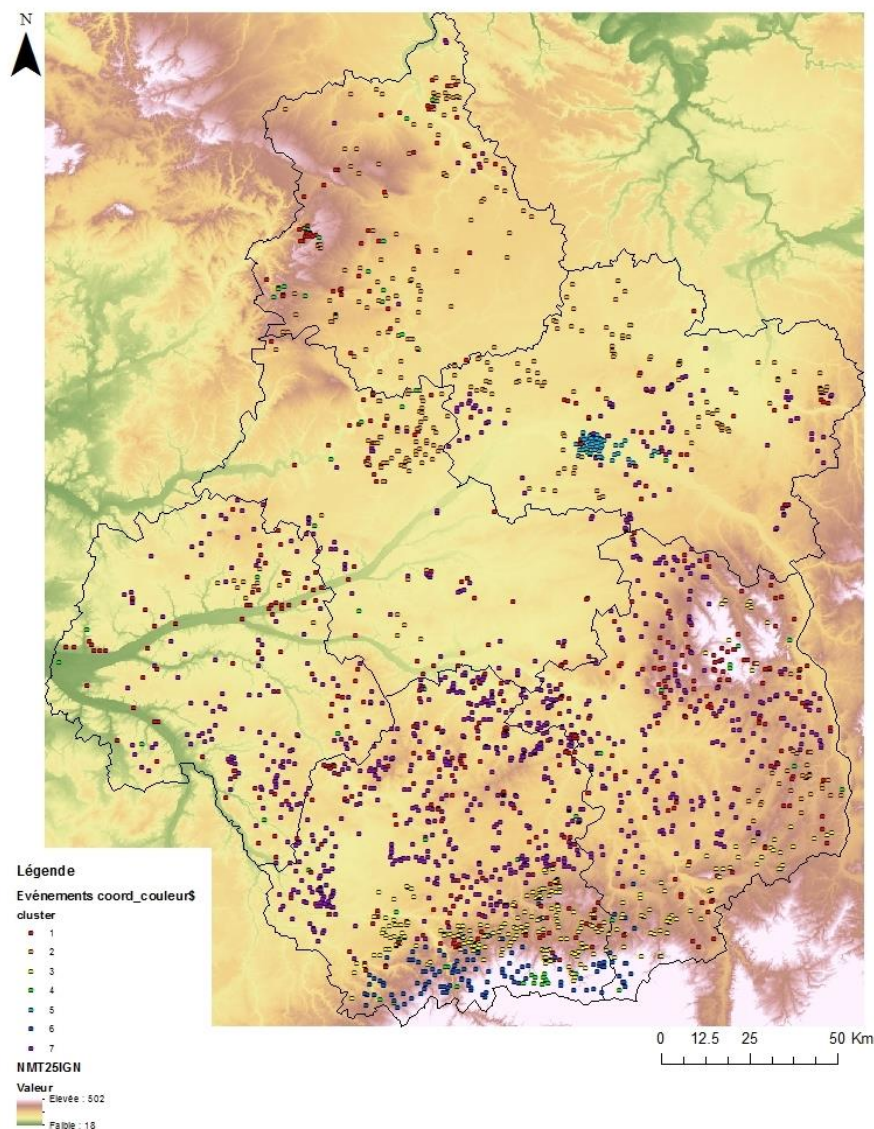
Submitted on 5 Jun 2020

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.

Rapport de stage 2014

Master 1 Statistiques et Probabilités Appliquées



Homosoil, une méthode pour l'extrapolation d'informations quantitatives du sol sur des zones inconnues

Christophe Rabillon

Liste des abréviations :

INRA : Institut National de la Recherche Agronomique

EPST : Etablissement Public à Caractère Scientifique

EPCS : Etablissement Public de Coopération Scientifique

CIRAD : Centre de Coopération Internationale en Recherche Agronomique pour le Développement

UMR : Unités Mixtes de Recherches

CNRS : Centre National de la Recherche Scientifique

US : Unité de Service

ONF : Office National des Forêts

AGPF : Amélioration Génétique et Physiologie Forestière

ZF : Zoologie Forestière

GBFor : Génétique et Biomasse Forestière

IGN : Institut Géographique National

IRD : Institut de Recherche pour le Développement

ADEME : Agence de l'Environnement et de la Maitrise de l'Energie

IUSS : Union Internationale des Sciences du Sol

MNT : Modèle Numérique de Terrain

SRTM : Shuttle Radar Topography Mission

ASTER : Advanced Spaceborne Thermal Emission & Reflexion

Remerciements

Je tiens à remercier l'ensemble de l'équipe de l'Unité de Service InfoSol de l'INRA d'Orléans pour son accueil et son aide qui a contribué au bon déroulement de ce stage.

Je remercie plus particulièrement Marion BARDY, directrice de l'US InfoSol pour m'avoir accueilli au sein de sa structure.

Je remercie Dominique ARROUAYS et Nicolas SABY, mes tuteurs de stage, d'avoir pris le temps de répondre à mes questions, ainsi que pour m'avoir guidé progressivement tout au long de ma mission.

Je remercie également Manuel MARTIN, Sébastien LEHMANN et Anne RICHER DE FORGES pour l'extraction des données et pour leur disponibilité tout au long de ce projet.

Je remercie aussi Titia MULDER pour ses explications à propos du clustering, ainsi que pour les codes R qu'elle m'a fournis.

Enfin, je remercie Didier CHAUVEAU, professeur encadrant du stage ainsi que toute l'équipe pédagogique du master « Statistiques et Probabilités Appliquées » du Collegium Sciences et Techniques d'Orléans pour leurs enseignements m'ayant permis d'effectuer ce stage.

Contenu

Introduction	6
1 Contexte et Objectifs	7
1.1 Contexte du stage	7
1.1.1 Présentation générale de l'INRA	7
1.1.2 Centre INRA Val de Loire, site d'Orléans	8
1.1.3 L'unité de service InfoSol.....	9
1.2 Contexte de la demande	10
1.3 Hypothèses	11
1.4 Objectifs	11
2 Matériel et méthodes	13
2.1 Données	13
2.2 Méthodes	15
2.2.1 Filtrage	15
2.2.2 ACP	20
2.2.3 Clustering	23
2.3 Analyse de la performance des modèles	26
3 Résultats	27
3.1 Filtrage	27
3.2 ACP	32
3.3 Clustering	34
4 Discussion	38
Conclusion.....	40
Bibliographie.....	41
Annexes	42

Introduction

Etudiant en première année de Master, Statistiques et Probabilités Appliquées à l'université d'Orléans, j'ai eu l'opportunité d'effectuer un stage de quatre mois au sein de l'Institut National de Recherche Agronomique d'Orléans, dans l'Unité de Service InfoSol.

Dans de nombreux endroits de la planète, les informations du sol sont difficiles à obtenir, voire inexistantes. Quand aucune carte détaillée ou observation du sol n'est disponible, une des solutions proposées (Homosoil, Mallavan et al. , 2010) est d'extrapoler à partir de données provenant d'autres régions du monde.

La méthode Homosoil part du principe que les sols et leurs propriétés- ne sont pas répartis au hasard dans l'espace, mais dépendent des grands facteurs de formation du sol (lithologie, topographie, végétation et occupation des sols, etc.). Elle pose l'hypothèse que l'homologie des facteurs de formation du sol permet un rattachement entre une zone de référence connue et une zone d'intérêt, constituant la « cible » inconnue. Dans ce cadre, les facteurs de formation des sols sont représentés par des « proxys » qui sont des variables disponibles en tout point de l'espace géographique et que l'on utilise pour effectuer ce rattachement.

L'objectif de mon stage est de tester la validité de cette méthode en réalisant une étude de cas sur la région Centre.

Dans ce rapport, je vais tout d'abord présenter le contexte de la demande ainsi que les objectifs, puis je présenterai le matériel et les méthodes utilisées durant ce stage. Ensuite j'exposerai les résultats obtenus. Je terminerai enfin par une discussion et interprétation des résultats.

1 Contexte et Objectifs

1.1 Contexte du stage

1.1.1 Présentation générale de l'INRA

Etablissement public à caractère scientifique et technologique (EPST), l'INRA est placé sous la double tutelle du ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt et du ministère de l'Enseignement Supérieur et de la Recherche. Premier institut de recherche agronomique en Europe, deuxième dans le monde, l'INRA mène des recherches finalisées pour une alimentation saine et de qualité, - pour une agriculture compétitive et durable, et pour un environnement préservé et valorisé.

Créé en 1946, dans un contexte de reconstruction du pays et de modernisation de l'agriculture française, ses objectifs ont évolué et ont désormais une dimension mondiale. En effet, les recherches de l'INRA sont motivées par l'évolution permanente des questionnements scientifiques ainsi que les défis planétaires posés à l'agronomie et l'agriculture tels que le changement climatique, l'épuisement des ressources fossiles et la nutrition humaine.

L'INRA compte 8 478 agents titulaires dont plus de 1800 chercheurs, 2500 ingénieurs, 4000 techniciens et administratifs, mais également près de 2000 doctorants. Environ 2000 stagiaires sont accueillis chaque année.

L'INRA fait partie d'Agreenium, qui est un Etablissement Public de Coopération Scientifique (EPCS) qui comprend aussi le Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) et les écoles nationales d'agronomie. Il entretient de nombreuses collaborations et échanges avec la communauté scientifique internationale avec de nombreux pays.

13 départements scientifiques et 18 centres régionaux composent l'INRA. On y trouve 213 Unités de Recherche dont 112 Unités Mixtes de Recherche (UMR) qui associent donc l'INRA à d'autres instituts de recherche (CNRS, CIRAD...) et d'enseignement (Universités et grandes écoles). L'INRA comprend aussi 49 unités expérimentales qui représentent une surface totale d'environ 10 000 ha ainsi que 94000 animaux en élevage. En 2013, l'INRA possède un budget prévisionnel de 881,61 millions d'Euros. Cet institut détient donc un dispositif unique de recherche dans les sciences du vivant.

1.1.2 Centre INRA Val de Loire, site d'Orléans

Le centre INRA Val de Loire compte 960 agents dont 660 titulaires, ce qui en fait le sixième centre de l'institut par son effectif. Il mène des recherches autour de quatre pôles répartis sur trois sites : Orléans, Tours et Bourges.

- Le pôle **Dynamique des sols et gestion de l'environnement** comprend 2 unités du département Environnement et Agronomie. Les enjeux de ce pôle concernent la production agricole et la protection des sols et de l'environnement.
- Le pôle **Biologie intégrative des arbres et organismes associés** comprend 3 unités du département écologie forestière, Prairies et Milieux Aquatiques et deux unités sous contrat avec l'Université d'Orléans et l'Office National des Forêts. Les recherches de ce pôle concernent l'étude des populations d'insectes vivant au sein de ces écosystèmes, le processus responsable de la formation du bois ainsi que la tolérance des arbres aux maladies et au stress hydrique.
- Le pôle **Biologie intégrative animale et gestion durable des productions animales** comprend 5 unités du département Physiologie Animale et Système d'Élevage et de génétique animale. L'objectif de ce pôle consiste à améliorer la durabilité des systèmes d'élevage dans une dimension économique, sociale et environnementale.
- Le pôle **Santé animale et santé publique** comprend une unité et une plateforme d'infectiologie expérimentale des départements de Santé Animale et de Microbiologie de la Chaîne Alimentaire.

Les deux premiers pôles sont sur le site d'Orléans.

C'est en 1972 que le site d'Ardon fut choisi pour y implanter un centre de l'INRA. Ce centre était, au tout début composé d'un simple bungalow ; il est composé aujourd'hui de 3 domaines de recherches et de 5 unités différentes. Ces domaines sont les suivants :

- **La sélection des arbres forestiers** dont les différents critères sont une croissance optimale, une certaine qualité du bois et une bonne résistance aux parasites. Dans ce domaine, diverses connaissances sont développées telles que la génétique, la physiologie des arbres, les biotechnologies et perfectionnement des techniques de diffusion des variétés améliorées.
- **La biologie des insectes forestiers ravageurs** ainsi que leur épidémiologie et leur relation avec les arbres-hôtes.

- **La cartographie et la surveillance des sols et des recherches finalisées portant par exemple sur la maîtrise des érosions et pollutions, l'évaluation des risques agro climatiques** en se basant sur la modélisation des comportements hydriques et physiques des sols, en relation avec leur histoire, leur gestion et leurs propriétés.

Ces différentes thématiques donnent au centre d'Orléans une orientation en matière d'environnement et de développement durable. Il a par ailleurs déployé une politique d'accueil et de partenariat avec différents organismes (ONF, ArboCentre) qui favorise le lien entre la recherche, le développement et l'aide à la décision publique.

Le centre est composé de 5 unités :

- 3 unités de recherche : Amélioration, Génétique et Physiologie Forestières (AGPF), Zoologie Forestière (ZF) et Science du Sol
- 1 unité de service : InfoSol
- 1 unité expérimentale : Génétique et Biomasse Forestière (GBFor)

Le centre d'Orléans compte environ 200 agents titulaires dont 38% sont des chercheurs et ingénieurs, 19% des techniciens et 13% des administratifs. Environ 50 non titulaires sont accueillis chaque année, beaucoup sont des étudiants Doctorants ou Masters mais aussi des post-doctorants et chercheurs étrangers. Le budget du centre en 2005, hors salaires, s'élevait à 5,9 millions d'Euros dont les origines sont diverses (Ministère, collectivités territoriales, Europe, autres organismes de recherche, partenaires,...).

1.1.3 L'unité de service InfoSol

Dirigée par Marion Bardy, l'Unité de Service (US) InfoSol fait partie du Département Environnement et Agronomie. La mission principale de cette unité est de constituer et de gérer un système d'information à vocation nationale sur les sols, par rapport à leur distribution spatiale, leurs propriétés et l'évolution de leurs qualités.

L'US InfoSol travaille dans le cadre du groupement d'intérêt scientifique Sol (GIS Sol), en collaboration avec le ministère de l'agriculture et de la pêche, le ministère de l'environnement et du développement durable, l'Institut Géographique National (IGN), l'Institut de Recherche pour le Développement (IRD), et l'Agence de l'Environnement et de la Maitrise de l'Energie (ADEME).

Le GIS SOL propose un ensemble de programmes nationaux pour faciliter et encourager une gestion patrimoniale et durable des sols.

Les principales activités de l'unité sont d'une part, la réalisation et la coordination de l'acquisition des données nécessaires à la constitution d'un système d'information sur les sols de France ainsi que le contrôle de la qualité de ces données et d'autre part, la création et l'alimentation des bases de données permettant l'archivage et l'exploitation des informations. Enfin, la dernière activité principale est la diffusion et la valorisation des connaissances sur les sols et des outils d'exploitation thématique produits par les recherches de l'institut.

1.2 Contexte de la demande

Les grands enjeux planétaires actuels liés au sol concernent la sécurité alimentaire, le changement climatique, l'accaparement des terres, l'urbanisation, la gestion de l'eau et la protection de la biodiversité. Afin de répondre à ces enjeux, il est indispensable de connaître et de protéger les sols. Il est donc primordial de mettre en place des outils permettant de prendre en compte les propriétés des sols à l'échelle mondiale.

C'est ainsi qu'en 2006 a été lancé le projet GlobalSoilMap, à l'initiative du groupe de travail « Digital Soil Mapping » de l'Union Internationale des Sciences du Sol (IUSS).

L'objectif est donc d'obtenir une base de données librement accessible de quelques propriétés des sols d'intérêt majeur sur la totalité du monde. Cependant, dans de nombreux endroits de la planète, ces informations sont difficiles à obtenir ou inexistantes, soit parce qu'aucune étude pédologique ne les a concernées soit par ce que ces études ont été perdues.

Ce problème est un verrou important pour le déploiement de cette base de données. Une nouvelle approche, basée sur l'existence de co-variables spatialement exhaustives, pourrait être une réponse à ce problème. Elle consiste à rattacher les sites dépourvus de données pédologiques proches à des sites qui, tout en pouvant être très éloignés, partageraient des caractéristiques très semblables, si bien que leur pédogénèse –et les propriétés de sols qui en résultent- seraient elles aussi semblables.

Cette approche a été évoquée par Minasny et McBratney (2010) sous le nom d'« homosol » mais n'a jamais fait l'objet d'implémentation numérique ni de validation à ce jour.

1.3 Hypothèses

Notre hypothèse est que le raisonnement analogique, souvent sollicité dans les études pédologiques de terrain, pourrait pallier ce manque de données, pour peu qu'il soit implémenté dans des procédures numériques s'appuyant sur les données spatiales de caractérisation des paysages

Le concept d'Homosoil est assez voisin du concept d'extrapolation à partir de zones connues. Il part du principe que si les covariables sont identiques entre le point connu et le point inconnu, ce point inconnu doit posséder les mêmes propriétés.

Homosoil se distingue des techniques d'apprentissage déjà éprouvées par le fait qu'il :

- i) Vise à prédire des profils de propriétés et non des unités cartographiques ou typologiques de sol
- ii) Ne se restreint pas à extrapoler une zone connue à une zone limitrophe, mais peut aller chercher dans l'ensemble du « connu » (la région, la France, le Monde) le point *a priori* le plus semblable

1.4 Objectifs

L'enjeu est ici de tester une innovation méthodologique, dont les concepts ont été fondés par la communauté scientifique mondiale, mais qui n'a jamais été éprouvée.

L'objectif général du projet est de tester une nouvelle méthode de prédiction spatiale des propriétés des sols, adaptée à l'absence de données pédologiques sur un secteur donné. Nous utiliserons pour ce faire les bases de données DoneSol de la Région Centre.

On attend de ces résultats une première validation du potentiel de la méthode homosoil pour générer des propriétés de sol dans des zones entièrement vierges.

La méthode Homosoil est représentée par l'algorithme présenté par la figure 1

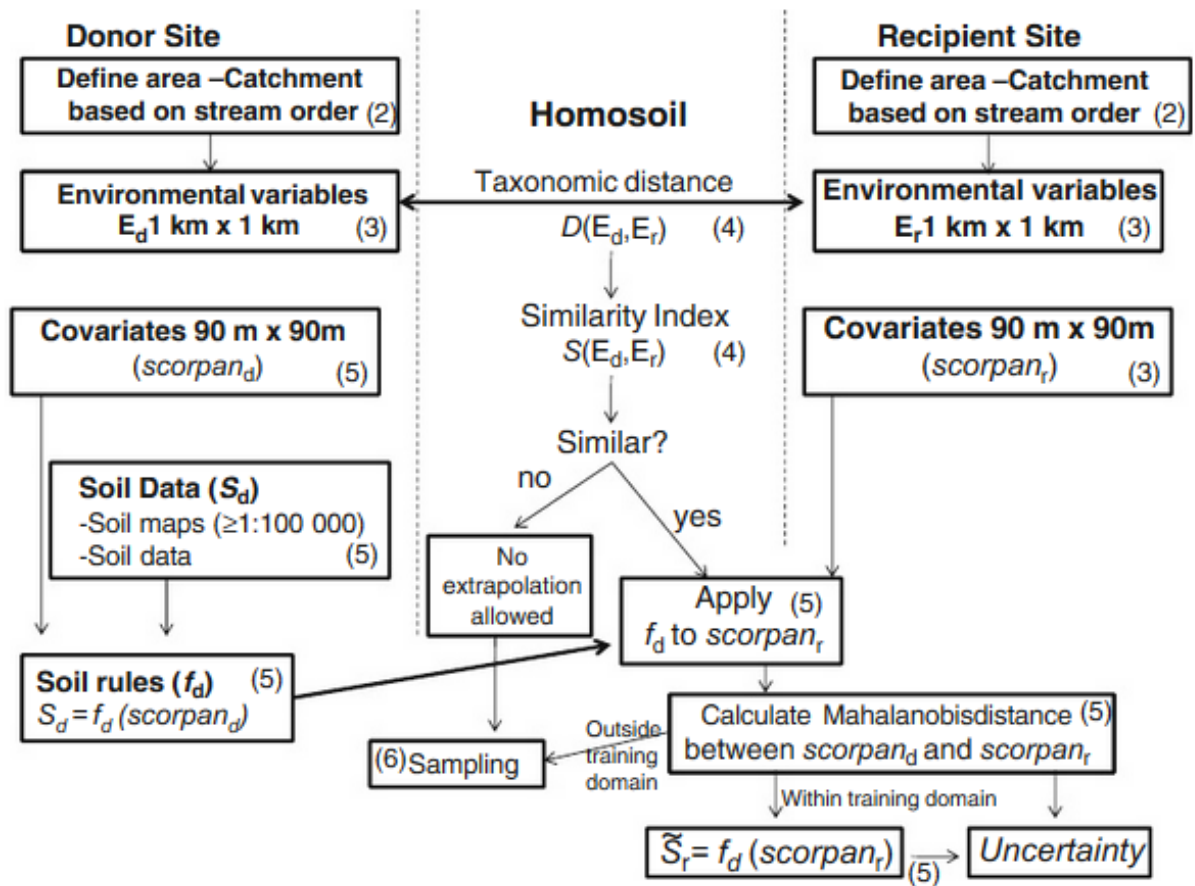


Figure 1 : Algorithme proposé d'après Minasny et McBratney (2010)

2 Matériel et méthodes

2.1 Données

Les données se composent de deux ensembles de variables décrivant des propriétés au droit de sites de mesures géolocalisés. Elles sont regroupées au sein d'un tableau de 2503 observations de 111 variables dont la liste complète est donnée en annexes. Premièrement, un ensemble de variables pédologiques ont été extraites de la base de données DoneSol Ces données représentent des informations sur les sols et leur composition. Ces variables représentent les données que l'on veut prédire. Deuxièmement, un ensemble de covariables a été extrait par croisement géographique avec des bases de données spatialisées et exhaustives : le type de géologie, la topographie (différents indices issus de modèles numériques d'altitudes nommés MNT SRTM ou ASTER), l'occupation des sols et les données d'émission en rayonnement gamma obtenus par une prospection aéroportée sur la région Centre.

-On utilisera la variable `id_profil` comme identifiant pour nos sites de mesures, que nous appellerons pixels.

-les variables `x` et `y` représentent la position géographique des pixels. Les coordonnées sont fournies dans la projection Lambert 93 (un mode de projection conique).

-l'ensemble des variables dont le nom commence par `srtm` et `ast` représentent la topographie :

- `slp`: slope, c'est la pente locale ou dérivée première de l'altitude
- `asp`: aspect direction : l'angle du pixel avec le nord (l'orientation)
- `prf`: profile curvature : dérivée seconde de l'altitude dans le plan
- `pln`: plan curvature : dérivée seconde de l'altitude dans le plan
- `lng`: longitudinal curvature : dérivée seconde de l'altitude dans le plan
- `crs`: cross-sectionnal curvature : moyenne

-les variables `uranium`, `ti`, `th`, `potassium` et `total` sont les données d'émission en rayonnement gamma

-les variables `crl00`, `vig00`, `olea00`, `bois00`, `jach00` représentent l'occupation humaine du territoire

-les variables dont le nom commence par `argile`, `limon` ou `sable` représentent la composition granulométrique des sols. Ces données ont été obtenues par la méthode dite "equal area quadratic splines" qui permet de générer une fonction continue des granulométries en fonction de la profondeur à partir de données discrètes obtenues sur des horizons (ou des couches) prélevés à des profondeurs variables (Ciampalini et al., 2014).

Ces fonctions "spline" permettent ainsi de générer des valeurs pour les épaisseurs standards recommandées dans le cadre de GlobalSoilMap.

-la variable geo_50k représente les informations de la carte géologique au 50 millièm.

Avant tout, on affiche les diagrammes de dispersion (scatterplots) pour les différentes variables pédologiques et les covariables qui doivent les expliquer (variables topologiques, d'occupation et gamma). Ces scatterplots permettent de repérer d'éventuelles relations entre les variables mais on s'intéressera principalement aux graphiques qui concernent les covariables argile, limon et sable.

Sur la diagonale on trouve le nom de chaque covariable ainsi que son histogramme.

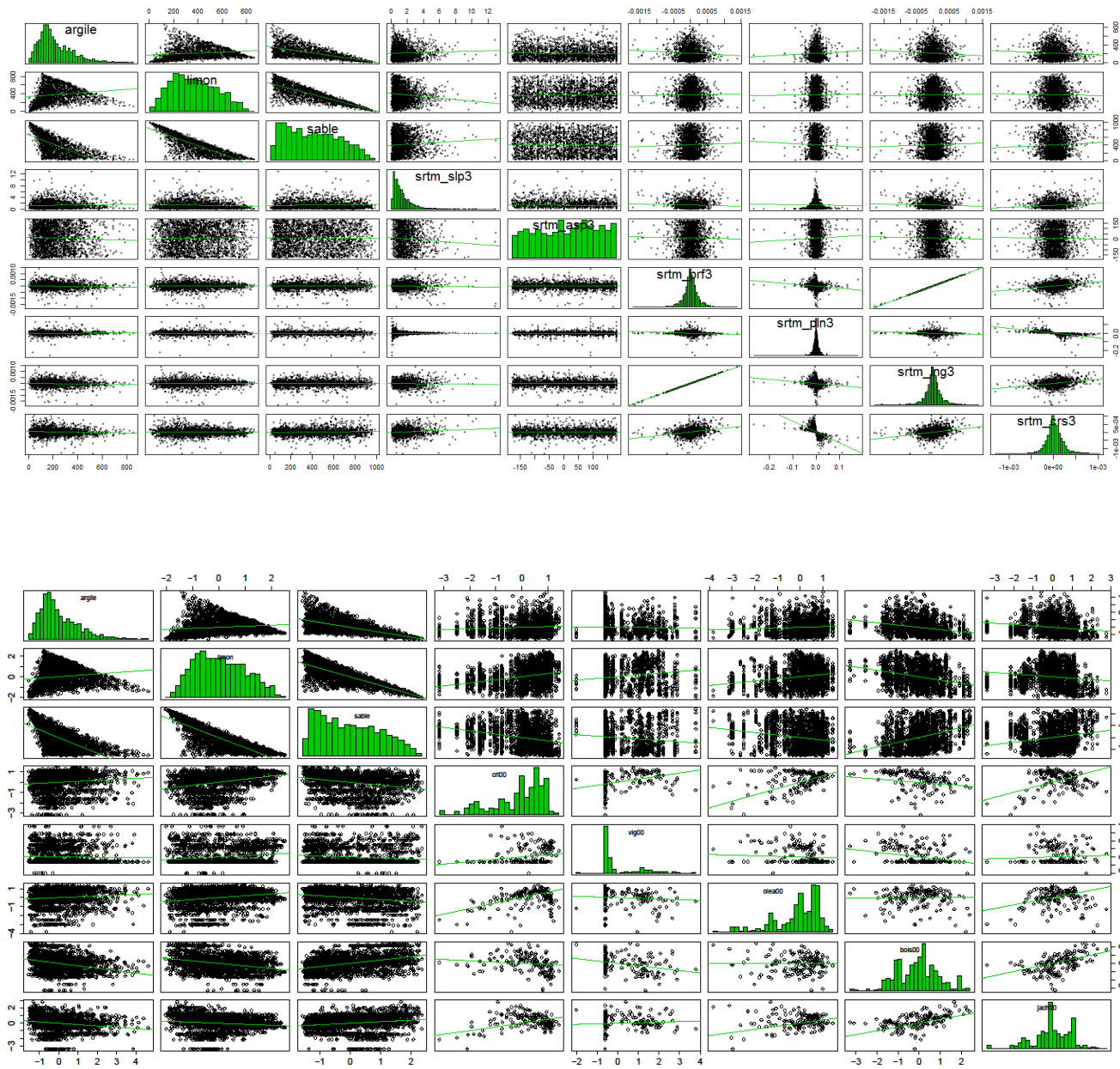


Figure 2 : Scatterplot des variables externes

On constate qu'une partie de ces variables ne suit pas une distribution gaussienne, on va donc procéder à un log-normalisation des variables dont le code est fourni en annexes. Nous disposerons ainsi de deux jeux de données pour comparer l'efficacité de nos méthodes.

2.2 Méthodes

La première opération consiste à découper le jeu de données en deux parties selon la variable x qui représente l'abscisse dans le système de coordonnées Lambert 93. On obtient ainsi une partie Ouest et une partie Est de la région Centre que l'on va recoller pour obtenir une région Centre dont les parties Est et Ouest sont bien séparées. On choisit de découper dans le sens Est/Ouest car ceci nous permet d'obtenir 2 jeux de données de taille similaire et possédant une répartition des sols à priori plus uniforme que si nous avions fait un découpage Nord/Sud.

Pour l'étude, on va supposer inconnues les variables pédologiques de la partie Ouest et on va tenter de les approcher à partir de celles de la partie Est.

On implémente alors trois différentes méthodes pour déterminer les valeurs pédologiques de la partie Ouest : le filtrage, l'affectation à la suite d'une ACP et un clustering suivi de filtrage.

2.2.1 Filtrage

La méthode de filtrage s'appuie sur l'indice de dissimilarité de Gower S ; cet indice est compris entre 0 et 1, plus celui-ci est proche de 1, plus les entrées du jeu de données sont différentes. Inversement, plus cet indice est proche de 0, plus les observations correspondantes sont proches l'une de l'autre. C'est un indice multi varié qui permet de mélanger des variables qualitatives et quantitatives. Il se calcule de cette façon :

$$S_{ij} = \frac{1}{p} \sum_{k=1}^p \left(1 - \frac{|x_{ik} - x_{jk}|}{\text{range } k} \right)$$

Équation 1: Indice de dissimilarité de Gower

p : le nombre de variables que l'on décide d'observer.

i et j : les différents individus du data frame

x_{ik} : valeur de la variable k pour l'individu i

La fonction R associée au calcul de cet indice de dissimilarité est la fonction daisy. Celle-ci calcule l'ensemble des indices de dissimilarité entre les observations d'un data-frame. Pour interpréter le résultat de cette fonction, il est nécessaire de le placer sous forme de matrice.

La fonction de filtrage s'appelle `filtrage_similarites.R`

Notre fonction de filtrage prend en arguments un niveau de filtre et un pourcentage de rejet ; elle s'appuie aussi sur une autre fonction : `filtre.R` qui renvoie une liste de vecteurs. Chaque vecteur est composé des valeurs de `geo_50k` (la composition géologique) présentant une similarité supérieure ou égale au niveau de filtre choisi en paramètre.

Par exemple, si le niveau de filtre est égal à 0.5, le vecteur `a2` contiendra les valeurs `a2` et `a3`. Si le niveau de filtre est égal à 0.1, le vecteur `a2` contiendra les valeurs `a2`, `a3`, `a4`, `a5`, `a10` et `a11`.

La similarité géologique est extraite d'une carte géologique régionale qui a été simplifiée et les indices de similarité entre les différentes unités géologiques ont été établis par expertise des pédologues de l'Unité InfoSol, sur une échelle allant de 0 à 1. Cette matrice est fournie en annexes.

La fonction de filtrage commence par découper le jeu de données selon l'axe Est/Ouest

On calcule ensuite plusieurs matrices d'indices de dissimilarité selon différents groupes de variables : les variables `srtm`, les variables `ast`, les variables `gamma` et les variables d'occupation. On sélectionne la partie supérieure gauche de ces matrices. Ces sous-matrices correspondent aux indices de dissimilarité entre la partie Ouest (lignes) et la partie Est (colonnes).

Puis, on ordonne les lignes de ces matrices dans un ordre décroissant, afin de repérer, pour chaque pixel Ouest, quels sont les pixels Est les plus proches. On tronque ensuite ces matrices selon le pourcentage choisi en paramètre (par exemple on ne garde que 40% des colonnes).

On teste trois qualités de filtre, pour cela on ne va pas utiliser les mêmes parmi ces matrices :

-filtre 1 : Matrices de dissimilarité selon les variables `srtm`, `ast`, `gamma` et occupation ; on effectue ensuite un filtrage supplémentaire selon l'indice `geo_50k` (variable qui représente la géologie du sol).

-filtre 2 : Matrices de dissimilarité selon les variables `srtm`, `gamma` et occupation; on effectue ensuite un filtrage supplémentaire selon l'indice `geo_50k`

-filtre 3 : Matrice de dissimilarité selon les variables `srtm`, `gamma` et occupation

Ces filtres sont classés du plus au moins fin. Les figures suivantes nous montrent le nombre de pixels n'ayant récupéré aucun voisin à la fin du filtrage selon de niveau de filtre et le pourcentage choisi :

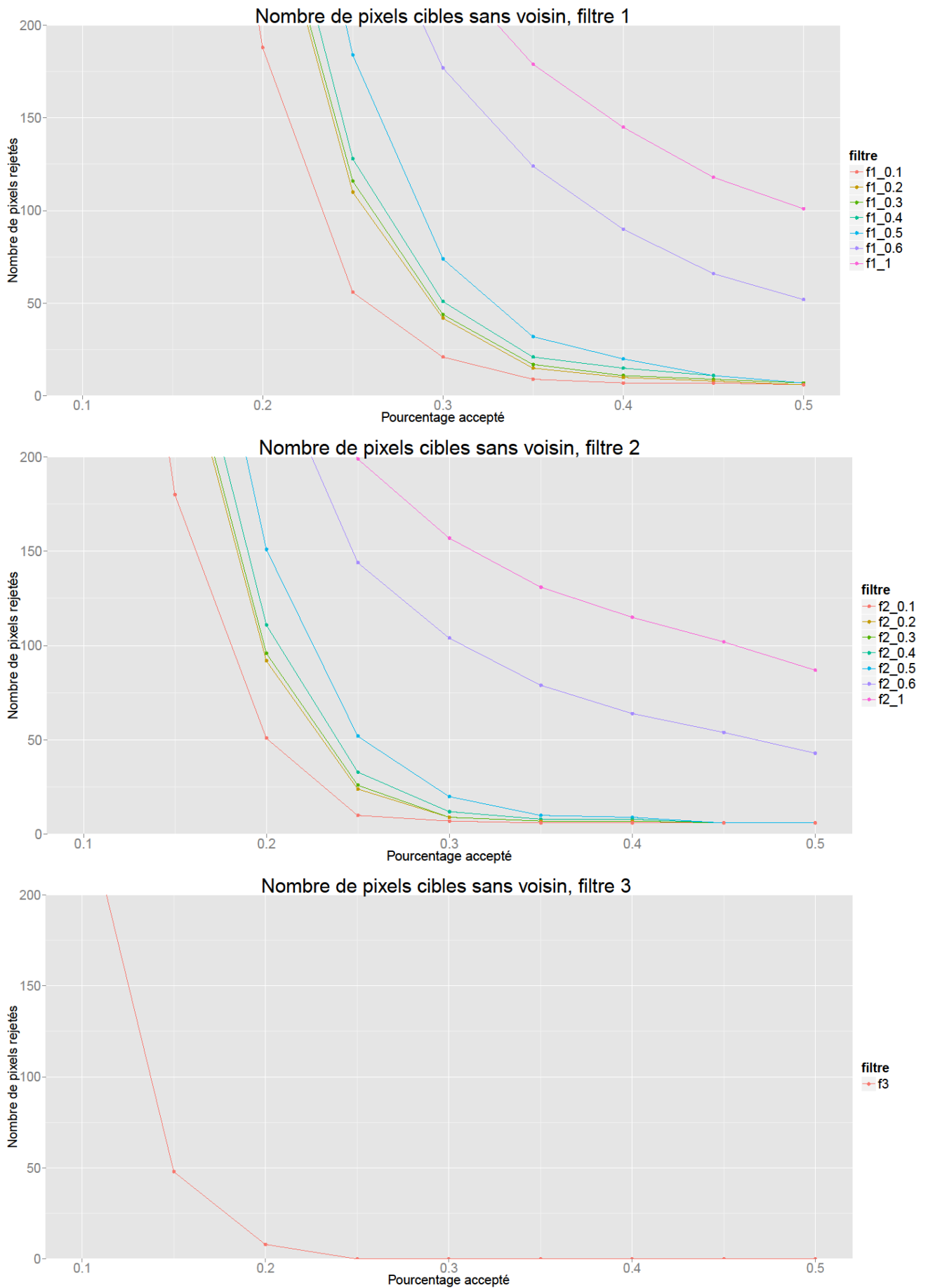


Figure 3 : Nombre de pixels sans voisins en fonction du filtre, du niveau de filtre et du pourcentage de points conservés

On choisit donc d'utiliser la qualité de filtre 1, avec le niveau de filtre 0.3 et le pourcentage 0.4.

On crée ensuite quatre matrices contenant en première colonne les id_profils de la partie Ouest et les id_profils des pixels de la partie Est correspondant.

On procède maintenant à l'intersection des lignes de ces matrices, ainsi, pour chaque pixel Ouest, un pixel Est qui en sera proche selon les quatre groupes de variables, sera conservé. Un pixel trop éloigné pour l'un de ces groupes sera éliminé.

Enfin, on applique un dernier filtre, provenant de notre fonction filtre.R : parmi les pixels renvoyés, ceux qui ont un profil géologique trop éloigné de celui du pixel cible sont supprimés.

On obtient donc une matrice que l'on appelle p dont les premières lignes sont comme ceci :

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
1	66	90957	899	90911	127734	90927	2202	74941	127527	90841	90787	90641	90722	2201	127546	90778	2280	NA	NA	NA
2	70	1497	2466	90857	1441	1367	1317	90496	1498	2147	1438	1255	1256	1382	993	1635	1230	1448	1426	NA

La première colonne de cette matrice est constitué de l'id_profil du pixel cible, pour chaque ligne on a donc les id_profil des pixels Est qui sont suffisamment proches du pixel cible pour tous les groupes de variables observés.

On constate que tous les pixels cibles n'ont pas le même nombre de pixels voisins, certains peuvent même ne pas en avoir, ceci est du à un pourcentage ou un niveau de filtre trop strict ; la quantité de pixels n'ayant aucun voisin par rapport au filtre choisi, au niveau de filtre et au pourcentage accepté est montrée dans les graphiques précédents.

A partir de cette matrice p, on crée 2 autres matrices :

- m une matrice composée des moyennes des indices de dissimilarité.
- ped une matrice composé des indices de dissimilarité selon les variables pédologiques.

Le pixel Est sélectionné comme plus proche voisin est celui présentant la moyenne des indices de dissimilarité la plus faible.

La stratégie du filtrage est résumée dans le schéma suivant :

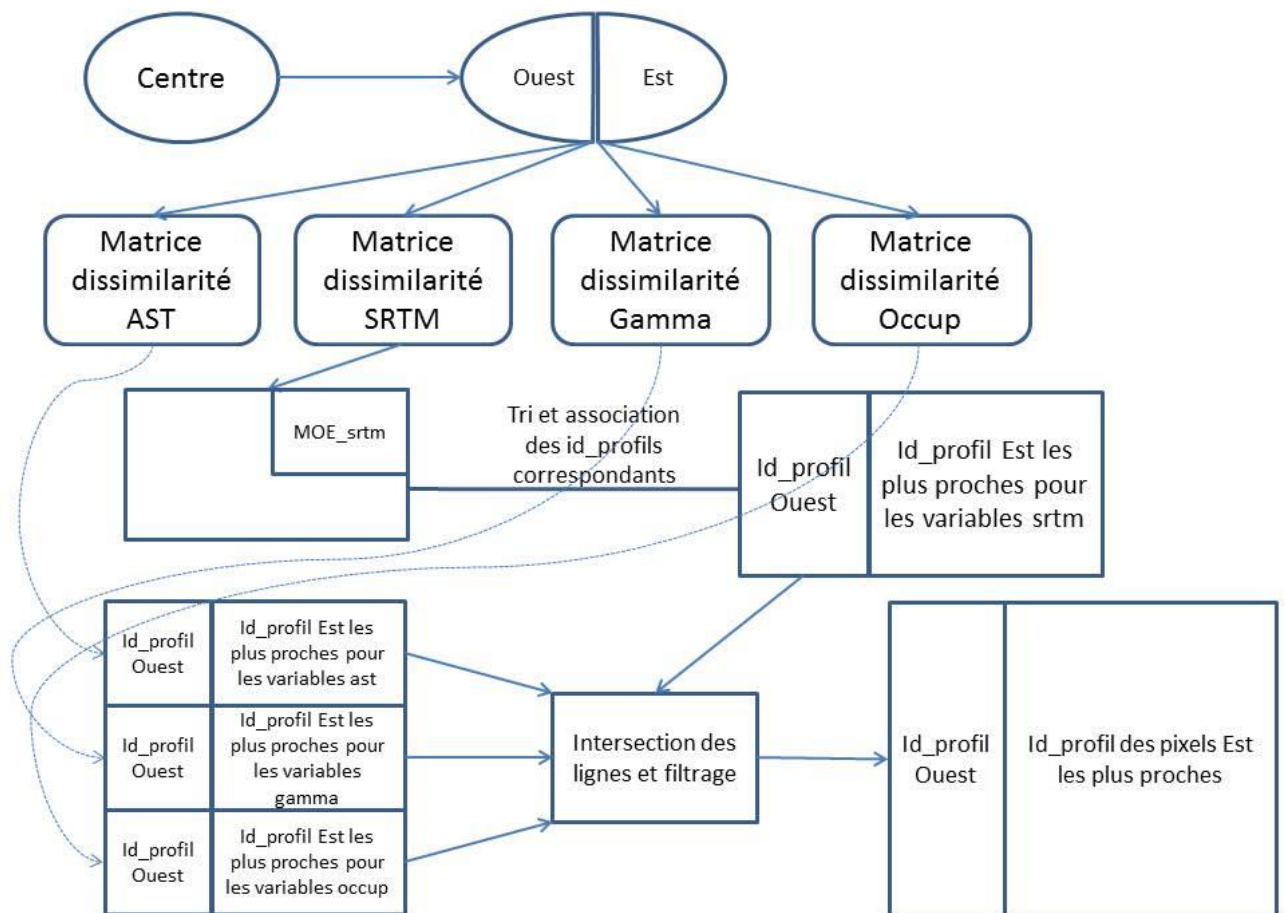


Figure 4 : Algorithme de la méthode de filtrage

2.2.2 ACP

Une deuxième méthode d'affectation du plus proche voisin passe par une Analyse en Composantes Principales sur l'ensemble des variables externes quantitatives.

La commande R pour effectuer une ACP est PCA

```
acp ← PCA(donnees[c(10:83,85:89)],ncp=n, graph=FALSE)
```

Les colonnes sélectionnées pour le calcul de l'ACP correspondent à l'ensemble des variables externes.

D'après l'éboulis des valeurs propres, on décide de conserver 20 composantes principales (n=20).

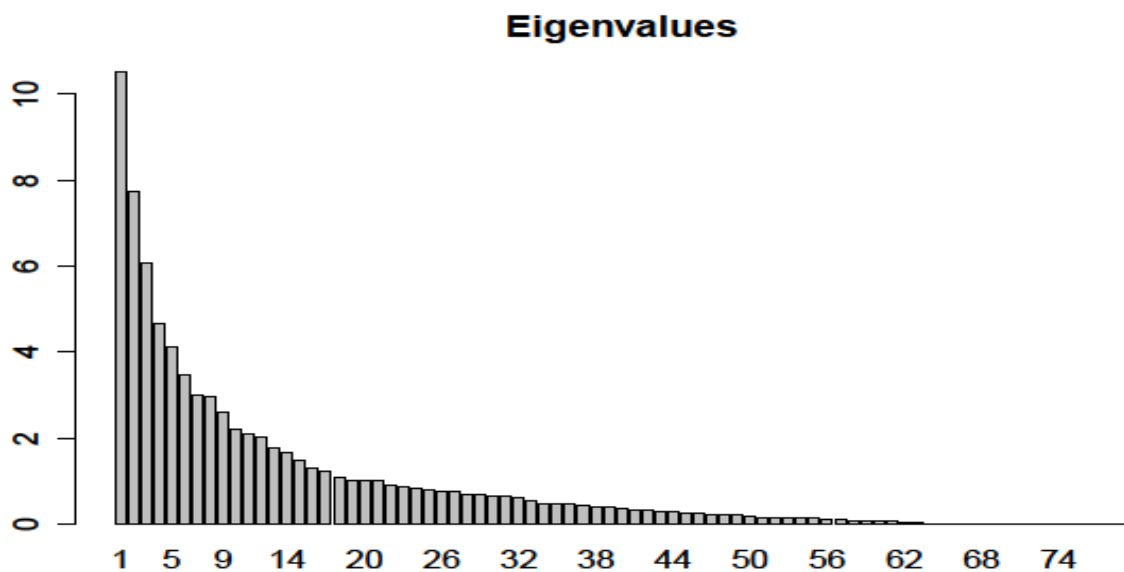


Figure 5 : Eboulis des valeurs propres de l'ACP

Les 5 premières composantes principales sont constituées uniquement de variables topographiques (la projection des variables sur les plans principaux est fournie en annexe) ; il faut attendre la sixième composante principale pour voir apparaître les émissions globales en rayonnement gamma.

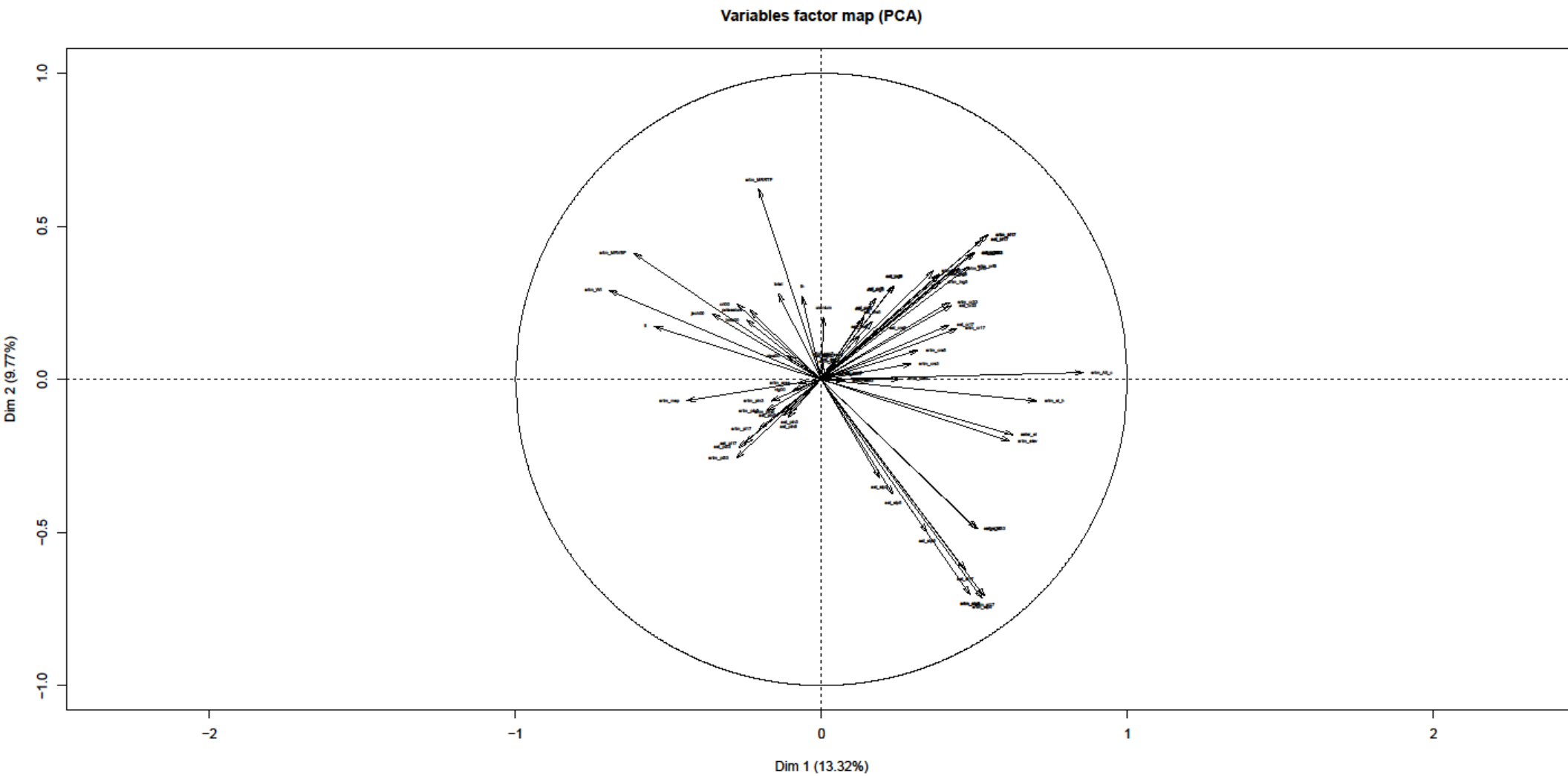


Figure 6 : Projection des variables externes sur les 2 premières composantes principales

L'ACP réalisant une projection orthogonale, il est possible de calculer la distance euclidienne entre les pixels : on associe chaque pixel Ouest au pixel Est qui lui est le plus proche. Pour cela on calcule une matrice des distances euclidiennes entre les pixels et on conserve la partie de cette matrice qui correspond à la distance entre les pixels Ouest et Est, - de la même façon que lors du filtrage, cette sous matrice correspond à la partie supérieure gauche de la matrice.

Pour chaque pixel Ouest, on sélectionne alors le pixel Est qui présente la distance minimale par rapport au pixel cible.

On récupère ainsi une matrice ($\text{nb_de_pixels_Ouest} * 3$) avec en première colonne : l'id_profil du pixel Ouest, en deuxième colonne : l'id_profil du plus proche pixel Est et en troisième colonne : la distance euclidienne entre ces deux pixels.

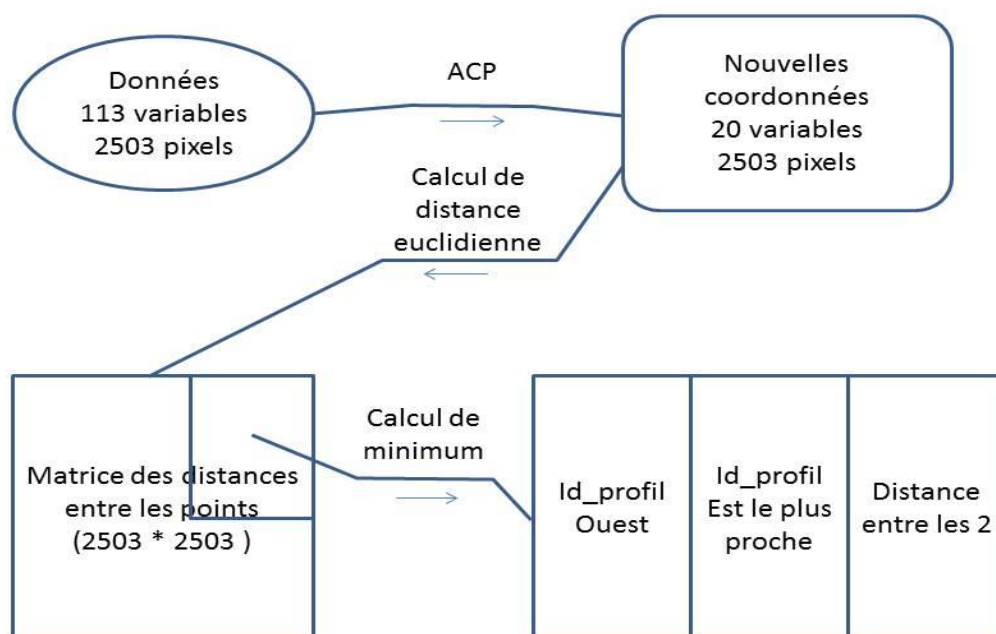


Figure 7 : Algorithme de l'affectation par ACP

2.2.3 Clustering

Le clustering est la dernière méthode d'affectation ; cette méthode consiste à diviser les données en différents groupes homogènes (les clusters) afin d'appliquer la méthode de filtrage au sein de ces groupes. Les données de chaque cluster partagent des caractéristiques communes.

Pour obtenir un bon partitionnement, on doit :

- minimiser l'écart entre les données d'un même cluster.
- maximiser l'écart entre les différents clusters.

Le clustering s'avère être plus efficace sur les données normalisées, ce sont donc celles-ci que l'on va utiliser.

Le code R pour procéder au clustering m'a été fourni par Titia Mulder ; la fonction procédant au clustering est `mixmodCluster`. On lui donne en paramètre les données, le nombre minimal et maximal de clusters, le type des données, une stratégie (définie par la fonction `mixmodStrategie`) et un ou plusieurs critères.

Le nombre de clusters a été obtenu en utilisant une version stochastique de l'algorithme Expectation Maximisation (SEM) (maximum de vraisemblance) de Dempster et al. (1977). Le nombre retenu est celui présentant une valeur du critère NEC (Normalized Entropy Criterion) minimale : ici 7 (séparer les données en 2 clusters ne propose pas une séparation suffisante).

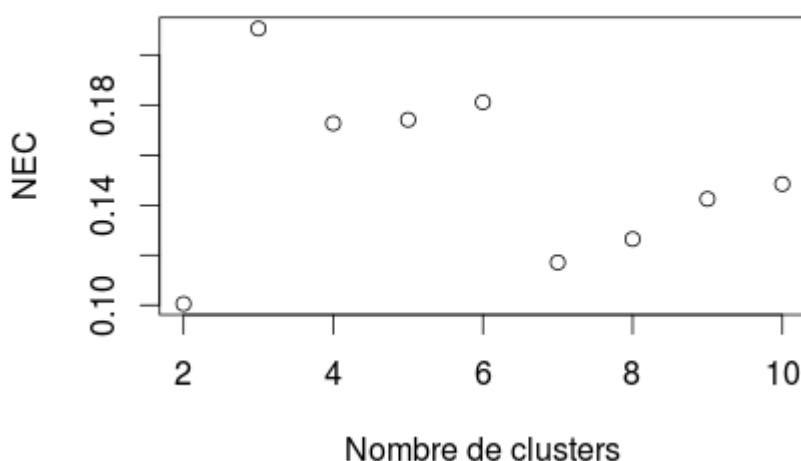


Figure 8 : Valeur du critère NEC en fonction du nombre de clusters

Le critère NEC mesure la capacité d'un modèle à fournir des groupes bien séparés. Un critère NEC supérieur à 1 indique une absence de structure de classification dans les données ; ce qui n'est pas le cas ici.

On remarque une assez bonne séparation géographique de la plupart des clusters qui sont répartis comme ceci :

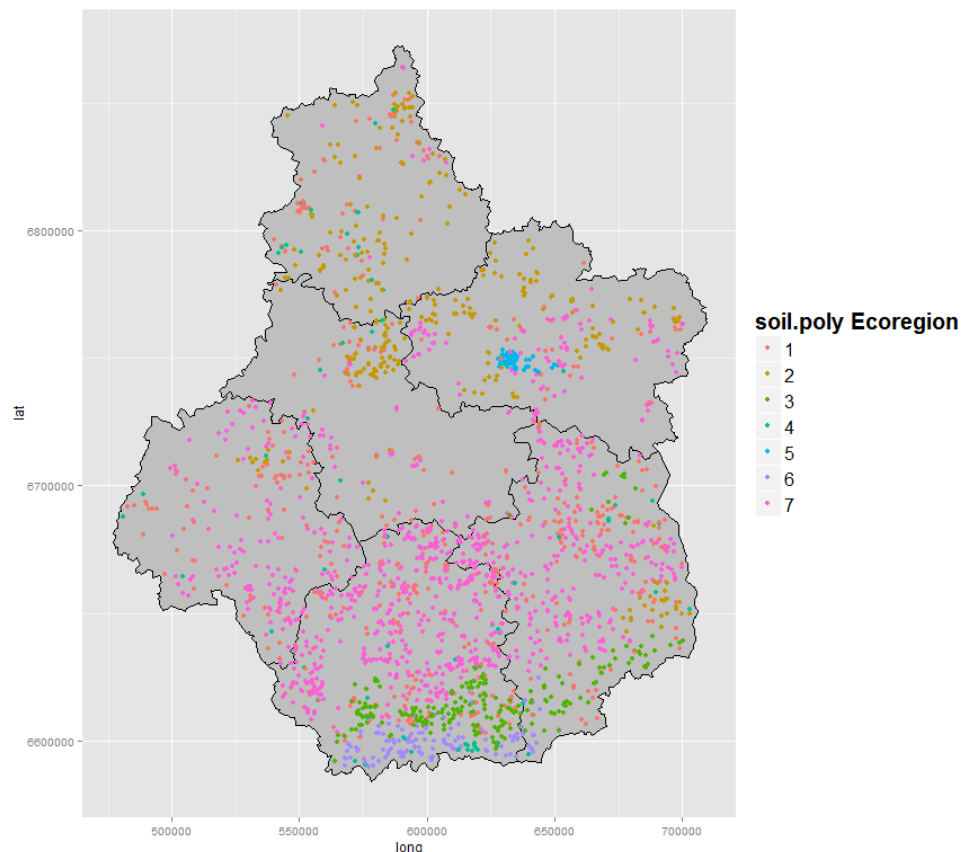


Figure 9 : Répartition des clusters au sein de la région

Le cluster 2 correspond à des zones de très faible pente et très faible courbure, avec une teneur en limon élevée ; il correspond à la Beauce.

Le cluster 3 correspond à des zones d'altitude moyennement élevées de faible courbure; il correspond au début du massif central.

Le cluster 4 présente les situations de pentes les plus fortes en bordure de vallons.

Le cluster 5 correspond au val de Loire avec l'altitude la plus basse et la teneur en sable la plus élevée.

Le cluster 6 présente une altitude plus élevée que le 3, il correspond à la partie de la région qui appartient déjà au massif central.

Les clusters 1 et 7 regroupent toutes les zones qui ne peuvent pas être classées dans les autres clusters ; ces deux clusters se différencient par l'orientation de leur pente.

Les histogrammes de toutes les variables, selon les clusters, sont fournis en annexes. On remarque que certaines variables pédologiques présentent des distributions bimodales au sein de certains clusters, ce qui signifie que notre clustering n'est probablement pas optimal pour ces variables.

Après avoir classifié les données, on applique à nouveau notre stratégie de filtrage au sein de chaque cluster. Ainsi, on approche chaque pixel Ouest par le pixel Est qui lui est le plus proche et qui appartient au même cluster. On obtient donc, pour chaque cluster, les matrices p , m et ped .

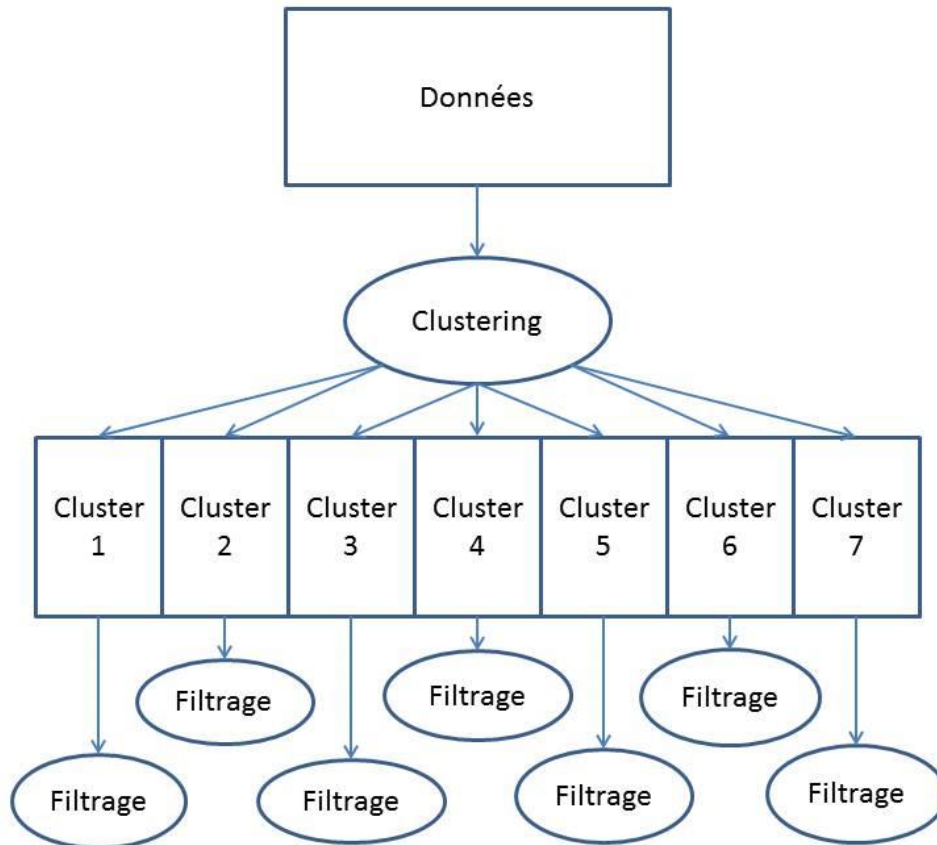


Figure 10 : Algorithme du clustering suivi du filtrage

2.3 Analyse de la performance des modèles

Afin de comparer les performances du filtrage, du clustering et de l'ACP, on utilise des indicateurs de qualité. Ces indicateurs se calculent en fonction de x , les valeurs observées (valeurs des variables pédologiques du pixel cible) et de y , les valeurs prédites (valeurs des variables pédologiques du plus proche voisin).

On utilise 4 indicateurs différents :

-le **Mean Prediction Error** : il mesure le biais et doit être proche de 0, il se calcule de la manière suivante :

$$MPE(x, y) = \text{mean}(x - y)$$

-le **Root Mean Square Error** : il permet de mesurer la précision du modèle ; plus il est faible, plus le modèle est de bonne qualité. Il se calcule de la manière suivante :

$$RMSE(x, y) = \sqrt{\text{mean}((x - y)^2)}$$

-le **Ratio of Performance to Deviation** : il s'agit de la performance moyenne de prédiction. Plus celui-ci est élevé, plus le modèle est de bonne qualité. Il se calcule de la manière suivante :

$$RPD(x, y) = \frac{sd(x)}{RMSE(x, y)}$$

-le **coefficient de détermination (R^2)**, il permet de juger de la qualité de la prédiction. Le R^2 est compris entre 0 et 1, plus il est proche de 1, plus la qualité du modèle est élevée. Il se calcule de la manière suivante :

$$R^2(x, y) = \text{cor}(x, y)^2$$

Le coefficient de détermination sera l'indicateur le plus utilisé ici.

3 Résultats

Après avoir implémenté les différentes méthodes, on peut procéder au traitement afin de déterminer si deux pixels proches auront une composition pédologique similaire.

3.1 Filtrage

Afin de vérifier la validité des résultats de la méthode de filtrage, on commence par afficher les valeurs de la matrice m par rapport à celles de la matrice ped :

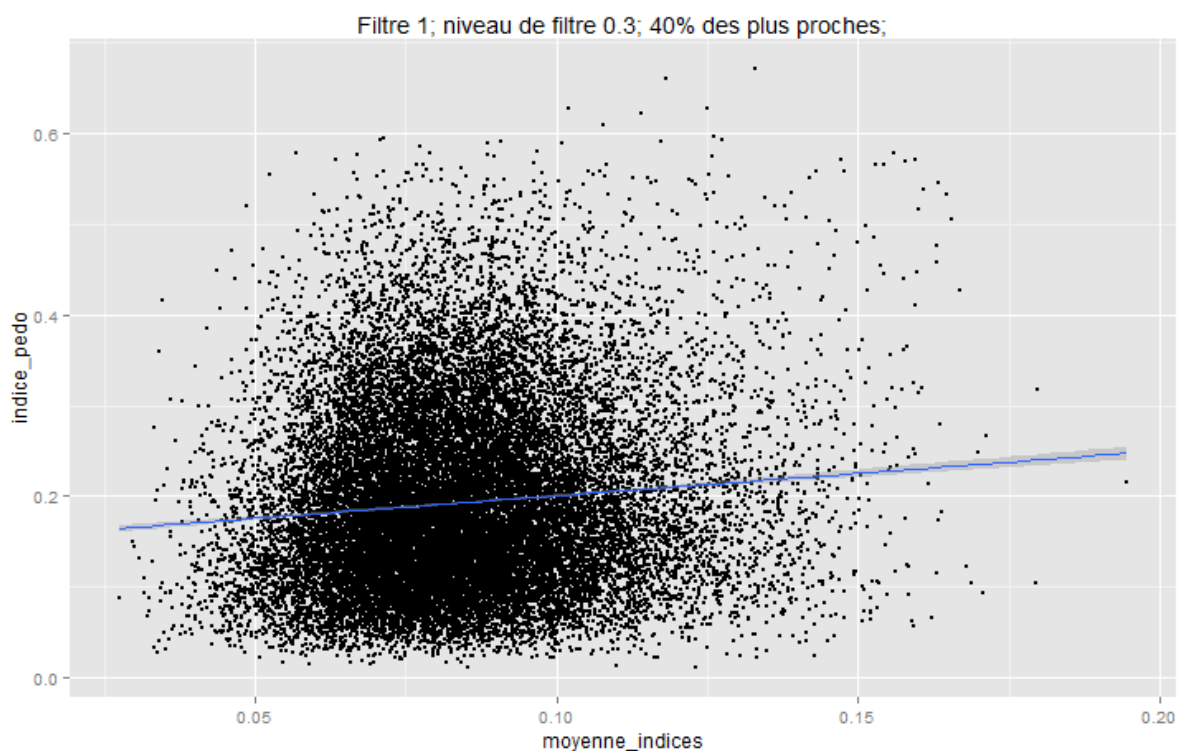


Figure 11 : Valeur de l'indice de dissimilarité pédologique en fonction de la moyenne des indices de dissimilarité des variables externes

La forme du nuage de points ne permet pas de déterminer s'il existe une relation entre la moyenne des indices srtm, ast, gamma et occup et l'indice de dissimilarité pédologique.

On décide alors de colorier le nuage de points en fonction de leur concentration afin de voir s'il s'en trouve une majorité autour de la droite de régression. Ceci se fait avec l'aide de la fonction R `densCols` :

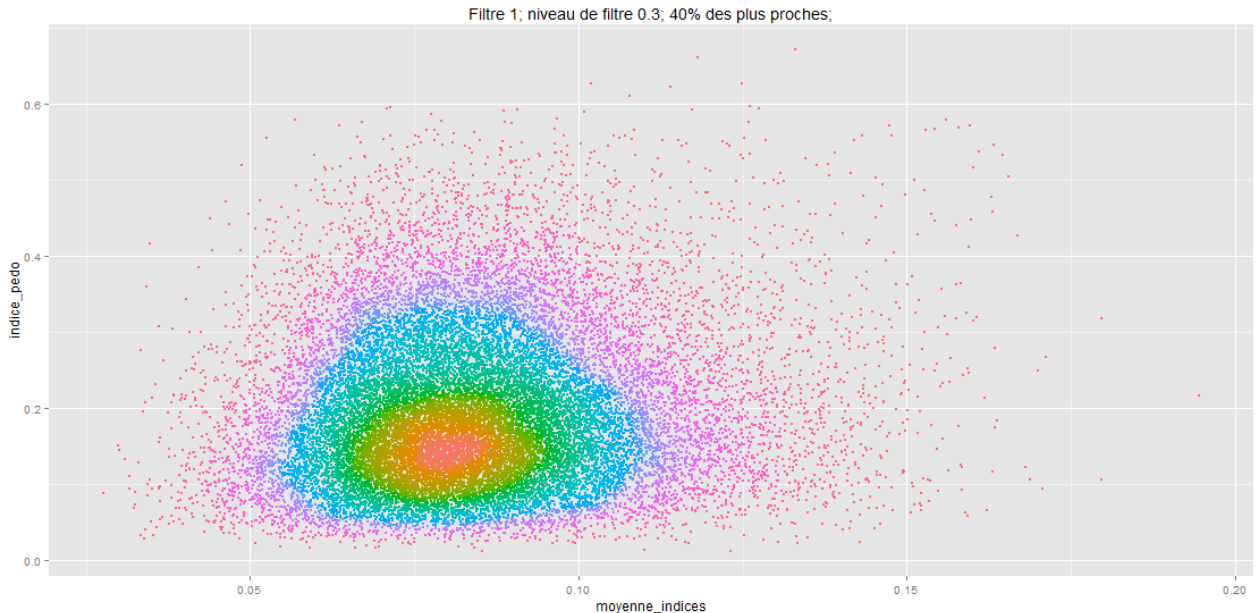


Figure 12 : Concentration de la figure 11

On constate donc que ce n'est pas le cas ; il est alors nécessaire de vérifier si notre méthode de filtrage renvoie bien un pixel effectivement proche au niveau pédologique.

Parmi la liste des pixels renvoyés par le filtrage, on regarde alors s'il s'en trouve qui soient effectivement proches selon les variables pédologiques. Pour cela, on va sélectionner, pour chaque pixel Ouest, les 10% de pixels Est qui ont un indice de Gower associé aux variables pédologiques le plus faible. On va ensuite vérifier si les valeurs correspondent à celles de la matrice p renvoyée.

On obtient une matrice comme ceci :

	V1	V2	V3	V4	V5	V6	V7
1	66	74941	NA	NA	NA	NA	NA
2	70	1317	NA	NA	NA	NA	NA
3	264	90615	NA	NA	NA	NA	NA
4	265	1367	985	1330	90490	NA	NA
5	549	NA	NA	NA	NA	NA	NA
6	703	91092	90970	90571	1231	90673	NA

Figure 13 : Intersection des 10% plus proche pixels avec le résultat du filtrage

Près de 90% des pixels Ouest ont, parmi leurs voisins, un voisin qui est effectivement proche. Cependant, pour déterminer duquel il s'agit, il est nécessaire de connaître les valeurs des variables pédologiques, ce qui ne sera pas le cas dans les zones pour lesquelles la méthode homosol est conçue : on sait qu'on a trouvé le bon pixel mais on ne sait pas encore comment déterminer duquel il s'agit. On choisit donc d'affecter à chaque pixel Ouest, les valeurs des variables pédologiques du pixel Est renvoyé ayant la moyenne des indices de dissimilarité srtm, ast, gamma et occup la plus faible. Ensuite on compare les valeurs effectives à celles qui ont été attribuées.

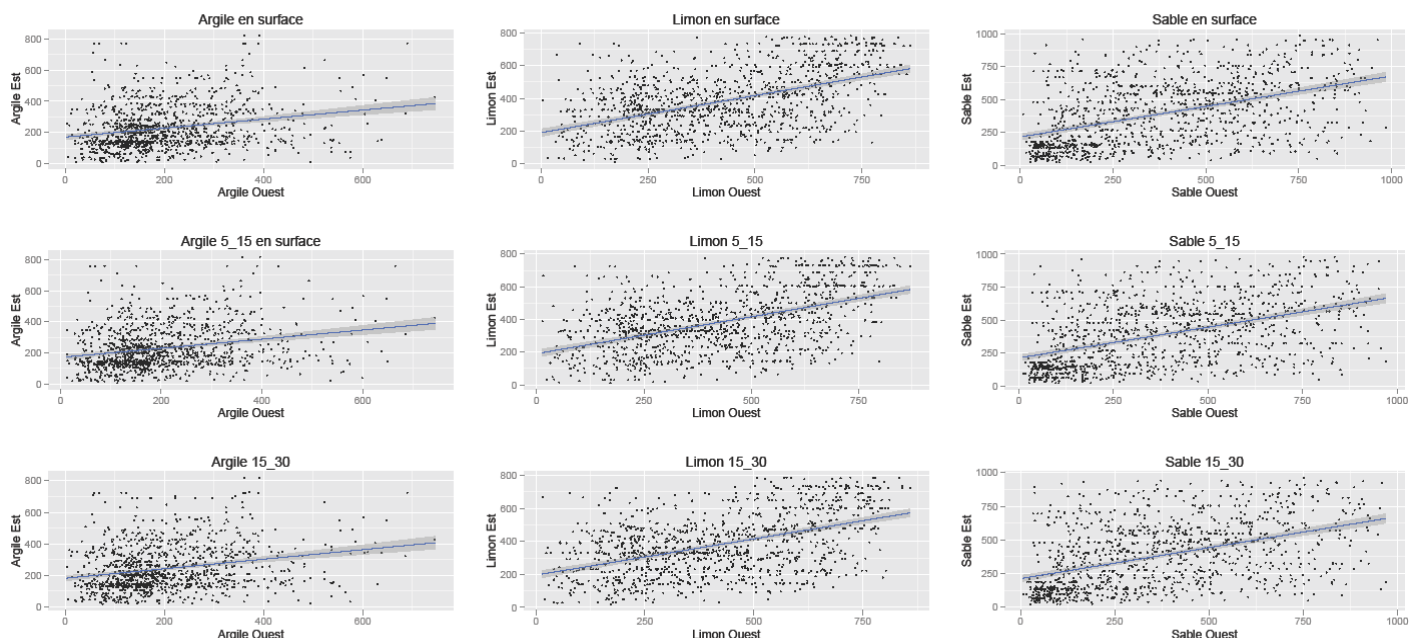


Figure 14 : Rapports entre les variables pédologiques prédites et observées par filtrage

On peut alors constater que selon cette méthode de décision (choisir le pixel proche avec la plus faible moyenne) ne renvoie pas des résultats très satisfaisants. On va alors afficher nos indicateurs de qualité :

MPE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	-0,19	-0,19	-0,17	-0,15	-0,12	-0,07
Limon	0,35	0,35	0,36	0,34	0,27	0,22
Sable	-0,17	-0,17	-0,17	-0,15	-0,1	-0,11
Ph	0,08					

RMSE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	1,23	1,24	1,29	1,37	1,38	1,37
Limon	1,1	1,1	1,13	1,21	1,27	1,29
Sable	1,01	1,02	1,04	1,12	1,21	1,26
Ph	0,3					

RPD						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,871	0,971	0,846	0,791	0,748	0,751
Limon	0,858	0,851	0,835	0,785	0,798	0,76
Sable	0,944	0,945	0,935	0,876	0,799	0,786
Ph	0,788					

R ²						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,0481	0,0066	0,2221	0,1312	0,2073	0,1127
Limon	0,0455	0,0013	0,2185	0,0896	0,2059	0,0469
Sable	0,0285	0,0037	0,1978	0,0521	0,188	0,0316
Ph	0,0205					

Le coefficient de détermination est trop faible pour nous permettre de valider ce modèle (en pédologie, et compte tenu du nombre de données dont nous disposons, un R² de 0.3 est jugé acceptable).

On décide alors de vérifier si la méthode de filtrage est plus efficace qu'une affectation aléatoire à chaque pixel cible d'un pixel Est.

On compare alors le boxplot des indices de dissimilarité pédologiques entre chaque pixel cible et son plus proche voisin attribué par filtrage et le boxplot des indices de dissimilarité pédologiques entre chaque pixel cible et le pixel qui lui a été attribué aléatoirement.

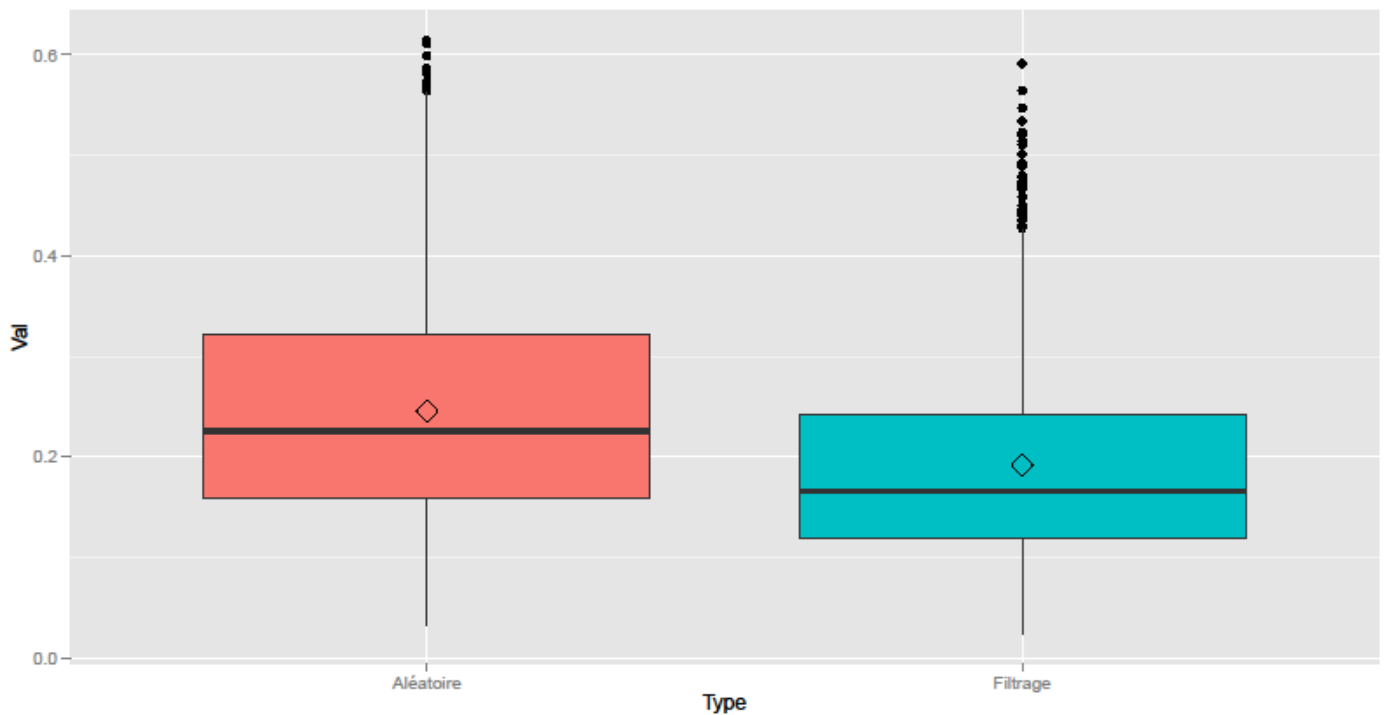


Figure 15 : Comparaison des indices pédologiques entre filtrage et affectation aléatoire

Un test d'égalité de la moyenne nous indique que les indices de dissimilarité pédologiques ne suivent pas la même distribution s'ils sont obtenus par filtrage ou par affectation aléatoire. On peut donc affirmer que la méthode de filtrage nous apporte quand même des informations sur la composition du sol, même si ces informations sont faibles.

La faiblesse des informations apportées peut provenir d'un manque de diversité des sols de la région Centre. Afin de vérifier si le sol manque vraiment de diversité, on effectue un clustering selon les variables pédologiques : argile, limon et sable, aux différentes profondeurs :

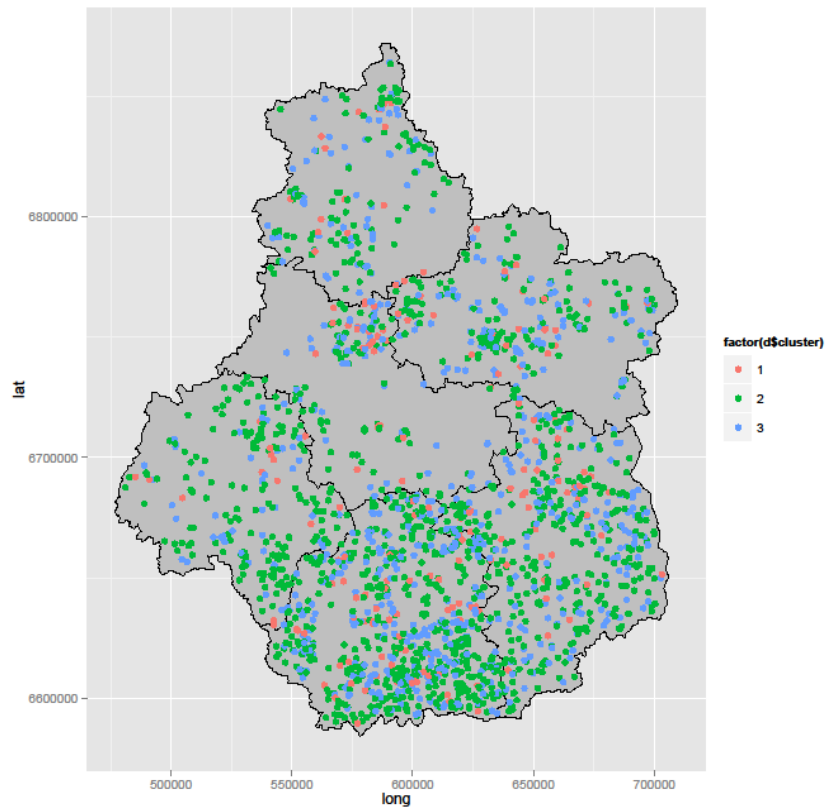


Figure 16 : Répartition des clusters pédologiques au sein de la région

On voit alors que ces données sont réparties en seulement 3 clusters qui ne présentent pas de répartition géographique bien définie.

On va maintenant regarder les résultats renvoyés par l'ACP

3.2 ACP

Notre méthode d'affectation par ACP ne renvoie qu'un seul voisin pour chaque pixel cible, on peut donc directement comparer les valeurs pédologiques du pixel cible avec celles de son plus proche voisin :

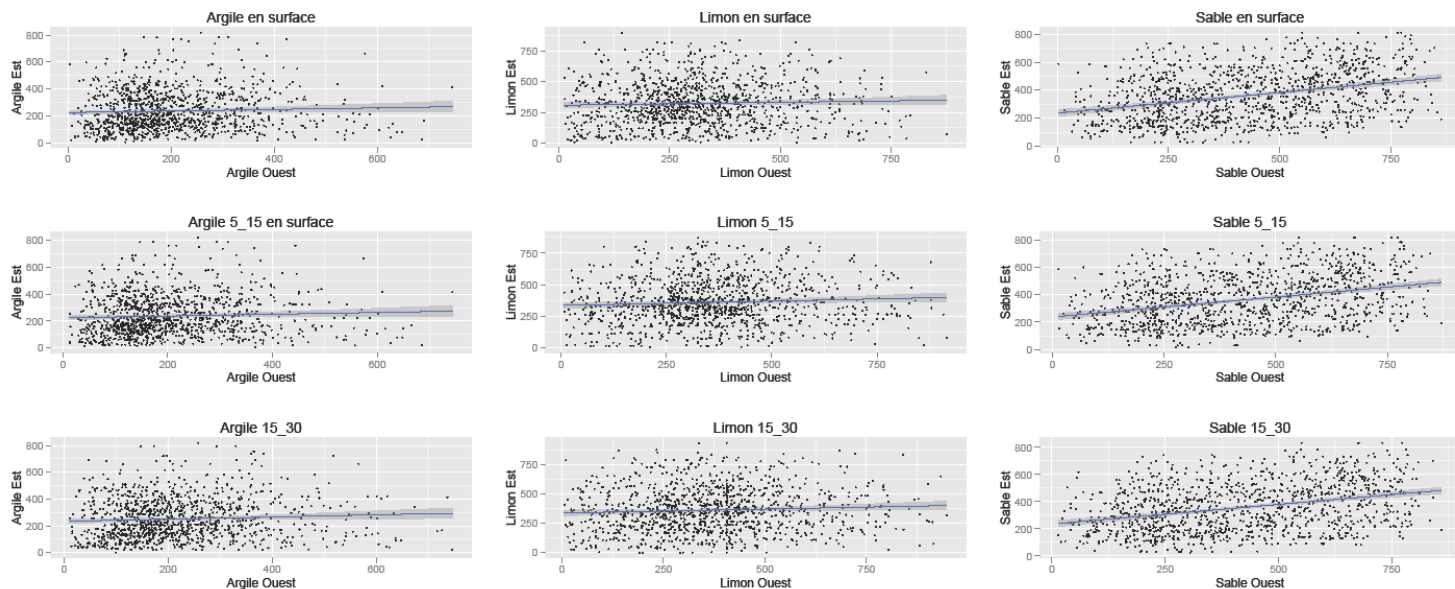


Figure 17 : Rapport entre les variables pédologiques prédites et observées par l'affectation par ACP

On constate alors que les résultats sont moins bons que pour la méthode de filtrage, particulièrement pour l'argile et le limon, quelle que soit la profondeur.

Voyons maintenant les indices de qualité pour cette méthode :

MPE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	-0,25	-0,25	-0,22	-0,1	-0,07	-0,07
Limon	0,35	0,36	0,35	0,3	0,19	0,12
Sable	-0,13	-0,13	-0,14	-0,15	-0,09	-0,04
Ph	0,01					

RMSE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	1,35	1,35	1,35	1,38	1,35	1,35
Limon	1,14	1,15	1,16	1,22	1,27	1,28
Sable	1,13	1,13	1,13	1,17	1,17	1,18
Ph	1,36					

RPD						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,818	0,821	0,891	0,779	0,775	0,762
Limon	0,78	0,773	0,766	0,731	0,733	0,725
Sable	0,867	0,869	0,874	0,855	0,828	0,811
Ph	0,792					

R2						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,0106	0,003	0,1538	0,0895	0,1075	0,0837
Limon	0,0115	0,0082	0,1492	0,0524	0,1085	0,0738
Sable	0,013	0,0083	0,1386	0,0327	0,1097	0,0605
Ph	0,0115					

On peut voir que les valeurs du coefficient de détermination sont inférieures à celles du filtrage ; nous allons donc écarter cette méthode et regarder l'efficacité du clustering suivi du filtrage :

3.3 Clustering

Après avoir procédé au clustering, on dispose de 7 clusters sur lesquels procéder à notre stratégie de filtrage. On procède donc à la vérification de la même façon que pour cette méthode : on compare les valeurs pédologiques des pixels cibles à celles de leur plus proche voisin.

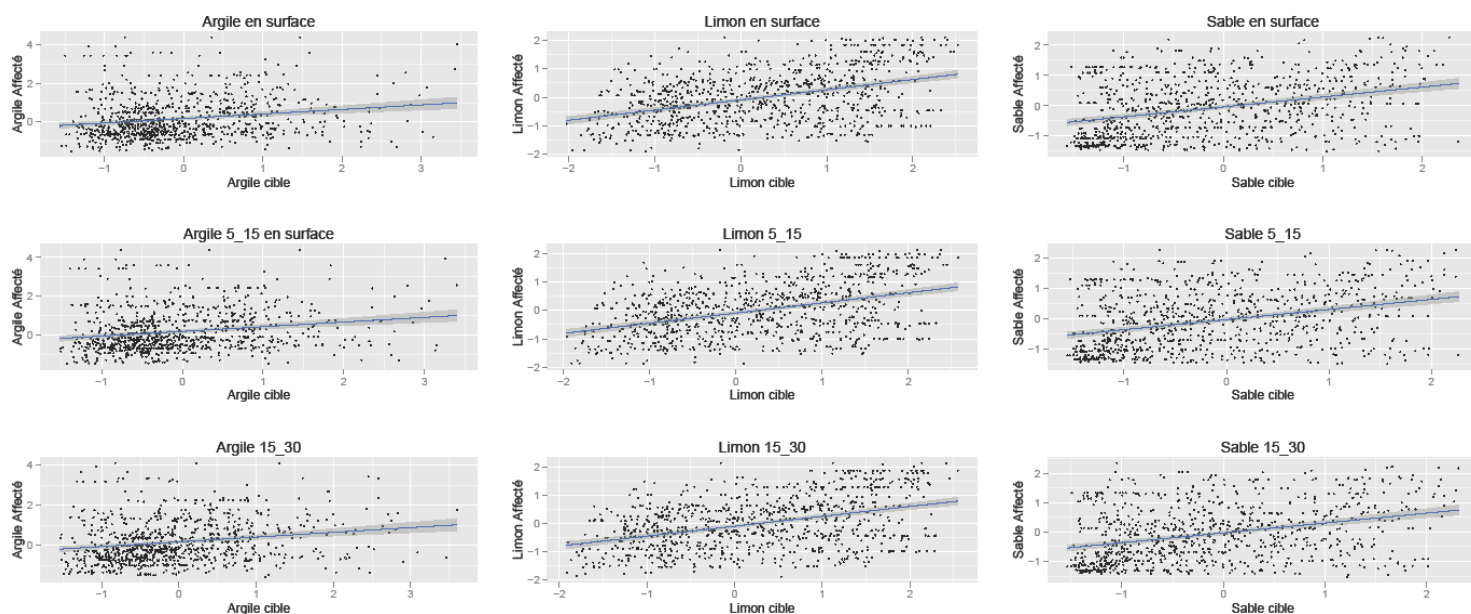


Figure 18 : Rapport entre les variables pédologiques prédites et observées par le clustering suivi du filtrage

Une fois encore, les valeurs pédologiques exactes des pixels cibles sont éloignées de celles de leurs plus proches voisins. On va, une fois encore, regarder les indicateurs de qualité de ce modèle :

MPE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	-0,23	-0,23	-0,21	-0,17	-0,14	-0,07
Limon	0,26	0,26	0,26	0,27	0,22	0,2
Sable	-0,07	-0,07	-0,08	-0,08	-0,06	-0,09
Ph	0,12					

RMSE						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	1,28	1,27	1,27	1,31	1,35	1,33
Limon	1,12	1,12	1,13	1,2	1,27	1,29
Sable	1,1	1,1	1,1	1,15	1,24	1,16
Ph	1,3					

RDP						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,853	0,862	0,848	0,837	0,766	0,8
Limon	0,816	0,813	0,806	0,742	0,677	0,666
Sable	0,885	0,89	0,899	0,896	0,791	0,793
Ph	0,733					

R ²						
Profondeur	0_5	5_15	15_30	30_60	60_100	100_200
Argile	0,0358	0,0221	0,1665	0,1081	0,1117	0,0786
Limon	0,0375	0,0123	0,169	0,0561	0,114	0,028
Sable	0,0377	0,0122	0,1646	0,0259	0,1128	0,0162
Ph	0,0219					

Les valeurs du coefficient de détermination sont une nouvelle fois trop faibles pour nous permettre de valider notre modèle.

Comparaison du coefficient de détermination des différentes méthodes

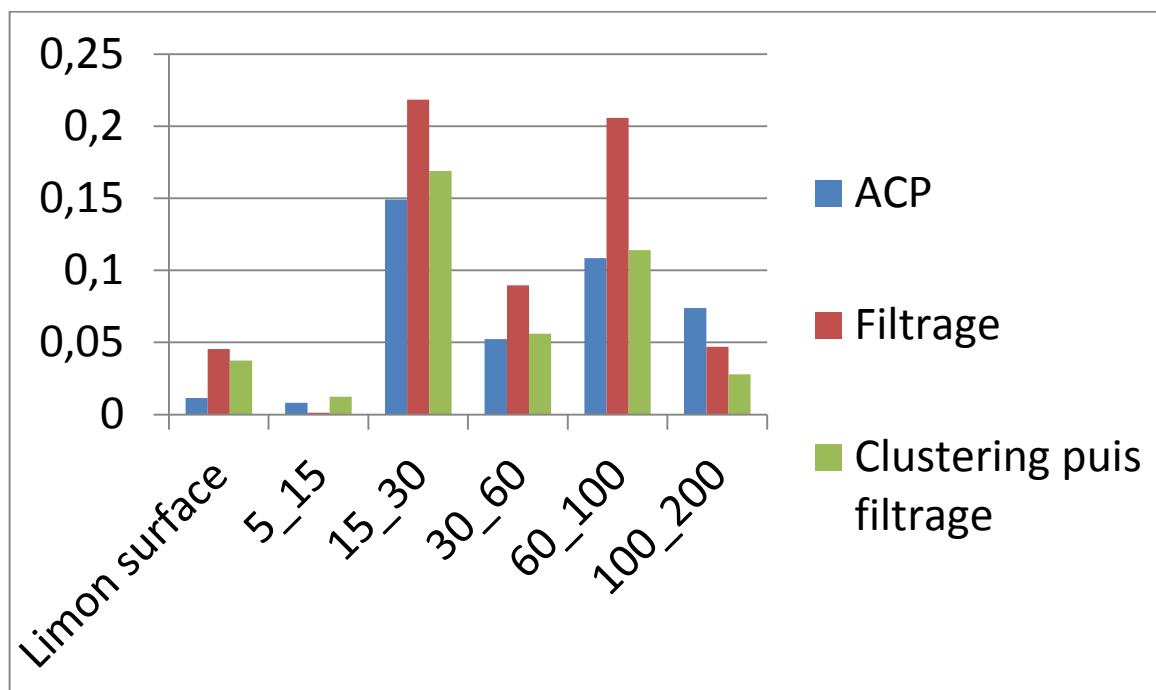


Figure 19 : Coefficient de détermination pour le limon

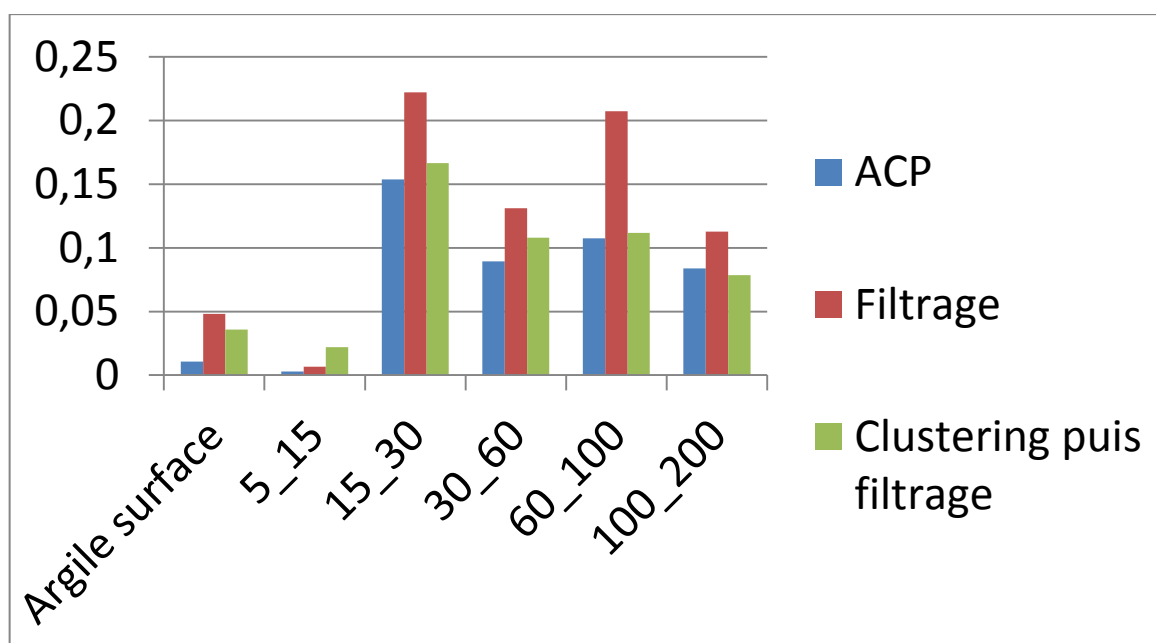


Figure 20 : Coefficient de détermination pour l'argile

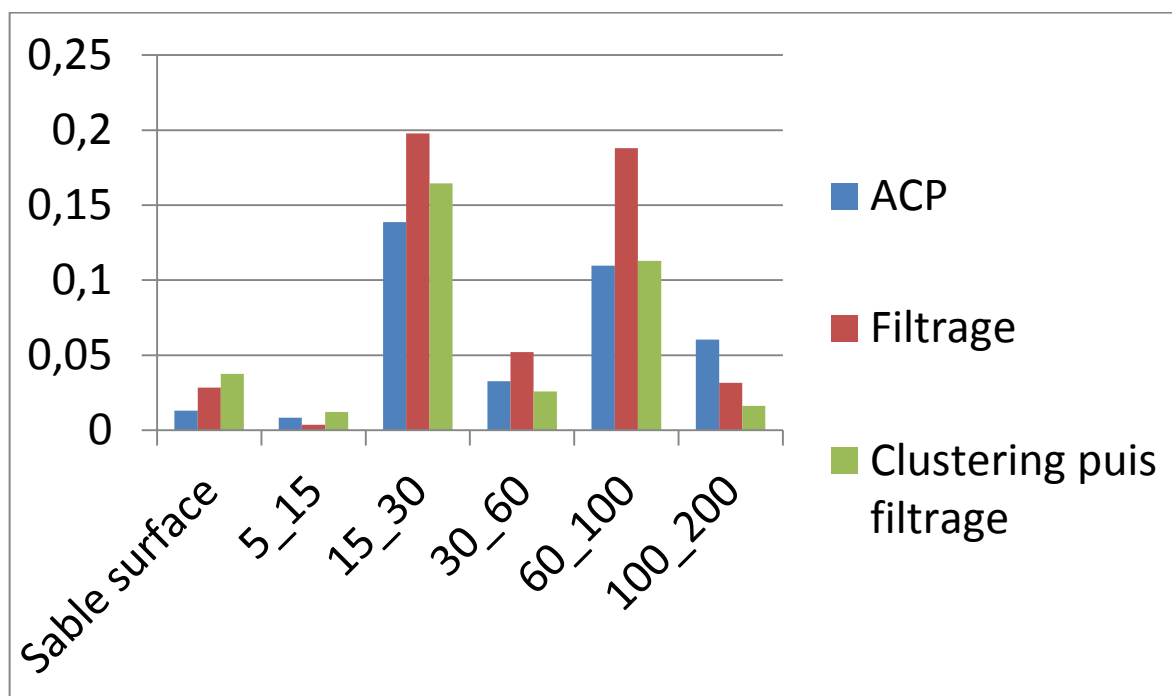


Figure 21 : Coefficient de détermination pour le sable

Aucune parmi les trois méthodes n'a renvoyé de résultats satisfaisants mais on constate que les résultats sont biens meilleurs pour les couches 15_30 et 60_100.

4 Discussion

Les raisons pour lesquelles on n'a pas pu établir de relation satisfaisante entre une proximité des variables pédologiques et une proximité des autres variables peuvent être variées :

- les variables externes, comme les variables pédologiques, ne présentent peut-être pas une étendue suffisante. En d'autres termes, notre jeu de données n'est pas suffisamment « contrasté ».
- les variables externes ne traduisent pas véritablement la répartition des sols, d'autres variables externes devraient donc être recherchées.
- Le jeu de données pédologiques est « biaisé » par la stratégie d'échantillonnage du pédologue. En effet, le pédologue, qui est contraint par des considérations financières, cherche à maximiser la différence entre ses points de prélèvements. En d'autres termes, il réalise rarement des prélèvements dans des sols en tous points identiques. Autrement dit, la probabilité de trouver le voisin « parfait » est réduite par sa stratégie.
- La méthode des "splines" est connue pour créer des valeurs peu fiables aux extrémités de la courbe d'ajustement (soit dans les couches les plus en surfaces ou les plus en profondeur) ou en cas de changement abrupt de granulométrie (Arrouays et al., 2014). Ceci explique probablement :
 - 1) que les valeurs des couches les plus superficielles et des couches les plus profondes soient systématiquement mal prédites.
 - 2) que la valeur 30-60 le soit également car à ces profondeurs il y a souvent dans notre cas des changements brutaux provoqués soit par le labour, soit par l'apparition d'horizons d'accumulation d'argile.

De par ces biais méthodologiques, il est logique que nous ayons de meilleurs résultats dans les couches 15-30 et 60-100. Ces couches sont moins affectées par les biais liés aux extrêmes de la courbe d'ajustement et par des changements brutaux. En particulier, la couche 15-30 est en général homogénéisée par le labour et correspond le plus souvent à un point de prélèvement. La couche 60-100 est également le plus souvent constituée d'un horizon relativement homogène en termes de granulométrie.

Le fait que nous ayons des résultats légèrement meilleurs pour la couche 15-30 pourrait s'expliquer par le fait que nos variables prédictives traduisent mieux des effets de surface que des effets de profondeur, en particulier pour les émissions gamma.

Sur la région Centre, et avec le jeu de données dont nous disposons, la méthode « Homosoil » a donc conduit à un échec. Pour aller plus loin, une piste serait d'élargir la variabilité au sein du jeu de données et de travailler sur un échantillon non biaisé. Une possibilité serait de tester la méthode sur la France entière, et sur les données du réseau systématique de mesure de la qualité des sols construit selon une maille systématique de 16x16-km. Nous aurions alors beaucoup plus de contraste dans certaines variables externes et nous pourrions introduire d'autres facteurs comme par exemple le climat.

Nous pourrions aussi tester la pertinence de la méthode sur d'autres variables pédologiques et en particulier intégrer la profondeur du sol.

Conclusion

Ce stage m'a permis d'apprendre les bases d'une nouvelle méthode statistique : le clustering et de mettre en application la méthode de l'Analyse en Composantes Principales. Il m'a également rappelé l'importance du prétraitement des données : la visualisation des histogrammes et la log-normalisation des données.

J'ai pu approfondir mes connaissances en programmation R par la création de plusieurs fonctions. J'ai également découvert de nouveaux outils pour ce logiciel, principalement avec l'interface RStudio et le package ggplot2.

J'ai aussi beaucoup appris sur la façon de mener une étude en équipe, principalement l'importance de l'interaction avec le reste de cette équipe et la présentation régulière des résultats obtenus.

J'ai également appris que lors d'une étude de ce genre, l'absence de résultat est en soi, un résultat. Il s'agissait de fait d'un test méthodologique qui ne s'est pas révélé positif dans le cadre de la région Centre et avec les variables que nous avons utilisées.

Plus généralement, les fonctions que j'ai mises en place pourront permettre de tester Homosoil sur des bases de données plus larges.

Bibliographie

- Mallavan, B.P., McBratney, A.B. & Minasny, B. (2010). Homosoil, a methodology for quantitative extrapolation of soil information across the globe. In: Boettinger, J.L., Howell, D., Moore, A.C., Hartemink, A.E., and Kienast-Brown, S. (eds), *Digital soil mapping - bridging research, environmental application, and operation*. Springer: Dordrecht.

- Arrouays D., Hempel J., McBratney A.B. McKenzie N., Richer de Forges A., (2014). GlobalSoilMap, Basis of the global spatial soil information system, CRC Press, Taylor&Francis, London, 478 p.

- Arrouays D., Ciampalini R., Martin M., Richer de Forges A. & Saby N. (2014) Soil texture GlobalSoilMap products for the French region « Centre ». In GlobalSoilMap, Basis of the global spatial soil information system p121-126

- Sites internet :
 - <http://www.inra.fr/>
 - <http://www.val-de-loire.inra.fr/>
 - <http://www.cookbook-r.com/>
 - <http://www.jybaudot.fr/General/indexstats.html>

Annexes

1 Présentation des variables.....	38
1-1 Dictionnaire des variables.....	38
1-2 Corrélacion entre les variables externes et pédologiques en surface.....	41
2 Code R.....	47
2-1 Fonction de filtre.....	47
2-2 Fonction de filtrage.....	48
2-3 Code de l'affectation par ACP.....	50
2-4 Normalisation et standardisation des variables.....	51
2-5 Multiplot.....	52
2-6 Code du clustering.....	53
2-7 Affichage des pixels sur la carte, groupés par cluster.....	54
2-8 Indices de qualité des modèles.....	55
3 Résultats.....	59
3-1 Densité des variables par cluster.....	59
3-2 Centre des clusters.....	91
3-3 Répartition des clusters par rapport à l'altitude.....	99
3-4 Répartition des clusters par rapport à la pente.....	100
3-5 Valeurs pédologiques prédites en fonction des valeurs exactes par le filtrage.....	101
3-6 Comparaison du R ² des méthodes	

1 Présentation des variables

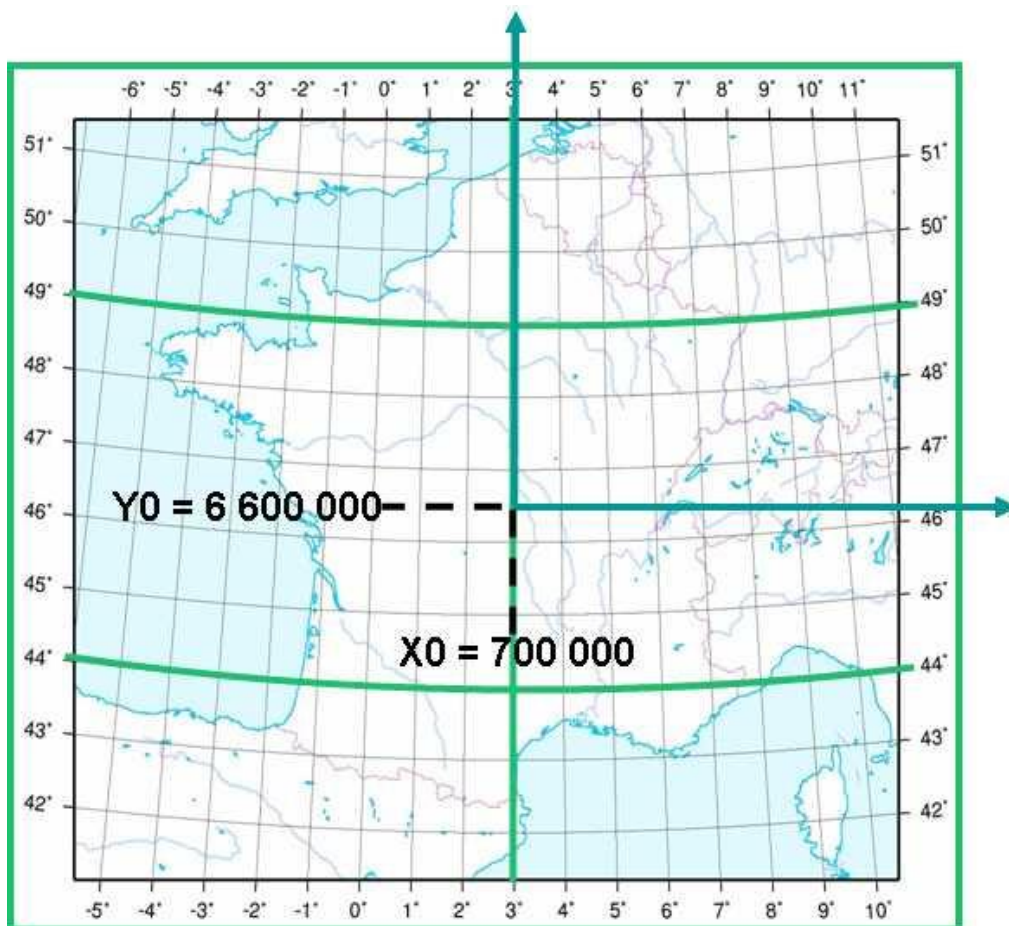
1-1 Dictionnaire des variables

Variable	Type	Description
id_profil	qualitatif (2503 niveaux)	identifiant du pixel
x		Coordonnées du pixel dans le système
y		Lambert 93
argile	quantitatif	Taux d'argile du pixel en surface
limon	quantitatif	Taux de limon du pixel en surface
sable	quantitatif	Taux de sable du pixel en surface
sum	quantitatif	sum of argile limon & sable
srtm_elev	quantitatif	elevation
	calculés sur une fenêtre de 3x3 pixels	
srtm_slp3	quantitatif	slope
srtm_asp3	quantitatif	aspect direction
srtm_prf3	quantitatif	profile curvature
srtm_pln3	quantitatif	plan curvature
srtm_lng3	quantitatif	longitudinal curvature
srtm_crs3	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 5x5 pixels	
srtm_slp5	quantitatif	slope
srtm_asp5	quantitatif	aspect direction
srtm_prf5	quantitatif	profile curvature
srtm_pln5	quantitatif	plan curvature
srtm_lng5	quantitatif	longitudinal curvature
srtm_crs5	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 9x9 pixels	
srtm_slp9	quantitatif	slope
srtm_asp9	quantitatif	aspect direction
srtm_prf9	quantitatif	profile curvature
srtm_pln9	quantitatif	plan curvature
srtm_lng9	quantitatif	longitudinal curvature
srtm_crs9	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 17x17 pixels	

srtm_sl17	quantitatif	slope
srtm_as17	quantitatif	aspect direction
srtm_pf17	quantitatif	profile curvature
srtm_pl17	quantitatif	plan curvature
srtm_ln17	quantitatif	longitudinal curvature
srtm_cr17	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 33x33 pixels	
srtm_sl33	quantitatif	slope
srtm_as33	quantitatif	aspect direction
srtm_pr33	quantitatif	profile curvature
srtm_pl33	quantitatif	plan curvature
srtm_ln33	quantitatif	longitudinal curvature
srtm_cr33	quantitatif	cross-sectionnal curvature
aster_el	quantitatif	elevation
	calculés sur une fenêtre de 3x3 pixels	
ast_slp3	quantitatif	slope
ast_asp3	quantitatif	aspect direction
ast_prf3	quantitatif	profile curvature
ast_pln3	quantitatif	plan curvature
ast_lng3	quantitatif	longitudinal curvature
ast_crs3	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 5x5 pixels	
ast_slp5	quantitatif	slope
ast_asp5	quantitatif	aspect direction
ast_prf5	quantitatif	profile curvature
ast_pln5	quantitatif	plan curvature
ast_lng5	quantitatif	longitudinal curvature
ast_crs5	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 9x9 pixels	
ast_slp9	quantitatif	slope
ast_asp9	quantitatif	aspect direction
ast_prf9	quantitatif	profile curvature
ast_pln9	quantitatif	plan curvature
ast_lng9	quantitatif	longitudinal curvature
ast_crs9	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 17x17 pixels	

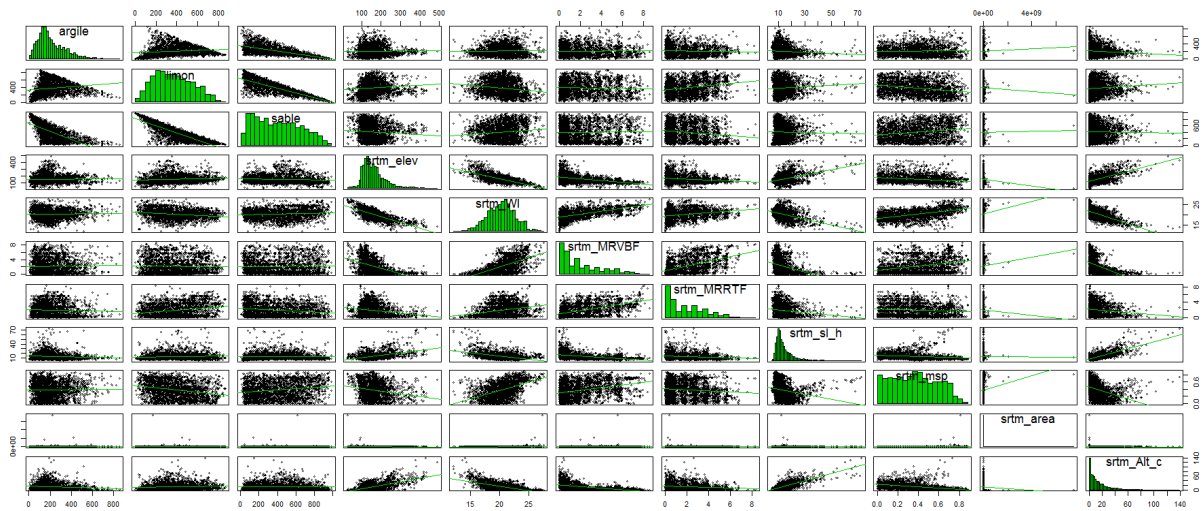
ast_sl17	quantitatif	slope
ast_as17	quantitatif	aspect direction
ast_pf17	quantitatif	profile curvature
ast_pl17	quantitatif	plan curvature
ast_ln17	quantitatif	longitudinal curvature
ast_cr17	quantitatif	cross-sectionnal curvature
	calculés de façon identique puis moyennés sur une fenêtre de 33x33 pixels	
ast_sl33	quantitatif	slope
ast_as33	quantitatif	aspect direction
ast_pf33	quantitatif	profile curvature
ast_pl33	quantitatif	plan curvature
ast_ln33	quantitatif	longitudinal curvature
ast_cr33	quantitatif	cross-sectionnal curvature
srtm_WI	quantitatif	Wetness index
srtm_MRVPF	quantitatif	Multiresolution valley bottom flatness
srtm_MRRTF	quantitatif	Multiresolution ridge top flatness
srtm_sl_h	quantitatif	slope height
srtm_msp	quantitatif	Mid-Slope Position
srtm_area	quantitatif	catchment area
srtm_Alt_c	quantitatif	altitude above channel network
uranium	quantitatif	uranium from gamma ray
ti	quantitatif	titane from gamma ray
th	quantitatif	thorium from gamma ray
potassuim	quantitatif	potassium from gamma ray
total	quantitatif	total gamma ray
cr100	quantitatif	% of cereals (yr2000)
vig00	quantitatif	% of vineyards(yr2000)
olea00	quantitatif	% of oleaginous (yr2000)
bois00	quantitatif	% of woods(yr2000)
jach00	quantitatif	% of fallow (yr2000)
geo_50k	qualitatif (16 niveaux)	geological map 50k
argile_5_15	quantitatif	Taux d'argile entre 5 et 15 cm
limon_5_15	quantitatif	Taux de limon entre 5 et 15 cm
sable_5_15	quantitatif	Taux de sable entre 5 et 15 cm

argile_15_30	quantitatif	Taux d'argile entre 15 et 30 cm
limon_15_30	quantitatif	Taux de limon entre 15 et 30 cm
sable_15_30	quantitatif	Taux de sable entre 15 et 30 cm
argile_30_60	quantitatif	Taux d'argile entre 30 et 60 cm
limon_30_60	quantitatif	Taux de limon entre 30 et 60 cm
sable_30_60	quantitatif	Taux de sable entre 30 et 60 cm
argile_60_100	quantitatif	Taux d'argile entre 60 et 100 cm
limon_60_100	quantitatif	Taux de limon entre 60 et 100 cm
sable_60_100	quantitatif	Taux de sable entre 60 et 100 cm
argile_100_200	quantitatif	Taux d'argile entre 100 et 200 cm
limon_100_200	quantitatif	Taux de limon entre 100 et 200 cm
sable_100_200	quantitatif	Taux de sable entre 100 et 200 cm

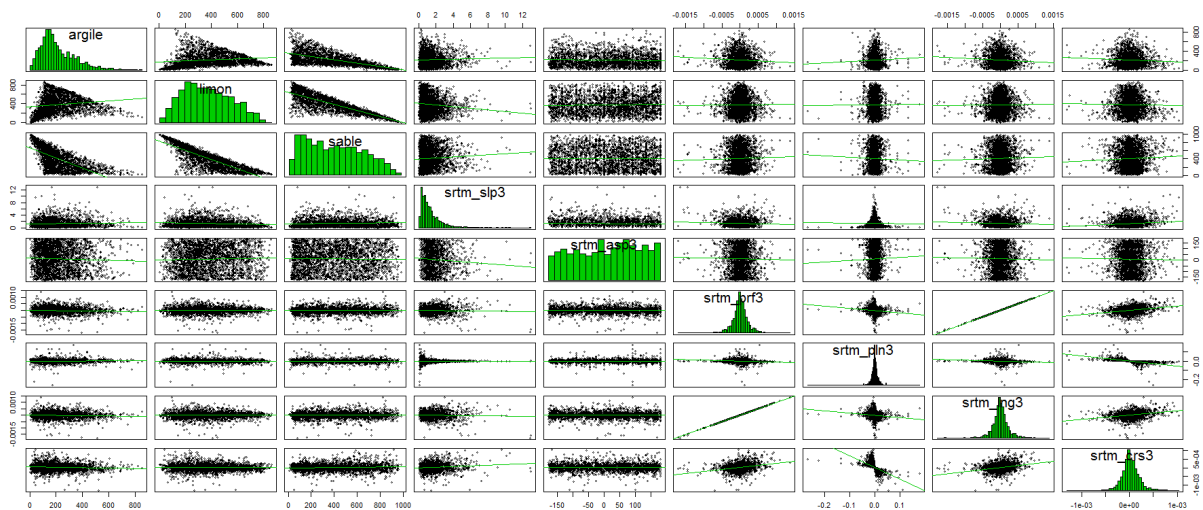


Annexe 1 : représentation de la France en Lambert 93

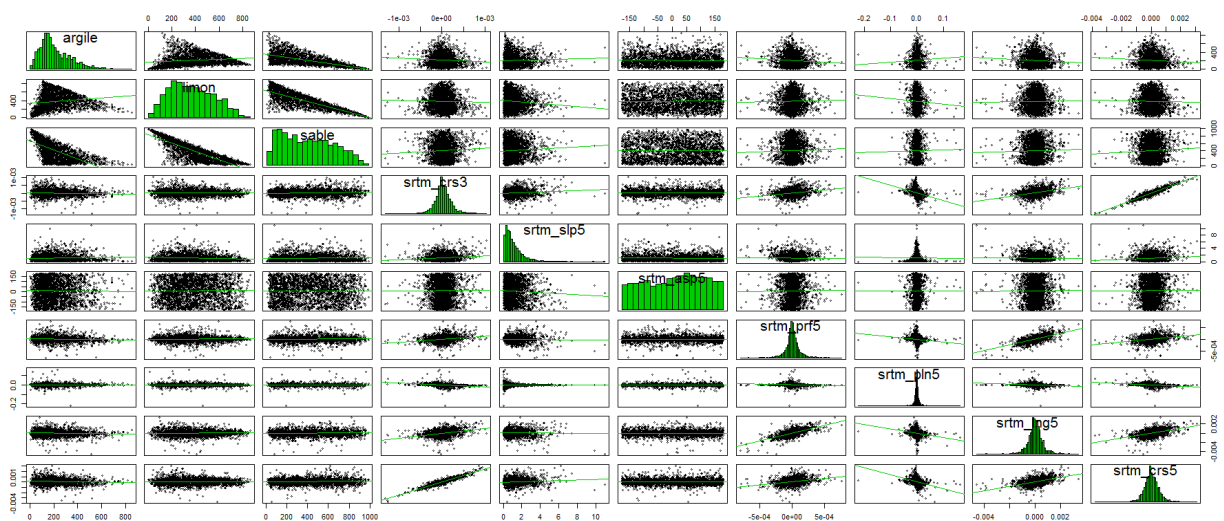
1-2 Corrélation entre les variables externes et pédologiques en surface :



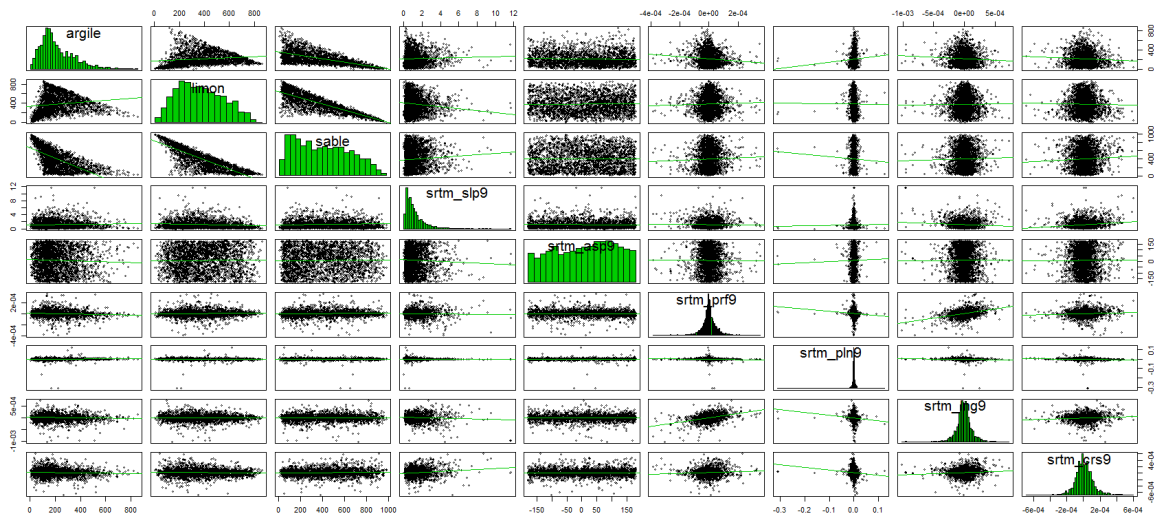
Scatterplot 1 : variables srtm



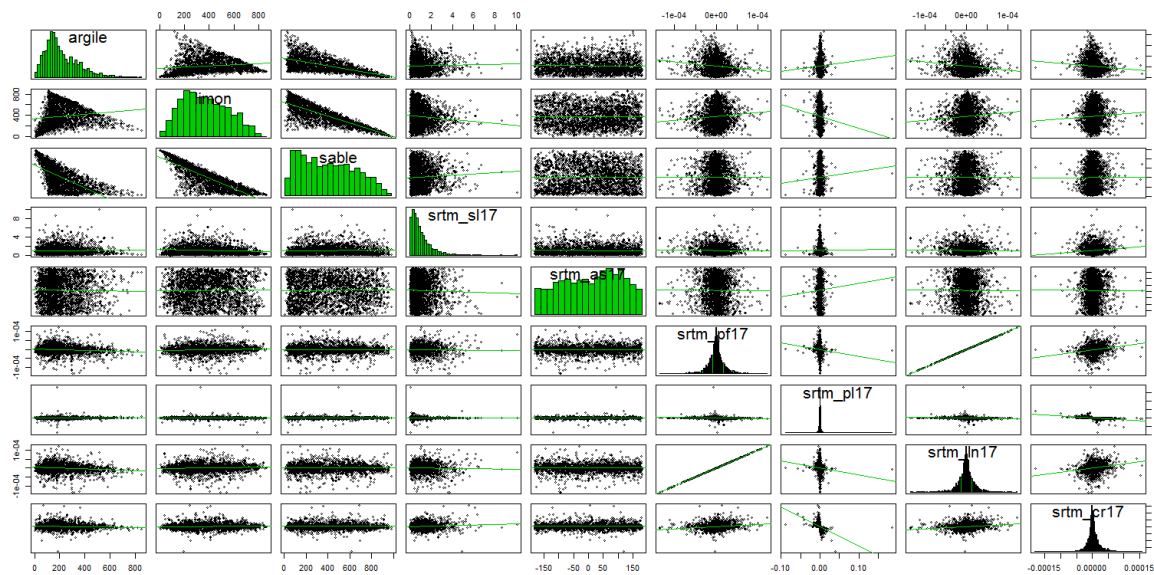
Scatterplot 2: variables srtm 3



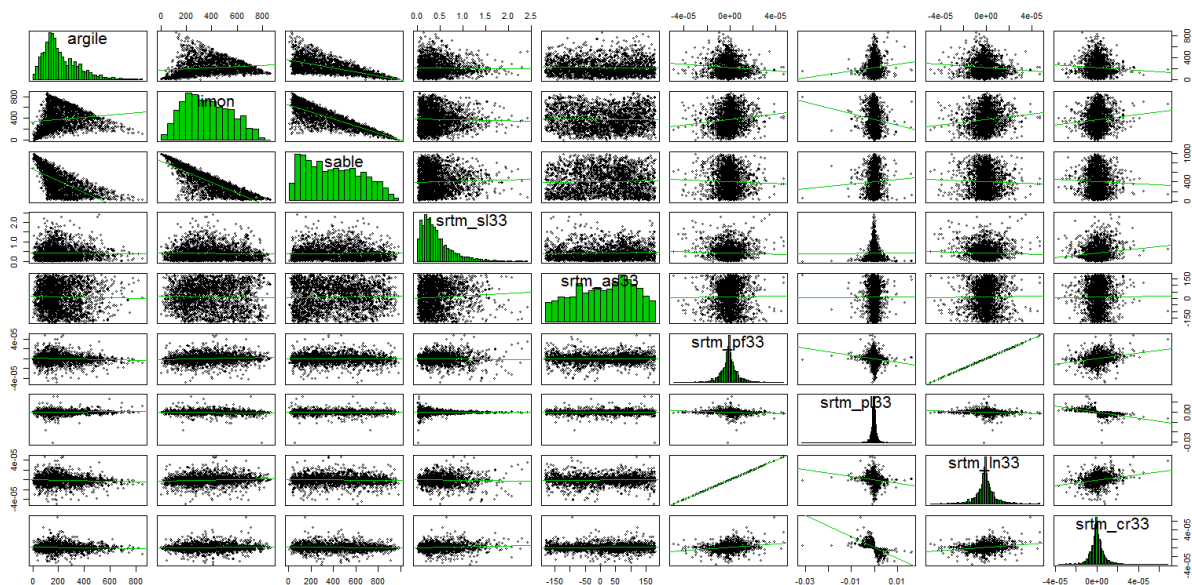
Scatterplot 3: variables srtm 5



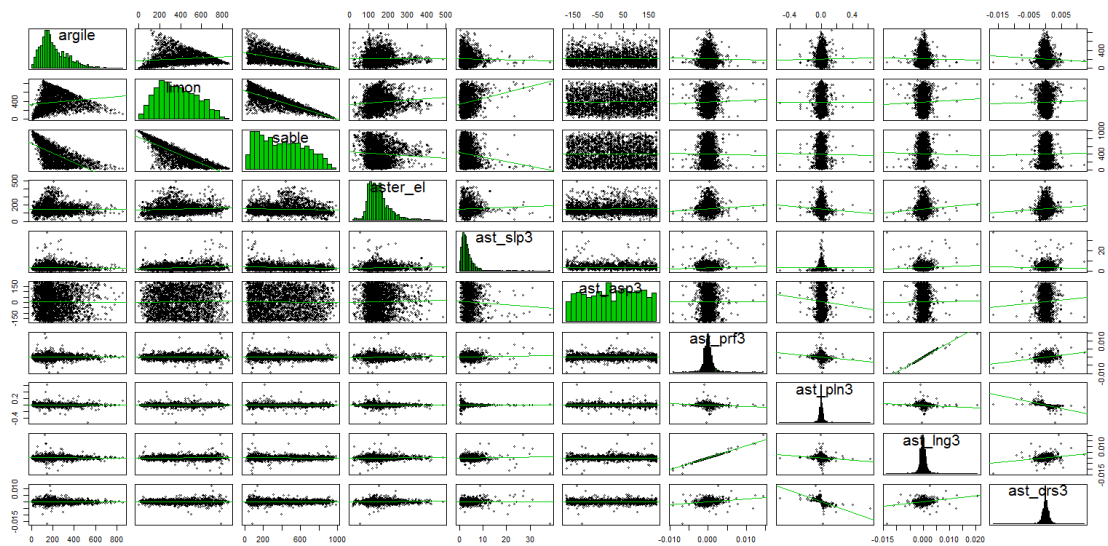
Scatterplot 4: variables srtm 9



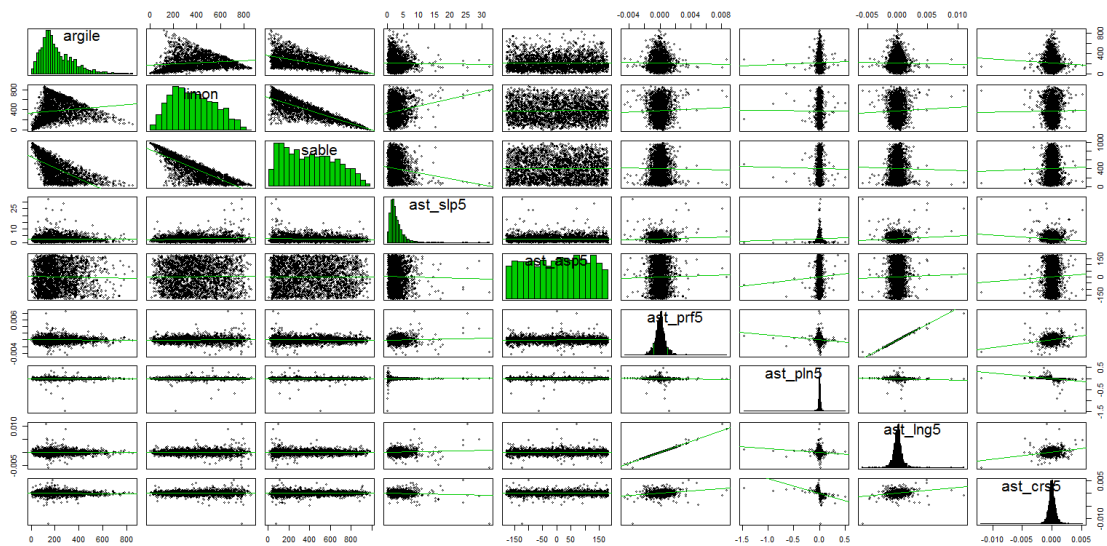
Scatterplot 5 : variables srtm 17



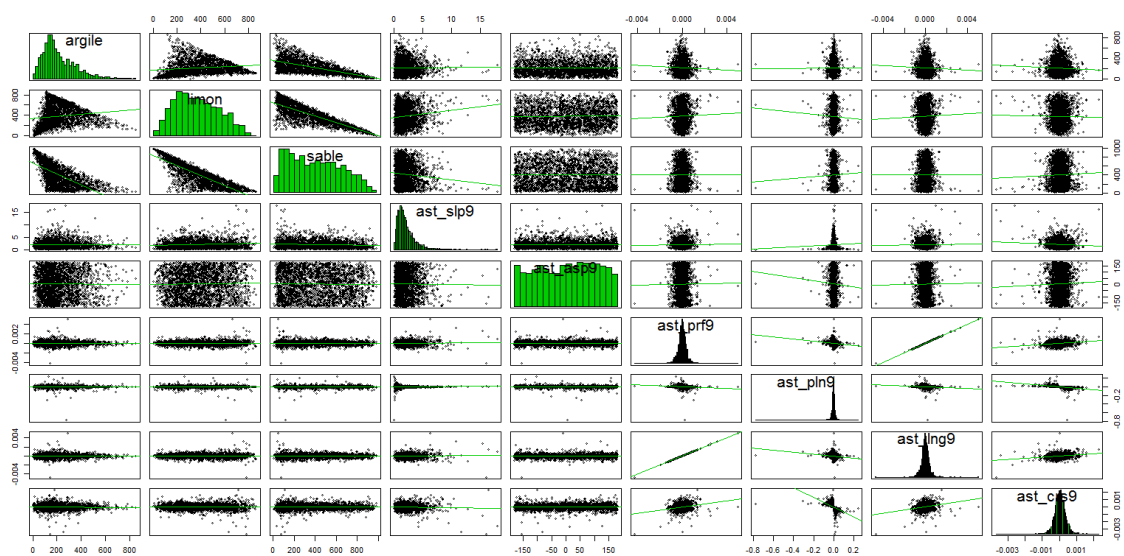
Scatterplot 6 : variables srtm 33



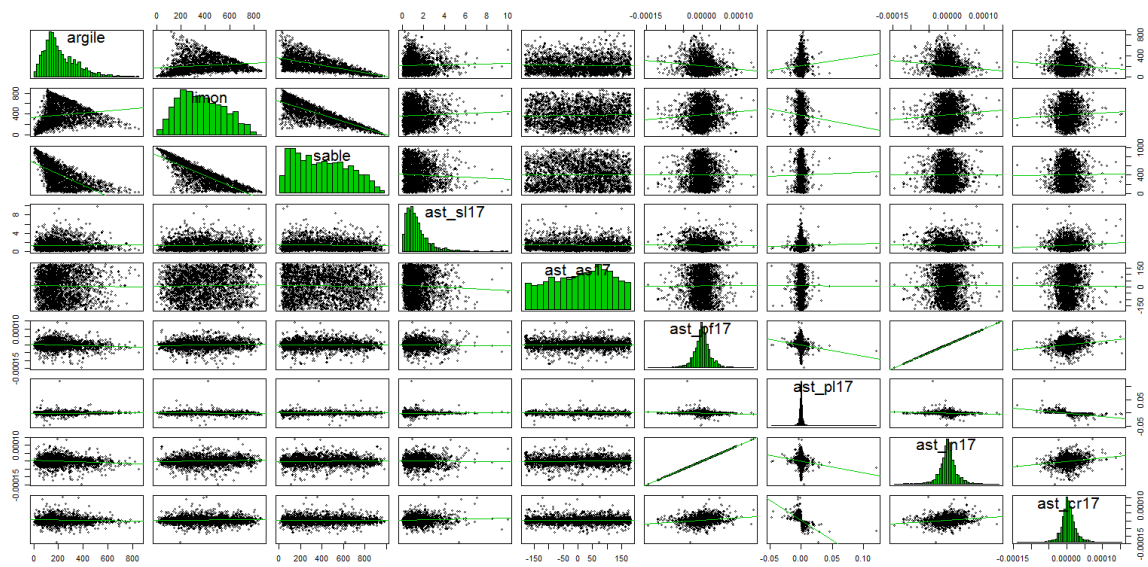
Scatterplot 7 : variables ast 3



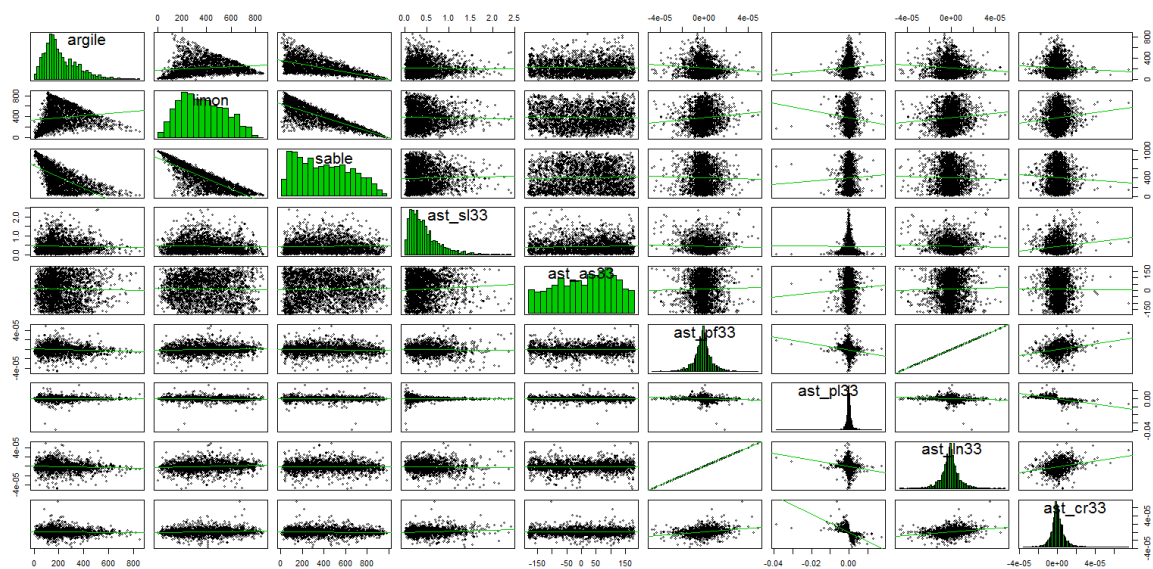
Scatterplot 8 : variables ast 5



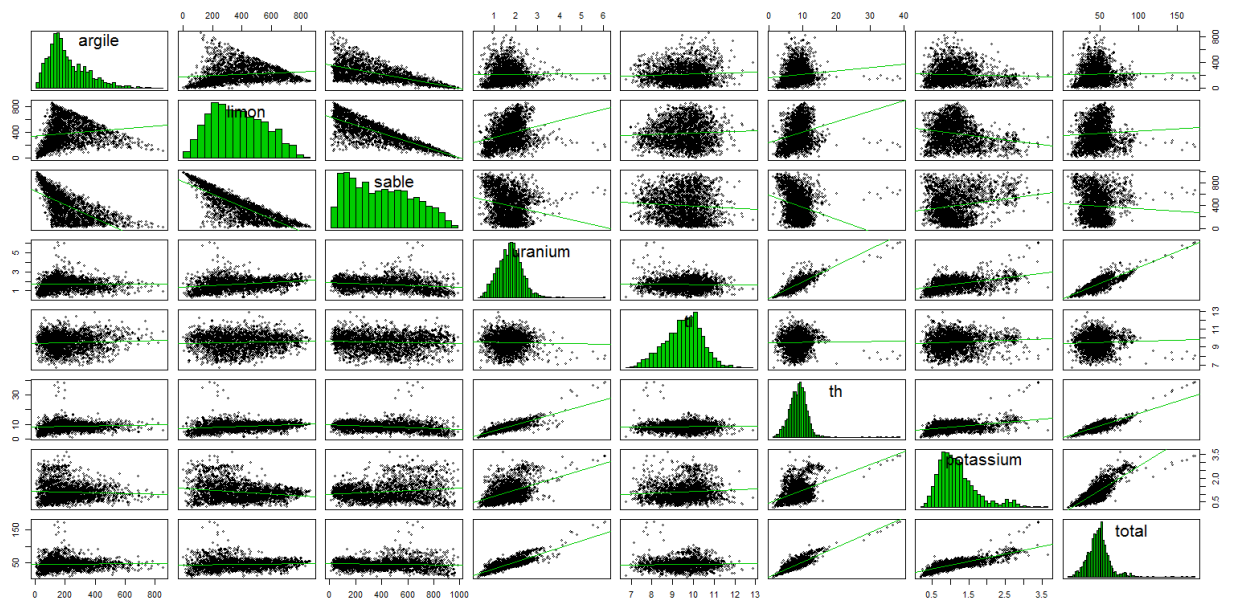
Scatterplot 9 : variables ast 9



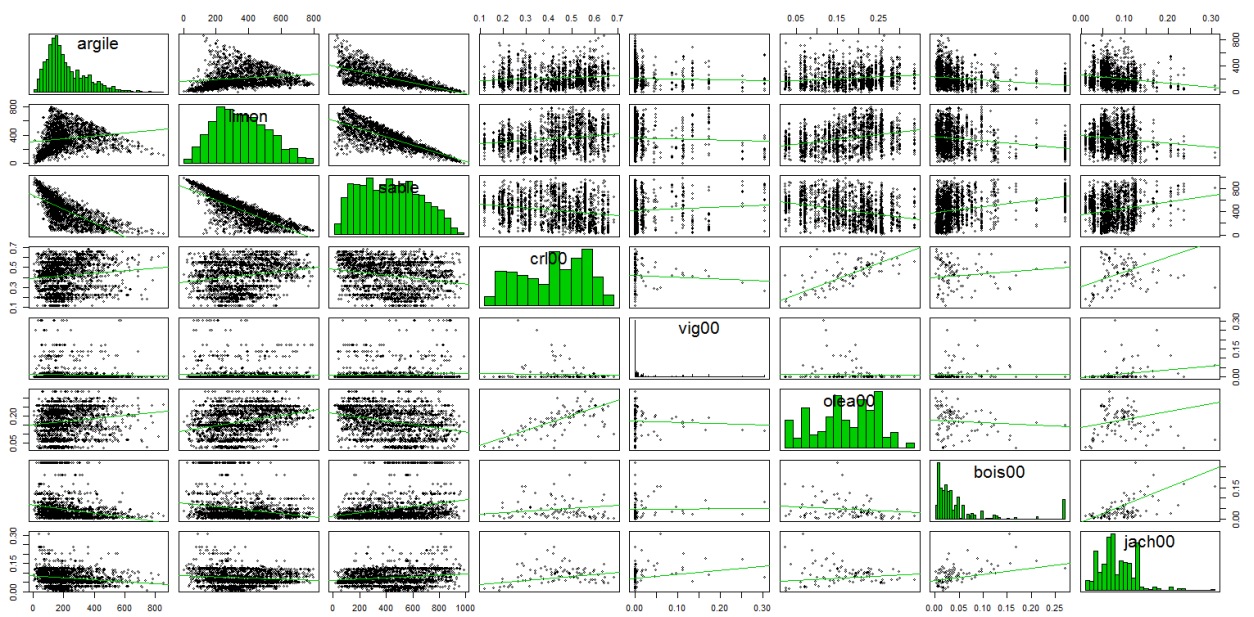
Scatterplot 10 : variables ast 17



Scatterplot 11 : variables ast 33



Scatterplot 12 : variables gamma



Scatterplot 13 : variables d'occupation

2 Code R

2-1 Fonction de filtre

On choisit un niveau de filtre parmi 1, 0.6, 0.5, 0.4, 0.3, 0.2 et 0.1 ; cette fonction renvoie la liste des identifiants des pixels qui présentent une similarité géologique inférieure ou égale à ce niveau. La similarité géologique est présentée après cette fonction

niv_filtre : niveau de similarité géologique minimum qu'on accepte

Centre : data frame qu'on observe

```
filtre <- fonction (niv_filtre, Centre){
  #créé un filtre pour la variable geo_50k plus ou moins
  #précis selon la similitude qu'on est prêt à accorder
  #le filtre est basé sur le fichier "distances geol_50k.xlsx"

  a1<-Centre$id_profil[Centre$geo_50k=="a1"]
  a2<-Centre$id_profil[Centre$geo_50k=="a2"]
  a3<-Centre$id_profil[Centre$geo_50k=="a3"]
  a4<-Centre$id_profil[Centre$geo_50k=="a4"]
  a5<-Centre$id_profil[Centre$geo_50k=="a5"]
  a6<-Centre$id_profil[Centre$geo_50k=="a6"]
  a7<-Centre$id_profil[Centre$geo_50k=="a7"]
  a8<-Centre$id_profil[Centre$geo_50k=="a8"]
  a9<-Centre$id_profil[Centre$geo_50k=="a9"]
  a10<-Centre$id_profil[Centre$geo_50k=="a10"]
  a11<-Centre$id_profil[Centre$geo_50k=="a11"]
  a12<-Centre$id_profil[Centre$geo_50k=="a12"]
  a13<-Centre$id_profil[Centre$geo_50k=="a13"]
  a14<-Centre$id_profil[Centre$geo_50k=="a14"]
  a15<-Centre$id_profil[Centre$geo_50k=="a15"]
  a16<-Centre$id_profil[Centre$geo_50k=="a16"]

  #correspond au niveau de filtre 1
  a1<-a1[!is.na(a1)];      a2<-a2[!is.na(a2)];      a3<-a3[!is.na(a3)]
  a4<-a4[!is.na(a4)];      a5<-a5[!is.na(a5)];      a6<-a6[!is.na(a6)]
  a7<-a7[!is.na(a7)];      a8<-a8[!is.na(a8)];      a9<-a9[!is.na(a9)]
  a10<-a10[!is.na(a10)];    a11<-a11[!is.na(a11)];    a12<-a12[!is.na(a12)]
  a13<-a13[!is.na(a13)];    a14<-a14[!is.na(a14)];    a15<-a15[!is.na(a15)]
  a16<-a16[!is.na(a16)]

  A1<-a1; A2<-a2; A3<-a3; A4<-a4; A5<-a5; A6<-a6; A7<-a7; A8<-a8; A9<-a9;
  A10<-a10; A11<-a11; A12<-a12; A13<-a13; A14<-a14; A15<-a15; A16<-a16

  if(niv_filtre==0.1){
    a2<-c(A2,A3,A4,A5,A10,A11);      a3<-c(A3,A2,A4,A5,A10,A11,A16)
    a4<-c(A4,A11,A2,A3,A13,A8,A14);  a5<-c(A5,A16,A3,A2)
    a6<-c(A6,A7,A9,A10);              a7<-c(A7,A9,A6)
    a8<-c(A8,A10,A11,A4);              a9<-c(A9,A7,A6)
    a10<-c(A10,A6,A8,A11,A16,A2,A3);  a11<-c(A11,A4,A8,A10,A13,A2,A3,A14)
    a13<-c(A13,A4,A11,A14);           a14<-c(A14,A4,A13,A11)
    a16<-c(A16,A5,A10,A3)
  }
  if(niv_filtre==0.2){
    a2<-c(A2,A3,A4);                  a3<-c(A3,A2,A4,A5)
    a4<-c(A4,A11,A2,A3,A13,A8,A14);  a5<-c(A5,A16,A3)
    a6<-c(A6,A7,A9,A10);              a7<-c(A7,A9,A6)
    a8<-c(A8,A10,A11,A4);              a9<-c(A9,A7,A6)
    a10<-c(A10,A6,A8,A11,A16);        a11<-c(A11,A4,A8,A13)
    a13<-c(A13,A4,A11,A14);           a14<-c(A14,A4,A13)
    a16<-c(A16,A5,A10)
  }
  if(niv_filtre==0.3){
    a2<-c(A2,A3,A4);                  a3<-c(A3,A2,A4,A5)
```

```

a4<-c(A4,A11,A2,A3,A13);
a6<-c(A6,A7,A9,A10);
a8<-c(A8,A10,A11);
a10<-c(A10,A6,A8,A11);
a13<-c(A13,A4);
}
if(niv_filtre==0.4){
a2<-c(A2,A3,A4);
a4<-c(A4,A11,A2);
a6<-c(A6,A7,A9,A10);
a9<-c(A9,A7,A6);
a11<-c(A11,A4);
}
if(niv_filtre==0.5){
a2<-c(A2,A3);
a4<-c(A4,A11);
a7<-c(A7,A9);
a11<-c(A11,A4);
}
if(niv_filtre==0.6){
a7<-c(A7,A9);
}

return( list( a1=a1, a2=a2, a3=a3, a4=a4, a5=a5, a6=a6, a7=a7, a8=a8, a9=a9,
a10=a10, a11=a11, a12=a12, a13=a13, a14=a14, a15=a15, a16=a16) )
}

```

	Pas de sol	Alluvions	Colluvions	Sables	Limons	Marnes	Calcaires	Molasses	Craie	Argiles	Sables et argiles	Tourbe	Roches cristallines	Roches métamorphiques	Autres	Limons sur argile ou silex
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16
A1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	1	0,5	0,4	0,1	0	0	0	0	0,1	0,1	0	0	0	0	0
A3	0	0,5	1	0,3	0,3	0	0	0	0	0,1	0,1	0	0	0	0	0,1
A4	0	0,4	0,3	1	0	0	0	0,2	0	0	0,5	0	0,3	0,2	0	0
A5	0	0,1	0,3	0	1	0	0	0	0	0	0	0	0	0	0	0,5
A6	0	0	0	0	0	1	0,4	0	0,4	0,4	0	0	0	0	0	0
A7	0	0	0	0	0	0,4	1	0	0,6	0	0	0	0	0	0	0
A8	0	0	0	0,2	0	0	0	1	0	0,3	0,3	0	0	0	0	0
A9	0	0	0	0	0	0,4	0,6	0	1	0	0	0	0	0	0	0
A10	0	0,1	0,1	0	0	0,4	0	0,3	0	1	0,3	0	0	0	0	0,2
A11	0	0,1	0,1	0,5	0	0	0	0,3	0	0,3	1	0	0,2	0,1	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A13	0	0	0	0,3	0	0	0	0	0	0	0,2	0	1	0,2	0	0
A14	0	0	0	0,2	0	0	0	0	0	0	0,1	0	0,2	1	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5	0
A16	0	0	0,1	0	0,5	0	0	0	0	0,2	0	0	0	0	0	1

Annexe 2 : Similarité géologique entre les valeurs de la variable geo_50k

2-2 Fonction de filtrage :

filtrage des similarités selon les variables srtm, ast, gamma occup et geo_50k

Données : data frame qu'on observe

niv_filtre_ : niveau de similarité géologique minimum qu'on accepte

perc : pourcentage de profils proches qu'on conserve

```
filtrage_similarites <- fonction(donnees, niv_filtre, perc) {
```

```
library(grDevices)
```

```
library(cluster)
```

```
library(ggplot2)
```

```
xsep <- 600993#mesure de séparation entre la partie Ouest et la partie Est
```

```
Est <- subset( donnees,x >= xsep ) #on réordonne le jeu de données en une
```

```
Ouest <- subset( donnees, x < xsep ) # partie Est et Ouest bien séparées
```

```
Centre <- rbind( Ouest, Est )
```

```
fi <- filtre(niv_filtre, Centre)
```

```
n0 <- dim(Ouest)[1]
```

```

nE <- dim(Est)[1]

d_srtm <- daisy( Centre[ , c(10:40) ], metric = "gower" ) #calcul des indices
de dissimilarité selon les données srtm
M_srtm <- as.matrix( d_srtm ) #on met ce résultat sous forme de
matrice
MOE_srtm <- M_srtm[ c( 1:nO ), c( (nO+1):dim(Centre)[1] ) ] #on
sélectionne la sous-matrice des indices entre l'ouest et l'est
Msrtm <- MOE_srtm #on copie cette matrice pour pouvoir récupérer
les valeurs qui nous intéressent après le tri
MOE_srtm <- t( apply( MOE_srtm, 1, order ) ) #on trie cette matrice par
ligne, en ordre croissant

d_ast <- daisy( Centre[ , c( 41:71 ) ], metric = "gower" ) #calcul des indices
de dissimilarité selon les données ast
M_ast <- as.matrix( d_ast )
MOE_ast <- M_ast[ c( 1:nO ), c( (nO+1):dim(Centre)[1] ) ]
Mast <- MOE_ast
MOE_ast <- t( apply( MOE_ast, 1, order ) )

d_gamma <- daisy( Centre[ , c( 79 : 83 ) ], metric = "gower" ) #calcul des
indices de dissimilarité selon les données gamma
M_gamma <- as.matrix( d_gamma )
MOE_gamma <- M_gamma[ c( 1:nO ), c( (nO+1):dim(Centre)[1] ) ]
Mgamma <- MOE_gamma
MOE_gamma <- t( apply( MOE_gamma, 1, order ) )

d_occup <- daisy( Centre[ , c( 85:89 ) ], metric = "gower" ) #calcul des
indices de dissimilarité selon les données d'occupation
M_occup <- as.matrix( d_occup )
MOE_occup <- M_occup[ c( 1:nO ), c( (nO+1):dim(Centre)[1] ) ]
Moccup <- MOE_occup
MOE_occup <- t( apply( MOE_occup, 1, order ) )

d_pedo <- daisy( Centre[ , c(4:6, 97:110) ], metric = "gower" ) #calcul des
indices de dissimilarité selon les données pédologiques
M_pedo <- as.matrix( d_pedo )
M_pedo <- M_pedo[ c( 1:nO ), c( (nO+1):dim(Centre)[1] ) ]

col <- dim( MOE_srtm )[2] #nombre de colonnes (le même pour toutes les matrices
MOE_ )
lim <- round( col*perc, 0 ) #nombre de colonnes correspondant au pourcentage
choisi

#on récupère les perc % des parcelles les plus proches pour chaque groupe de
variables
MOE_p_srtm <- cbind( Ouest$id_profil, MOE_srtm[ , 1:lim ] ) #on associe
l'id_profil des pixels Ouest à l'indice de dissimilarité?
#entre ce pixel et
les pixels Est
MOE_p_srtm[ , -1 ] <- Est$id_profil[ MOE_p_srtm[ , -1 ] ] #on remplace
l'indice par l'id_profil du pixel auquel il correspond

MOE_p_ast <- cbind( Ouest$id_profil, MOE_ast[ , 1:lim ] )
MOE_p_ast[ , -1 ] <- Est$id_profil[ MOE_p_ast[ , -1 ] ]

MOE_p_gamma <- cbind( Ouest$id_profil, MOE_gamma[ , 1:lim ] )
MOE_p_gamma[ , -1 ] <- Est$id_profil[ MOE_p_gamma[ , -1 ] ]

MOE_p_occup <- cbind( Ouest$id_profil, MOE_occup[ , 1:lim ] )
MOE_p_occup[ , -1 ] <- Est$id_profil[ MOE_p_occup[ , -1 ] ]

p <- matrix(NA, nrow = dim( MOE_p_srtm )[1], ncol = dim( MOE_p_srtm )[2] )
m <- p # on crée 3 matrices vides de la même taille que MOE_p_srtm
ped <- p

for(i in 1:dim( MOE_p_srtm )[1] ){ #on fait l'intersection des lignes de
chacune des 4 matrices,

```

```

v1 <- intersect( MOE_p_srtm[ i, ], MOE_p_ast[ i, ] ) #ce qui va nous donner
les parcelles qui seront
v2 <- intersect( MOE_p_gamma[ i, ], MOE_p_occup[ i, ] )# proches pour chaque
groupe de variables
int <- intersect( v1, v2 )
f <- Centre$geo_50k[ which( Centre$id_profil == int[1] ) ]#on récupère le
type de filtre

if(f == "a1" )fil = fi$a1;   if(f == "a2" )fil = fi$a2;
if(f == "a3" )fil = fi$a3;   if(f == "a4" )fil = fi$a4
if(f == "a5" )fil = fi$a5;   if(f == "a6" )fil = fi$a6;
if(f == "a7" )fil = fi$a7;   if(f == "a8" )fil = fi$a8
if(f == "a9" )fil = fi$a9;   if(f == "a10" )fil = fi$a10;
if(f == "a11" )fil = fi$a11; if(f == "a12" )fil = fi$a12
if(f == "a13" )fil = fi$a13; if(f == "a14" )fil = fi$a14;
if(f == "a15" )fil = fi$a15; if(f == "a16" )fil = fi$a16
p[ i, 1 ] <- int[1] # numéros des pixels proches
m[ i, 1 ] <- int[1] # moyennes des indices srtm, ast, gamma et occup
ped[ i, 1 ] <- int[1] # indice de Gower pour les variables pédologiques
int2 <- intersect( int, fil ) #int2 est la liste des pixels proches d'un
pixel Ouest dont la valeur geo_50k correspond
if( length(int2) > 1){ #on rentre dans la boucle si on a de quoi remplir
la matrice
  for(j in 2:length( int2 ) ){
    p[ i, j ] <- int2[j] #on place ces valeurs dans p
    ind <- which( Est$id_profil == p[ i, j ] )
    m[ i, j ] <- mean( c( Msrtm[ i, ind ], #dans m on place la moyenne des 4
indices de dissimilarité
                        Mast[ i, ind ],
                        Mgamma[ i, ind ],
                        Moccup[ i, ind ] ), na.rm = TRUE)
    ped[ i, j ] <- M_pedo[ i, ind ] #dans ped on place l'indice de
dissimilarité entre les données pédologiques
  }
}
}
write.table(p, "results/pixel_filtre/pixels_voisin.csv", sep=";")
write.table(m, "results/pixel_filtre/moyenne_indices.csv", sep=";")
write.table(ped, "results/pixel_filtre/indices_pedo.csv", sep=";")
return( list(p=p, m=m, ped=ped) )
}

```

2-3 Code de l'affectation par ACP

```

#nombre de composantes principales qu'on veut garder
n <- 20

library(FactoMineR)
acp <- PCA(donnees[c(10:83, 85:89)], ncp=n, graph=FALSE)
#les variables concernées par l'ACP sont les variables externes quantitatives

#affichage du cercle de corrélation selon les différents axes principaux
plot(acp, axes=c(1, 2), choix="var", cex=0.3)
plot(acp, axes=c(1, 3), choix="var", cex=0.3)
plot(acp, axes=c(1, 4), choix="var", cex=0.3)
plot(acp, axes=c(1, 5), choix="var", cex=0.3)
plot(acp, axes=c(1, 6), choix="var", cex=0.3)
plot(acp, axes=c(1, 7), choix="var", cex=0.3)
plot(acp, axes=c(1, 8), choix="var", cex=0.3)
plot(acp, axes=c(1, 9), choix="var", cex=0.3)
plot(acp, axes=c(1, 10), choix="var", cex=0.3)
plot(acp, axes=c(1, 11), choix="var", cex=0.3)
plot(acp, axes=c(1, 12), choix="var", cex=0.3)
plot(acp, axes=c(1, 13), choix="var", cex=0.3)
plot(acp, axes=c(1, 14), choix="var", cex=0.3)

```



```

plot(acp, axes=c(1, 15), choix="var", cex=0.3)
plot(acp, axes=c(1, 16), choix="var", cex=0.3)
plot(acp, axes=c(1, 17), choix="var", cex=0.3)
plot(acp, axes=c(1, 18), choix="var", cex=0.3)
plot(acp, axes=c(1, 19), choix="var", cex=0.3)
plot(acp, axes=c(1, 20), choix="var", cex=0.3)

#éboulis de valeurs propres
barplot(acp$seig[, 1], main="Eigenvalues", names.arg=1:nrow(acp$seig))

#l'ACP projette les observations sur un plan orthonormé on peut donc calculer
leur distance euclidiennes

individus <- cbind(donnees$id_profil, acp$ind$coord)

Est <- individus[ which( donnees$x >= 600993), ] #on réordonne le jeu de données
Ouest <- individus[ which( donnees$x < 600993), ] # en une partie Est et une
Centre <- rbind(Ouest, Est) # partie Ouest distincte

M <- dist(Centre[, -1], method="euclidean") # calcul de la distance entre les
M <- as.matrix(M) # points
M <- M[1:1143, 1144:2503] #on sélectionne la partie Est/Ouest de cette matrice

PProche <- matrix(NA, dim(Ouest)[1], 3) #matrice du résultat
PProche[, 1] <- Ouest[, 1] #on y place l'id_profil des pixels Ouest
for (i in 1:dim(Ouest)[1]) { #pour chaque pixel Ouest on récupère le plus proche
  mi <- which(M[i, ] == min(M[i, ])) [1] #pixel Est ainsi que la distance entre
  PProche[i, 2] <- Est[mi, 1] # ces 2 pixels
  PProche[i, 3] <- M[i, mi]
}
colnames(PProche) <- c("id_Ouest", "id_Est", "distance")

```

2-4 Normalisation et standardisation des variables

```

setwd("D:/crabillon/Homosoil/traitement/data")
data <- read.csv("source/profil_covariables3.csv", sep=';', header=TRUE)
colnames(data) #on repere quelle variables ont trop de NA et on les supprime
N <- rep(0, length(data))
for(i in 1:length(data))
  N[i] <- sum(is.na(data[, i])); N

data <- data[, -94]
data <- data[complete.cases(data), ]

non_norm <- c(11, 12, 17, 18, 23, 24, 29, 30, 35, 36, #var à log_normaliser
             41, 42, 43, 48, 49, 54, 55, 60, 61, 66, 67,
             73, 74, 76, 77, 78,
             85, 86, 87, 88, 89)
var_quali <- c(1, 2, 3, 84, 93, 94, 95) #var pas touché
data_norm <- data.frame( data[1:3])

for( i in 4:length( data ) ) { #pour chaque variable
  b <- data[i] #
  if (!any( var_quali == i )) { #si elle n'est pas normale
    if( any( non_norm == i ) ) { #et pas qualitative
      c <- 0 #
      if( any( b < 0 ) ) c <- -min(b) # si elle prend des valeurs négatives
      b <- log(b + c) # on ajoute le min dans le log
    }
    aver <- mean( b[ b != "-Inf" ] , na.rm=TRUE) #on calcule la moyenne
    std <- sd( b[ b != "-Inf" ] , na.rm=TRUE) #l'écart type
    b <- ( b - aver ) / std #et on normalise
  }
}

```

```

data_norm <- data.frame(data_norm, b) #on place cette variable
                                     # log_normalisée dans le nouveau data
frame
}
data_norm[data_norm=="-Inf"]<-NA      #le log peut générer des -Inf donc
colnames(data_norm ) <- colnames(data) #on les remplace par des NA
write.table(data_norm, "source/log_norm.csv", sep=";")

#on affiche les différents scatterplots pour voir si les différentes variables
#ont bien été normalisées.
library(car)

#srtm
scatterplotMatrix(data_norm[,c(4:6,10,72:78)],
                  diagonal="histogram",      smooth=FALSE)
#srtm 3
scatterplotMatrix(data_norm[,c(4:6,11:16)],
                  diagonal="histogram",      smooth=FALSE)
#srtm 5
scatterplotMatrix(data_norm[,c(4:6,17:22)],
                  diagonal="histogram",      smooth=FALSE)
#srtm 9
scatterplotMatrix(data_norm[,c(4:6,23:28)],
                  diagonal="histogram",      smooth=FALSE)
#srtm 17
scatterplotMatrix(data_norm[,c(4:6,29:34)],
                  diagonal="histogram",      smooth=FALSE)
#srtm 33
scatterplotMatrix(data_norm[,c(4:6,35:40)],
                  diagonal="histogram",      smooth=FALSE)
#ast 3
scatterplotMatrix(data_norm[,c(4:6,41:47)],
                  diagonal="histogram",      smooth=FALSE)
#ast 5
scatterplotMatrix(data_norm[,c(4:6,48:53)],
                  diagonal="histogram",      smooth=FALSE)
#ast 9
scatterplotMatrix(data_norm[,c(4:6,54:59)],
                  diagonal="histogram",      smooth=FALSE)
#ast 17
scatterplotMatrix(data_norm[,c(4:6,60:65)],
                  diagonal="histogram",      smooth=FALSE)
#ast 33
scatterplotMatrix(data_norm[,c(4:6,66:71)],
                  diagonal="histogram",      smooth=FALSE)
#gamma
scatterplotMatrix(data_norm[,c(4:6,79:83)],
                  diagonal="histogram",      smooth=FALSE)
#occup
scatterplotMatrix(data_norm[,c(4:6,85:89)],
                  diagonal="histogram",      smooth=FALSE)

```

2-5 Multiplot :

Fonction qui permet d'afficher plusieurs graphiques dans la même fenêtre

```

multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

```

```

# If layout is NULL, then use 'cols' to determine layout
if (is.null(layout)) {
  # Make the panel
  # ncol: Number of columns of plots
  # nrow: Number of rows needed, calculated from # of cols
  layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                    ncol = cols, nrow = ceiling(numPlots/cols))
}

if (numPlots==1) {
  print(plots[[1]])
} else {
  # Set up the page
  grid.newpage()
  pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

  # Make each plot, in the correct location
  for (i in 1:numPlots) {
    # Get the i,j matrix positions of the regions that contain this subplot
    matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

    print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                     layout.pos.col = matchidx$col))
  }
}
}

```

2-6 Code du clustering

```

require(Rmixmod)
library(ggplot2)
library(FactoMineR)
data <- read.csv("source/log_norm.csv", sep=";", header=TRUE)
data <- data[,-c(111,112)] #on supprime les deux dernières variables qui sont
ajoutées par ce script

data$side <- ifelse(data$x < 600993, "O", "E")
cote <- cbind(data$id_profil, data$side)
ind_W <- cote[cote[,2]=="O",]
ind_E <- cote[cote[,2]=="E",]

n <- 20
acp <- PCA(data[c(10:83,85:89)],ncp=n, graph=FALSE)

data_clust <- data.frame(acp$ind$coord)
cc <- complete.cases(data_clust)
ind <- data[cc,1]
data_clust <- data_clust[complete.cases(data_clust),]
##### define strategy
strat.soil2 <- mixmodStrategy(algo="SEM",
                             nbTry=2,
                             initMethod="SEMMax",
                             nbTryInInit=150,
                             nbIterationInInit=100,
                             nbIterationInAlgo=200,
                             epsilonInAlgo=NaN,
                             seed=1)
#####perform clustering
SEM_soil1 <- mixmodCluster(data_clust,
                           nbCluster=2:10,
                           dataType="quantitative",
                           strategy=strat.soil2,
                           criterion=c("BIC", "NEC"))
)
gc()

```

```

sort_sem <- sortByCriterion(SEM_soil1,"NEC")

# on affiche le nombre de cluster en fonction du critère NEC
resu <- c()
for (i in 1:9) resu <- rbind(resu,c(sort_sem@results[[i]]@criterionValue[2],
                                sort_sem@results[[i]]@nbCluster))
plot(resu[,2],resu[,1], xlab="Nombre de clusters", ylab="NEC")

### extract clusters to data
clust_result <- sort_sem@results[[2]]@partition
mixmod_result <- data.frame(cbind(ind, clust_result))
names(mixmod_result) <- c("id_profil", "cluster")
d <- cbind(data[cc,], mixmod_result$cluster)
names(d) <- c(names(d)[-113], "cluster")

list of pixels grouped by cluster
clus1 <- d$id_profil[ d$cluster == 1]
clus2 <- d$id_profil[ d$cluster == 2]
clus3 <- d$id_profil[ d$cluster == 3]
clus4 <- d$id_profil[ d$cluster == 4]
clus5 <- d$id_profil[ d$cluster == 5]
clus6 <- d$id_profil[ d$cluster == 6]
clus7 <- d$id_profil[ d$cluster == 7]

```

2-7 Affichage des pixels sur la carte, groupés par cluster

```

setwd("D:/crabillon/Homosoil/traitement/data/source/departement")

require("rgdal")
require("maptools")
require("ggplot2")
require("plyr")
gpclibPermit() # required for fortify method

soil.poly <- readOGR(dsn=".", "centre")
ogrInfo(dsn=".", "centre")
soil.poly <- spTransform(soil, CRS("+init=epsg:2154"))

soil.poly@data$id = (soil.poly@data$CODE_DEPT)
soil.poly.points = fortify(soil.poly, region="id")
soil.poly.df = join(soil.poly.points, soil.poly@data, by="id")

for (i in 1:64) soil.poly.df <- rbind(soil.poly.df, rep(NA,18))

soil.poly.df <- cbind(soil.poly.df, d$x,d$y, d$cluster)

dept <- ggplot(soil.poly.df) +
  aes(long,lat,group=group) +
  geom_polygon(colour = "black", fill = "grey75") +
  coord_equal() +
  scale_color_discrete("soil.poly Ecoregion") +
  geom_point(aes(x=d$x, y=d$y,col=factor(d$cluster)))

plot(dept)

```

2-8 Indices de qualité des modèles

```

dat <- clus_i #on choisit le cluster sur lequel on travaille

#si un pixel n'a pas de valeur pour geo_50k, on le supprime
dat$side <- ifelse(dat$x < 600993, "O", "E")
if( sum( is.na( dat[ , 96 ] ) ) > 0 ) dat <- dat[ -which(is.na(dat[,96])), ]
dat_O <- dat[dat$side == "O",]
dat_E <- dat[dat$side == "E",]
dat <- rbind(dat_O,dat_E)
nO <- dim(dat_O)[1]+1; n <- dim(dat)[1]

#on récupère nos trois matrices: m, p et ped
res<-filtrage_similarites(dat, 0.3,0.4)

#on récupère le plus proche voisin
min_filtre <- rep(NA, dim(res$m)[1])
for(i in 1:dim(res$m)[1]){
  if (!is.na(res$m[i,2])) min_filtre[i] <- res$p[i,which.min(res$m[i,])]
}

#pour chaque variable pédologique, on récupère la valeur
#du plus proche voisin

#####argile surface
Arg_X <- dat_O$argile; Arg_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Arg_Y[i] <- dat_E$argile[which(dat_E$id_profil ==
min_filtre[i])]
}
#####limon surface
Lim_X <- dat_O$limon; Lim_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Lim_Y[i] <- dat_E$limon[which(dat_E$id_profil ==
min_filtre[i])]
}
#####sable surface
Sab_X <- dat_O$sable; Sab_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Sab_Y[i] <- dat_E$sable[which(dat_E$id_profil ==
min_filtre[i])]
}
#####argile 5_15
Arg_5_15_X <- dat_O$argile_5_15; Arg_5_15_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Arg_5_15_Y[i] <- dat_E$argile_5_15[which(dat_E$id_profil ==
min_filtre[i])]
}
#####limon 5_15
Lim_5_15_X <- dat_O$limon_5_15; Lim_5_15_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Lim_5_15_Y[i] <- dat_E$limon_5_15[which(dat_E$id_profil ==
min_filtre[i])]
}
#####sable 5_15
Sab_5_15_X <- dat_O$sable_5_15; Sab_5_15_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Sab_5_15_Y[i] <- dat_E$sable_5_15[which(dat_E$id_profil ==
min_filtre[i])]
}
#####argile 15_30
Arg_15_30_X <- dat_O$argile_15_30; Arg_15_30_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){

```

```

M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Arg_15_30_Y[i] <- dat_E$argile_15_30[which(dat_E$id_profil
== min_filtre[i])]
}
#####limon 15_30
Lim_15_30_X <- dat_O$limon_15_30; Lim_15_30_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Lim_15_30_Y[i] <- dat_E$limon_15_30[which(dat_E$id_profil ==
min_filtre[i])]
}
#####sable 15_30
Sab_15_30_X <- dat_O$sable_15_30; Sab_15_30_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Sab_15_30_Y[i] <- dat_E$sable_15_30[which(dat_E$id_profil ==
min_filtre[i])]
}
#####argile 30_60
Arg_30_60_X <- dat_O$argile_30_60; Arg_30_60_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Arg_30_60_Y[i] <- dat_E$argile_30_60[which(dat_E$id_profil
== min_filtre[i])]
}
#####limon 30_60
Lim_30_60_X <- dat_O$limon_30_60; Lim_30_60_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Lim_30_60_Y[i] <- dat_E$limon_30_60[which(dat_E$id_profil ==
min_filtre[i])]
}
#####sable 30_60
Sab_30_60_X <- dat_O$sable_30_60; Sab_30_60_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Sab_30_60_Y[i] <- dat_E$sable_30_60[which(dat_E$id_profil ==
min_filtre[i])]
}
#####argile 60_100
Arg_60_100_X <- dat_O$argile_60_100; Arg_60_100_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Arg_60_100_Y[i] <- dat_E$argile_60_100[which(dat_E$id_profil
== min_filtre[i])]
}
#####limon 60_100
Lim_60_100_X <- dat_O$limon_60_100; Lim_60_100_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Lim_60_100_Y[i] <- dat_E$limon_60_100[which(dat_E$id_profil
== min_filtre[i])]
}
#####sable 60_100
Sab_60_100_X <- dat_O$sable_60_100; Sab_60_100_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Sab_60_100_Y[i] <- dat_E$sable_60_100[which(dat_E$id_profil
== min_filtre[i])]
}
#####argile 100_200
Arg_100_200_X <- dat_O$argile_100_200; Arg_100_200_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
M <- res$m[i,!is.na(res$m[i,])]
if(length(M) > 1) Arg_100_200_Y[i] <-
dat_E$argile_100_200[which(dat_E$id_profil == min_filtre[i])]
}
#####limon 100_200

```

```

Lim_100_200_X <- dat_O$limon_100_200; Lim_100_200_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Lim_100_200_Y[i] <-
dat_E$limon_100_200[which(dat_E$id_profil == min_filtre[i])]
}
#####sable 100_200
Sab_100_200_X <- dat_O$sable_100_200; Sab_100_200_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) Sab_100_200_Y[i] <-
dat_E$sable_100_200[which(dat_E$id_profil == min_filtre[i])]
}
#####Ph
PH_X <- dat_O$ph; PH_Y <- rep(NA, dim(res$m)[1])
for (i in 1:dim(res$m)[1]){
  M <- res$m[i,!is.na(res$m[i,])]
  if(length(M) > 1) PH_Y[i] <- dat_E$ph[which(dat_E$id_profil == min_filtre[i])]
}
#####on place dans une grande matrice la valeur des
variables pédologiques de chaque pixel ainsi que celles de son plus proche
voisin
comp <- data.frame(
  Arg_X, Arg_Y, Arg_5_15_X, Arg_5_15_Y, Arg_15_30_X,
  Arg_15_30_Y, Arg_30_60_X, Arg_30_60_Y, Arg_60_100_X,
  Arg_60_100_Y, Arg_100_200_X, Arg_100_200_Y,
  Lim_X, Lim_Y, Lim_5_15_X, Lim_5_15_Y, Lim_15_30_X,
  Lim_15_30_Y, Lim_30_60_X, Lim_30_60_Y, Lim_60_100_X,
  Lim_60_100_Y, Lim_100_200_X, Lim_100_200_Y,
  Sab_X, Sab_Y, Sab_5_15_X, Sab_5_15_Y, Sab_15_30_X,
  Sab_15_30_Y, Sab_30_60_X, Sab_30_60_Y, Sab_60_100_X,
  Sab_60_100_Y, Sab_100_200_X, Sab_100_200_Y, PH_X, PH_Y
)
comp <- comp[complete.cases(comp),]

#matrice contenant les résultats du MPE
MPE <- matrix(c( "Profondeur","Argile","Limon","Sable",
  "0_5",
  mean(comp[,1] - comp[,2]),
  mean(comp[,13] - comp[,14]),
  mean(comp[,25] - comp[,26]),
  "5_15",
  mean(comp[,3] - comp[,4]),
  mean(comp[,15] - comp[,16]),
  mean(comp[,27] - comp[,28]),
  "15_30",
  mean(comp[,5] - comp[,6]),
  mean(comp[,17] - comp[,18]),
  mean(comp[,29] - comp[,30]),
  "30_60",
  mean(comp[,7] - comp[,8]),
  mean(comp[,19] - comp[,20]),
  mean(comp[,31] - comp[,32]),
  "60_100",
  mean(comp[,9] - comp[,10]),
  mean(comp[,21] - comp[,22]),
  mean(comp[,33] - comp[,34]),
  "100_200",
  mean(comp[,11] - comp[,12]),
  mean(comp[,23] - comp[,24]),
  mean(comp[,35] - comp[,36]) ),
  ncol=7)

#Matrice contenant les résultats du RMSE des données pédologiques
RMSE <- matrix( c( "Profondeur","Argile","Limon","Sable",
  "0_5",
  rmse(comp[,1], comp[,2]),

```

```

rmse(comp[,13], comp[,14]),
rmse(comp[,25], comp[,26]),
"5_15",
rmse(comp[,3], comp[,4]),
rmse(comp[,15], comp[,16]),
rmse(comp[,27], comp[,28]),
"15_30",
rmse(comp[,5], comp[,6]),
rmse(comp[,17], comp[,18]),
rmse(comp[,29], comp[,30]),
"30_60",
rmse(comp[,7], comp[,8]),
rmse(comp[,19], comp[,20]),
rmse(comp[,31], comp[,32]),
"60_100",
rmse(comp[,9], comp[,10]),
rmse(comp[,21], comp[,22]),
rmse(comp[,33], comp[,34]),
"100_200",
rmse(comp[,11], comp[,12]),
rmse(comp[,23], comp[,24]),
rmse(comp[,35], comp[,36]) ),
ncol=7)

```

#Matrice contenant les résultats du RPD des données pédologiques

```

RPD <- matrix( c( "Profondeur", "Argile", "Limon", "Sable",
"0_5",
sd(comp[,2], na.rm=TRUE)/rmse(comp[,1], comp[,2]),
sd(comp[,14], na.rm=TRUE)/rmse(comp[,13], comp[,14]),
sd(comp[,26], na.rm=TRUE)/rmse(comp[,25], comp[,26]),
"5_15",
sd(comp[,4], na.rm=TRUE)/rmse(comp[,3], comp[,4]),
sd(comp[,16], na.rm=TRUE)/rmse(comp[,15], comp[,16]),
sd(comp[,28], na.rm=TRUE)/rmse(comp[,27], comp[,28]),
"15_30",
sd(comp[,6], na.rm=TRUE)/rmse(comp[,5], comp[,6]),
sd(comp[,18], na.rm=TRUE)/rmse(comp[,17], comp[,18]),
sd(comp[,30], na.rm=TRUE)/rmse(comp[,29], comp[,30]),
"30_60",
sd(comp[,8], na.rm=TRUE)/rmse(comp[,7], comp[,8]),
sd(comp[,20], na.rm=TRUE)/rmse(comp[,19], comp[,20]),
sd(comp[,32], na.rm=TRUE)/rmse(comp[,31], comp[,32]),
"60_100",
sd(comp[,10], na.rm=TRUE)/rmse(comp[,9], comp[,10]),
sd(comp[,22], na.rm=TRUE)/rmse(comp[,21], comp[,22]),
sd(comp[,34], na.rm=TRUE)/rmse(comp[,33], comp[,34]),
"100_200",
sd(comp[,12], na.rm=TRUE)/rmse(comp[,11], comp[,12]),
sd(comp[,24], na.rm=TRUE)/rmse(comp[,23], comp[,24]),
sd(comp[,36], na.rm=TRUE)/rmse(comp[,35], comp[,36]) ),
ncol=7)

```

#matrice du R?

```

R2 <- matrix( c( "Profondeur", "Argile", "Limon", "Sable",
"0_5",
cor(comp[,1], comp[,2])^2,
cor(comp[,3], comp[,4])^2,
cor(comp[,5], comp[,6])^2,
"5_15",
cor(comp[,7], comp[,8])^2,
cor(comp[,9], comp[,10])^2,
cor(comp[,11], comp[,12])^2,
"15_30",
cor(comp[,13], comp[,14])^2,
cor(comp[,15], comp[,16])^2,
cor(comp[,17], comp[,18])^2,
"30_60",
cor(comp[,19], comp[,20])^2,

```



```

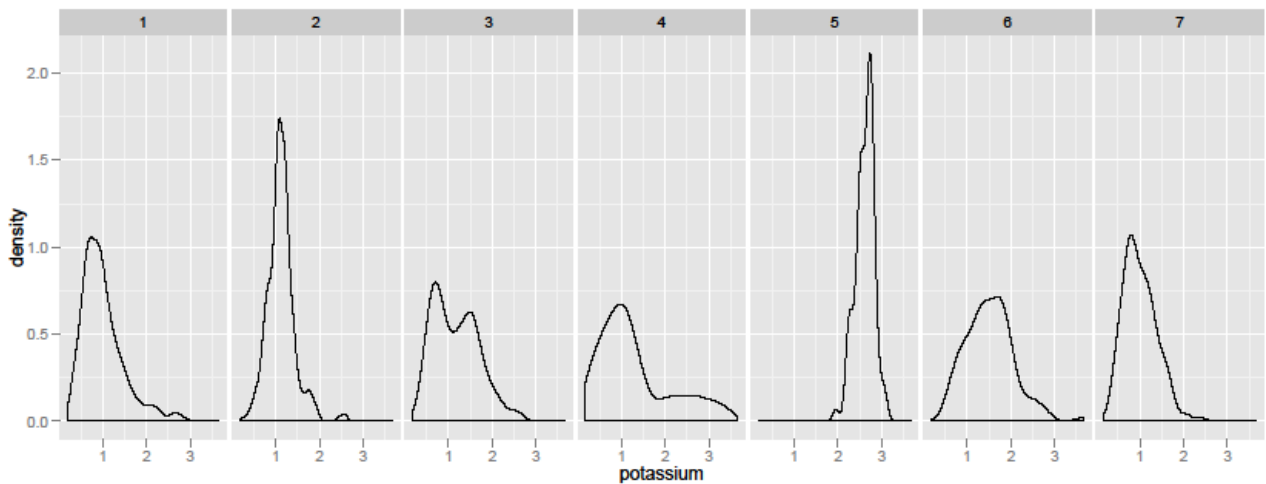
cor(comp[,21], comp[,22])^2,
cor(comp[,23], comp[,24])^2,
"60_100",
cor(comp[,25], comp[,26])^2,
cor(comp[,27], comp[,28])^2,
cor(comp[,29], comp[,30])^2,
"100_200",
cor(comp[,31], comp[,32])^2,
cor(comp[,33], comp[,34])^2,
cor(comp[,35], comp[,36])^2),
ncol=7)

##### Affichage de la valeur des variables pédologiques du pixel cible en
fonction de celle de son plus proche voisin
f1 <- ggplot(dat, aes( x=comp[,1], y=comp[,2]) ) +
  geom_point( size = 0.5 ) +
  labs(x="Argile Ouest",y="Argile Est") +
  theme(legend.position="null") +
  stat_smooth( method = lm ) +
  ggtitle("Argile en surface")

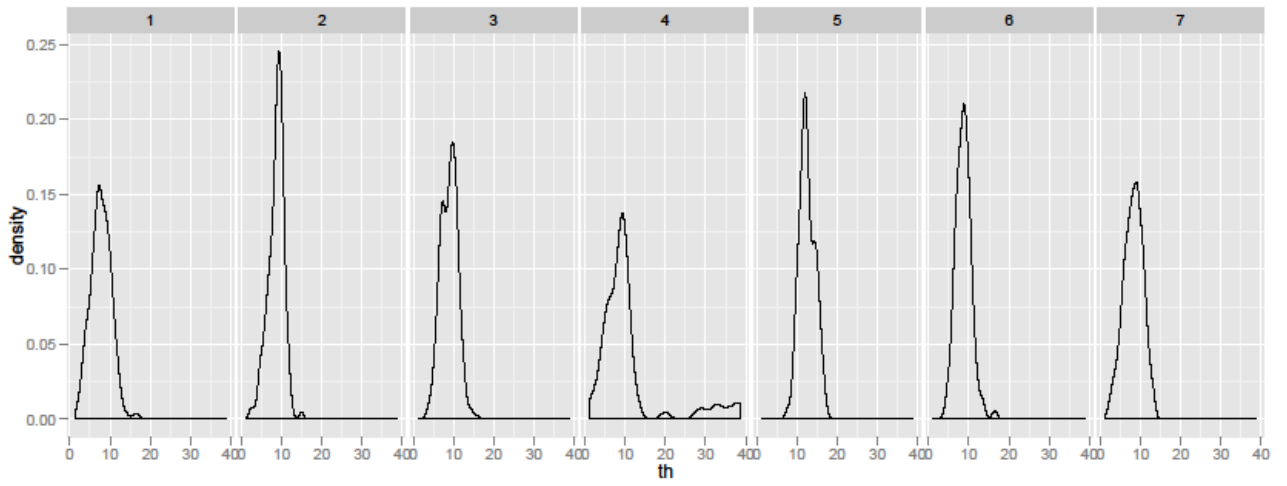
```

3 Résultats

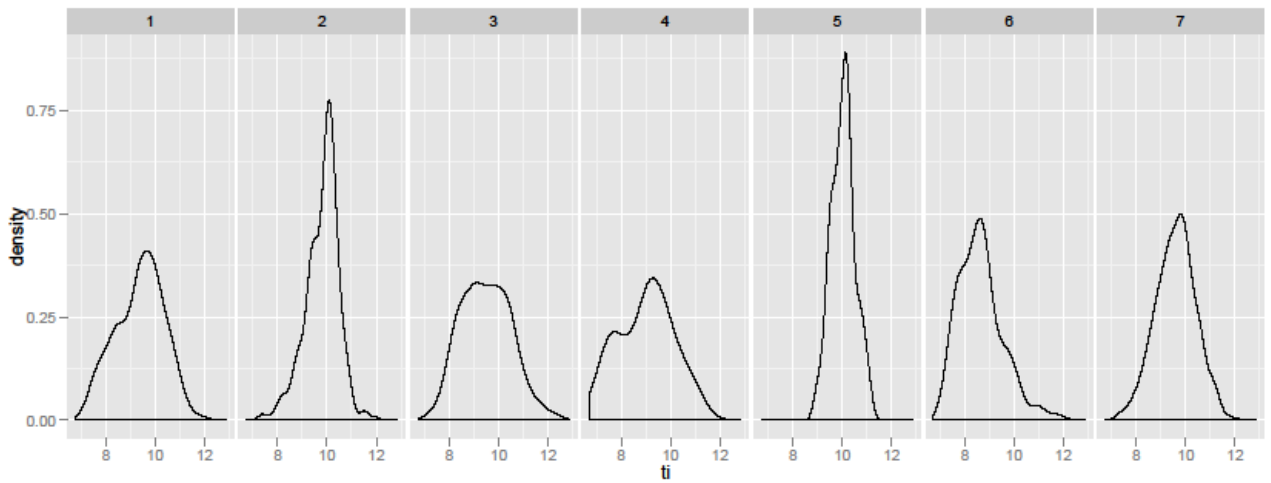
3-1 Densité des variables par cluster



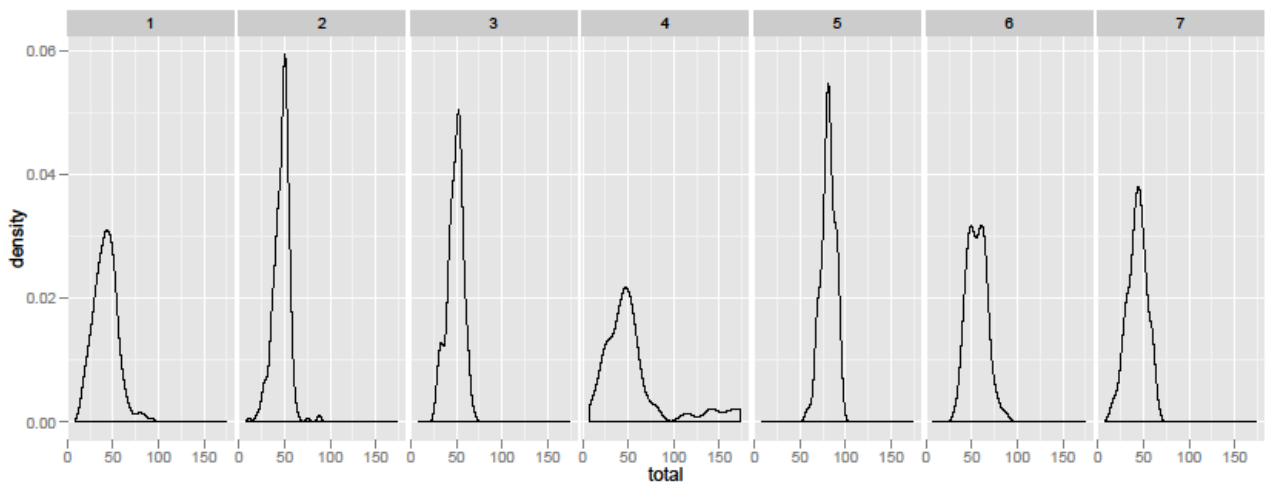
Variable 1 : Potassuim



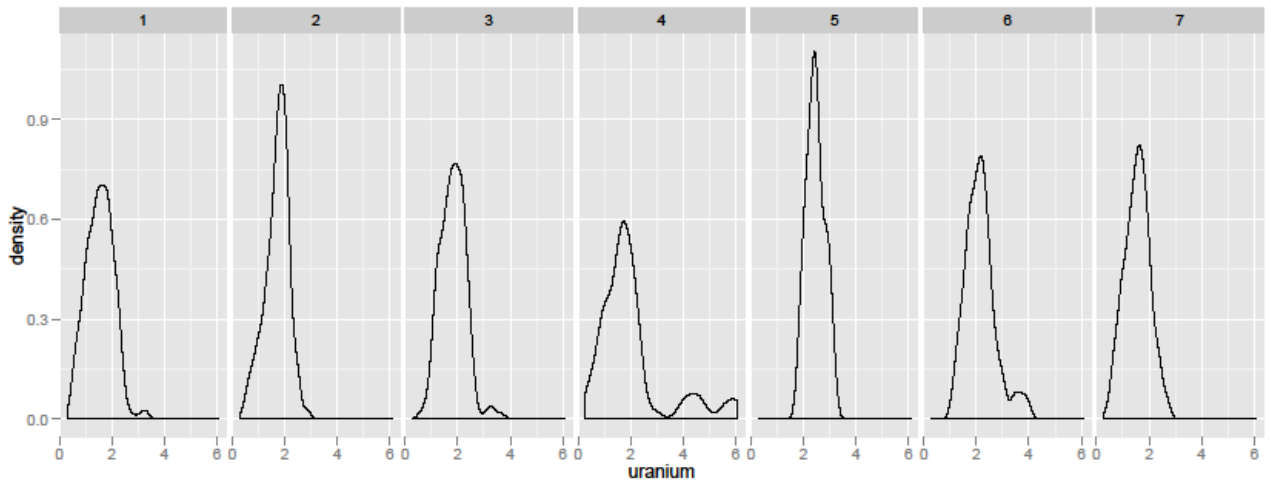
Variable 2: Thorium



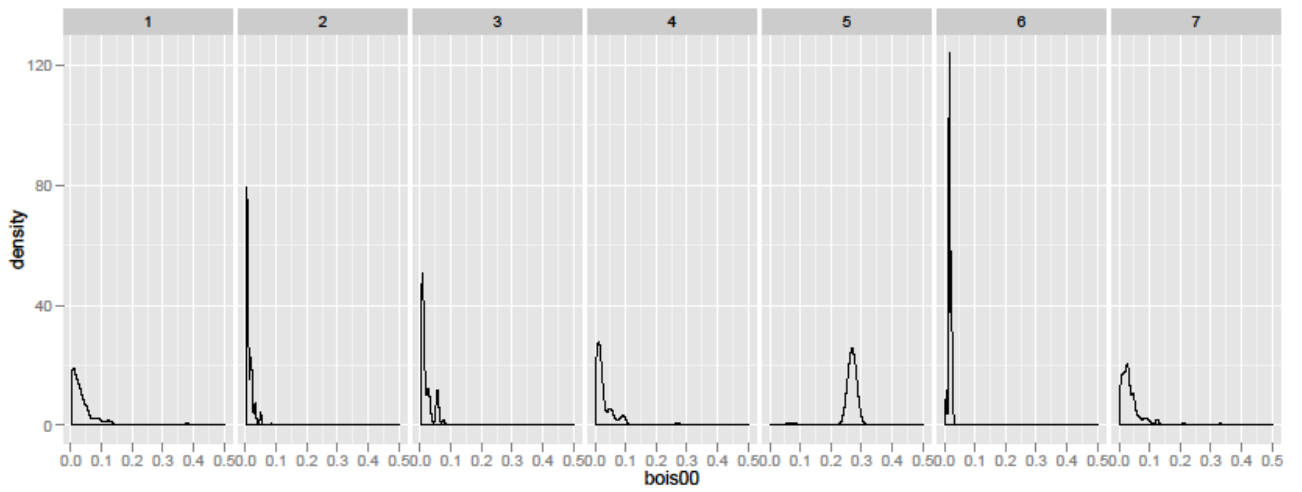
Variable 3 : Titane



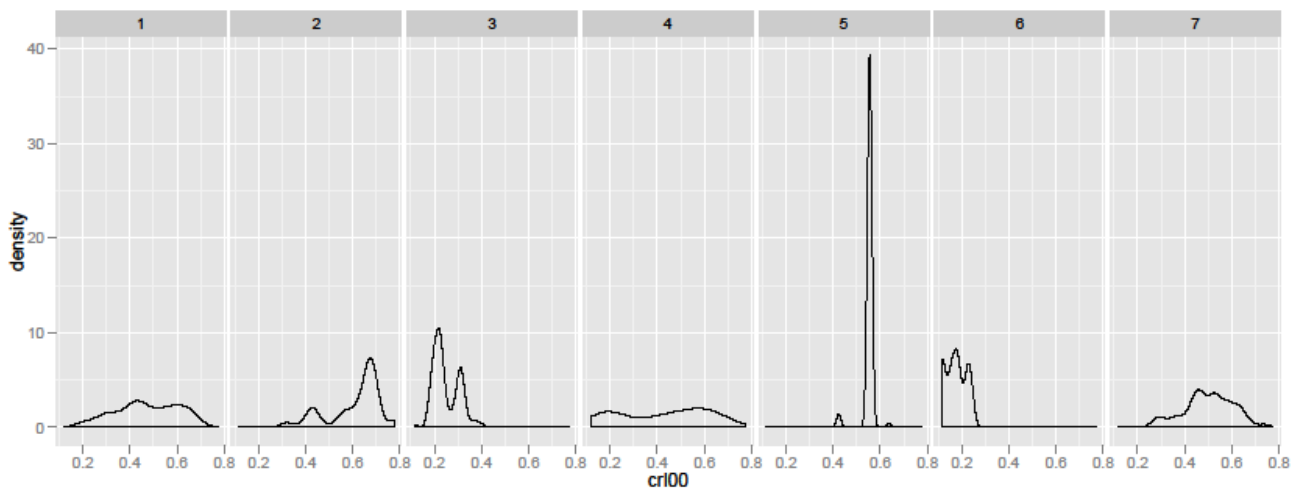
Variable 4 : Total d'émission gamma



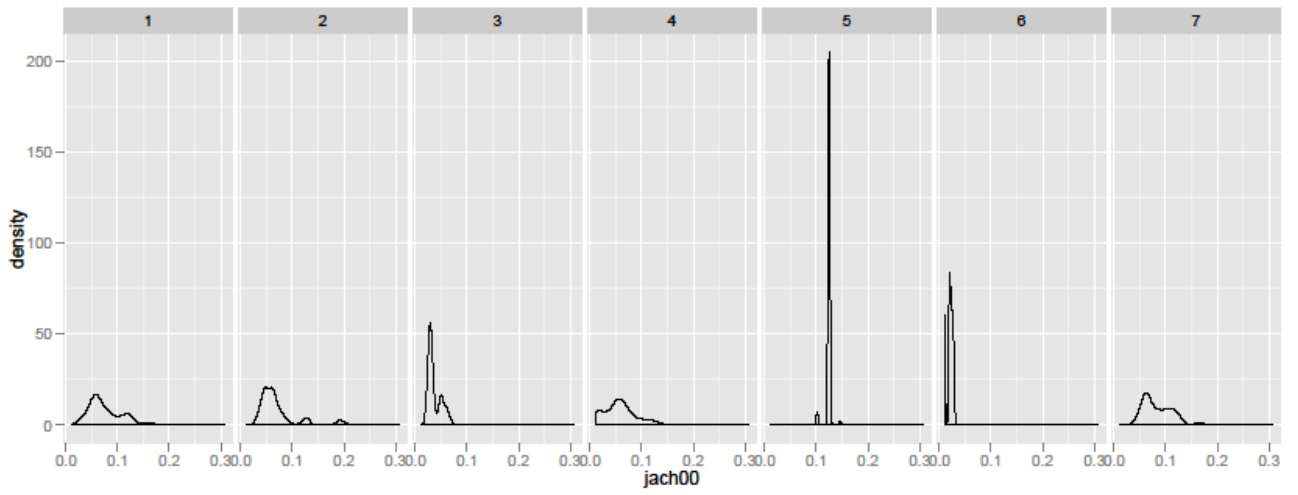
Variable 5 : Uranium



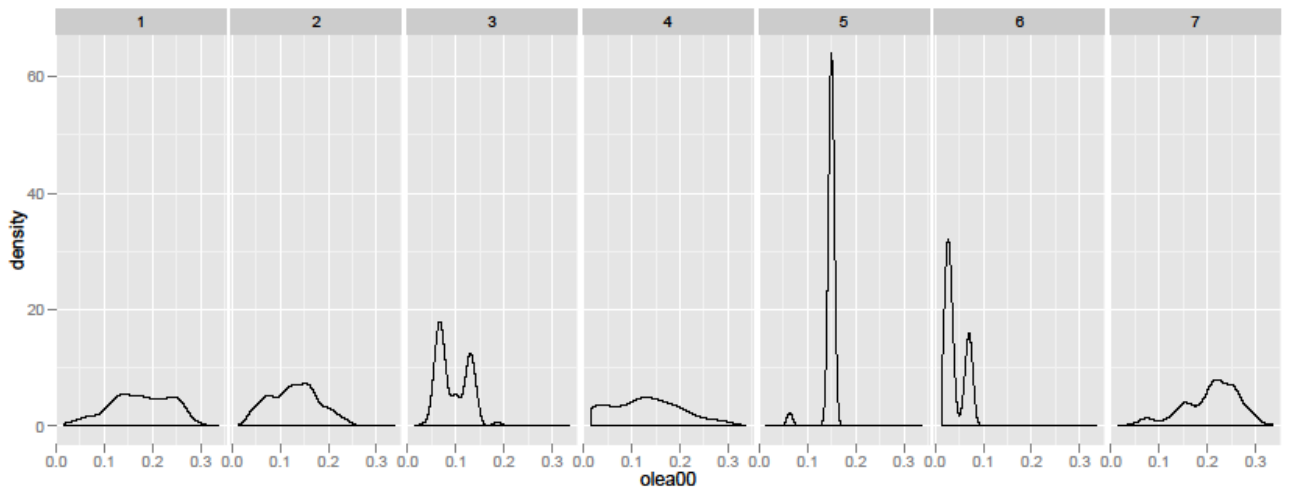
Variable 6 : Pourcentage de zones boisées



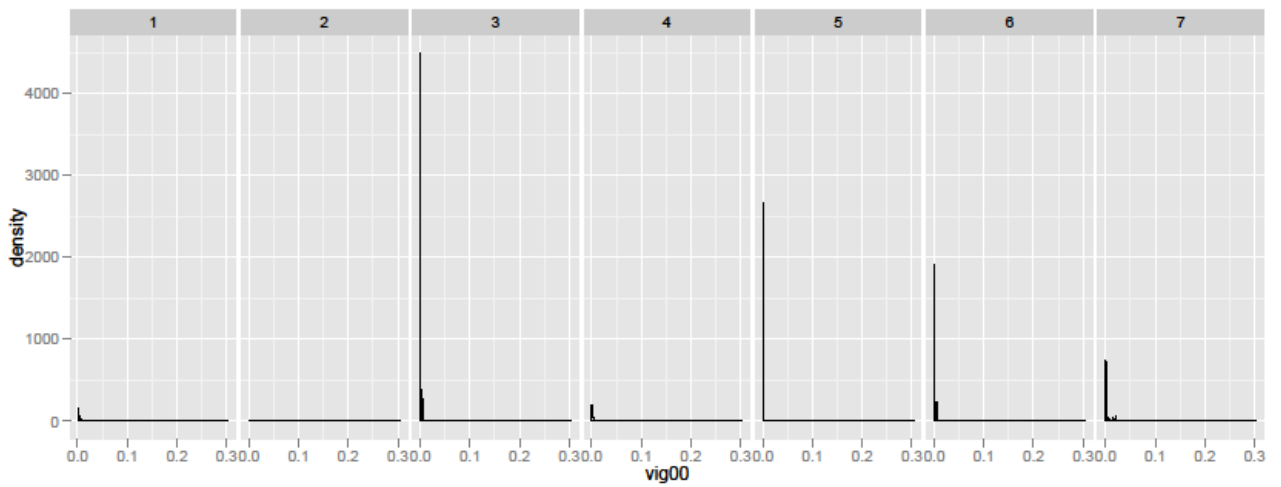
Variable 7 : Pourcentage de cultures céréalières



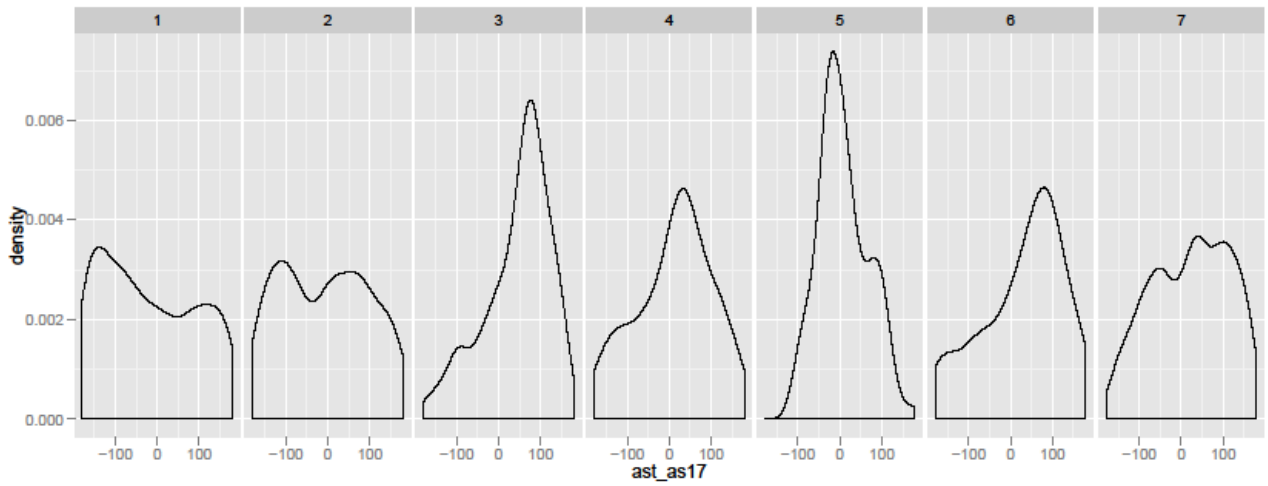
Variable 8 : Pourcentage de jachères



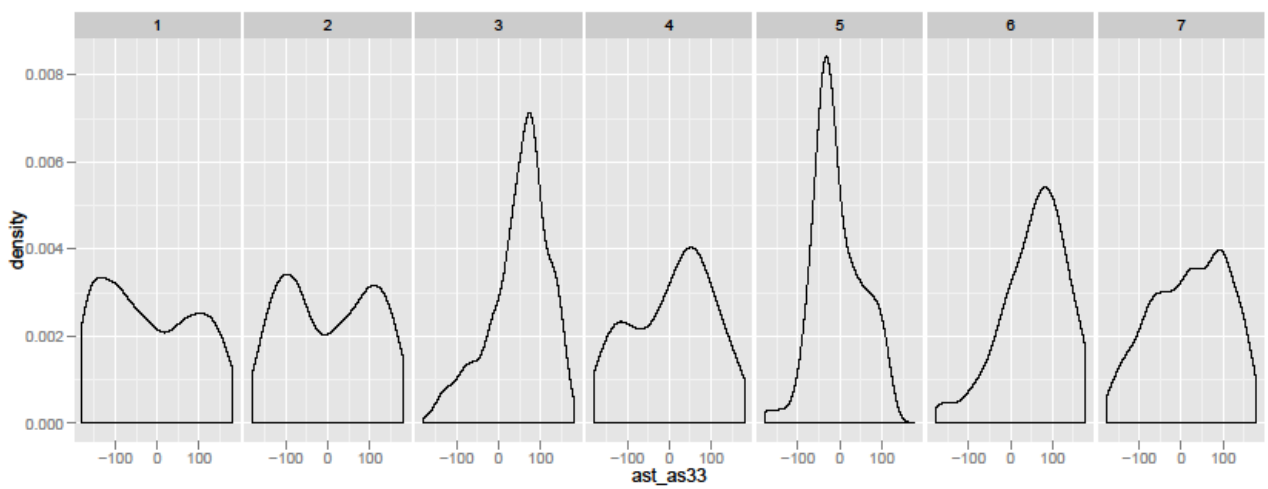
Variable 9 Pourcentage de cultures d'oléagineux



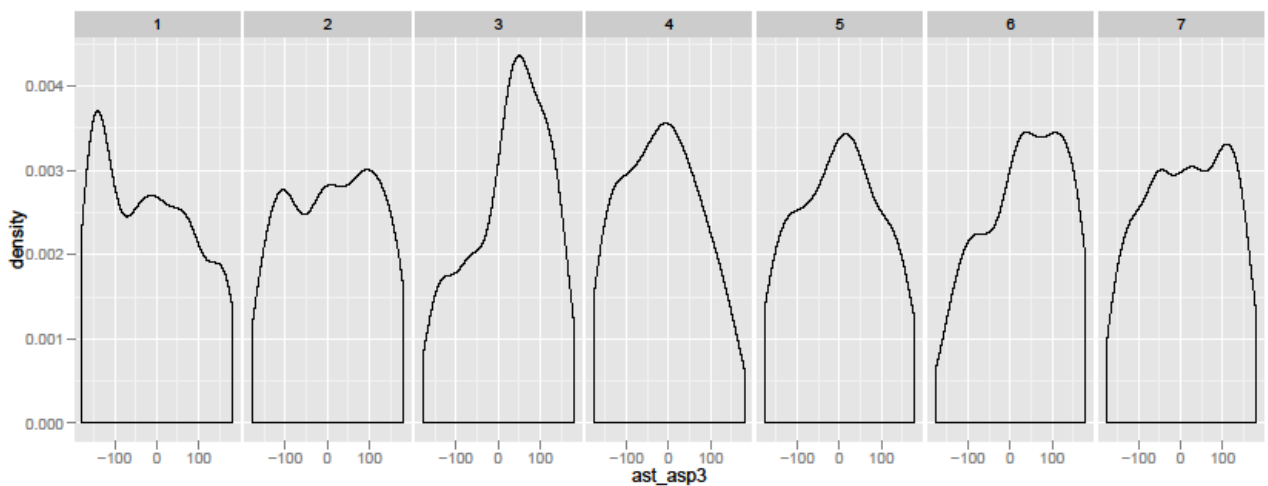
Variable 10 : Pourcentage de vignobles



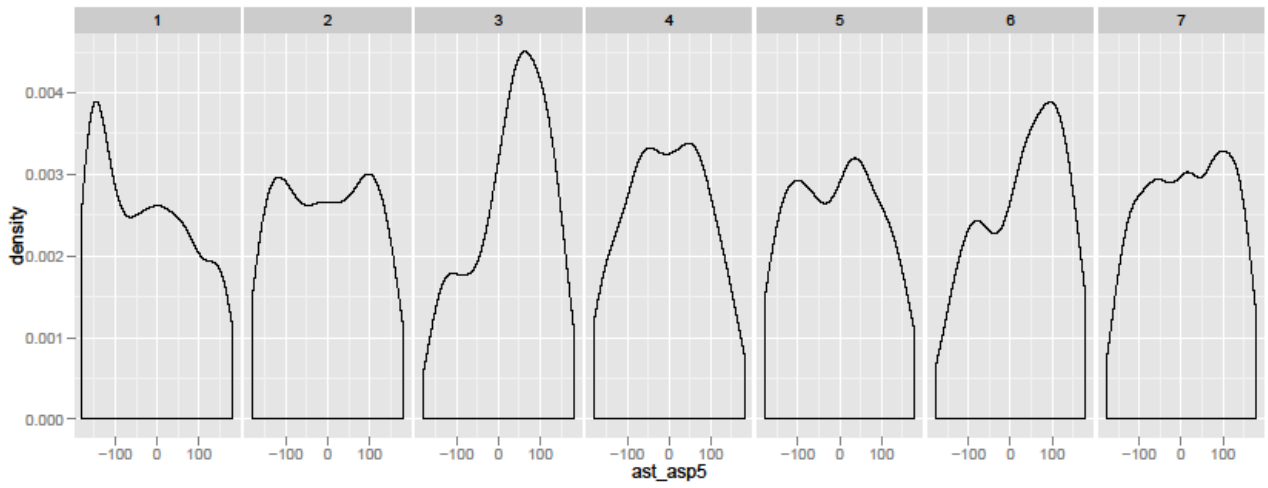
Variable 11 : Aspect direction pour le MNT ASTER à l'échelle 17



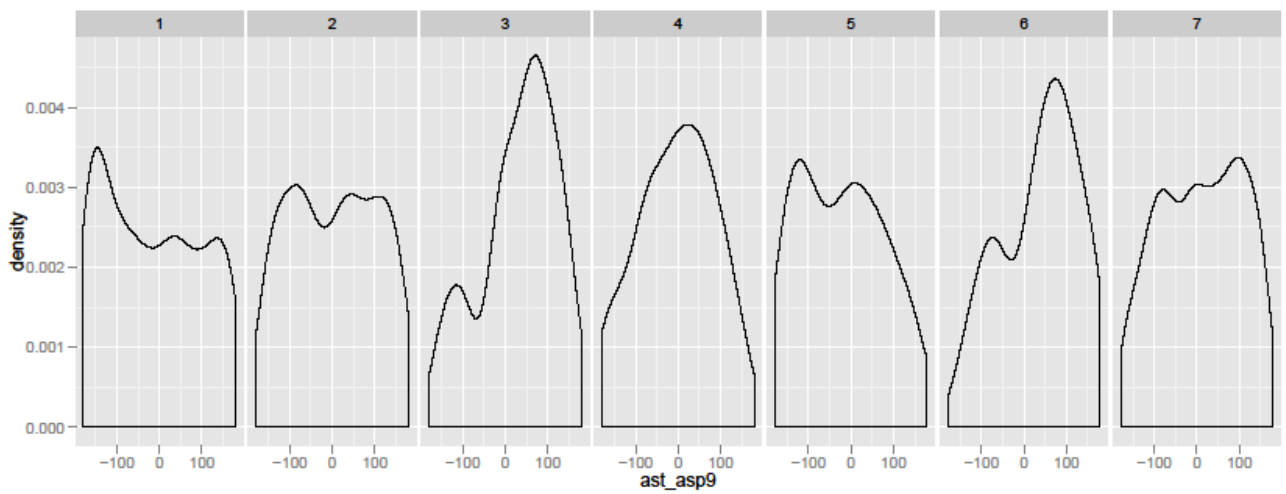
Variable 12 : Aspect direction pour le MNT ASTER à l'échelle 33



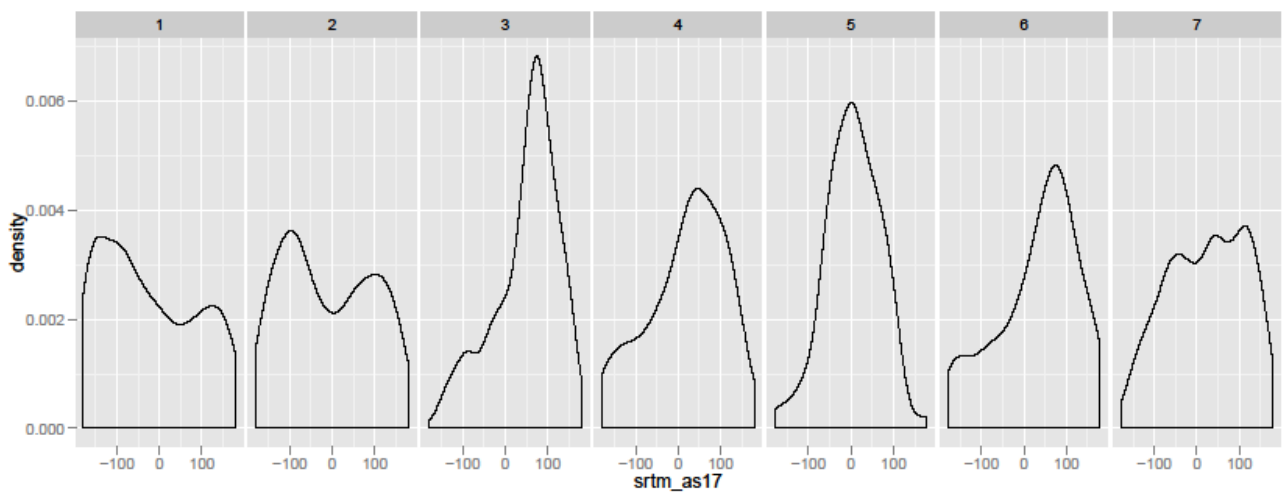
Variable 13 : Aspect direction pour le MNT ASTER à l'échelle 3



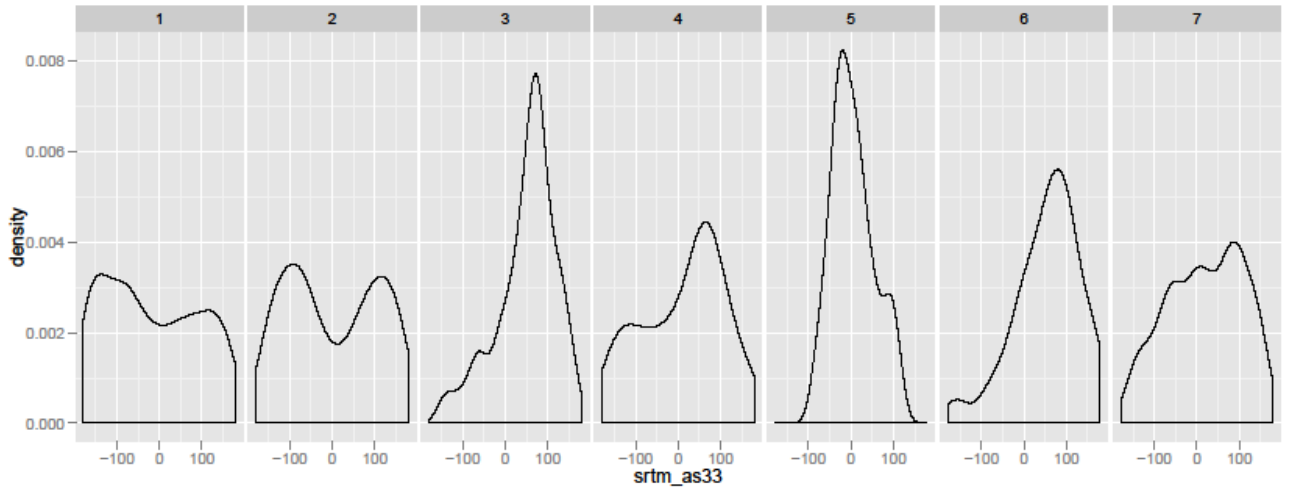
Variable 14 : Aspect direction pour le MNT ASTER à l'échelle 5



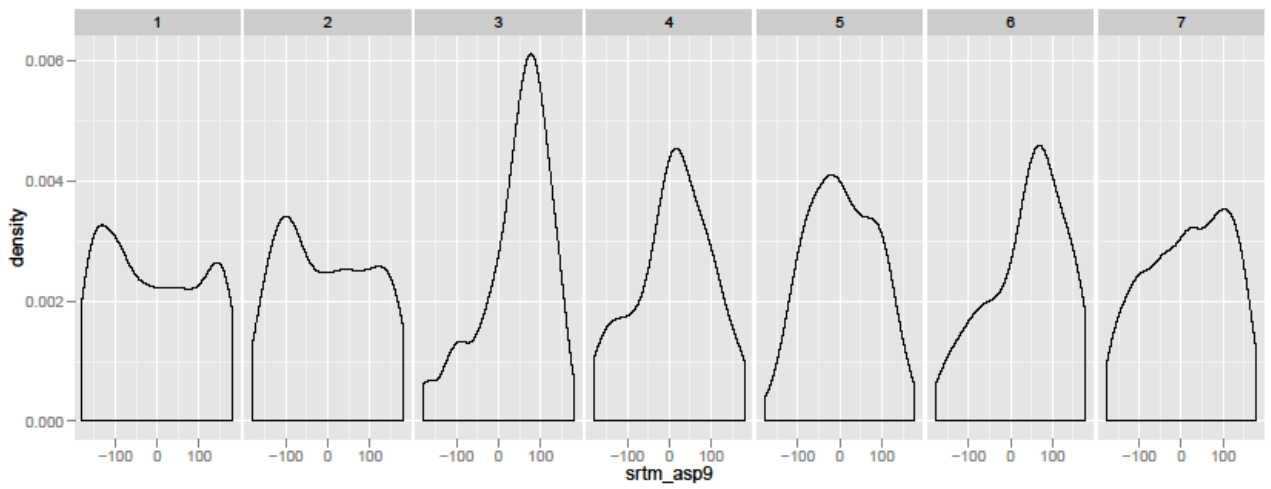
Variable 15 : Aspect direction pour le MNT ASTER à l'échelle 9



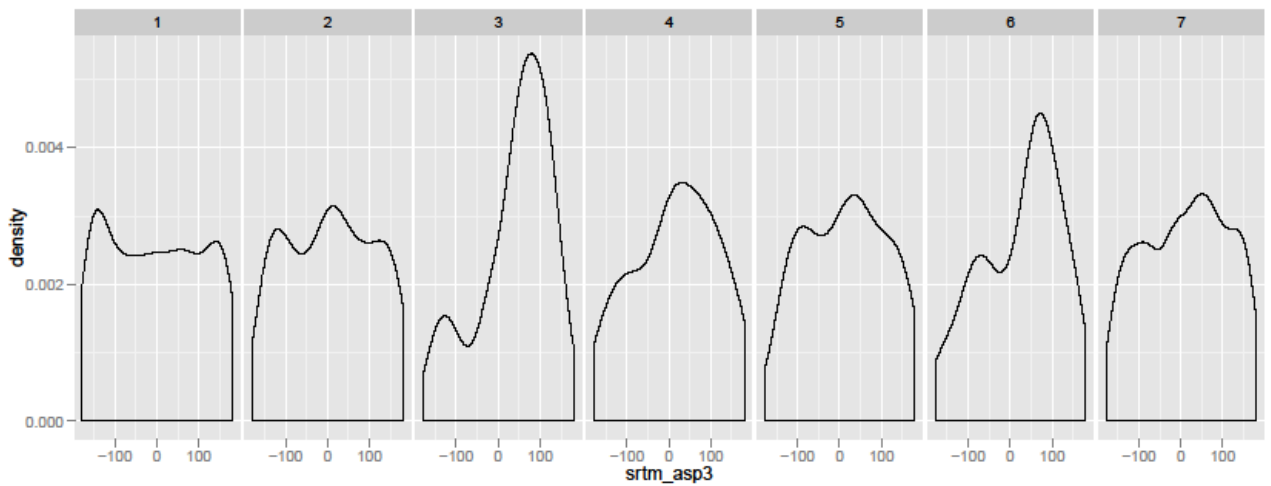
Variable 16 : Aspect direction pour le MNT SRTM à l'échelle 17



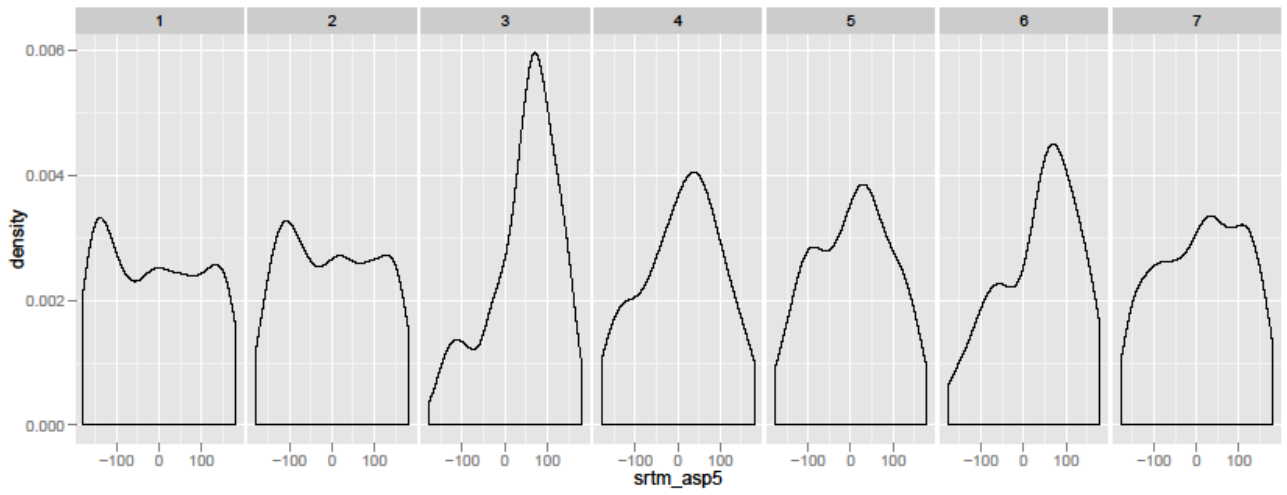
Variable 17 : Aspect direction pour le MNT SRTM à l'échelle 33



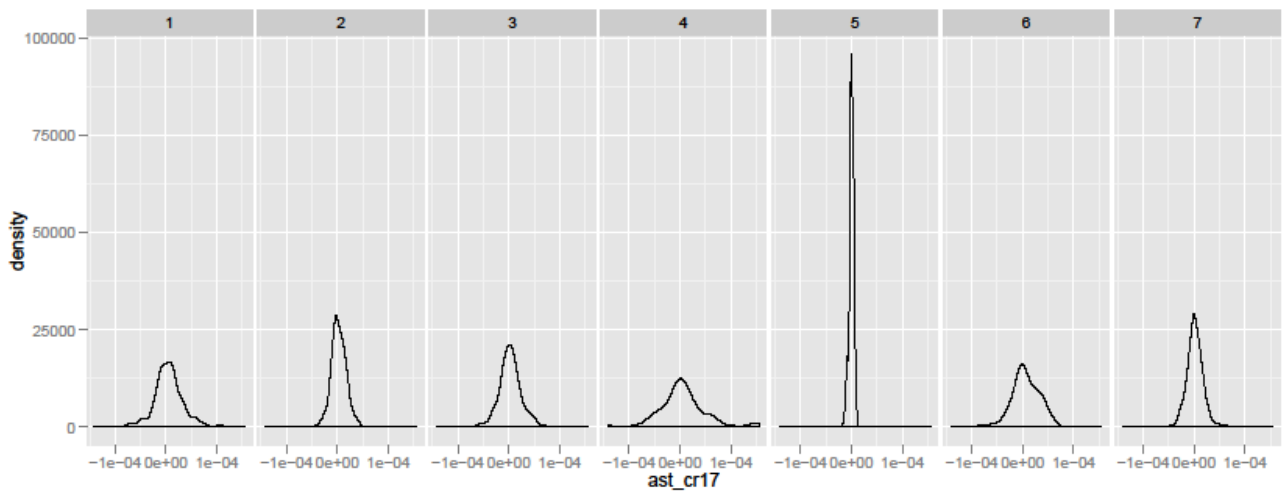
Variable 18 : Aspect direction pour le MNT SRTM à l'échelle 9



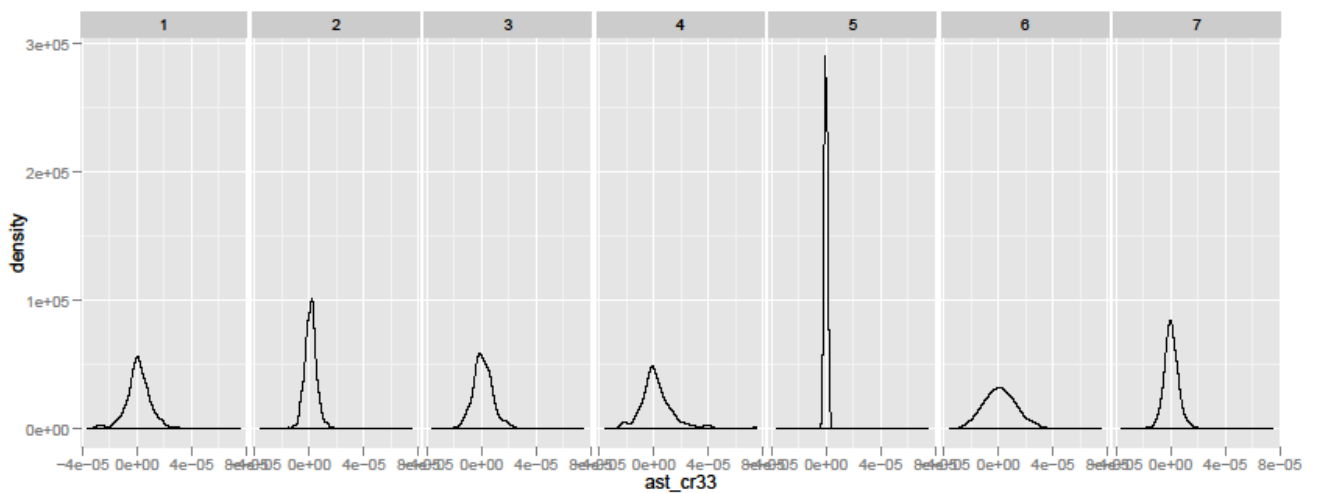
Variable 19 : Aspect direction pour le MNT SRTM à l'échelle 3



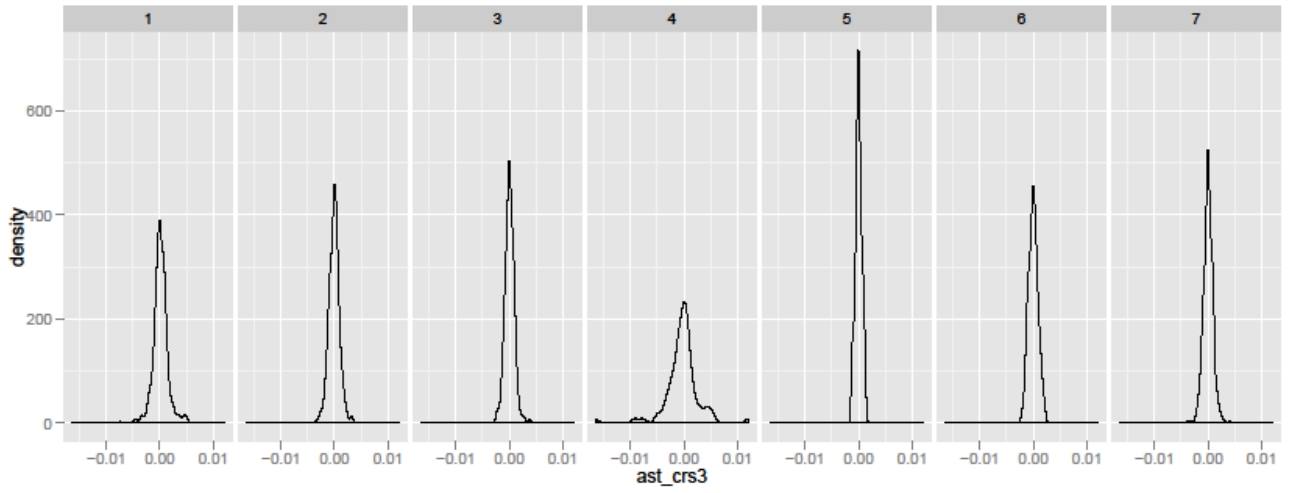
Variable 20 : Aspect direction pour le MNT SRTM à l'échelle 5



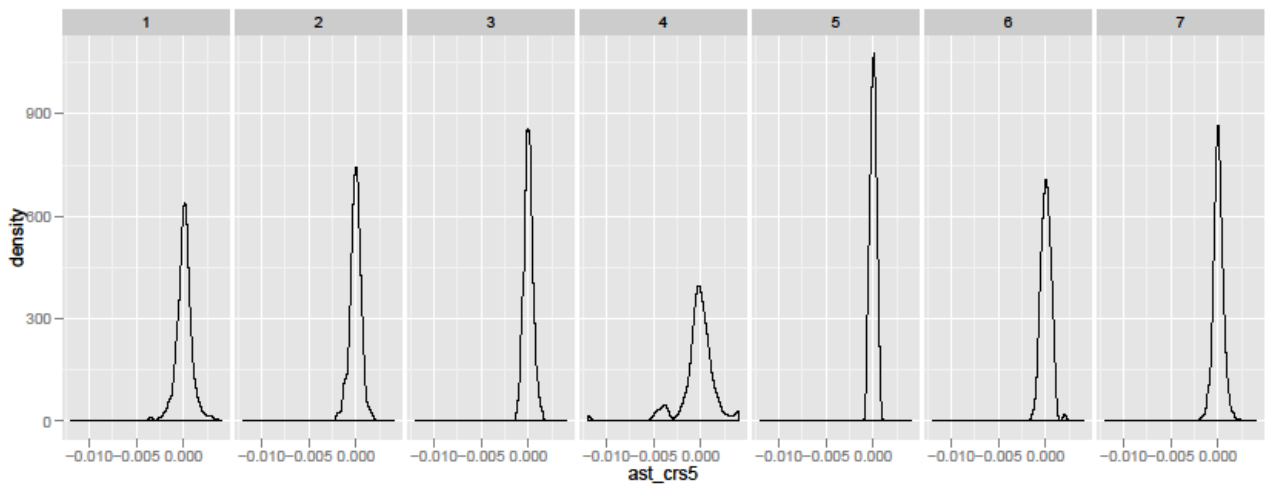
Variable 21 : Cross sectionnal curvature pour le MNT ASTER à l'échelle 17



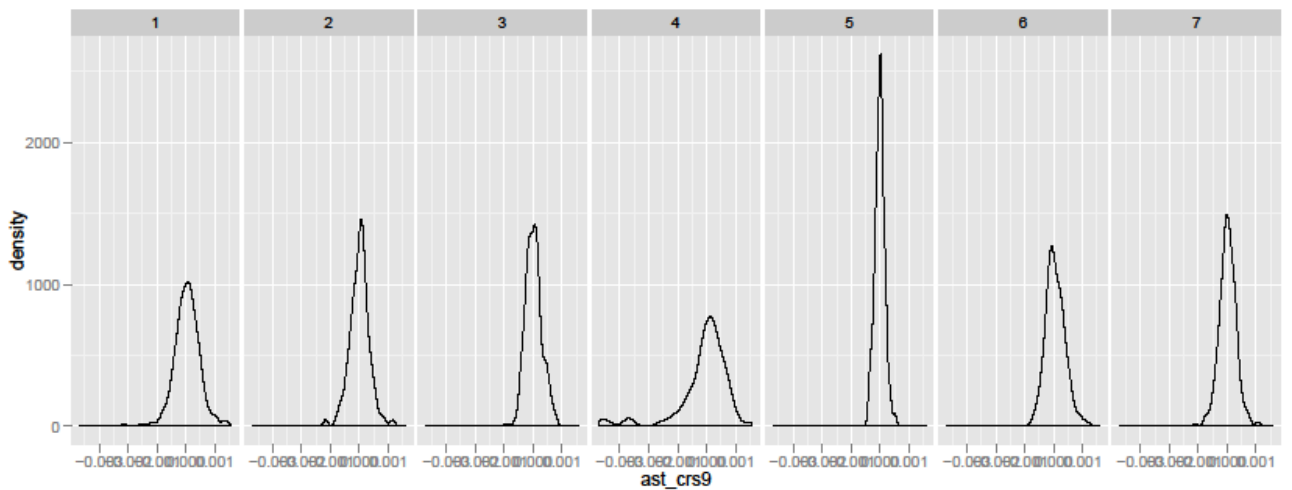
Variable 22 : Cross sectionnal curvature pour le MNT ASTER à l'échelle 33



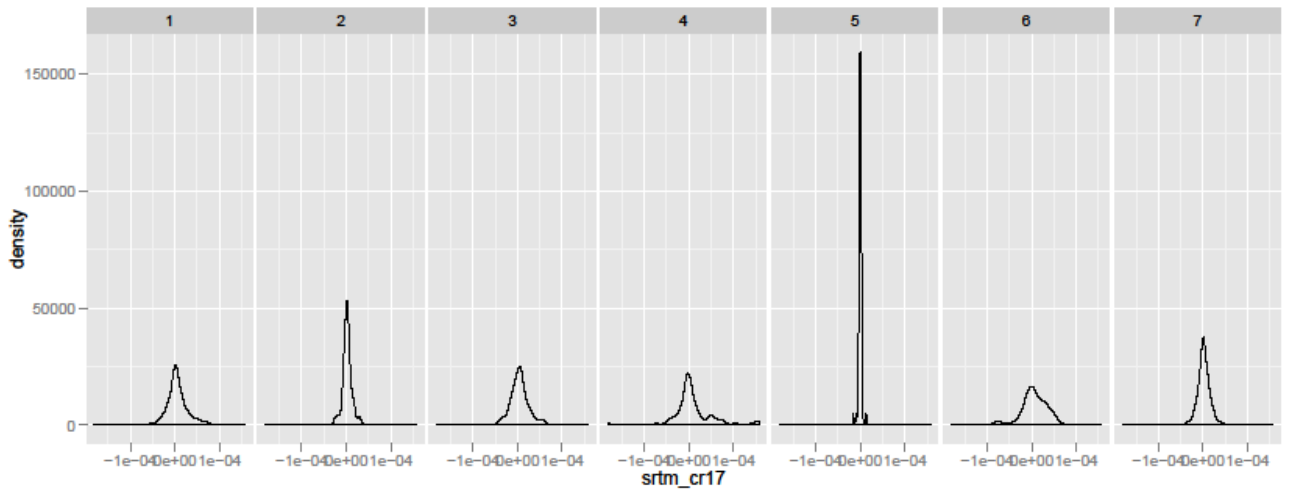
Variable 23 : Cross sectionnal curvature pour le MNT ASTER à l'échelle 3



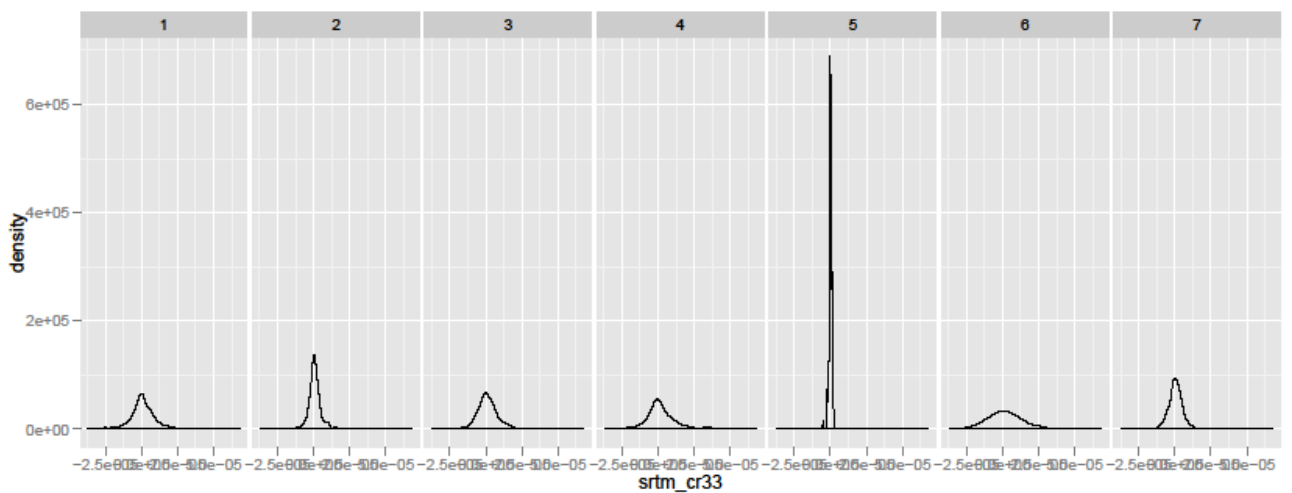
Variable 24 : Cross sectionnal curvature pour le MNT ASTER à l'échelle 5



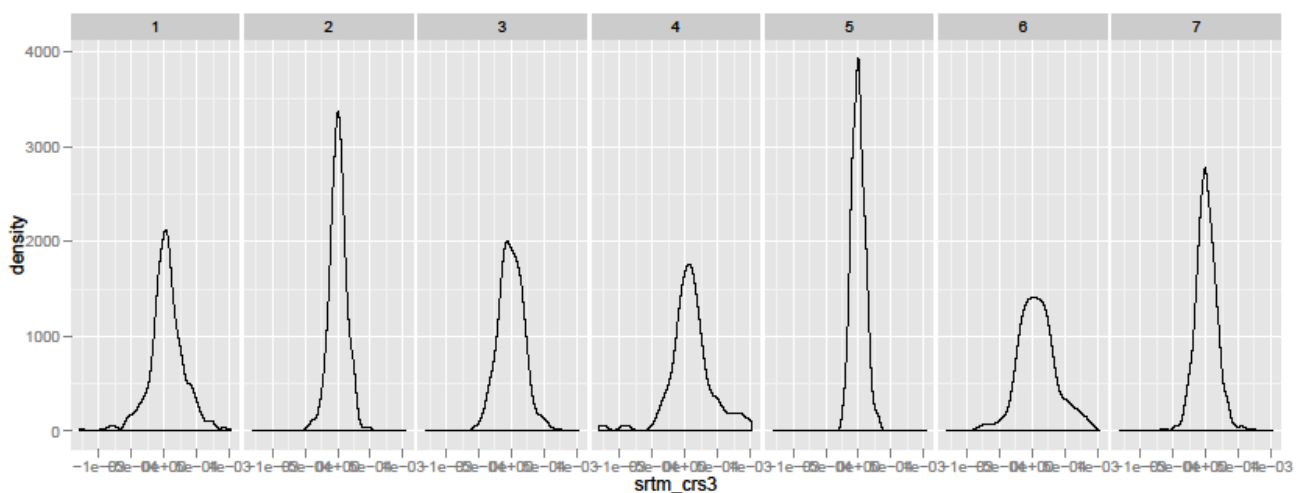
Variable 25 : Cross sectionnal curvature pour le MNT ASTER à l'échelle 9



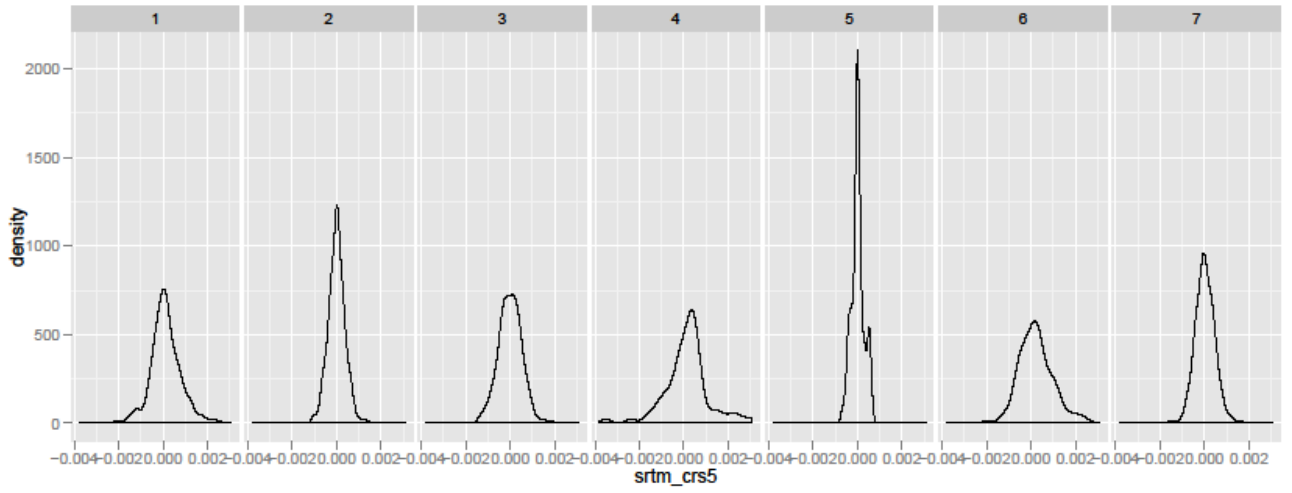
Variable 26 : Cross sectionnal curvature pour le MNT SRTM à l'échelle 17



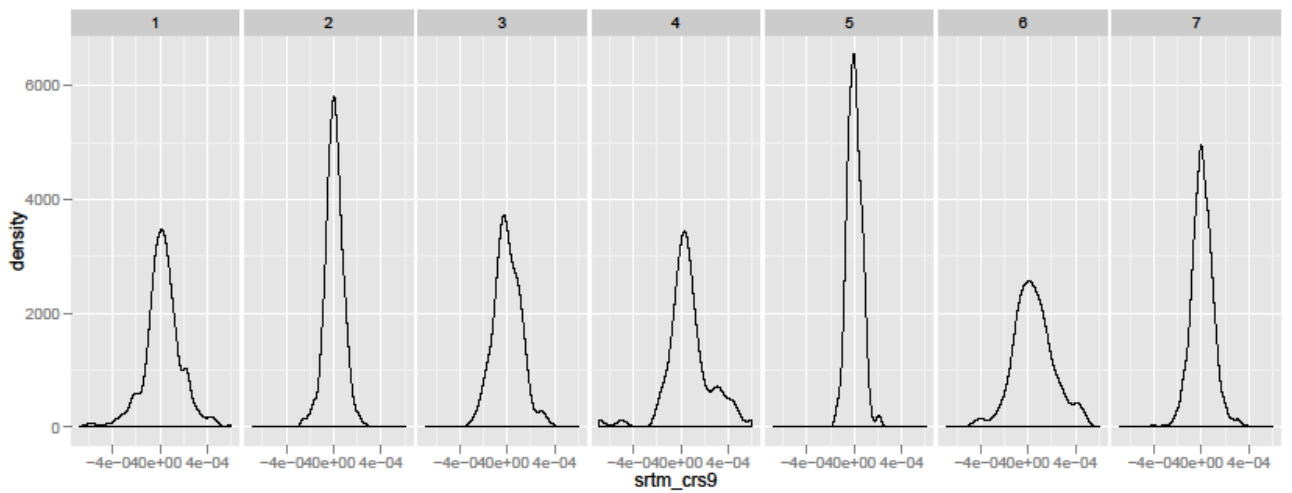
Variable 27 : Cross sectionnal curvature pour le MNT SRTM à l'échelle 33



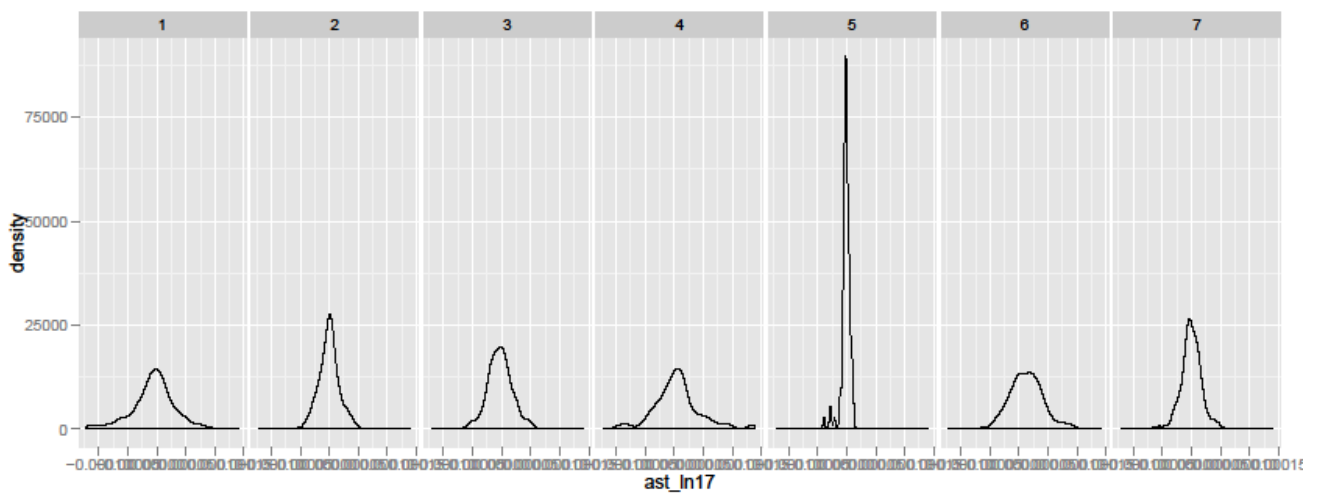
Variable 28 : Cross sectionnal curvature pour le MNT SRTM à l'échelle 3



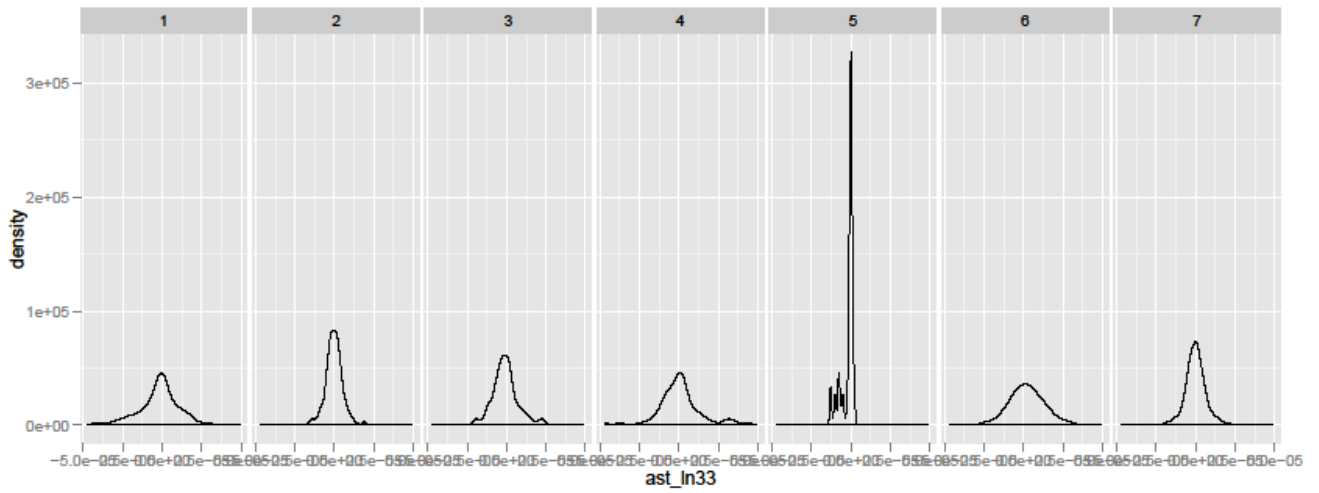
Variable 29 : Cross sectional curvature pour le MNT SRTM à l'échelle 5



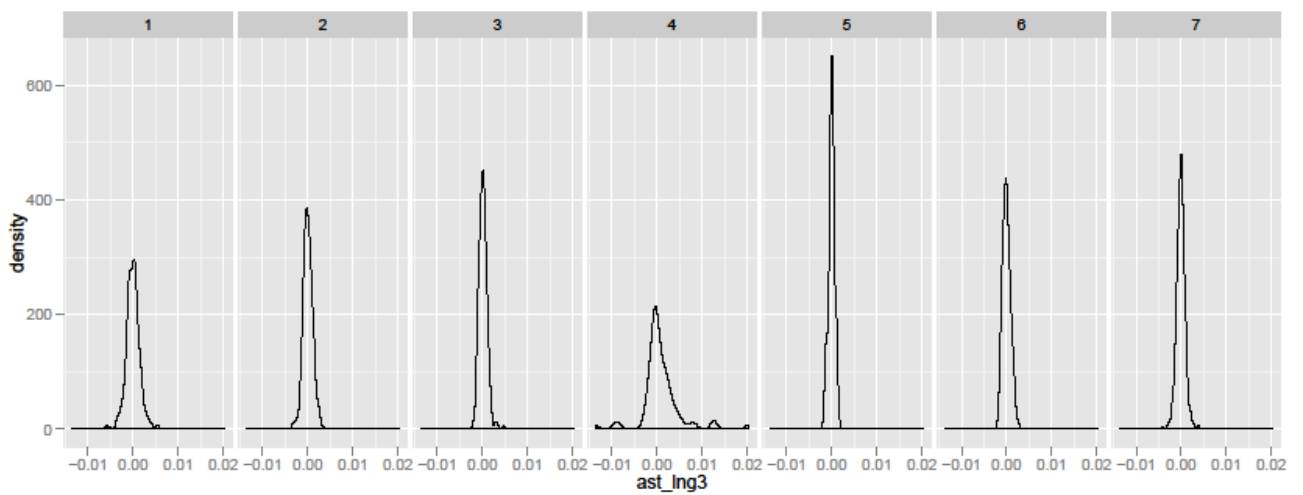
Variable 30 : Cross sectional curvature pour le MNT SRTM à l'échelle 9



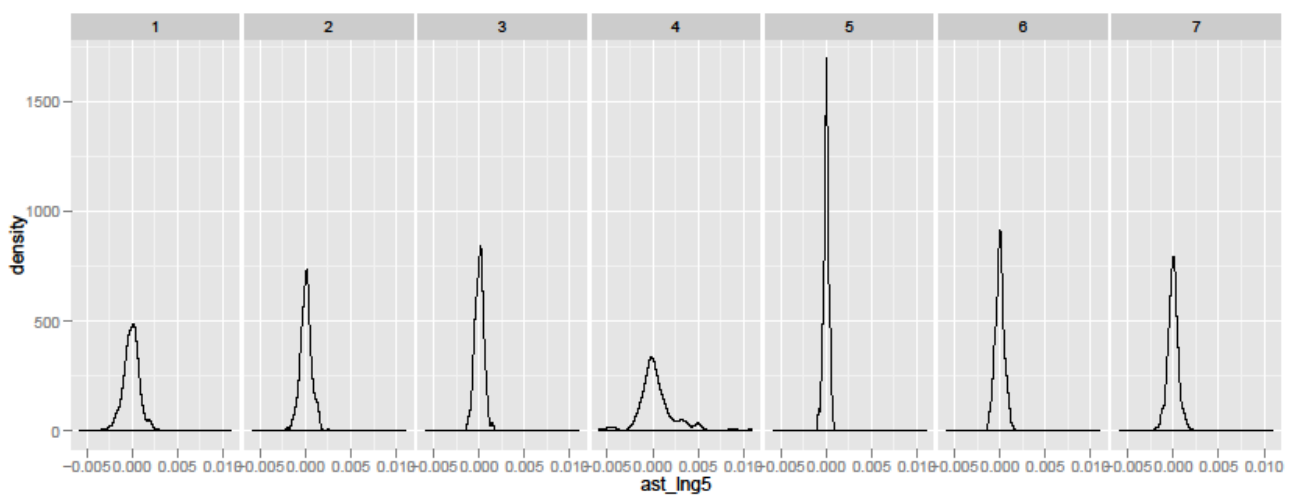
Variable 31 : Longitudinal curvature pour le MNT ASTER à l'échelle 17



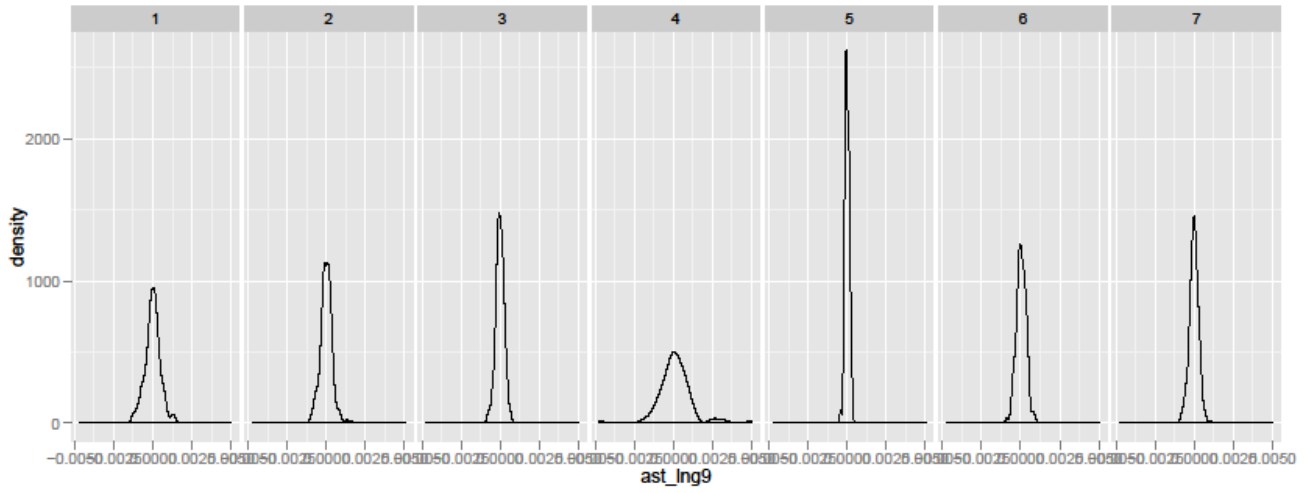
Variable 32 : Longitudinal curvature pour le MNT ASTER à l'échelle 33



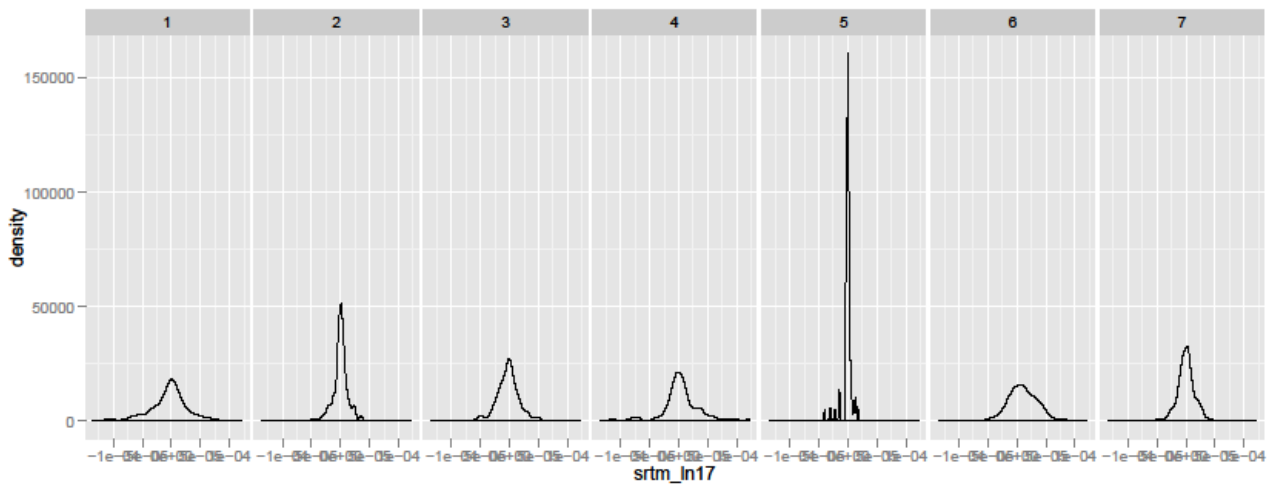
Variable 33 : Longitudinal curvature pour le MNT ASTER à l'échelle 3



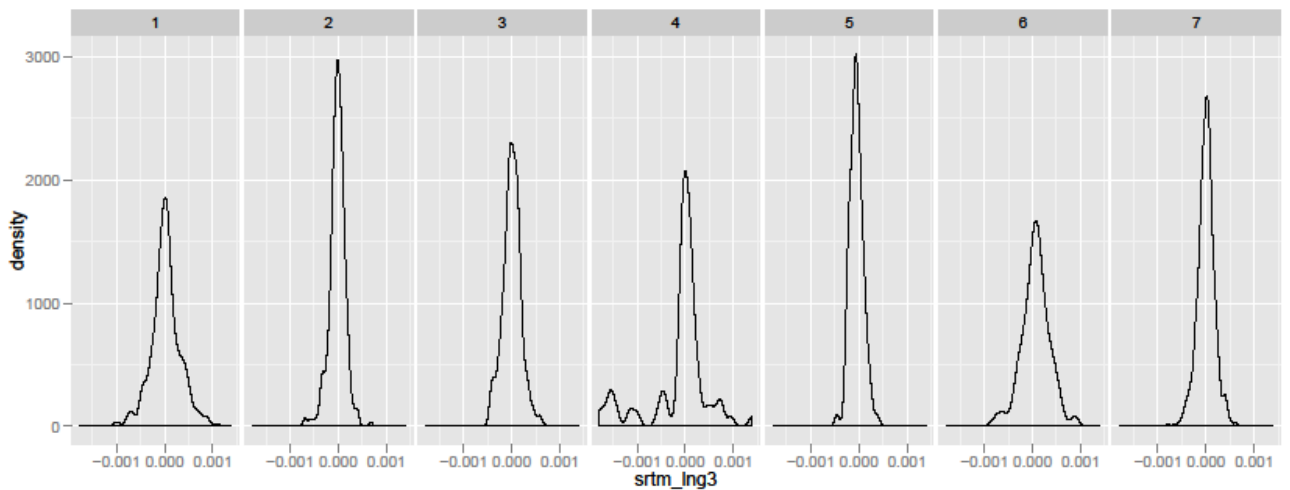
Variable 34 : Longitudinal curvature pour le MNT ASTER à l'échelle 5



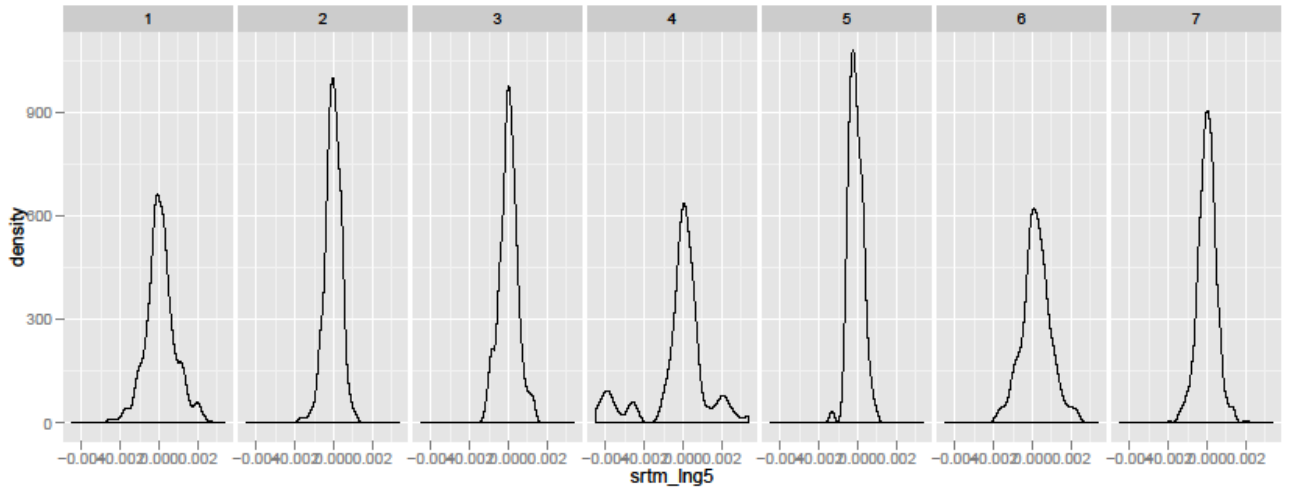
Variable 35 : Longitudinal curvature pour le MNT ASTER à l'échelle 9



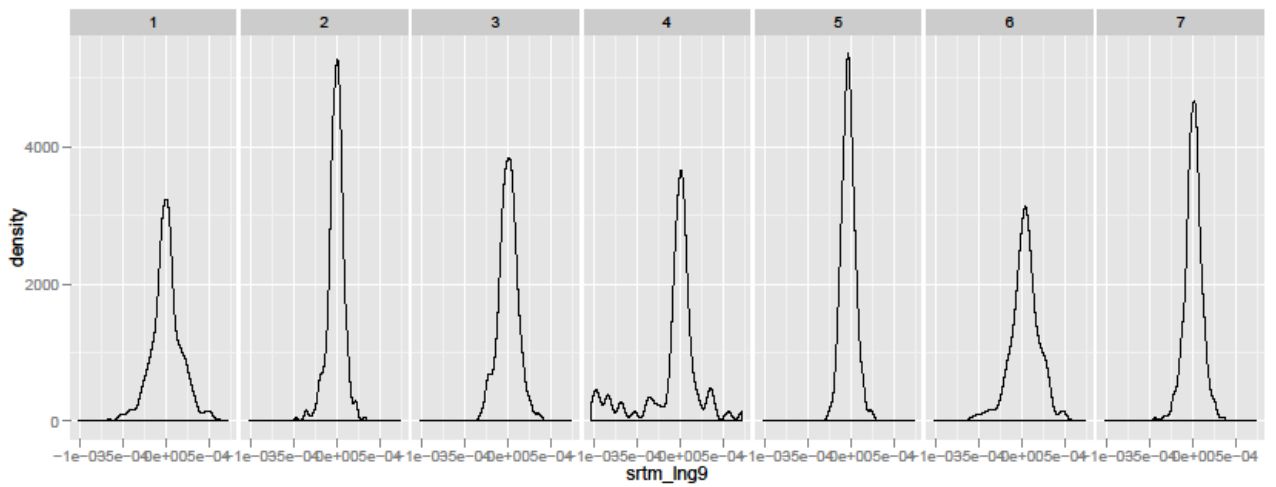
Variable 36 : Longitudinal curvature pour le MNT SRTM à l'échelle 17



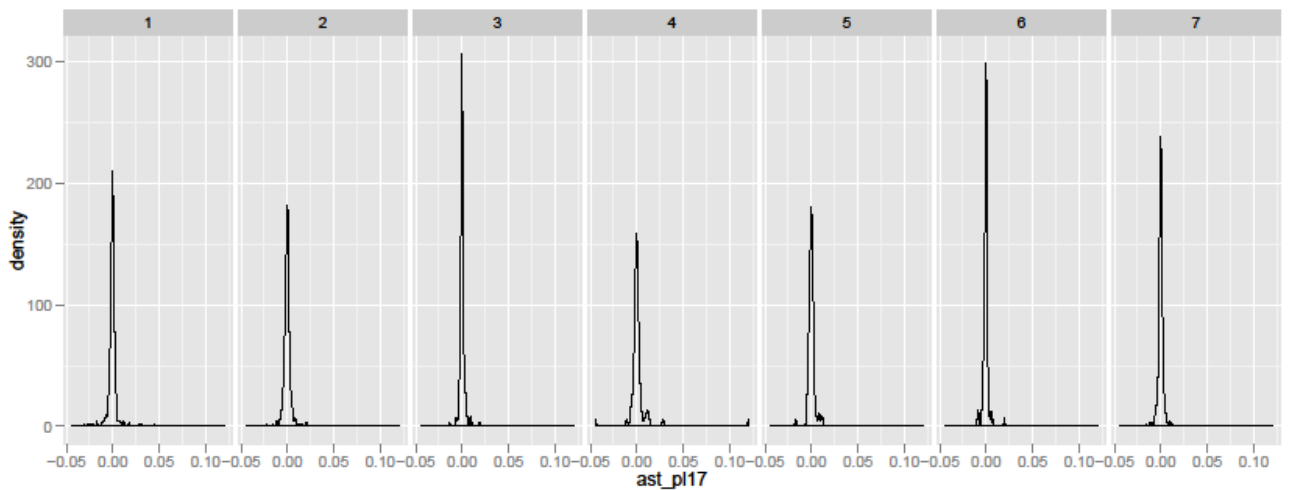
Variable 37 : Longitudinal curvature pour le MNT SRTM à l'échelle 3



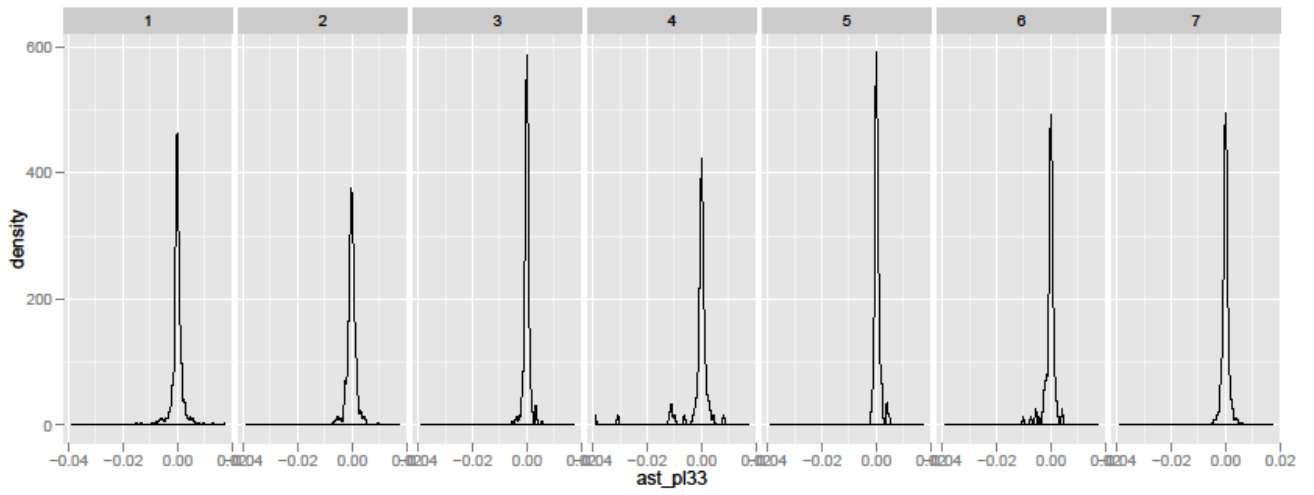
Variable 38 : Longitudinal curvature pour le MNT SRTM à l'échelle 5



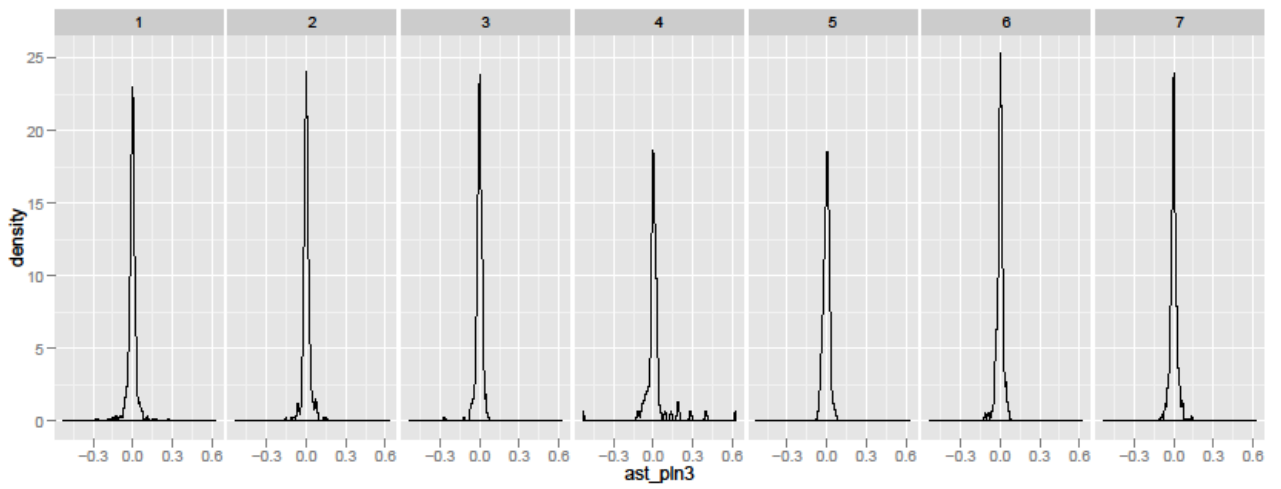
Variable 39 : Longitudinal curvature pour le MNT SRTM à l'échelle 9



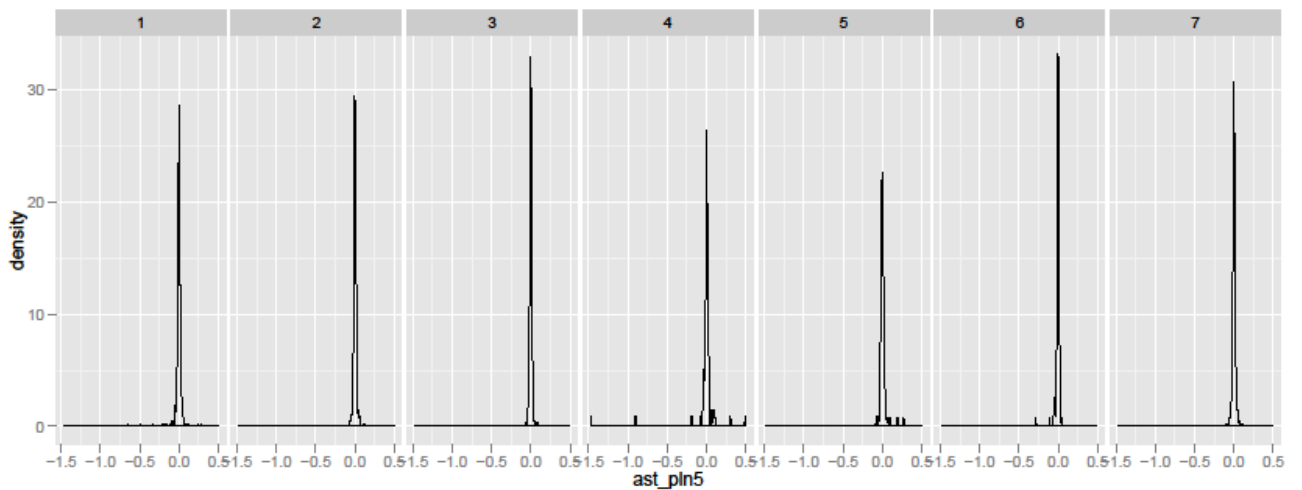
Variable 40 : Longitudinal curvature pour le MNT ASTER à l'échelle 17



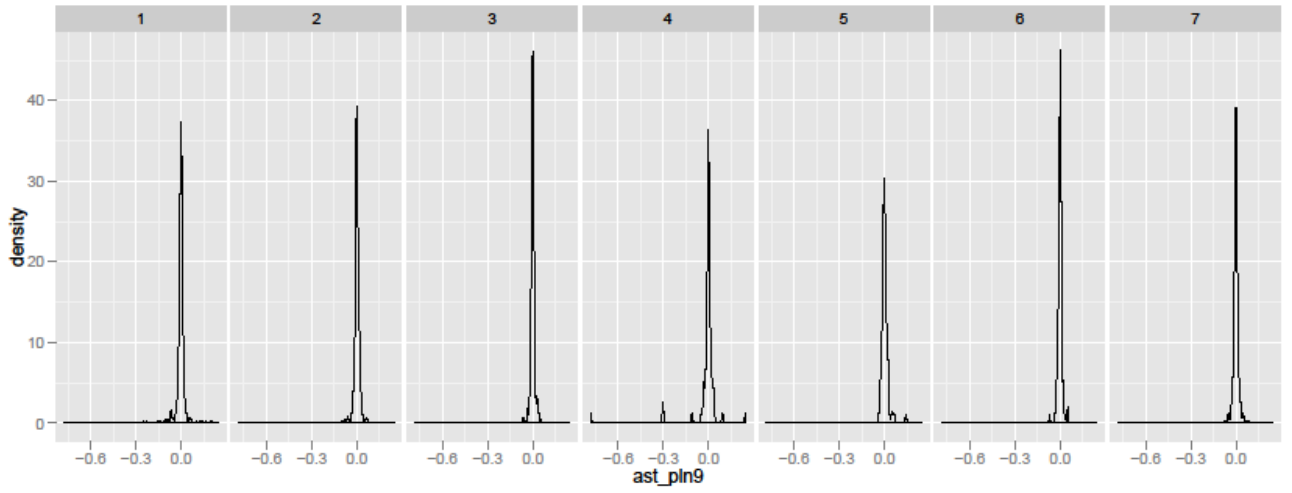
Variable 41 : Plan courvure pour le MNT ASTER à l'échelle 33



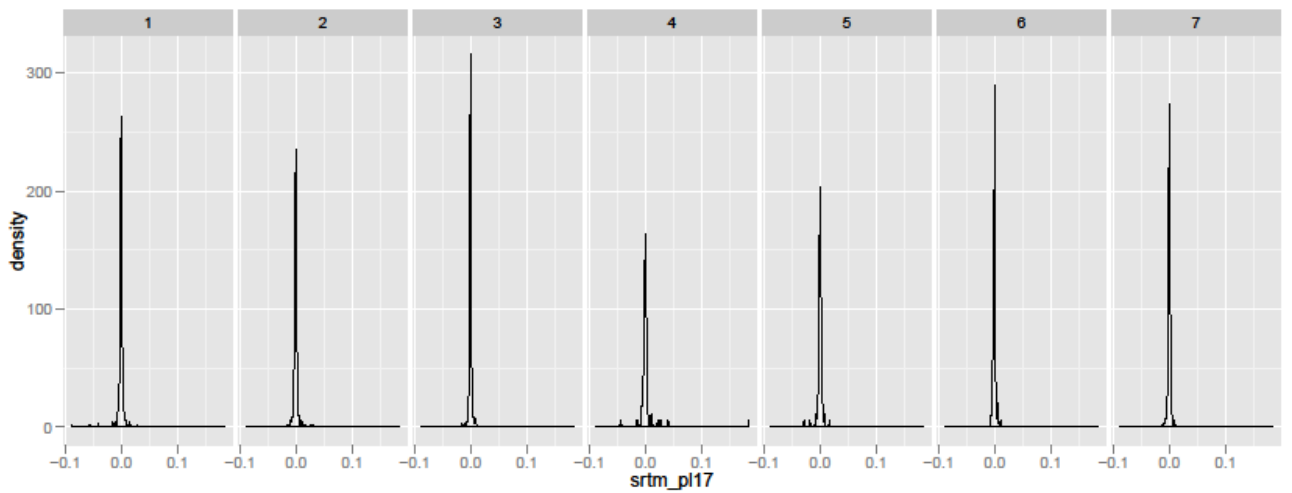
Variable 42 : Plan courvure pour le MNT ASTER à l'échelle 3



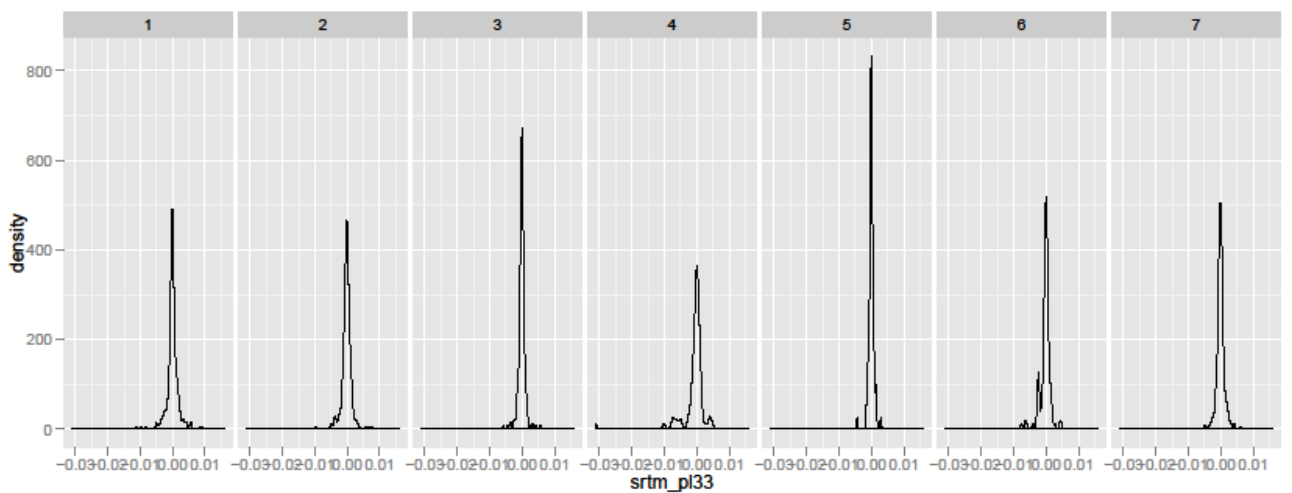
Variable 43 : Plan courvure pour le MNT ASTER à l'échelle 5



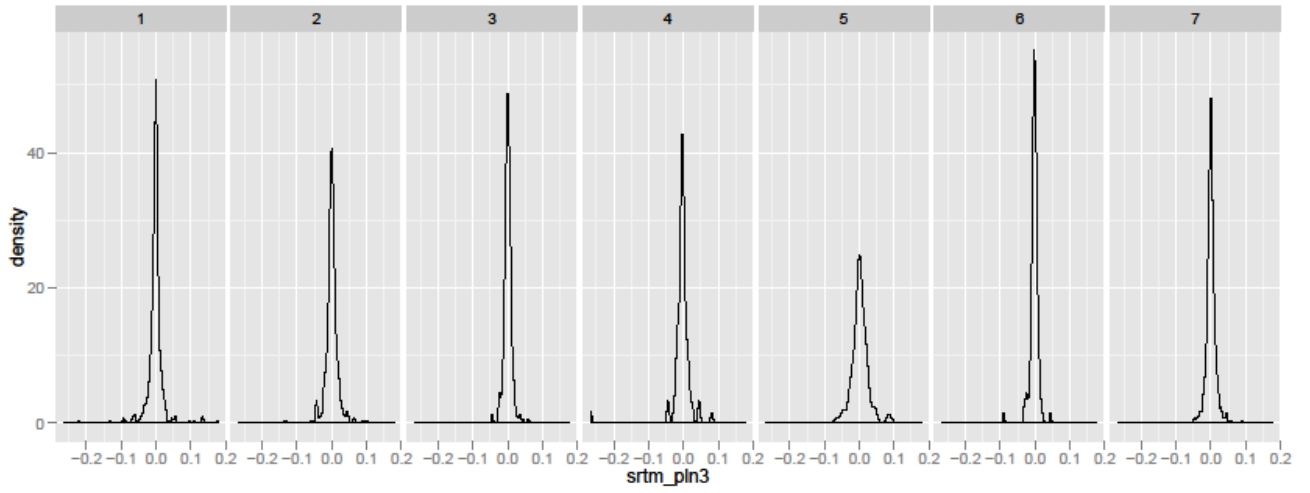
Variable 44 : Plan courvure pour le MNT ASTER à l'échelle 9



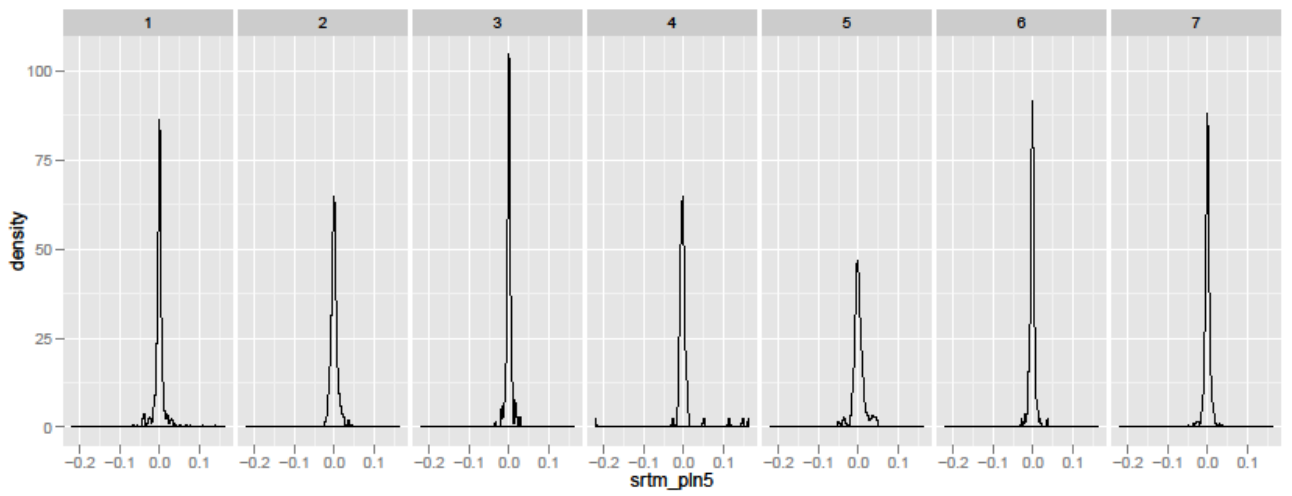
Variable 45 : Plan courvure pour le MNT SRTM à l'échelle 17



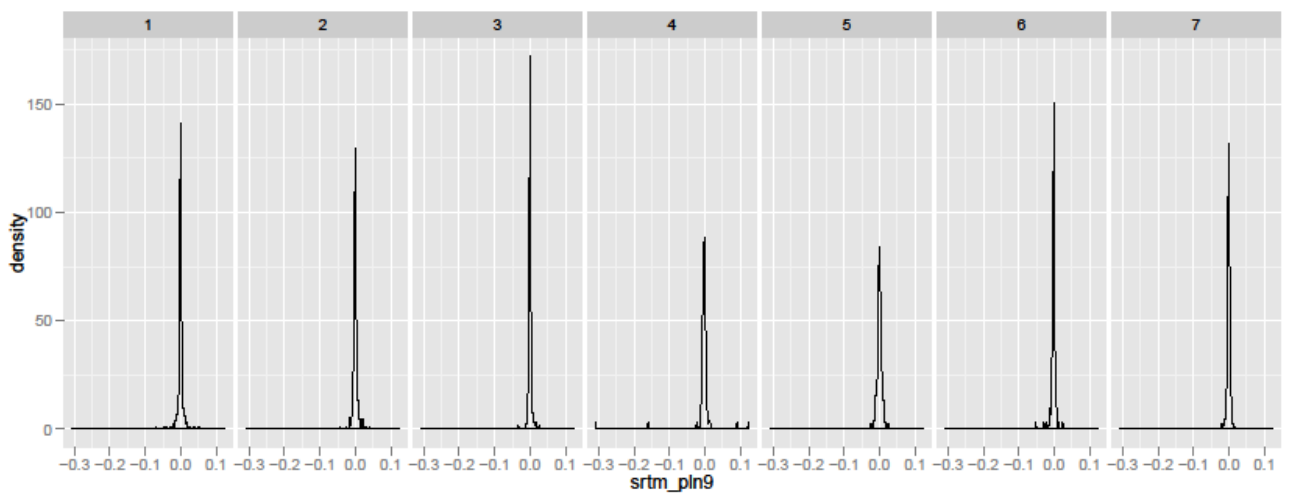
Variable 46 : Plan courvure pour le MNT SRTM à l'échelle 33



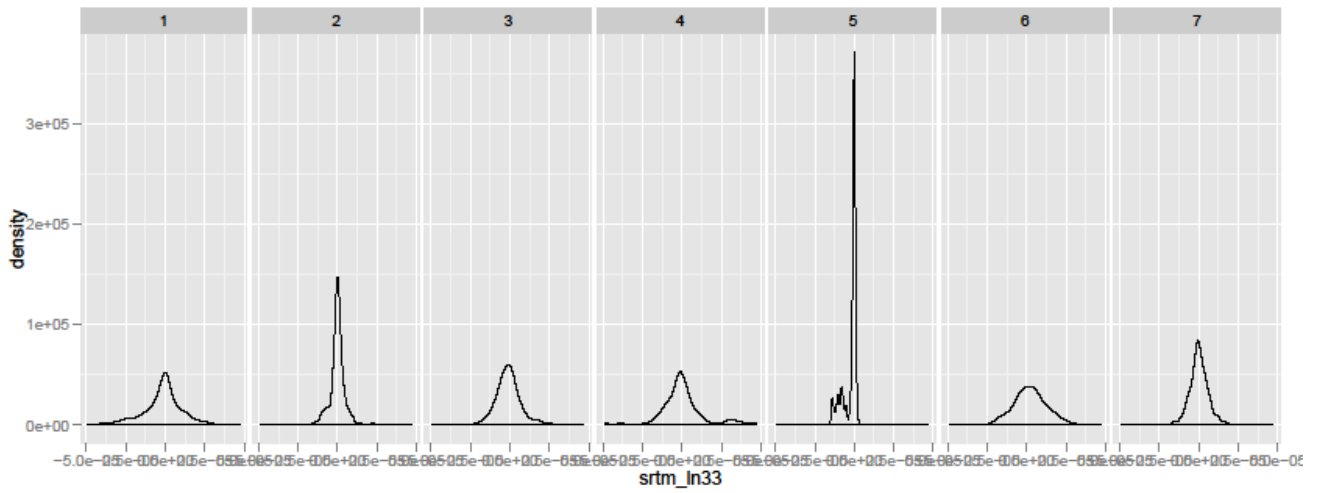
Variable 47 : Plan courvure pour le MNT SRTM à l'échelle 3



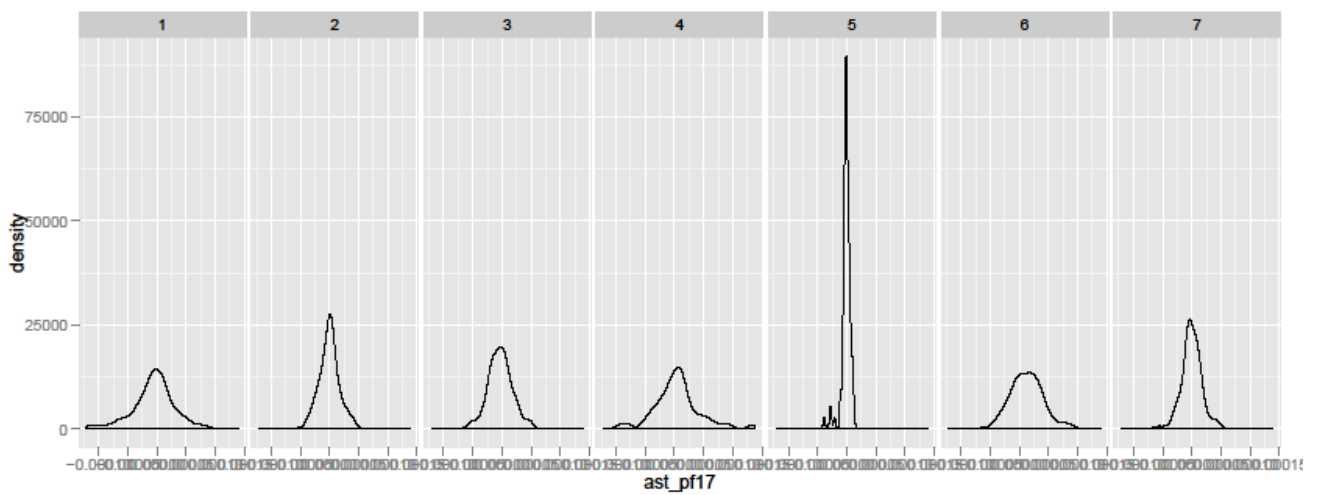
Variable 48 : Plan courvure pour le MNT SRTM à l'échelle 5



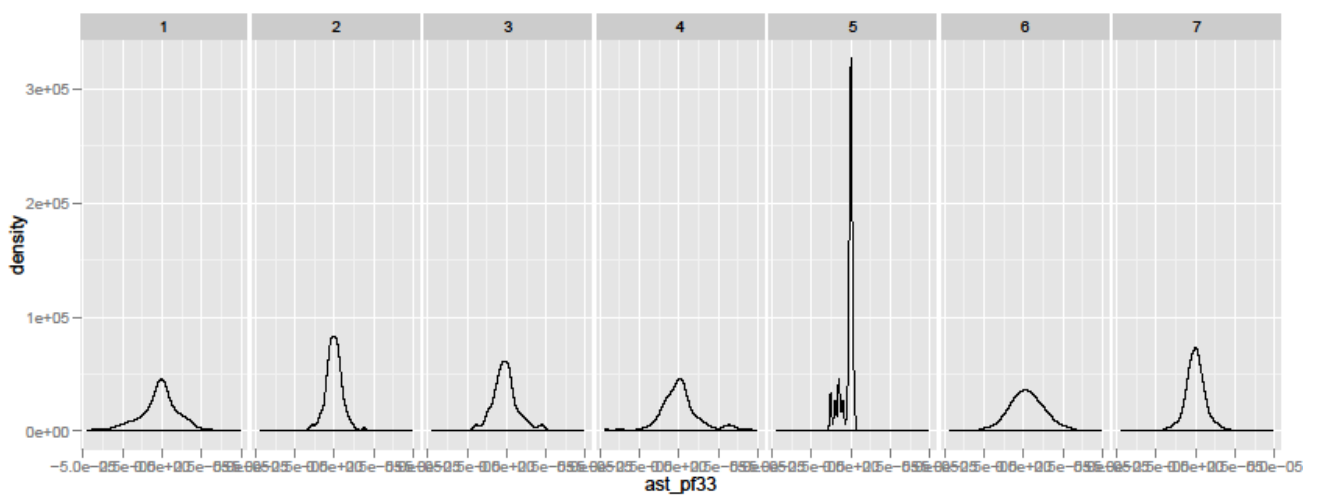
Variable 49 : Plan courvure pour le MNT SRTM à l'échelle 9



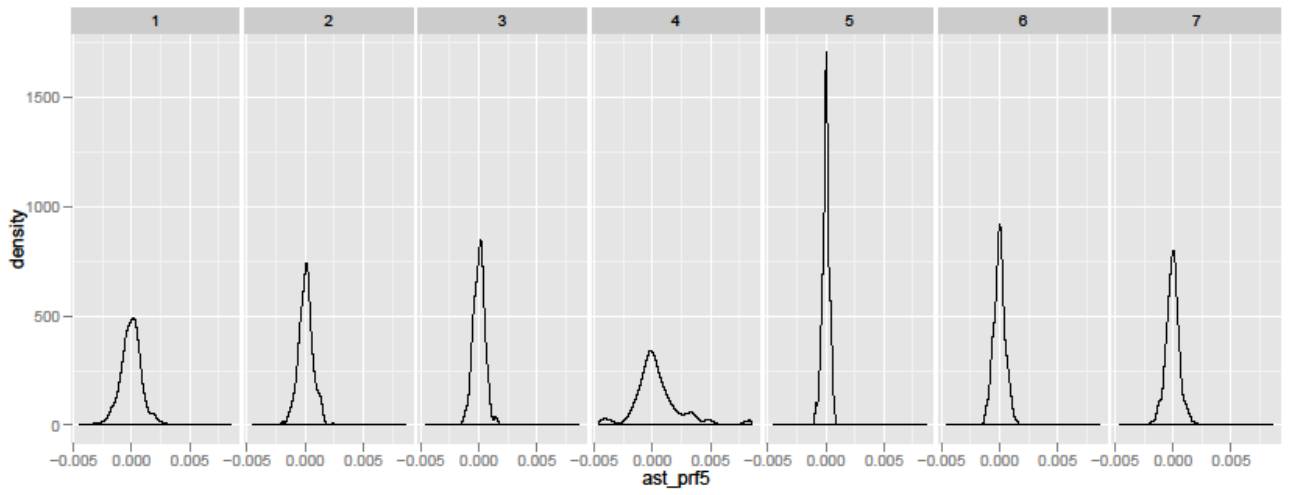
Variable 50 : Longitudinal curvature pour le MNT SRTM à l'échelle 33



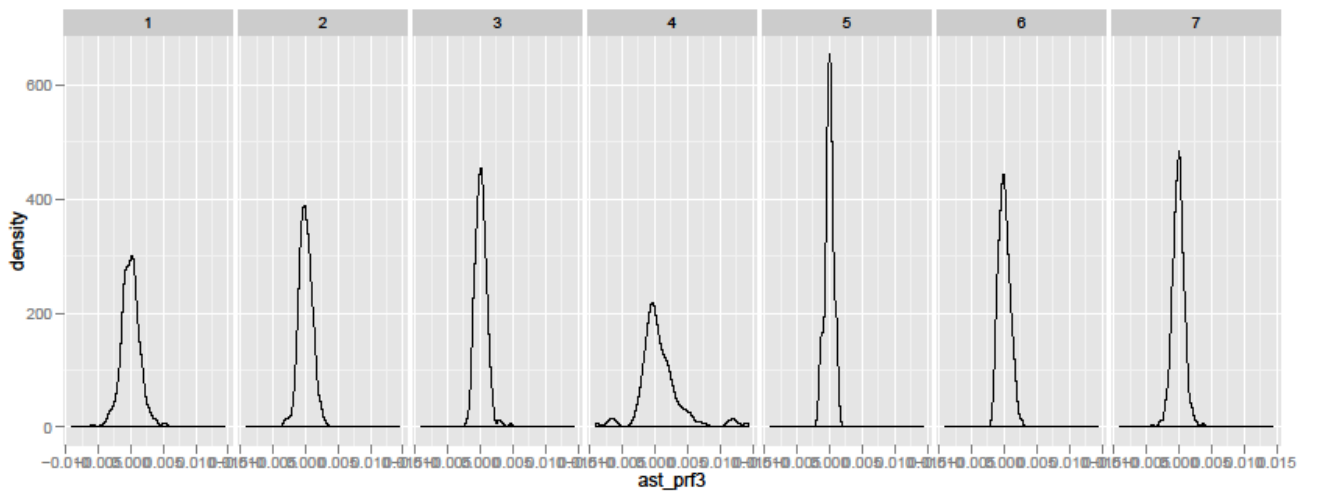
Variable 51 : Profile curvature pour le MNT ASTER à l'échelle 17



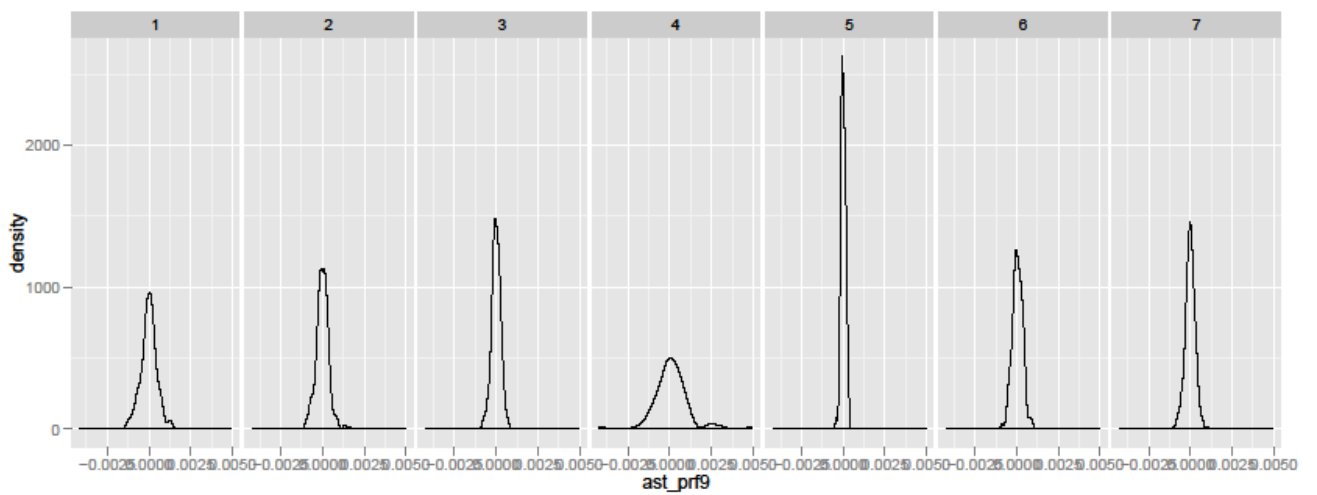
Variable 52 : Profile curvature pour le MNT ASTER à l'échelle 33



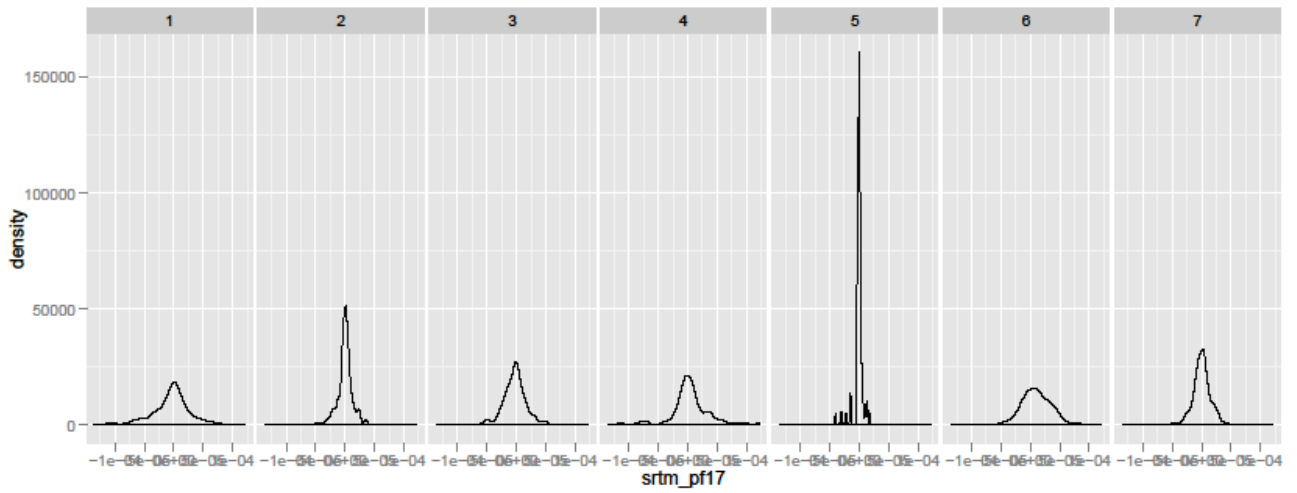
Variable 53 : Profile courvure pour le MNT ASTER à l'échelle 5



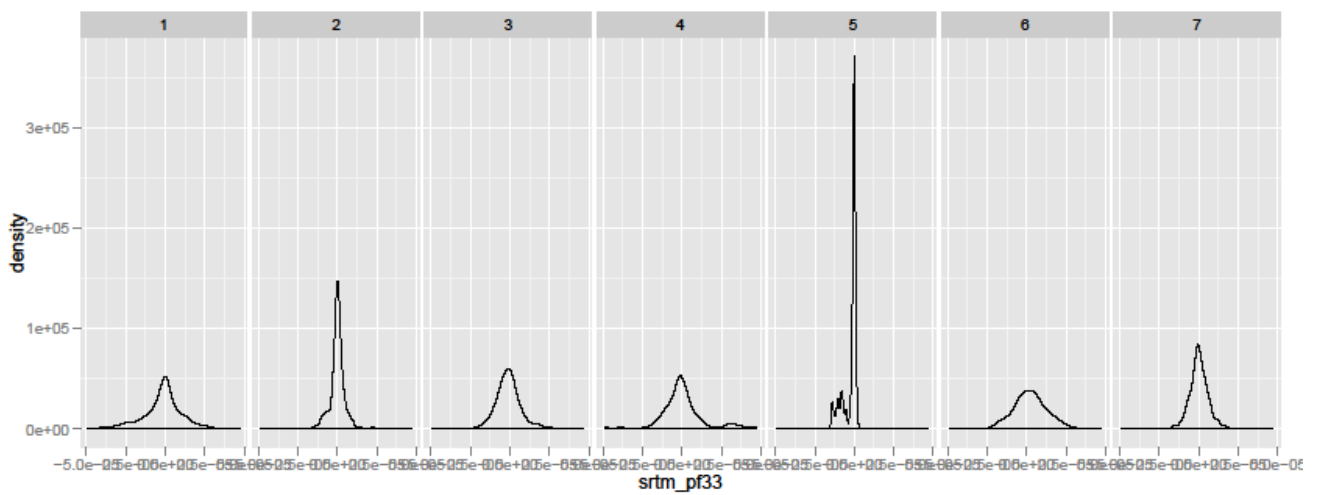
Variable 54 : Profile courvure pour le MNT ASTER à l'échelle 3



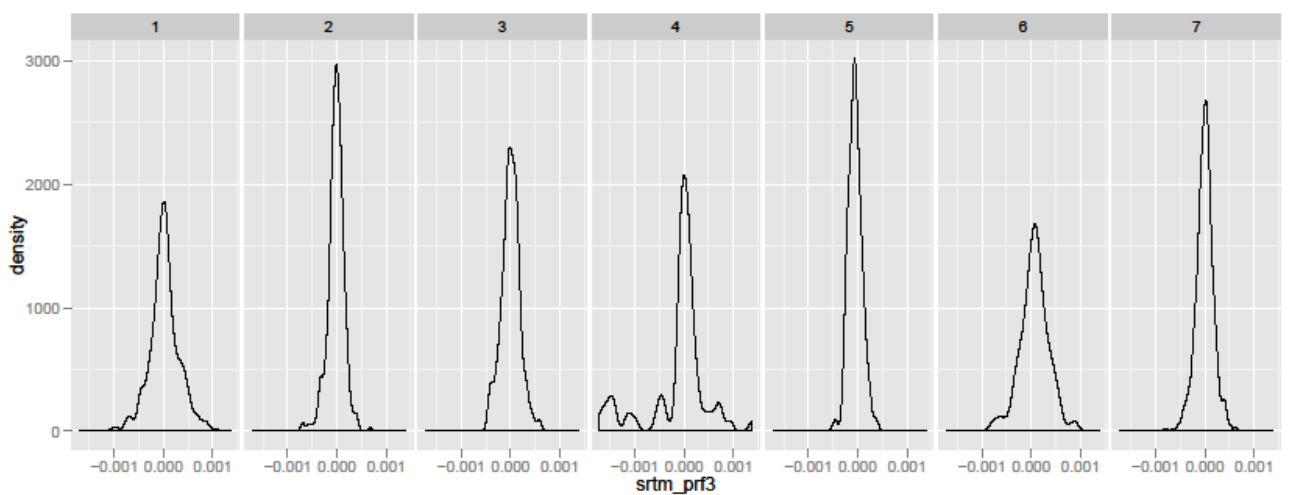
Variable 55 : Profile courvure pour le MNT ASTER à l'échelle 9



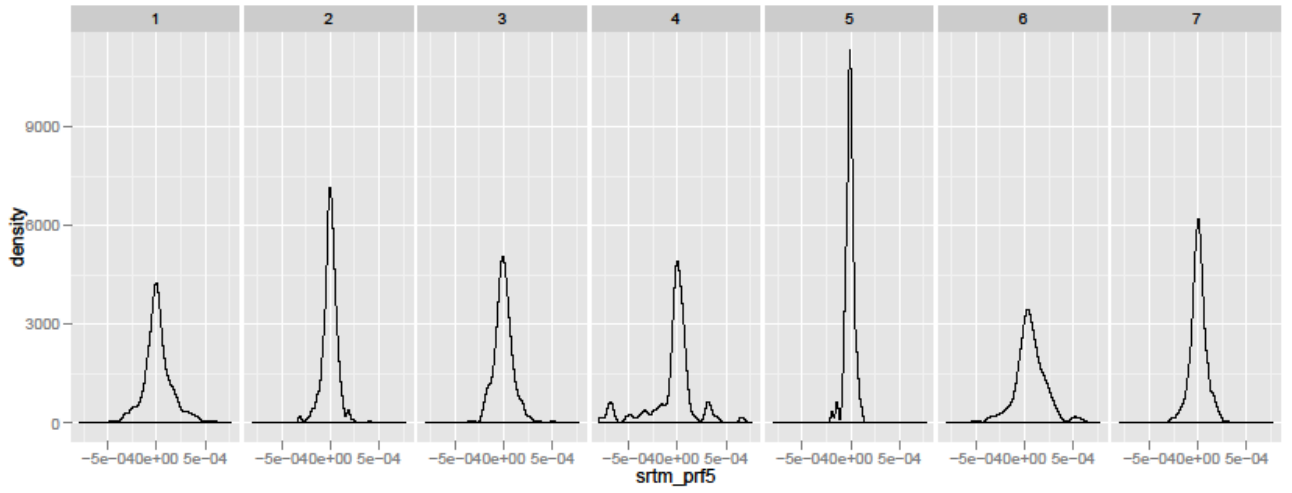
Variable 56 : Profile courvure pour le MNT SRTM à l'échelle 17



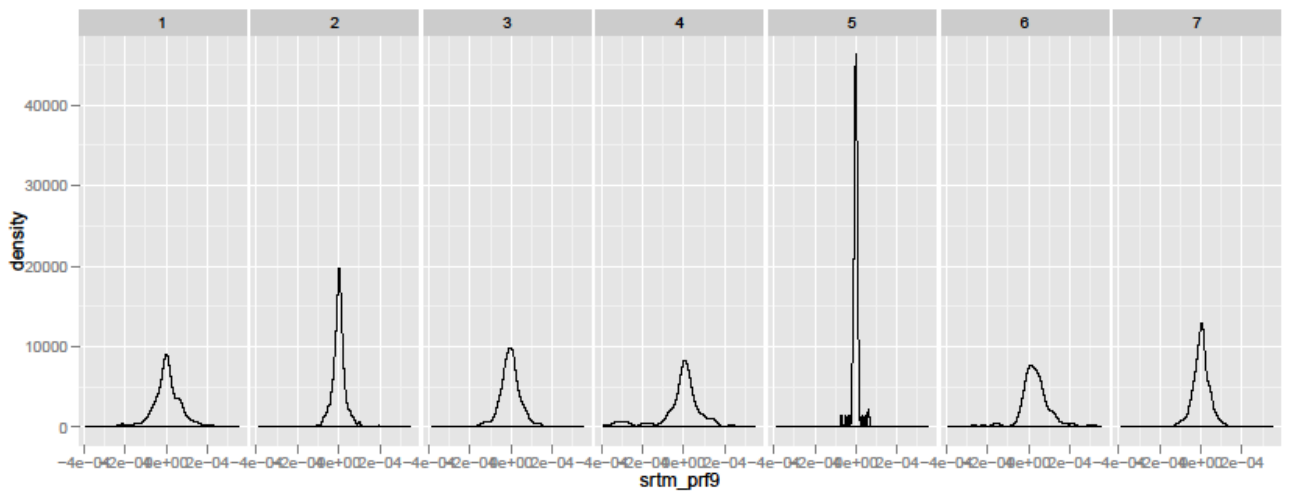
Variable 57 : Profile courvure pour le MNT SRTM à l'échelle 33



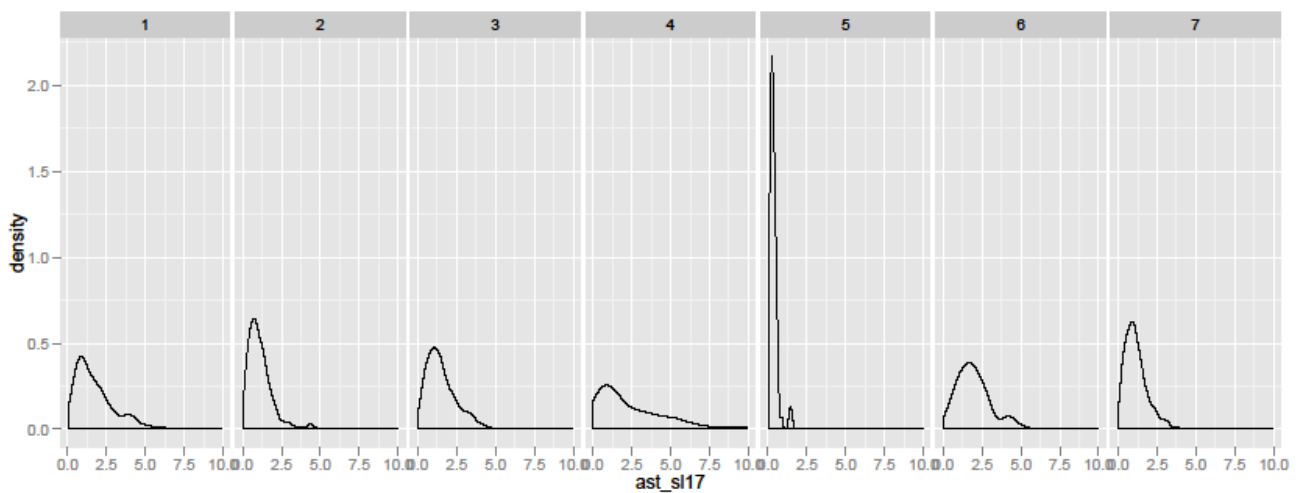
Variable 58 : Profile courvure pour le MNT SRTM à l'échelle 3



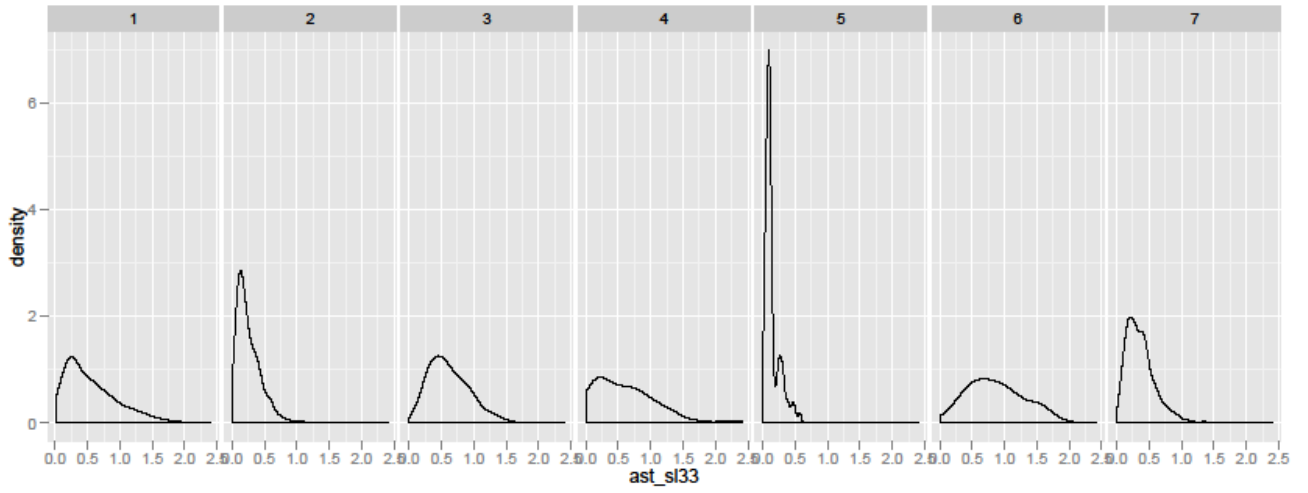
Variable 59 : Profile curvature pour le MNT SRTM à l'échelle 5



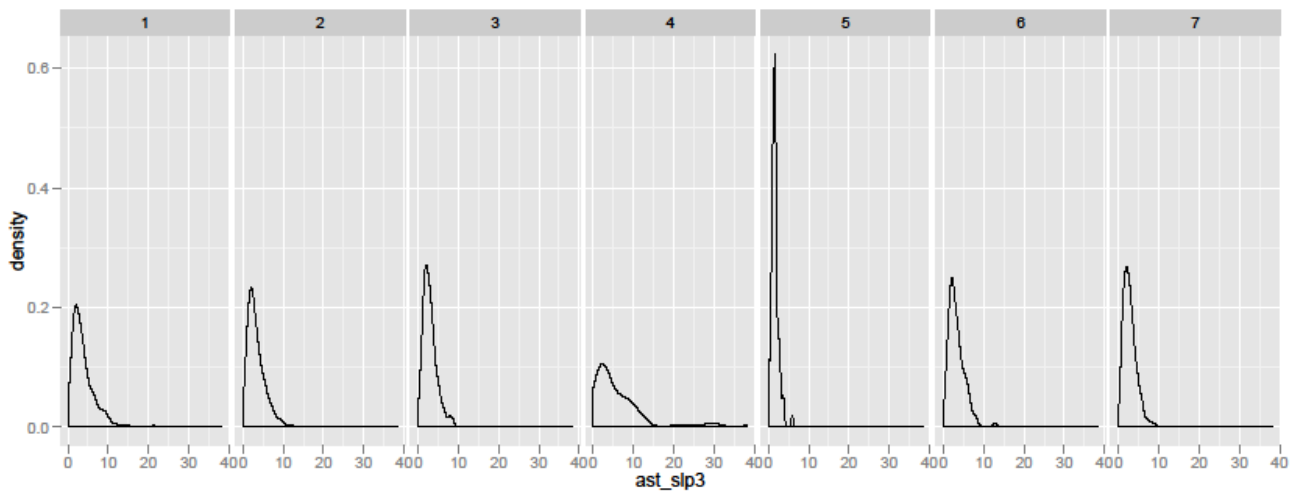
Variable 60 : Profile curvature pour le MNT SRTM à l'échelle 9



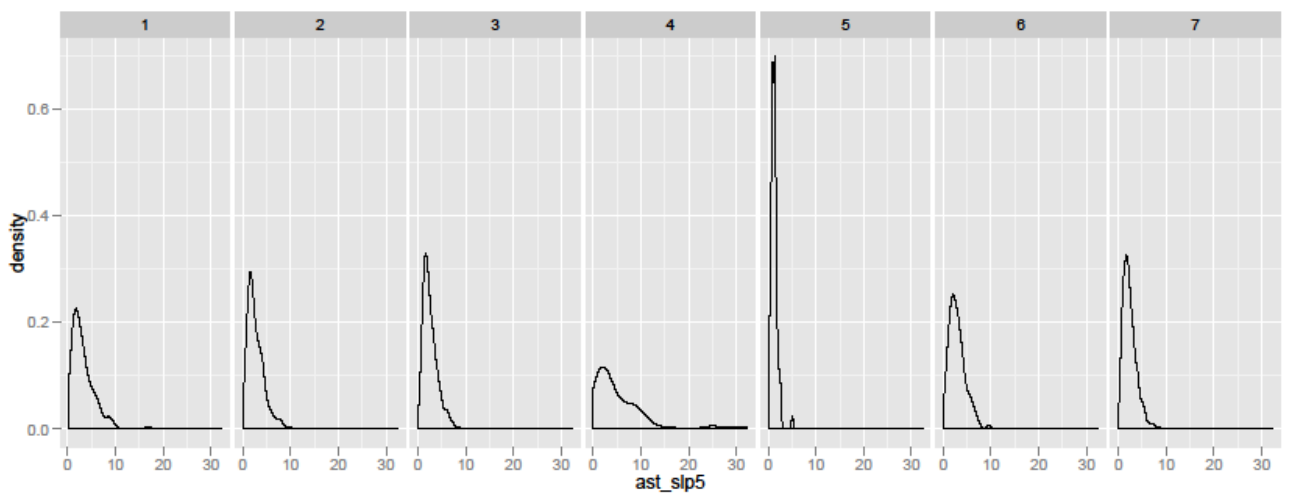
Variable 61 : Slope pour le MNT ASTER à l'échelle 17



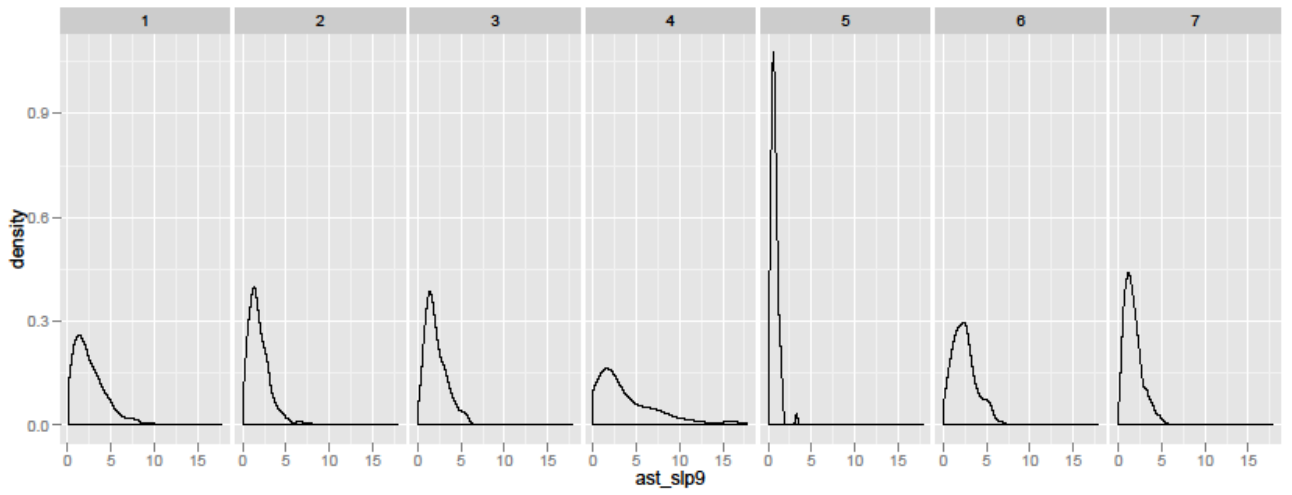
Variable 62 : Slope pour le MNT ASTER à l'échelle 33



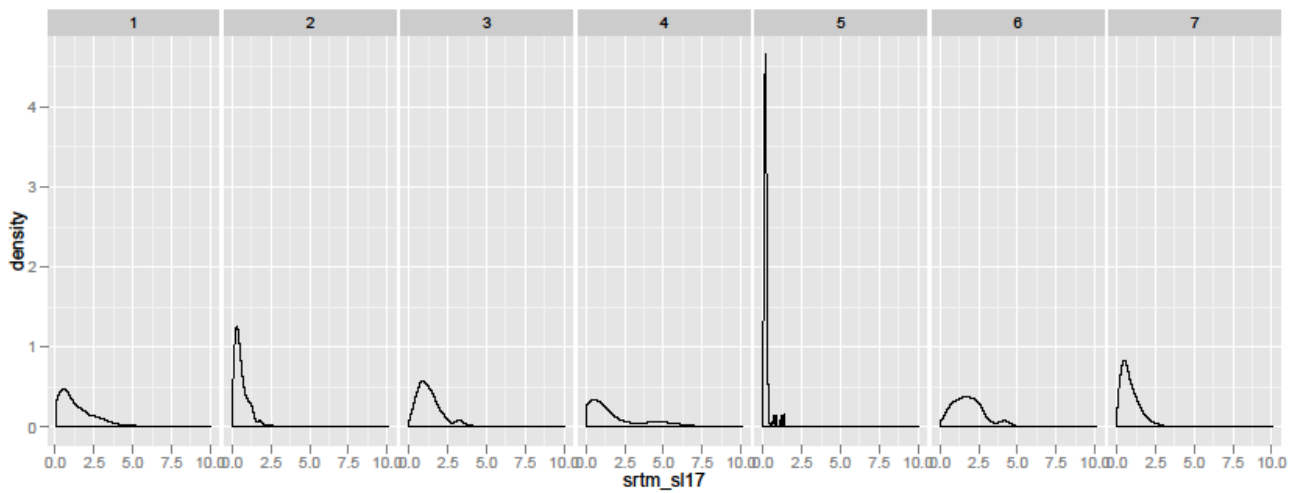
Variable 63 : Slope pour le MNT ASTER à l'échelle 3



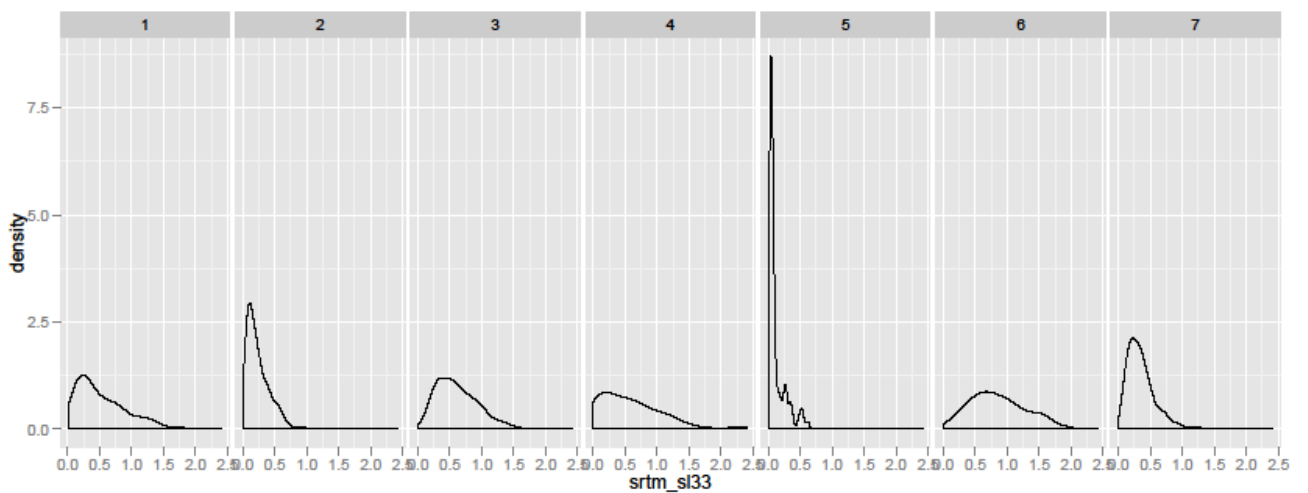
Variable 64 : Slope pour le MNT ASTER à l'échelle 5



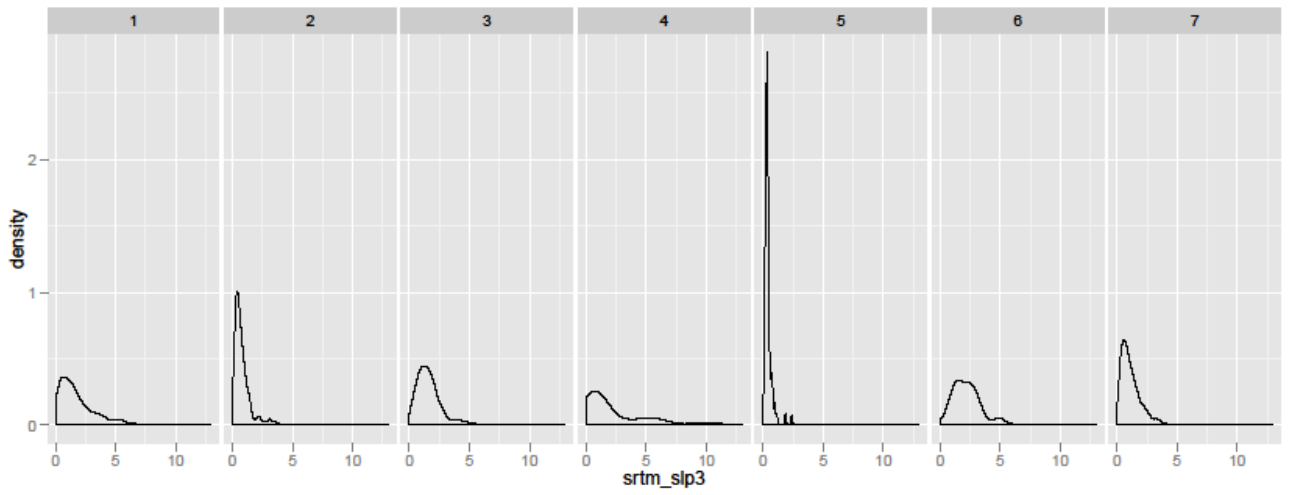
Variable 65 : Slope pour le MNT ASTER à l'échelle 9



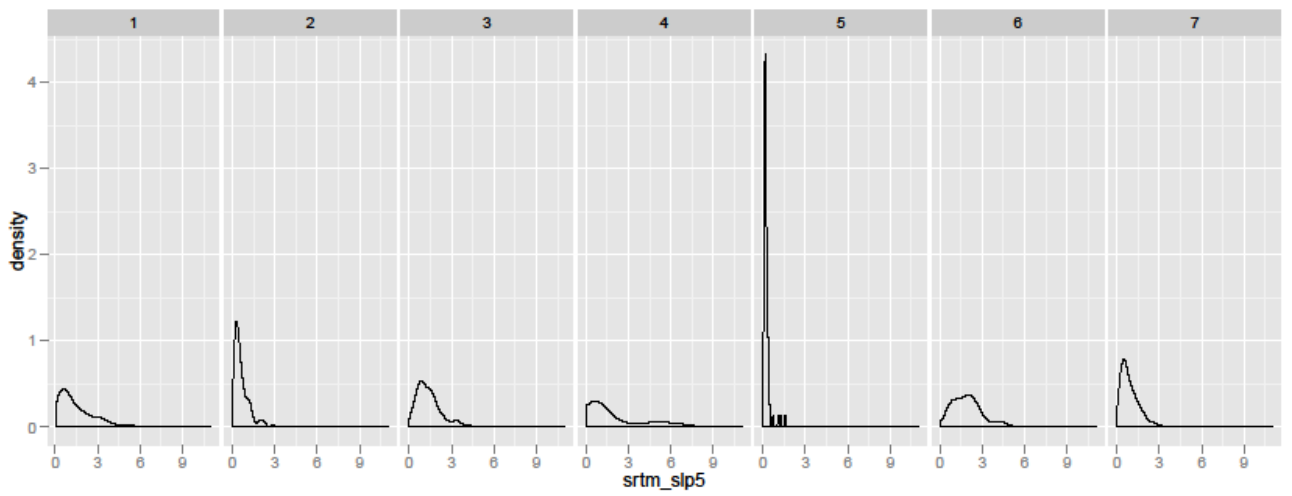
Variable 66 : Slope pour le MNT SRTM à l'échelle 17



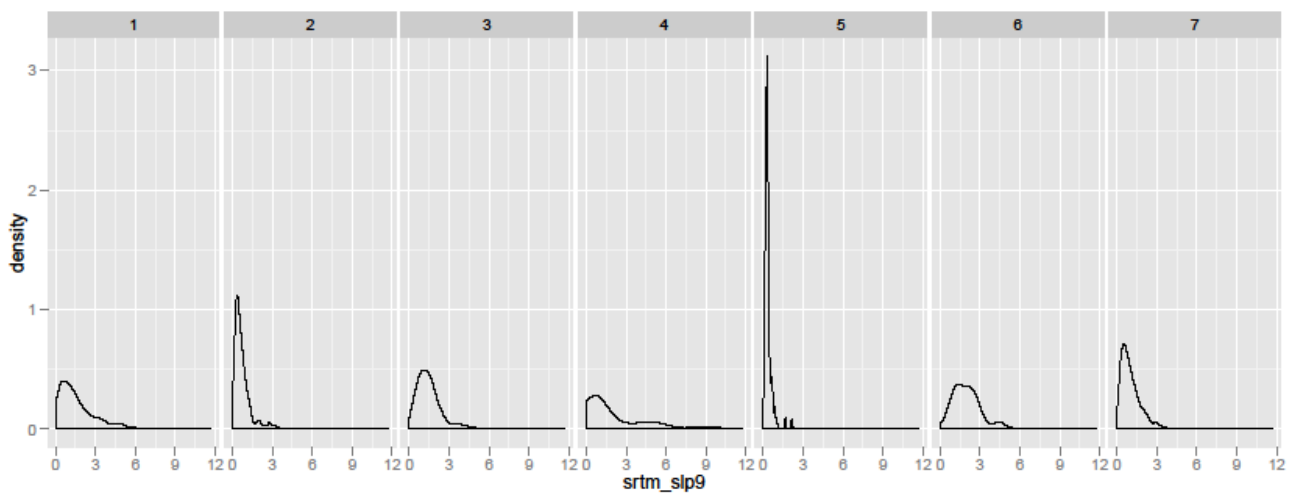
Variable 67: Slope pour le MNT SRTM à l'échelle 33



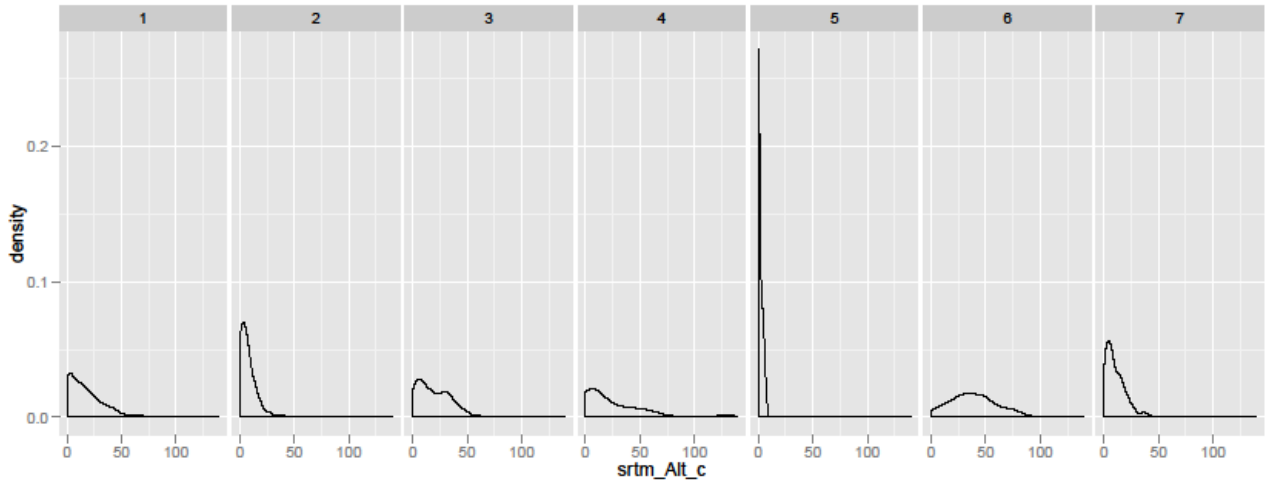
Variable 68: Slope pour le MNT SRTM à l'échelle 3



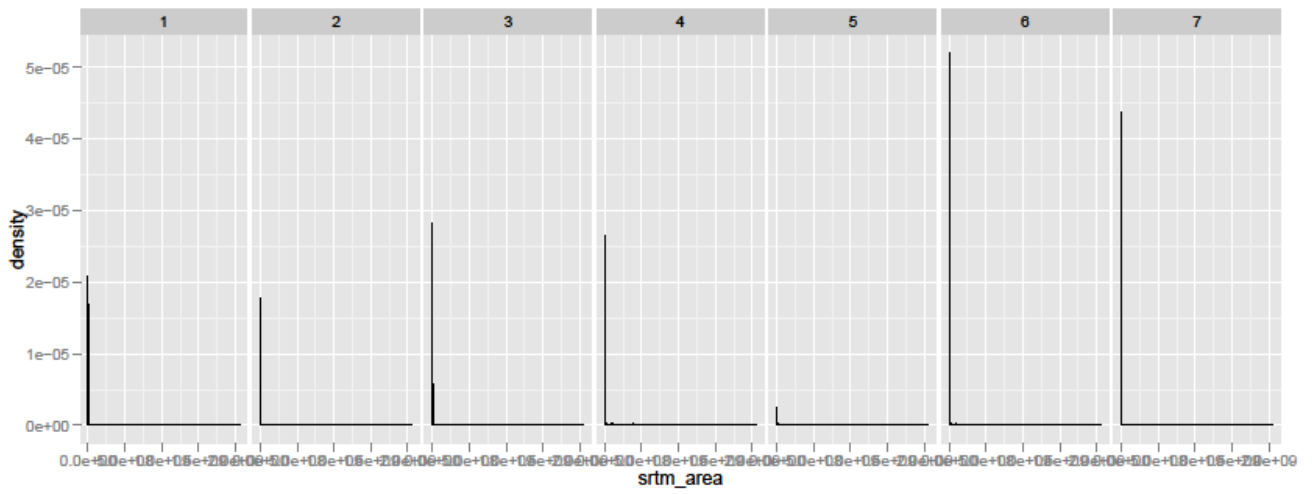
Variable 69: Slope pour le MNT SRTM à l'échelle 5



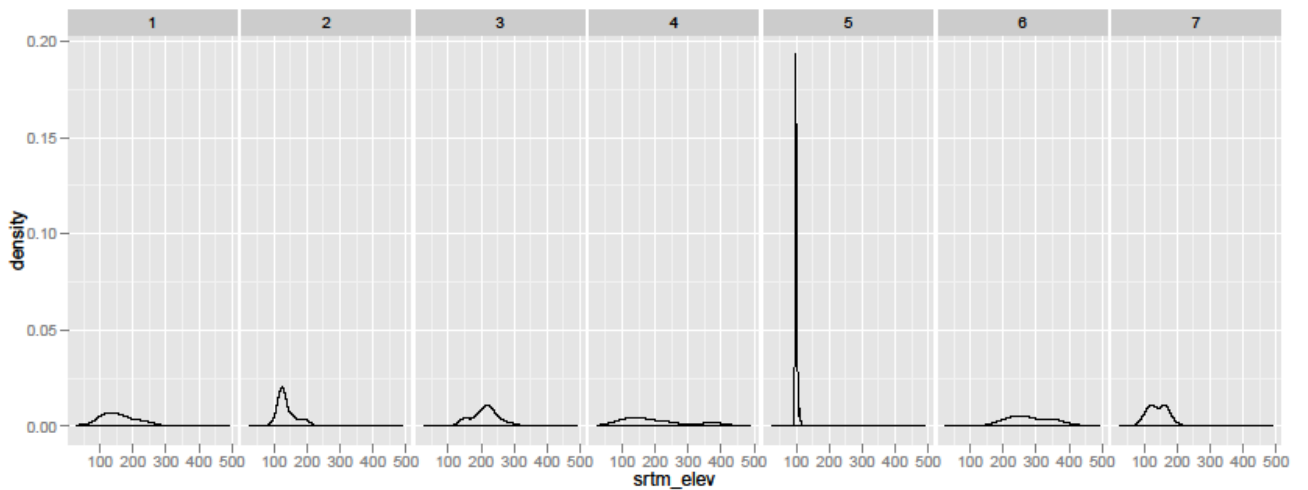
Variable 70: Slope pour le MNT SRTM à l'échelle 9



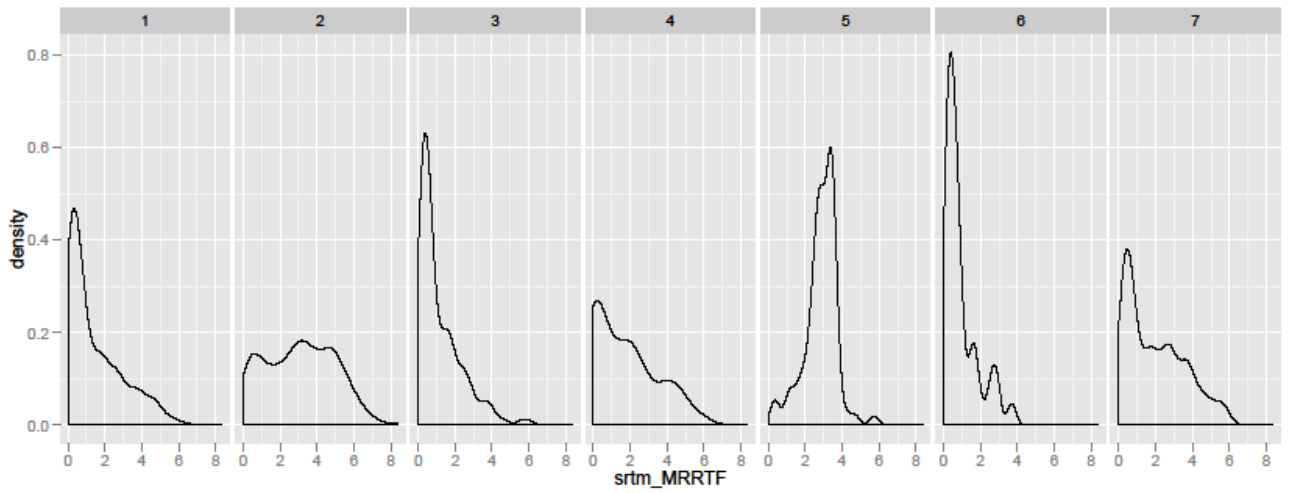
Variable 71 : Altitude above channel network pour le MNT SRTM



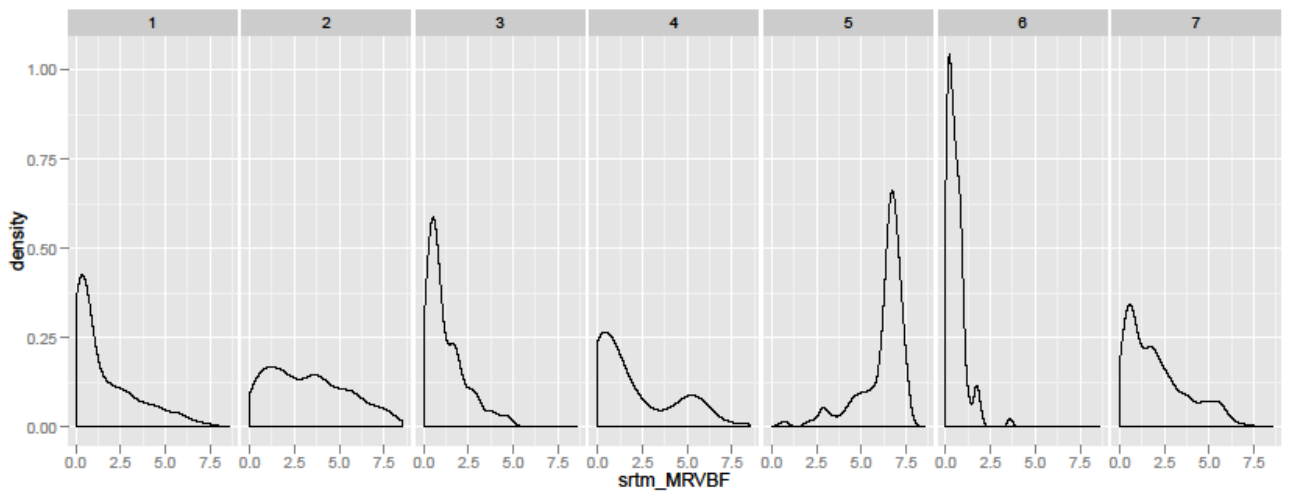
Variable 72 : Catchment area pour le MNT SRTM



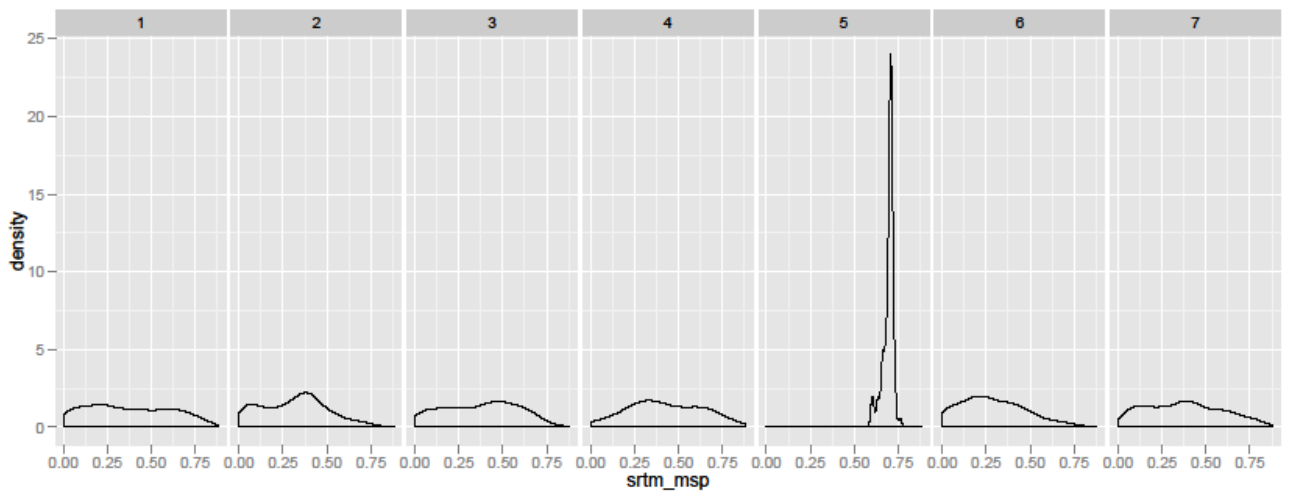
Variable 73 : Elevation pour le MNT SRTM



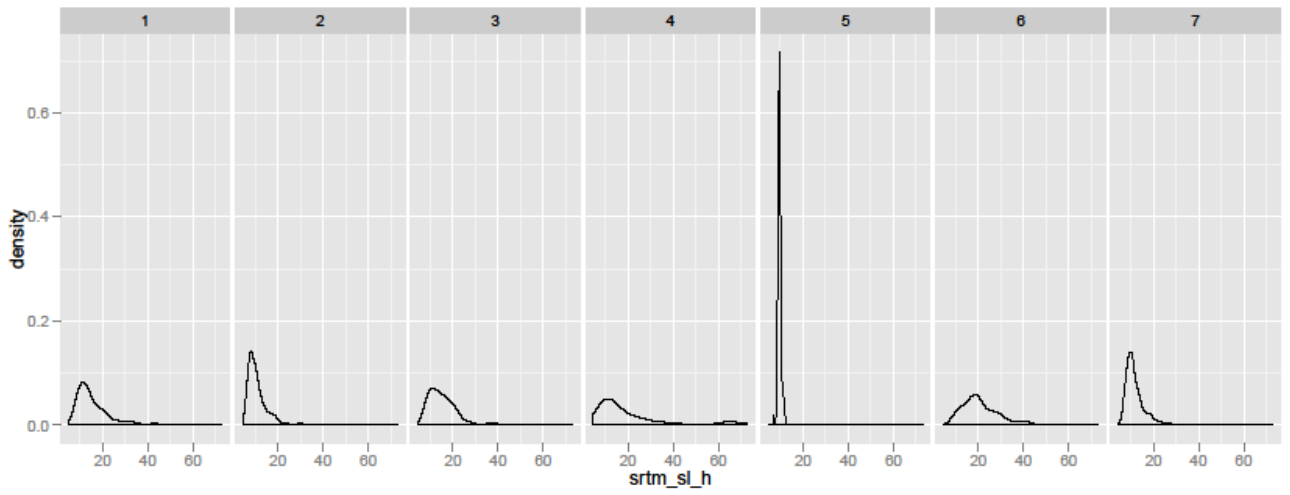
Variable 74 : Multi-resolution index of valley bottom flatness pour le MNT SRTM



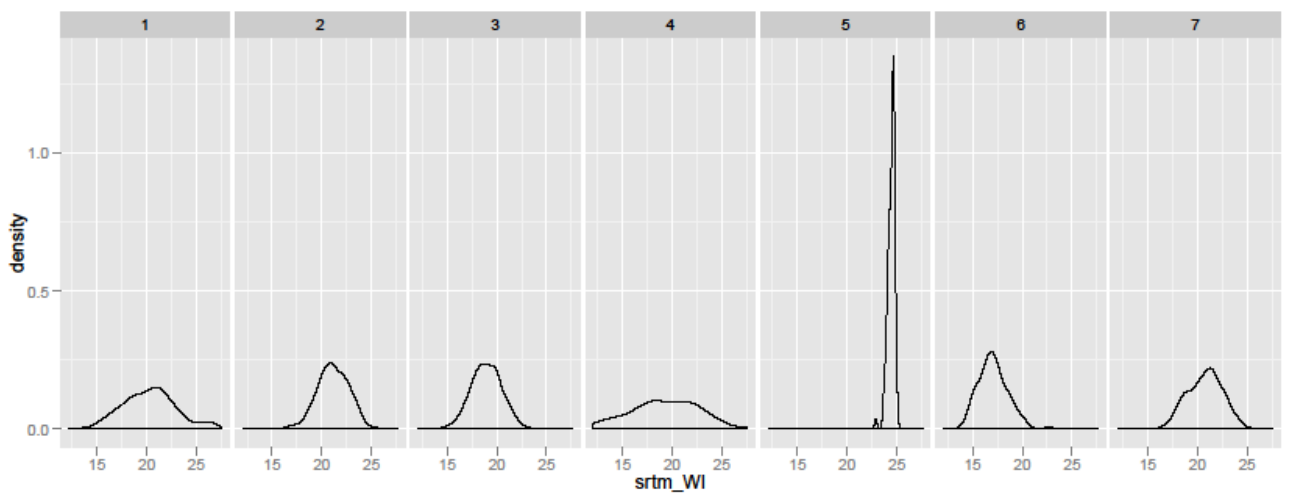
Variable 75 : Multi-resolution valley bottom flatness



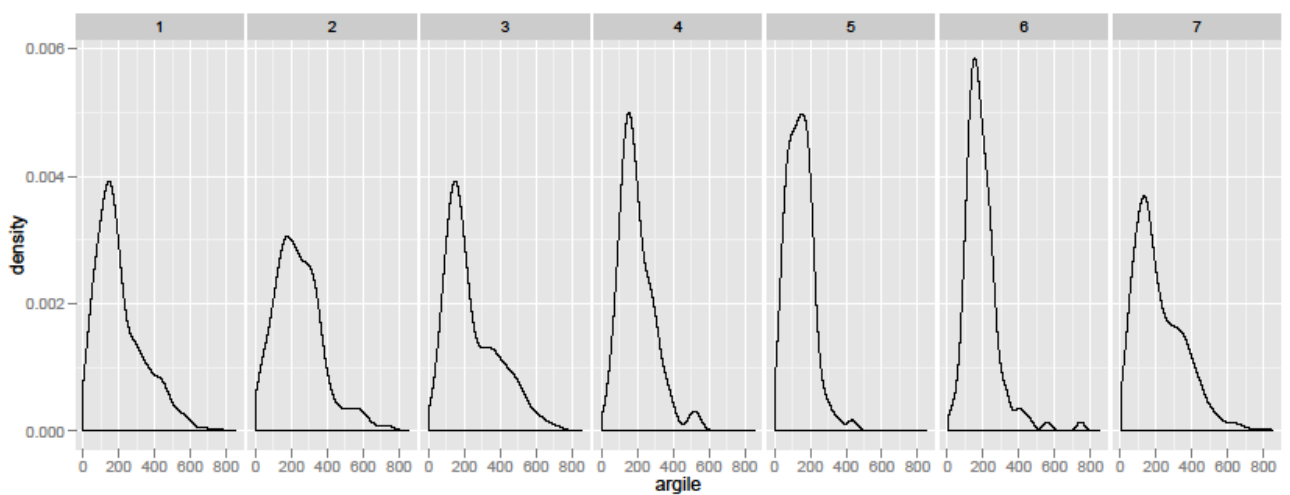
Variable 76 : Mid-Slope Resolution pour le MNT SRTM



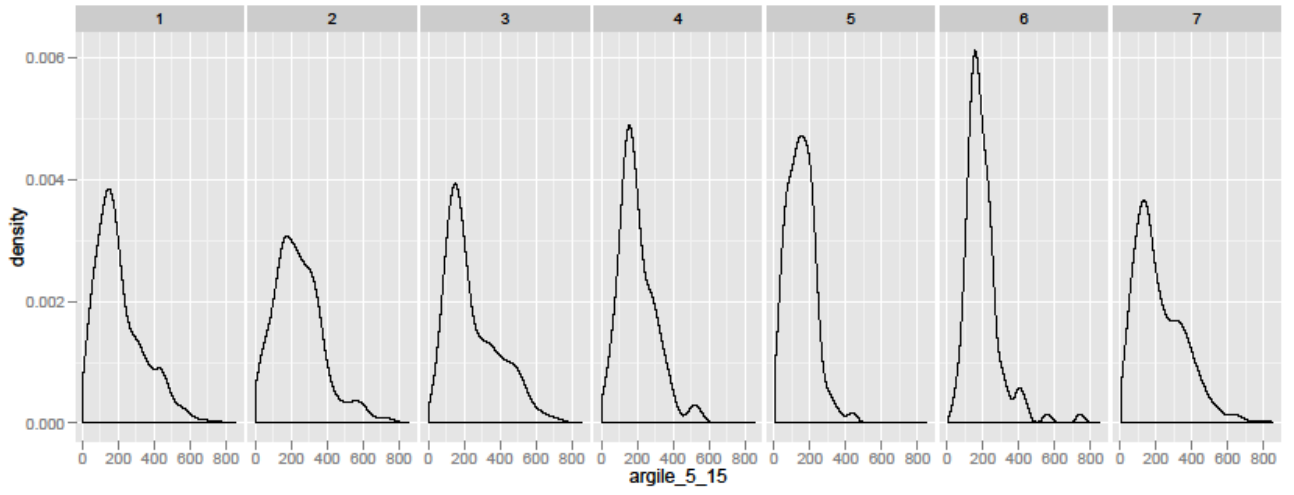
Variable 77 : Slope height pour le MNT SRTM



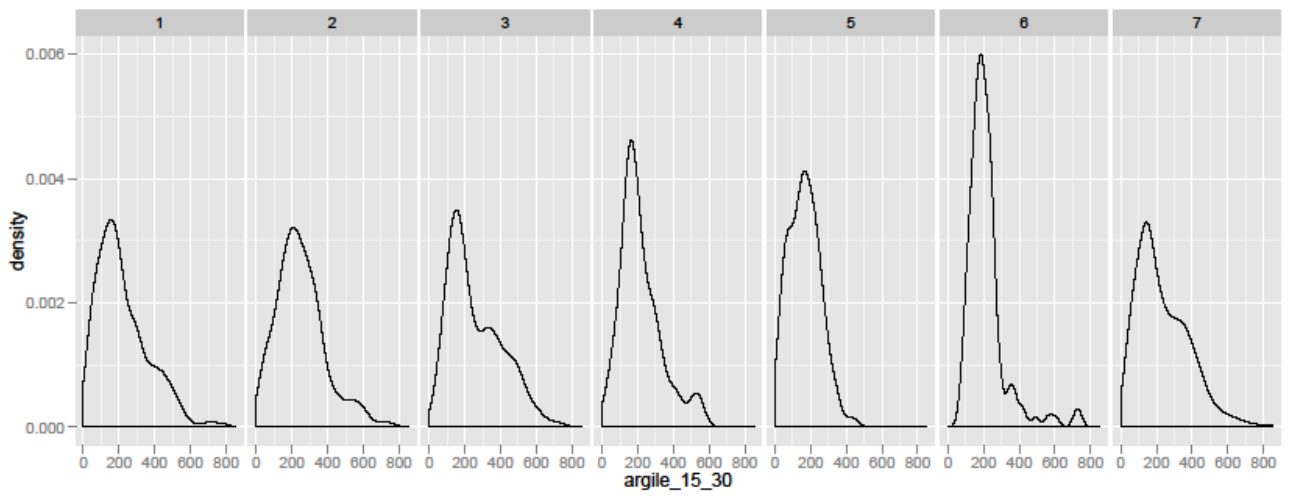
Variable 78 : Wetness index pour le MNT SRTM



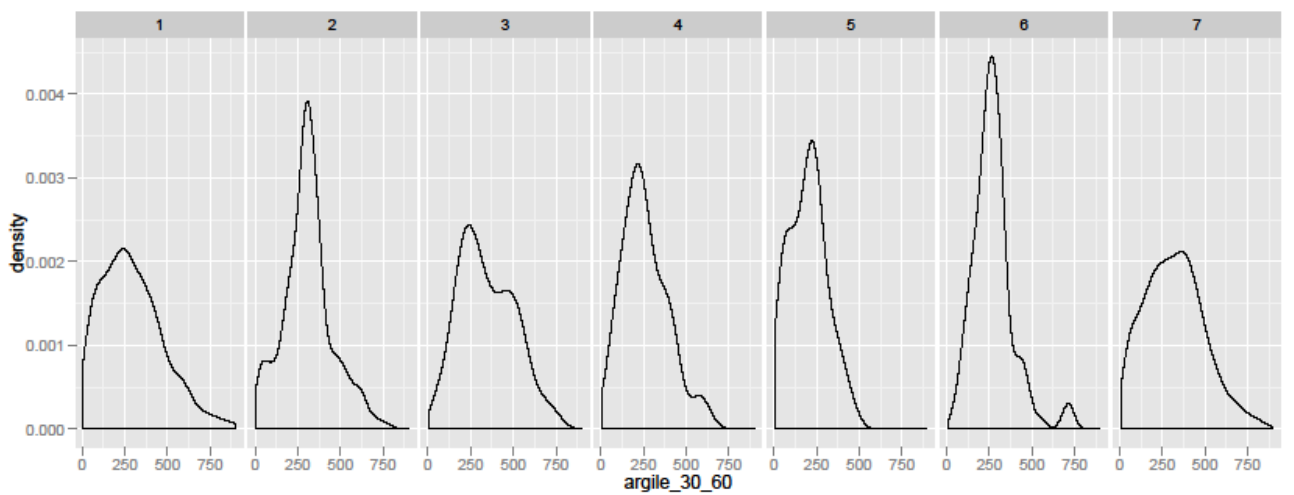
Variable 79 : Argile en surface



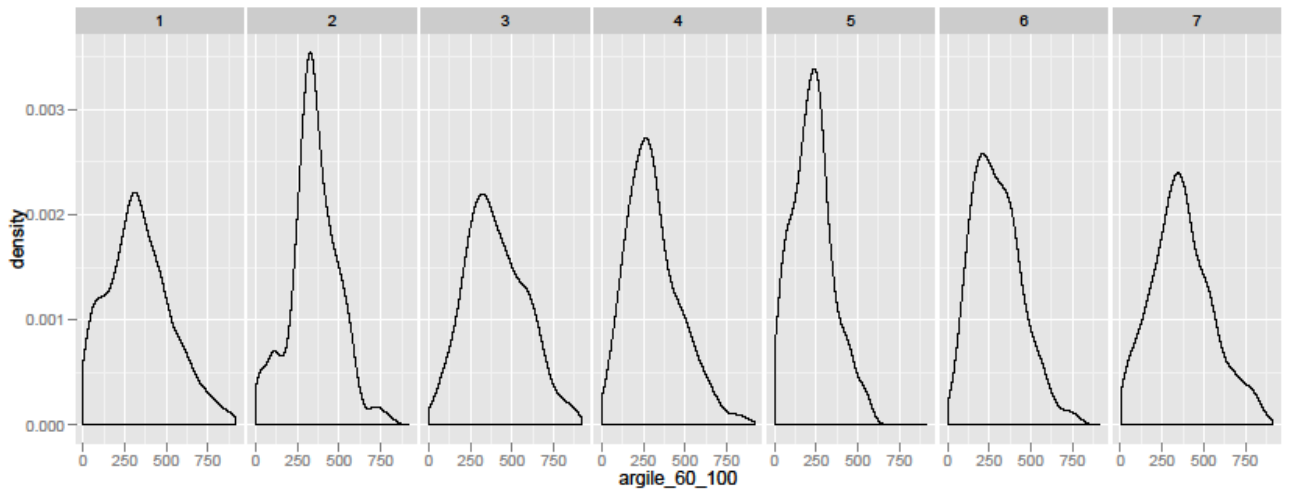
Variable 80 : Argile entre 5 et 15 cm



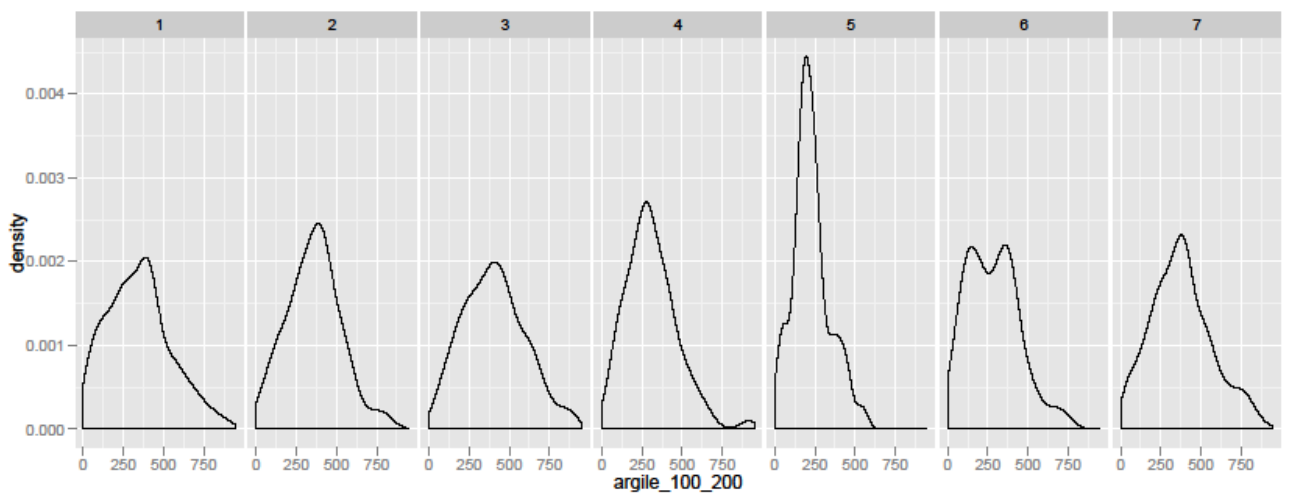
Variable 81 : Argile entre 15 et 30 cm



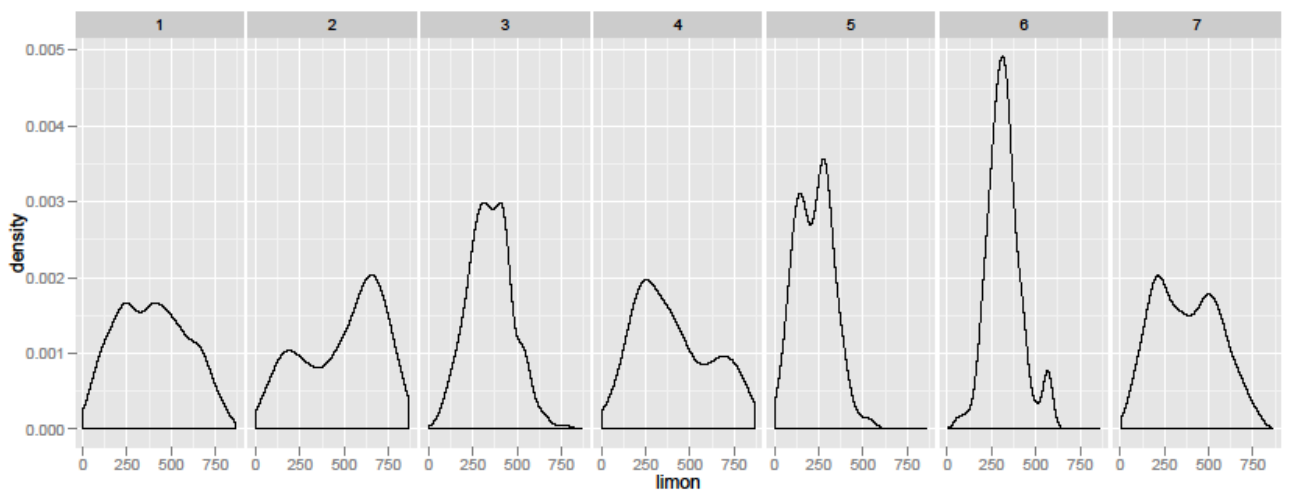
Variable 82 : Argile entre 30 et 60 cm



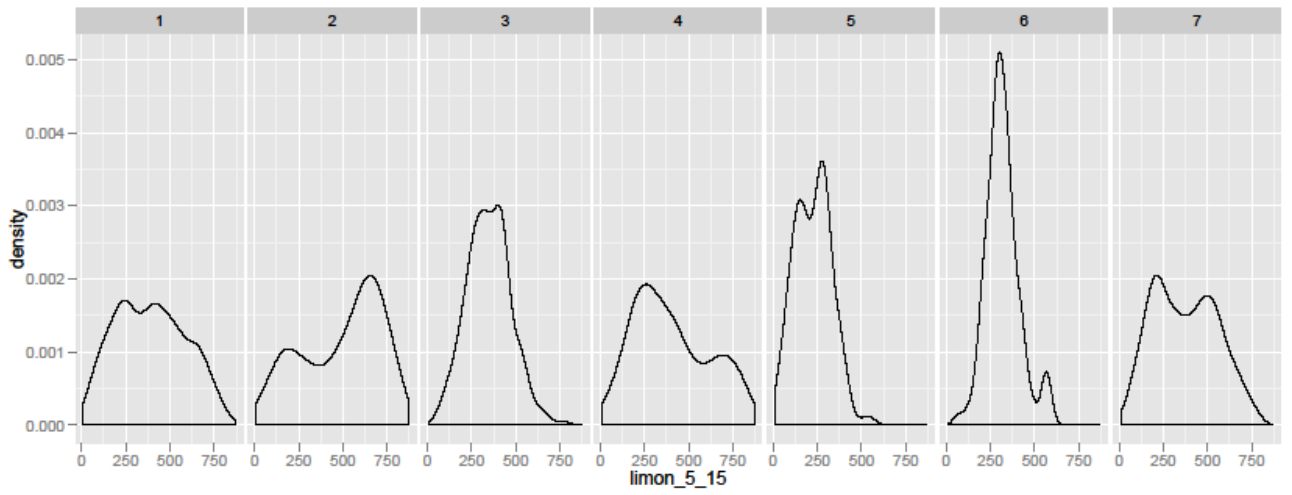
Variable 83 : Argile entre 60 et 100 cm



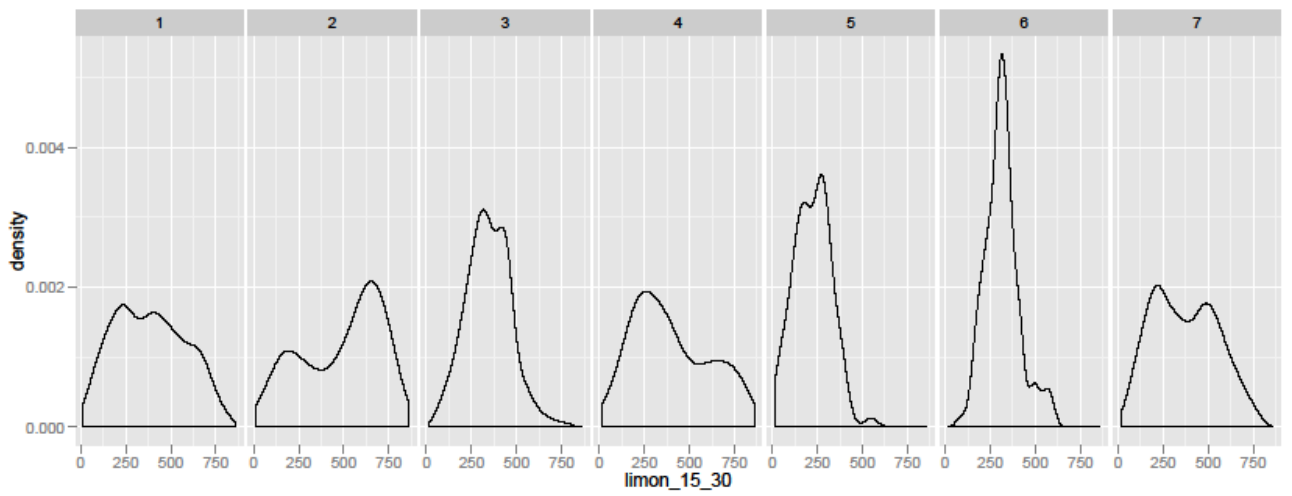
Variable 84 : Argile entre 100 et 200 cm



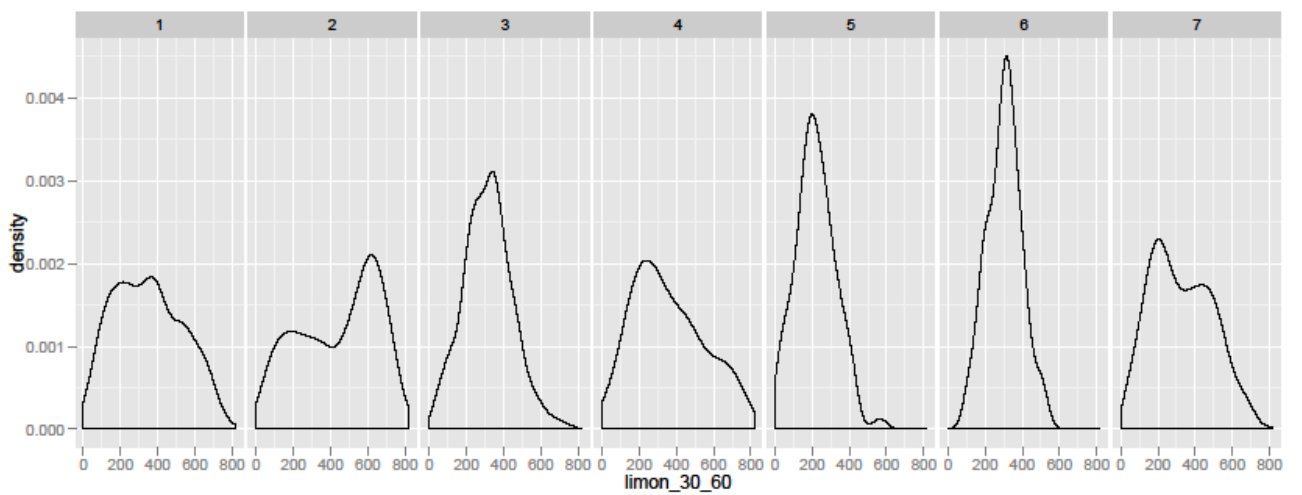
Variable 85 : Limon en surface



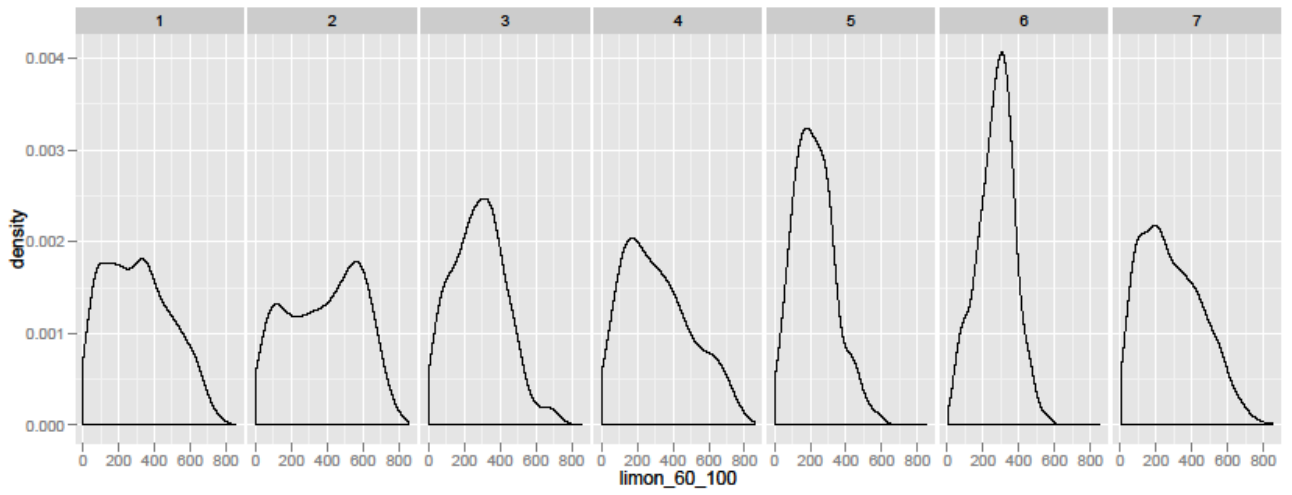
Variable 86 : Limon entre 5 et 15 cm



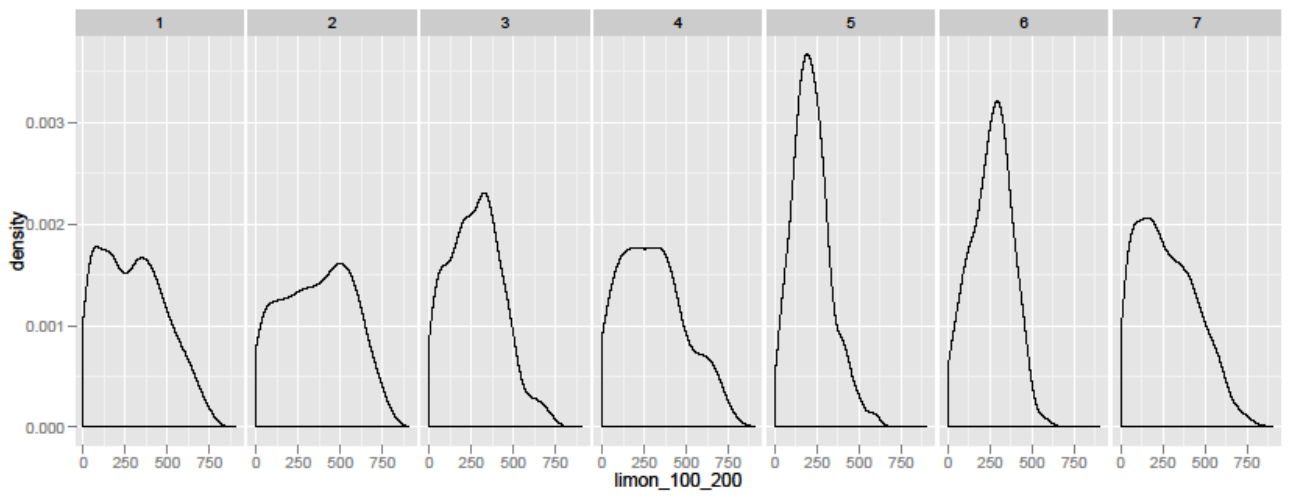
Variable 87 : Limon entre 15 et 30 cm



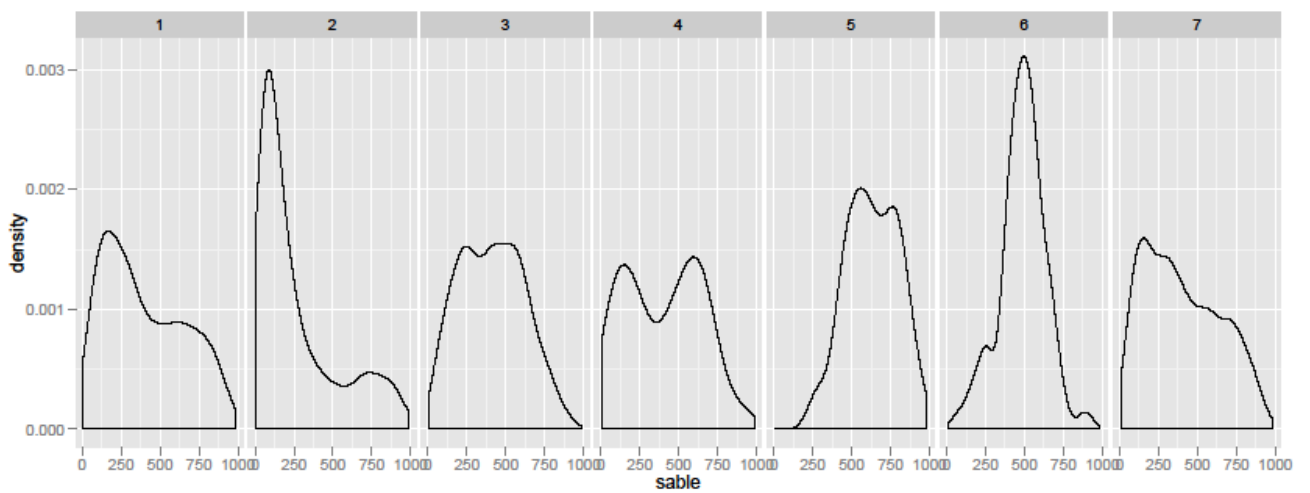
Variable 88 : Limon entre 30 et 60 cm



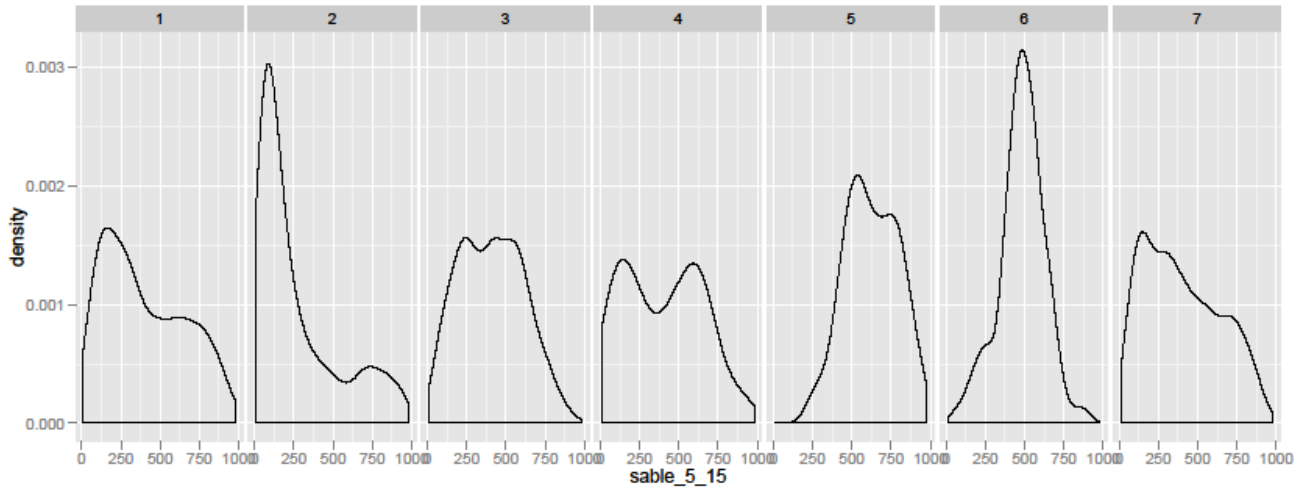
Variable 89 : Limon entre 60 et 100 cm



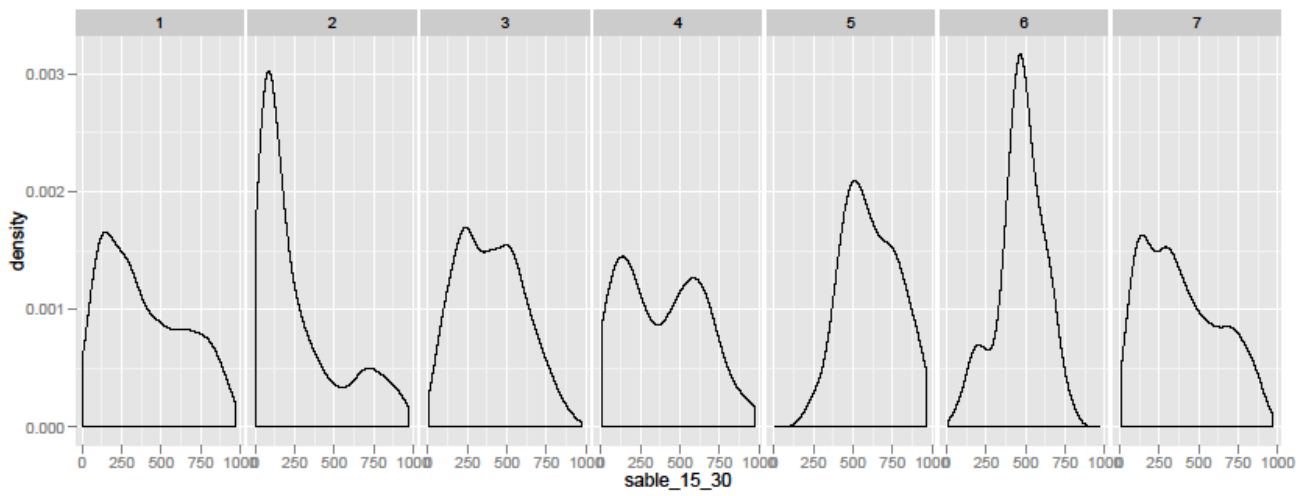
Variable 90 : Limon entre 100 et 200 cm



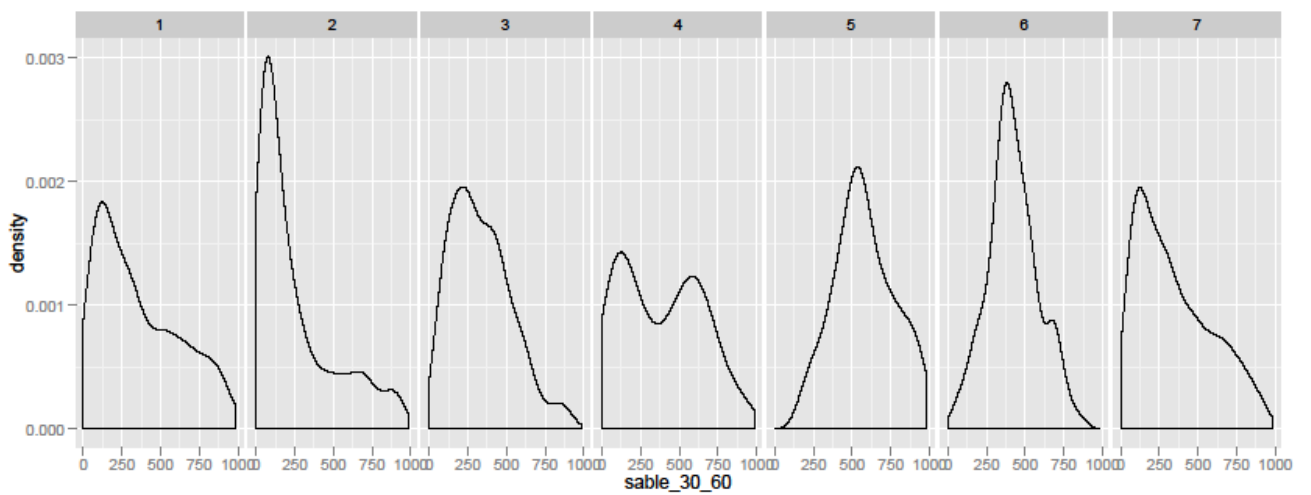
Variable 91 : Sable en surface



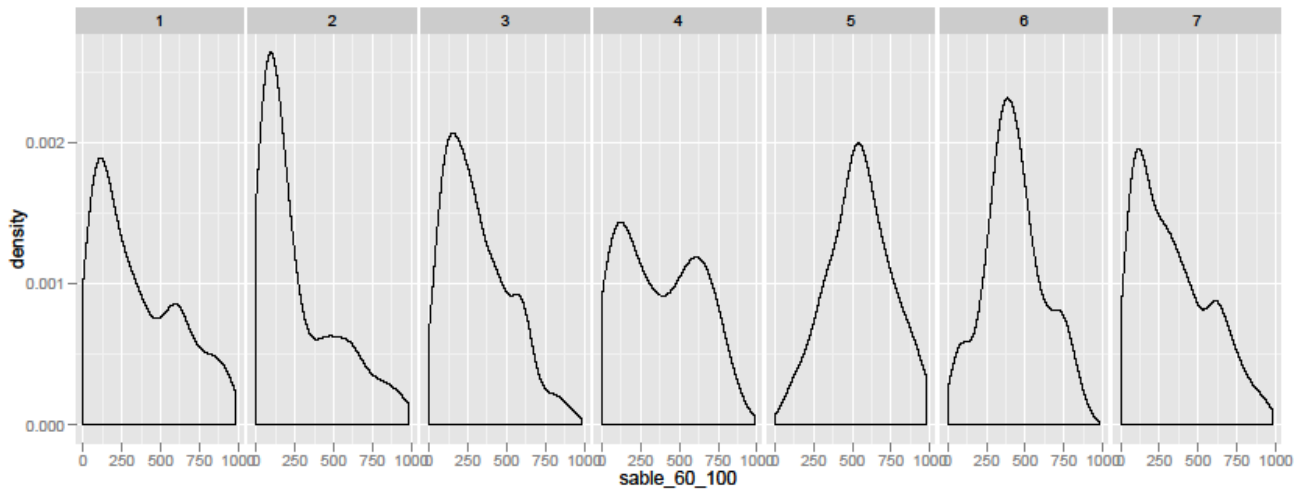
Variable 92 : Sable entre 5 et 15 cm



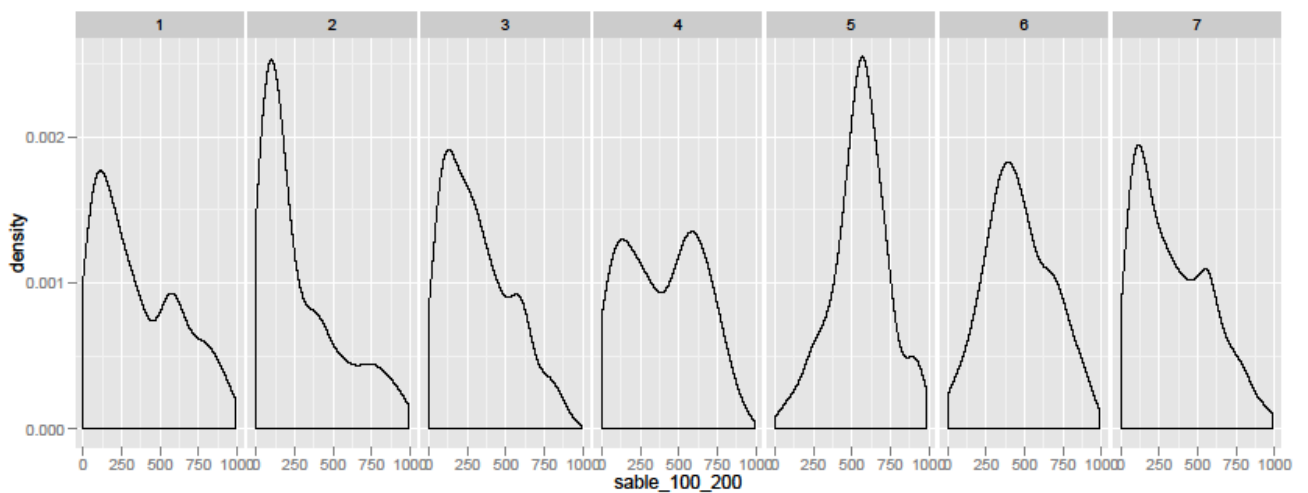
Variable 93 : Sable entre 15 et 30 cm



Variable 94 : Sable entre 30 et 60 cm



Variable 95 : Sable entre 60 et 100 cm



Variable 96 : Sable entre 100 et 200 cm

3-2 Centre des clusters

Cluster 1		Cluster 2		Cluster 3	
argile	2.064375e+02	argile	2.410228e+02	argile	2.408256e+02
limon	3.922288e+02	limon	4.928812e+02	limon	3.486930e+02
sable	4.013334e+02	sable	2.660949e+02	sable	4.104825e+02
srtm_elev	1.581376e+02	srtm_elev	1.396151e+02	srtm_elev	2.139146e+02
srtm_slp3	1.823639e+00	srtm_slp3	8.108921e-01	srtm_slp3	1.629144e+00
srtm_asp3	-2.507187e+00	srtm_asp3	7.006309e+00	srtm_asp3	3.952033e+01
srtm_prf3	2.500411e-05	srtm_prf3	-1.837539e-05	srtm_prf3	4.020325e-06
srtm_pln3	-2.137448e-03	srtm_pln3	-7.240379e-05	srtm_pln3	4.089512e-04

srtm_lng3	2.513552e-05	srtm_lng3	-1.838801e-05	srtm_lng3	4.065041e-06
srtm_crs3	3.799795e-05	srtm_crs3	5.268139e-06	srtm_crs3	-1.182927e-06
srtm_slp5	1.498530e+00	srtm_slp5	6.505148e-01	srtm_slp5	1.401133e+00
srtm_asp5	-9.394251e+00	srtm_asp5	3.880126e-01	srtm_asp5	4.393089e+01
srtm_prf5	1.451745e-05	srtm_prf5	-1.974763e-06	srtm_prf5	-1.512195e-06
srtm_pln5	-6.964333e-04	srtm_pln5	4.321767e-06	srtm_pln5	-2.334959e-05
srtm_lng5	7.512936e-05	srtm_lng5	-3.614511e-05	srtm_lng5	1.045935e-05
srtm_crs5	1.023203e-04	srtm_crs5	2.716088e-06	srtm_crs5	-5.577236e-06
srtm_slp9	1.641712e+00	srtm_slp9	7.298478e-01	srtm_slp9	1.466428e+00
srtm_asp9	-8.628337e+00	srtm_asp9	-3.129338e+00	srtm_asp9	4.203252e+01
srtm_prf9	3.940452e-06	srtm_prf9	9.873817e-07	srtm_prf9	-2.597561e-06
srtm_pln9	-5.395996e-04	srtm_pln9	-3.139495e-04	srtm_pln9	-2.781301e-05
srtm_lng9	1.058727e-05	srtm_lng9	-1.140694e-05	srtm_lng9	-2.357724e-07
srtm_crs9	2.259959e-05	srtm_crs9	4.504732e-06	srtm_crs9	1.731707e-06
srtm_sl17	1.402143e+00	srtm_sl17	6.008102e-01	srtm_sl17	1.333206e+00
srtm_as17	-2.531828e+01	srtm_as17	-1.031230e+01	srtm_as17	4.885772e+01
srtm_pf17	-4.765914e-06	srtm_pf17	1.955836e-07	srtm_pf17	-2.556911e-06
srtm_pl17	-8.058994e-04	srtm_pl17	-2.571861e-04	srtm_pl17	-2.747805e-04
srtm_ln17	-4.761807e-06	srtm_ln17	1.955836e-07	srtm_ln17	-2.552846e-06
srtm_cr17	6.030801e-06	srtm_cr17	2.293375e-06	srtm_cr17	4.447154e-06
srtm_sl33	5.426476e-01	srtm_sl33	2.386518e-01	srtm_sl33	6.283235e-01
srtm_as33	-1.836140e+01	srtm_as33	1.258675e+00	srtm_as33	5.004878e+01
srtm_pf33	-1.546201e-06	srtm_pf33	-5.678233e-08	srtm_pf33	-8.292683e-07
srtm_pl33	-5.241889e-05	srtm_pl33	-2.208612e-04	srtm_pl33	-1.044878e-04
srtm_ln33	-1.546201e-06	srtm_ln33	-5.678233e-08	srtm_ln33	-8.211382e-07
srtm_cr33	6.509240e-07	srtm_cr33	8.296530e-07	srtm_cr33	1.260163e-06
aster_el	1.511602e+02	aster_el	1.336498e+02	aster_el	2.072114e+02
ast_slp3	3.923506e+00	ast_slp3	3.319709e+00	ast_slp3	2.914375e+00
ast_asp3	-1.849897e+01	ast_asp3	9.725552e+00	ast_asp3	2.586992e+01
ast_prf3	-5.014374e-05	ast_prf3	4.282650e-05	ast_prf3	1.442561e-04
ast_pln3	-3.684345e-03	ast_pln3	2.132379e-03	ast_pln3	-3.350805e-03
ast_lng3	-5.927310e-05	ast_lng3	4.405678e-05	ast_lng3	1.456545e-04
ast_crs3	1.353060e-04	ast_crs3	-7.443218e-05	ast_crs3	8.616667e-05
ast_slp5	3.332188e+00	ast_slp5	2.738902e+00	ast_slp5	2.575513e+00
ast_asp5	-2.687885e+01	ast_asp5	-4.350158e+00	ast_asp5	2.949187e+01
ast_prf5	-5.673101e-05	ast_prf5	5.693060e-05	ast_prf5	3.432114e-05
ast_pln5	-6.271585e-03	ast_pln5	1.392694e-03	ast_pln5	9.221951e-05

ast_lng5	-5.927105e-05	ast_lng5	5.771609e-05	ast_lng5	3.445122e-05
ast_crs5	7.964682e-05	ast_crs5	-3.496530e-05	ast_crs5	1.897154e-05
ast_slp9	2.546107e+00	ast_slp9	1.918548e+00	ast_slp9	2.096802e+00
ast_asp9	-1.673511e+01	ast_asp9	2.372240e+00	ast_asp9	3.043902e+01
ast_prf9	-1.494867e-06	ast_prf9	-8.977918e-06	ast_prf9	2.004065e-06
ast_pln9	-3.352667e-03	ast_pln9	-5.123849e-04	ast_pln9	-3.918049e-04
ast_lng9	-1.552361e-06	ast_lng9	-8.981073e-06	ast_lng9	2.040650e-06
ast_crs9	3.578029e-05	ast_crs9	6.227129e-06	ast_crs9	2.567073e-05
ast_sl17	1.764166e+00	ast_sl17	1.130750e+00	ast_sl17	1.514320e+00
ast_as17	-2.057290e+01	ast_as17	-7.921136e+00	ast_as17	4.448780e+01
ast_pf17	-5.652977e-06	ast_pf17	-9.652997e-07	ast_pf17	-3.008130e-06
ast_pl17	-4.078501e-04	ast_pl17	-2.819621e-04	ast_pl17	3.367480e-05
ast_ln17	-5.642710e-06	ast_ln17	-9.652997e-07	ast_ln17	-3.008130e-06
ast_cr17	5.751540e-06	ast_cr17	2.826498e-06	ast_cr17	3.142276e-06
ast_sl33	5.705518e-01	ast_sl33	2.587834e-01	ast_sl33	6.360037e-01
ast_as33	-1.908008e+01	ast_as33	2.280757e+00	ast_as33	4.829675e+01
ast_pf33	-1.809035e-06	ast_pf33	-3.186120e-07	ast_pf33	-8.902439e-07
ast_pl33	-6.954004e-05	ast_pl33	-2.465931e-04	ast_pl33	-5.642683e-05
ast_ln33	-1.806982e-06	ast_ln33	-3.186120e-07	ast_ln33	-8.902439e-07
ast_cr33	9.630390e-07	ast_cr33	8.990536e-07	ast_cr33	1.056911e-06
srtm_WI	2.027606e+01	srtm_WI	2.114798e+01	srtm_WI	1.893626e+01
srtm_MRVPF	.623585e+00	srtm_MRVPF	.282226e+00	srtm_MRVPF	.232066e+00
srtm_MRRTF	1.445133e+00	srtm_MRRTF	3.074196e+00	srtm_MRRTF	1.158902e+00
srtm_sl_h	1.616710e+01	srtm_sl_h	1.087826e+01	srtm_sl_h	1.543972e+01
srtm_msp	3.725667e-01	srtm_msp	3.127129e-01	srtm_msp	3.684553e-01
srtm_area	1.446699e+07	srtm_area	1.770014e+06	srtm_area	4.716757e+05
srtm_Alt_c	1.687294e+01	srtm_Alt_c	7.773060e+00	srtm_Alt_c	1.926789e+01
uranium	1.544034e+00	uranium	1.704332e+00	uranium	1.825447e+00
ti	9.385862e+00	ti	9.810699e+00	ti	9.488400e+00
th	7.758184e+00	th	8.601284e+00	th	8.915513e+00
potassium	9.974421e-01	potassium	1.120274e+00	potassium	1.168789e+00
total	4.193946e+01	total	4.643675e+01	total	4.874429e+01
vig00	2.47077e-02	vig00	1.207691e-01	vig00	6.450449e-04
olea00	1.755349e-01	olea00	1.290463e-01	olea00	9.431347e-02
bois00	4.328042e-02	bois00	1.071348e-02	bois00	2.049782e-02
jach00	7.864291e-02	jach00	7.116143e-02	jach00	3.717491e-02
argile_5_15	2.078489e+02	argile_5_15	2.418403e+02	argile_5_15	2.439031e+02

limon_5_15	3.907979e+02	limon_5_15	4.919884e+02	limon_5_15	3.477890e+02
sable_5_15	4.013528e+02	sable_5_15	2.661706e+02	sable_5_15	4.083089e+02
argile_15_30	2.203661e+02	argile_15_30	2.495538e+02	argile_15_30	2.587317e+02
limon_15_30	3.852492e+02	limon_15_30	4.870558e+02	limon_15_30	3.446670e+02
sable_15_30	3.943847e+02	sable_15_30	2.633900e+02	sable_15_30	3.966017e+02
argile_30_60	2.924068e+02	argile_30_60	3.112712e+02	argile_30_60	3.451646e+02
limon_30_60	3.491367e+02	limon_30_60	4.392500e+02	limon_30_60	3.168364e+02
sable_30_60	3.584561e+02	sable_30_60	2.494802e+02	sable_30_60	3.379976e+02
argile_60_100	3.407930e+02	argile_60_100	3.505285e+02	argile_60_100	4.019966e+02
limon_60_100	3.071408e+02	limon_60_100	3.814388e+02	limon_60_100	2.817761e+02
sable_60_100	3.520662e+02	sable_60_100	2.680349e+02	sable_60_100	3.162262e+02
argile_100_200	3.472537e+02	argile_100_200	3.594749e+02	argile_100_200	4.054576e+02
limon_100_200	.930450e+02	limon_100_200	3.579886e+02	limon_100_200	2.750514e+02
sable_100_200	.597012e+02	sable_100_200	.825376e+02	sable_100_200	.194891e+02

Cluster 4		Cluster 5		Cluster 6	
argile	1.967562e+02	argile	1.421589e+02	argile	1.925266e+02
limon	4.030195e+02	limon	2.262949e+02	limon	3.214680e+02
sable	4.002228e+02	sable	6.315479e+02	sable	4.860050e+02
srtm_elev	2.063889e+02	srtm_elev	9.992941e+01	srtm_elev	2.798482e+02
srtm_slp3	2.514561e+00	srtm_slp3	4.073243e-01	srtm_slp3	2.239131e+00
srtm_asp3	1.440278e+01	srtm_asp3	1.488235e+01	srtm_asp3	2.383036e+01
srtm_prf3	-1.204306e-04	srtm_prf3	-4.965882e-05	srtm_prf3	6.646429e-05
srtm_pln3	-5.323764e-03	srtm_pln3	2.093306e-03	srtm_pln3	-2.786607e-03
srtm_lng3	-1.242639e-04	srtm_lng3	-4.967059e-05	srtm_lng3	6.679464e-05
srtm_crs3	9.454167e-05	srtm_crs3	-6.823529e-07	srtm_crs3	8.456250e-05
srtm_slp5	2.041395e+00	srtm_slp5	2.316903e-01	srtm_slp5	1.927011e+00
srtm_asp5	8.527778e+00	srtm_asp5	3.176471e+00	srtm_asp5	3.443750e+01
srtm_prf5	-4.823611e-05	srtm_prf5	-1.157647e-05	srtm_prf5	5.972321e-05
srtm_pln5	4.325278e-04	srtm_pln5	8.412118e-04	srtm_pln5	-1.793393e-03
srtm_lng5	-2.356389e-04	srtm_lng5	-1.062941e-04	srtm_lng5	1.986161e-04
srtm_crs5	1.439167e-04	srtm_crs5	1.963529e-05	srtm_crs5	2.231250e-04
srtm_slp9	2.265604e+00	srtm_slp9	3.665978e-01	srtm_slp9	2.015600e+00

srtm_asp9	9.958333e+00	srtm_asp9	3.835294e+00	srtm_asp9	3.832143e+01
srtm_prf9	-1.763889e-05	srtm_prf9	-1.364706e-06	srtm_prf9	3.217857e-05
srtm_pln9	-5.390917e-03	srtm_pln9	1.848824e-04	srtm_pln9	-2.225768e-03
srtm_lng9	-8.094444e-05	srtm_lng9	-2.815294e-05	srtm_lng9	3.186607e-05
srtm_crs9	6.530556e-05	srtm_crs9	1.658824e-06	srtm_crs9	4.774107e-05
srtm_sl17	1.863688e+00	srtm_sl17	1.982955e-01	srtm_sl17	1.840417e+00
srtm_as17	1.837500e+01	srtm_as17	5.011765e+00	srtm_as17	3.194643e+01
srtm_pf17	5.652778e-06	srtm_pf17	-1.141176e-06	srtm_pf17	1.136607e-05
srtm_pl17	2.885667e-03	srtm_pl17	-2.205176e-04	srtm_pl17	-5.205000e-04
srtm_ln17	5.708333e-06	srtm_ln17	-1.141176e-06	srtm_ln17	1.137500e-05
srtm_cr17	9.680556e-06	srtm_cr17	-3.294118e-07	srtm_cr17	6.544643e-06
srtm_sl33	5.914721e-01	srtm_sl33	1.212692e-01	srtm_sl33	8.573820e-01
srtm_as33	1.163889e+01	srtm_as33	5.058824e+00	srtm_as33	5.441964e+01
srtm_pf33	2.638889e-07	srtm_pf33	-2.376471e-06	srtm_pf33	2.187500e-06
srtm_pl33	-9.922083e-04	srtm_pl33	-5.082353e-05	srtm_pl33	-3.053661e-04
srtm_ln33	2.500000e-07	srtm_ln33	-2.376471e-06	srtm_ln33	2.196429e-06
srtm_cr33	3.055556e-06	srtm_cr33	5.882353e-08	srtm_cr33	1.553571e-06
aster_el	1.995000e+02	aster_el	9.428235e+01	aster_el	2.715714e+02
ast_slp3	6.529825e+00	ast_slp3	1.580799e+00	ast_slp3	3.221304e+00
ast_asp3	-1.762500e+01	ast_asp3	-2.352941e-02	ast_asp3	3.190179e+01
ast_prf3	7.906944e-04	ast_prf3	-3.265882e-05	ast_prf3	3.175893e-05
ast_pln3	1.593296e-02	ast_pln3	-2.487365e-03	ast_pln3	-2.135759e-03
ast_lng3	8.376944e-04	ast_lng3	-3.276471e-05	ast_lng3	3.105357e-05
ast_crs3	-3.747083e-04	ast_crs3	2.404706e-05	ast_crs3	2.062500e-06
ast_slp5	5.593373e+00	ast_slp5	1.208201e+00	ast_slp5	2.903996e+00
ast_asp5	-5.583333e+00	ast_asp5	-6.929412e+00	ast_asp5	2.699107e+01
ast_prf5	4.629583e-04	ast_prf5	1.176471e-06	ast_prf5	7.196429e-06
ast_pln5	-2.095447e-02	ast_pln5	5.538800e-03	ast_pln5	-5.891295e-03
ast_lng5	5.064583e-04	ast_lng5	1.141176e-06	ast_lng5	6.794643e-06
ast_crs5	-3.283472e-04	ast_crs5	1.078824e-05	ast_crs5	4.980357e-05
ast_slp9	4.005267e+00	ast_slp9	7.396911e-01	ast_slp9	2.451287e+00
ast_asp9	-4.305556e+00	ast_asp9	-2.202353e+01	ast_asp9	3.674107e+01
ast_prf9	1.207639e-04	ast_prf9	4.588235e-06	ast_prf9	3.853571e-05
ast_pln9	-1.710685e-02	ast_pln9	3.338765e-03	ast_pln9	-1.328491e-03
ast_lng9	1.210000e-04	ast_lng9	4.611765e-06	ast_lng9	3.871429e-05
ast_crs9	-1.038194e-04	ast_crs9	-8.670588e-06	ast_crs9	5.900000e-05
ast_sl17	2.439703e+00	ast_sl17	3.558333e-01	ast_sl17	1.943414e+00

ast_as17	1.158333e+01	ast_as17	8.458824e+00	ast_as17	2.940179e+01
ast_pf17	6.083333e-06	ast_pf17	-1.588235e-06	ast_pf17	1.253571e-05
ast_pl17	1.727750e-03	ast_pl17	-4.054118e-05	ast_pl17	-3.856786e-04
ast_ln17	6.097222e-06	ast_ln17	-1.588235e-06	ast_ln17	1.254464e-05
ast_cr17	6.694444e-06	ast_cr17	2.588235e-07	ast_cr17	6.035714e-06
ast_sl33	6.140121e-01	ast_sl33	1.423732e-01	ast_sl33	8.659617e-01
ast_as33	6.666667e+00	ast_as33	-6.305882e+00	ast_as33	5.568750e+01
ast_pf33	5.138889e-07	ast_pf33	-2.141176e-06	ast_pf33	2.276786e-06
ast_pl33	-1.422597e-03	ast_pl33	3.185882e-04	ast_pl33	-3.475714e-04
ast_ln33	5.138889e-07	ast_ln33	-2.141176e-06	ast_ln33	2.276786e-06
ast_cr33	3.069444e-06	ast_cr33	-1.647059e-07	ast_cr33	1.598214e-06
srtm_WI	1.923972e+01	srtm_WI	2.443729e+01	srtm_WI	1.708804e+01
srtm_MRVPF	1.918313e+00	srtm_MRVPF	6.145383e+00	srtm_MRVPF	5.464127e-01
srtm_MRRTF	1.699028e+00	srtm_MRRTF	2.839529e+00	srtm_MRRTF	9.177679e-01
srtm_sl_h	2.101097e+01	srtm_sl_h	9.590588e+00	srtm_sl_h	2.164804e+01
srtm_msp	4.245833e-01	srtm_msp	6.929412e-01	srtm_msp	2.841071e-01
srtm_area	7.865185e+06	srtm_area	1.061120e+06	srtm_area	1.159119e+06
srtm_Alt_c	2.613347e+01	srtm_Alt_c	1.794118e+00	srtm_Alt_c	3.804232e+01
uranium	2.037304e+00	uranium	2.462443e+00	uranium	2.165696e+00
ti	9.004634e+00	ti	1.003059e+01	ti	8.646408e+00
th	1.095588e+01	th	1.252365e+01	th	8.680881e+00
potassium	1.286238e+00	potassium	2.612781e+00	potassium	1.516209e+00
total	5.625303e+01	total	8.101922e+01	total	5.565682e+01
vig00	2.815969e-02	vig00	1.227765e-04	vig00	7.901250e-04
olea00	1.265677e-01	olea00	1.465798e-01	olea00	4.124906e-02
bois00	2.895857e-02	bois00	2.606562e-01	bois00	1.830007e-02
jach00	5.515040e-02	jach00	1.239839e-01	jach00	1.984805e-02
argile_5_15	1.997262e+02	argile_5_15	1.486119e+02	argile_5_15	1.977412e+02
limon_5_15	4.021091e+02	limon_5_15	2.254926e+02	limon_5_15	3.201872e+02
sable_5_15	3.981642e+02	sable_5_15	6.258968e+02	sable_5_15	4.820712e+02
argile_15_30	2.164499e+02	argile_15_30	1.648836e+02	argile_15_30	2.167421e+02
limon_15_30	3.964354e+02	limon_15_30	2.231307e+02	limon_15_30	3.163815e+02
sable_15_30	3.871150e+02	sable_15_30	6.119865e+02	sable_15_30	4.668762e+02
argile_30_60	2.595288e+02	argile_30_60	2.009566e+02	argile_30_60	2.721812e+02
limon_30_60	3.565743e+02	limon_30_60	2.198744e+02	limon_30_60	3.038470e+02
sable_30_60	3.838973e+02	sable_30_60	5.791696e+02	sable_30_60	4.239715e+02

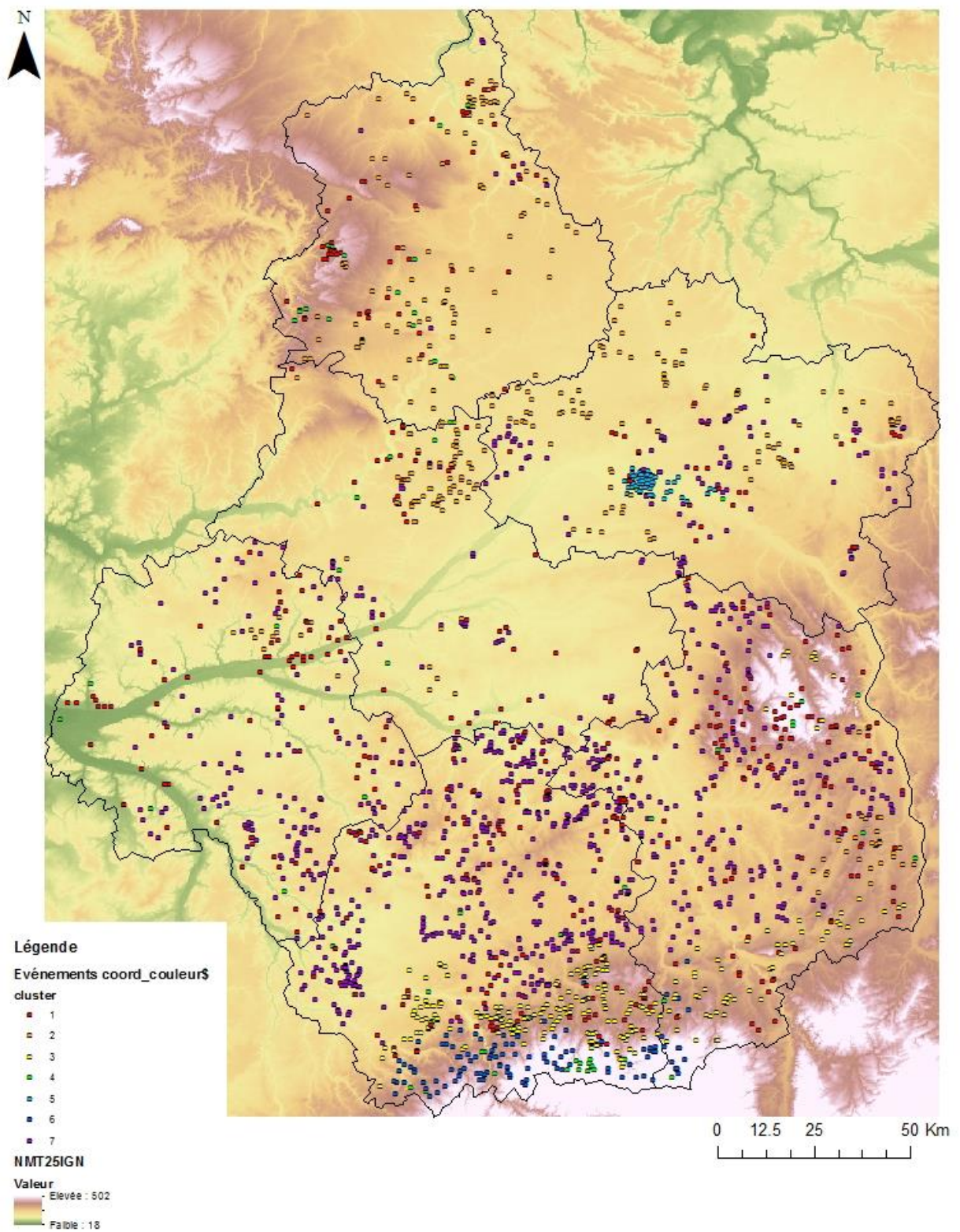
argile_60_100	3.118047e+02	argile_60_100	2.291623e+02	argile_60_100	2.983732e+02
limon_60_100	3.131714e+02	limon_60_100	2.239091e+02	limon_60_100	2.719847e+02
sable_60_100	3.750243e+02	sable_60_100	5.469272e+02	sable_60_100	4.296417e+02
argile_100_200	3.090171e+02	argile_100_200	2.275692e+02	argile_100_200	2.815425e+02
limon_100_200	2.968602e+02	limon_100_200	2.188576e+02	limon_100_200	2.563239e+02
sable_100_200	3.941238e+02	sable_100_200	5.535712e+02	sable_100_200	4.621334e+02

Cluster 7					
srtm_elev	1.424202e+02	aster_el	1.353491e+02	uranium	1.543403e+00
srtm_slp3	1.150687e+00	ast_slp3	2.799871e+00	ti	9.589456e+00
srtm_asp3	1.051390e+01	ast_asp3	1.192482e+01	th	8.187307e+00
srtm_prf3	-7.826982e-08	ast_prf3	-7.894130e-05	potassium	1.002553e+00
srtm_pln3	2.522132e-04	ast_pln3	-2.933450e-04		
srtm_lng3	-9.371782e-08	ast_lng3	-7.908445e-05	total	4.288654e+01
srtm_crs3	9.721936e-06	ast_crs3	4.401236e-05	vig00	5.32884e-3
				olea00	2.084654e-01
srtm_slp5	9.369993e-01	ast_slp5	2.384130e+00	bois00	4.523452e-02
srtm_asp5	1.058805e+01	ast_asp5	1.196292e+01	jach00	8.670905e-02
srtm_prf5	4.569516e-06	ast_prf5	-3.601957e-05		
srtm_pln5	-8.565314e-04	ast_pln5	3.698682e-04	argile	2.192181e+02
srtm_lng5	-9.433574e-07	ast_lng5	-3.608857e-05	limon	3.760124e+02
srtm_crs5	2.338414e-05	ast_crs5	3.810608e-05	sable	4.047694e+02
srtm_slp9	1.035700e+00	ast_slp9	1.744918e+00	argile_5_15	2.220308e+02
srtm_asp9	1.338723e+01	ast_asp9	9.060762e+00	limon_5_15	3.750160e+02
srtm_prf9	1.804325e-06	ast_prf9	-1.518435e-05	sable_5_15	4.029533e+02
srtm_pln9	-5.986818e-04	ast_pln9	-1.997703e-04		
srtm_lng9	-5.396498e-07	ast_lng9	-1.521318e-05	argile_15_30	2.362766e+02
srtm_crs9	5.574665e-06	ast_crs9	1.914315e-05	limon_15_30	3.714712e+02
				sable_15_30	3.922519e+02
srtm_sl17	8.725180e-01	ast_sl17	1.146934e+00		
srtm_as17	2.178888e+01	ast_as17	2.065911e+01	argile_30_60	3.234077e+02
srtm_pf17	1.915551e-07	ast_pf17	-3.892894e-07	limon_30_60	3.278024e+02
srtm_pl17	-2.877075e-04	ast_pl17	-2.051658e-04	sable_30_60	3.487901e+02
srtm_ln17	1.894954e-07	ast_ln17	-3.882595e-07		
srtm_cr17	1.866117e-06	ast_cr17	1.726056e-06	argile_60_100	3.716812e+02
				limon_60_100	2.876611e+02

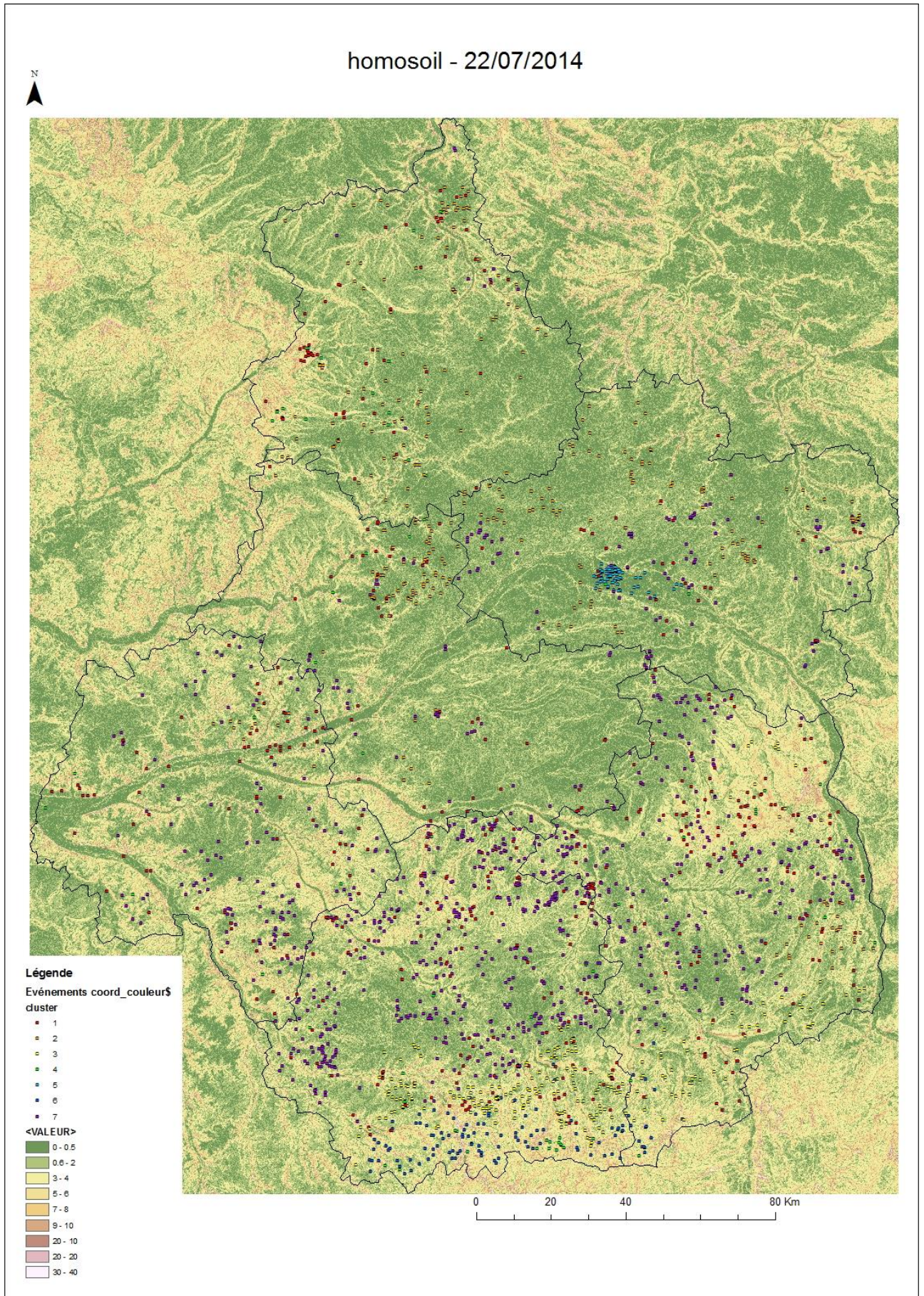
srtm_sl33	3.587694e-01	ast_sl33	3.763237e-01	sable_60_100	3.406593e+02
srtm_as33	1.857673e+01	ast_as33	1.908548e+01		
srtm_pf33	1.544799e-07	ast_pf33	-3.295572e-08	argile_100_200	3.791691e+02
srtm_pl33	-6.891349e-05	ast_pl33	-1.082904e-05	limon_100_200	2.758974e+02
srtm_ln33	1.544799e-07	ast_ln33	-3.089598e-08	sable_100_200	3.449349e+02
srtm_cr33	5.612770e-07	ast_cr33	4.994851e-07		
srtm_WI	2.076536e+01	srtm_msp	3.771164e-01		
srtm_MRVPF	2.066919e+00	srtm_area	1.218744e+06		
srtm_MRRTF	1.982029e+00	srtm_Alt_c	1.079136e+01		
srtm_sl_h	1.195519e+01				

3-3 Répartition des clusters par rapport à l'altitude

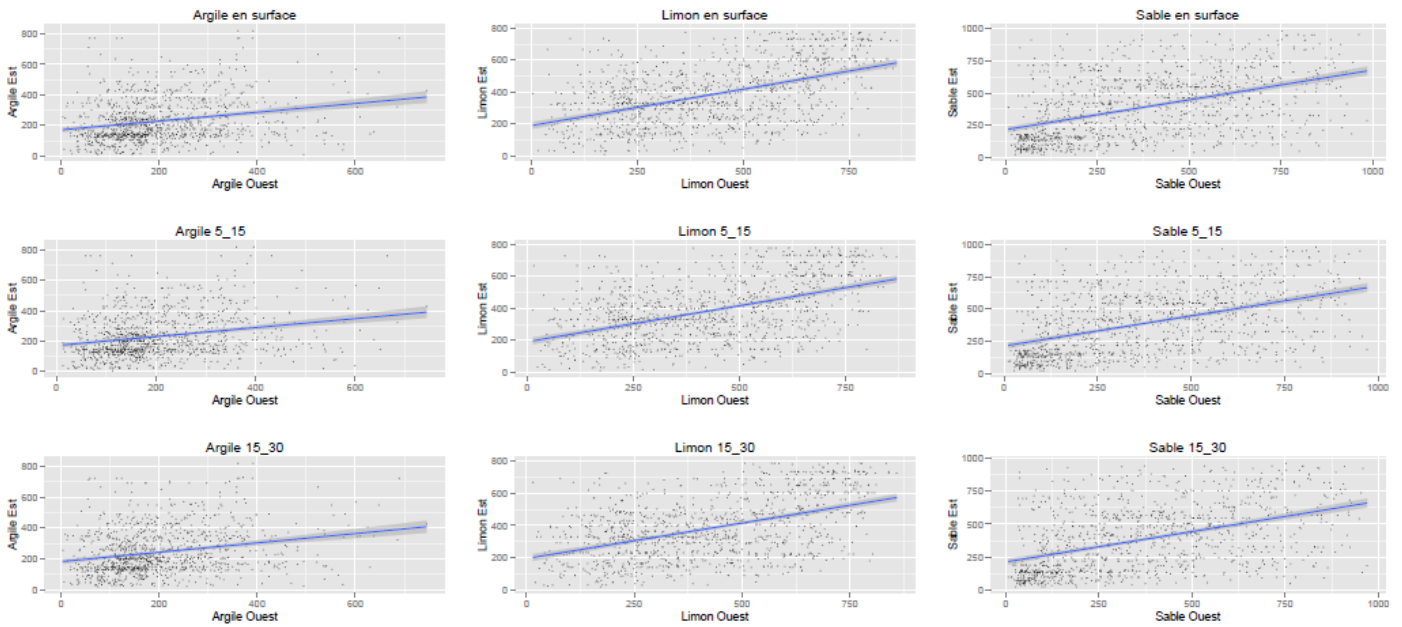
homosoil - 22/07/2014



3-3 Répartition des clusters par rapport à la pente



3-5 Valeurs pédologiques prédites en fonction des valeurs exactes par le filtrage



3-6 Comparaison du R² des différentes méthodes

