



SOMMAIRE

| | |
|---|----|
| 1. Introduction | 2 |
| 2. Méthode..... | 3 |
| 2.1. Ajustement du champ..... | 3 |
| 2.2. Appariement | 3 |
| 2.3. Calcul des distances entre exploitations..... | 4 |
| 2.4. Transfert | 5 |
| 3. Application transfert des UGB bovines | 6 |
| 4. Transfert sur un indicateur d'intensité agricole | 8 |
| 5. ANNEXES | 11 |
| 5.1. Ajustement des variables..... | 12 |
| 5.2. Représentativité..... | 14 |
| 5.3. Variables utilisées | 17 |
| 5.4. Choix de la distance utilisée et du caliper..... | 18 |
| 5.5. Performances par région | 23 |

1. Introduction

Le RICA est représentatif des principaux systèmes de production agricole (OTEX) par région. L'objectif de la désagrégation de données est de pouvoir utiliser un indicateur calculé sur le RICA, à un niveau infrarégional (commune, PRA, département, ...) en le transférant vers un référentiel géo-référencé exhaustif des exploitations agricoles (référentiel des cotisants non salariés MSA).

Le transfert se base sur des techniques de matching. En effet cela passe par un appariement des exploitations entre les deux bases de données, afin de retrouver les exploitations les plus similaires entre les deux bases. Des exploitations de la base d'enquête (RICA) sont alors identifiées comme ayant une jumelle dans la base sur laquelle doit se faire le transfert (**phase d'appariement**). Ensuite en utilisant le maximum de variables possibles sur la base à transférer, on va rechercher les exploitations identifiées les plus proches des exploitations non identifiées afin de procéder au transfert de l'indicateur (**phase de transfert**).

Dans le cas du RICA et de la MSA, la phase d'appariement va consister à trouver, pour chaque exploitation MSA, l'exploitation RICA la plus proche. Puis on pourra récupérer les indicateurs des exploitations appariées, dans la base à transférer. Ensuite on recherche pour chaque exploitation MSA non identifiée (non traitée) les exploitations MSA identifiées (traitées) les plus proches, qui serviront pour le transfert de l'indicateur. L'ensemble des indicateurs individuels obtenus n'ont pas pour but d'être traités tels quels mais plutôt agrégé au niveau géographique souhaité, inférieur à la région. Nous serons alors en mesure de désagréger les indicateurs de contexte que nous n'avions qu'au niveau régional par le RICA à un niveau plus fin, commune ou PRA. Le niveau à choisir sera étudié par la suite.

Nous présentons ainsi la méthode, puis nous observons les résultats sur le calcul des UGB bovins. Nous ferons également une illustration sur un indicateur d'intensité agricole calculé dans le cadre du projet de recherche FARMBIRD.

2. Méthode

2.1. Ajustement du champ

Le transfert se fera entre l'ensemble du RICA 2005 et la MSA 2006 (qui concerne les chefs d'exploitation actifs au 1^{er} janvier 2006) restreinte aux exploitations ayant une superficie positive et faisant partie du champ agricole. On exclut aussi la Corse, les exploitations composées uniquement de cotisants solidaires et les exploitations appartenant à des communes dont le code n'est pas reconnu. Par ailleurs il est fait un traitement des aberrants notamment sur le revenu (nous le prenons inférieur à 3 millions d'euros en valeur absolue). Au total 385 433 exploitations sont retenues dans le fichier 2006 de la MSA.

Pour prendre en compte le fait que les exploitations RICA sont des exploitations considérées comme professionnelles, nous devons nous rapprocher du champ de celles-ci. Ainsi nous ne prendrons que les exploitations MSA de plus de 35 hectares ou de revenu supérieur à 5000 euros. Il convient de s'assurer que le champ retenu est le bon. Ainsi nous vérifions l'ajustement de la variable SUPERFICIE, commune aux deux bases. Les graphiques d'ajustement sont présentés en annexe 5.1. Nous avons ainsi retenues 301 074 exploitations MSA en 2006. Le RICA compte lui 7363 exploitations qui une fois extrapolées en représentent 346 219 (source AGRESTE).

2.2. Appariement

Sachant qu'il n'existe pas d'identifiant commun aux deux bases (la version RICA utilisée par l'INRA étant anonyme), l'étape d'appariement consiste à trouver les « jumeaux » entre les deux bases de données, du RICA et de la MSA. Cet appariement se fait à travers les variables communes aux deux bases (commune, orientation, forme juridique, superficie, nombre d'exploitants, montant des salaires versés, revenu).

L'appariement repose sur deux distances (euclidiennes) D_1 et D_2 , calculées pour un couple d'exploitation (i, j) de la façon suivante :

- $D_1(i, j)$: sur les variables dont on est sûr que l'ajustement sera proche (commune, orientation, forme juridique, superficie, année de naissance, sexe).
- $D_2(i, j)$: sur l'ensemble des variables communes qui peuvent différer dans leur mode de calcul (commune, orientation, forme juridique, superficie, nombre d'exploitants, montant des salaires versés, revenu).

Nous calculons pour chaque exploitation les minimums des distances D_1 : $\min_j D_1(i, j)$.

Les exploitations i et j seront appariées si la condition (C) suivantes est remplie :

$$\min_i D_1(i, j) = \min_j D_1(i, j) = D_1(i, j)$$

Lorsque plusieurs exploitations MSA répondent à ce critère pour une exploitation RICA, nous gardons le couple qui minimise la distance D_2 .

Nous effectuons cet appariement sur les couples d'exploitations RICA et MSA qui sont dans le même canton et qui ont au moins deux variables égales parmi le sexe, l'année de naissance et la commune.

7341 couples RICA-MSA (99.7% de l'échantillon RICA) sont trouvés pour constituer le groupe de référence pour le transfert de l'indicateur : les exploitations MSA appariées récupèrent la valeur de l'indicateur de leur jumeau RICA. Les exploitations MSA non appariées vont alors constituer le groupe de contrôle auquel sera transféré l'indicateur. Pour information voici le classement des couples appariés selon l'égalité des critères commune, année de naissance d'un chef, forme juridique et orientation agricole.

| | Même commune et même année naissance | Même canton même année naissance, commune différente | Même commune, année naissance différente | Total |
|--|--------------------------------------|--|--|-------|
| Même orientation et même forme juridique | 4456 | 427 | 1026 | 5909 |
| Même orientation, forme juridique différente | 359 | 63 | 124 | 546 |
| Même forme juridique, orientation différente | 673 | 44 | 106 | 823 |
| Ni orientation, ni forme juridique | 35 | 9 | 19 | 63 |
| Total | 5523 | 543 | 1275 | 7341 |

2.3. Calcul des distances entre exploitations

A partir de ce stade, nous travaillons uniquement sur les données du référentiel (MSA), les variables de contexte définies au niveau communal fournies par des bases complémentaires et la valeur de l'indicateur pour les exploitations appariées (la liste des variables utilisées est présentée en annexe ??). Pour procéder au transfert d'indicateurs nous partons de l'hypothèse que deux exploitations identiques auront les mêmes indicateurs. Nous allons alors identifier les plus proches voisins pour chaque exploitation du groupe de contrôle (non appariées), parmi le groupe de traitement (appariées). Ceci est fait grâce au calcul d'une matrice des distances entre les exploitations. Dans la mesure où le RICA est représentatif des OTEX par région, nous calculons une matrice des distances pour chaque orientation / région en se basant sur l'hypothèse qu'une exploitation n'aura de voisins qu'au sein d'une même orientation et dans la même région. Cependant, compte tenu des faibles effectifs des exploitations RICA dans certains cas, il paraît pertinent de regrouper les régions en fonction de l'orientation. En annexe 5.2. se trouvent les effectifs par orientation / région puis les regroupements effectués sur les régions viticoles, les bassins laitiers, les régions de culture ou les grandes régions. Ainsi on calcule une matrice de distance entre exploitations pour chaque orientation et chaque regroupement géographique.

La distance utilisée est la distance de Mahalanobis qui permet de calculer la distance entre deux individus, tout en tenant compte des variations de densité à l'intérieur du nuage de points. Une exploitation i est représentée par le vecteur $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ où p est le nombre de variables mobilisées pour le transfert (la base MSA auxquels sont ajoutées des variables de contexte), Σ est la matrice de variance-covariance. La matrice de distance entre 2 individus i et j individus est alors :

$$D(X_i, X_j) = (X_i - X_j)^T \Sigma^{-1} (X_i - X_j)$$

Nous avons également testé une distance basée sur le score de propension mais au vu des résultats nous avons préféré utiliser la distance de Mahalanobis (c.f. annexe 5.4).

Dans la mesure où cette méthode va permettre de transférer différents indicateurs, nous pouvons penser qu'utiliser des variables choisies en fonction de l'indicateur pour le transfert sera plus pertinent qu'utiliser toutes les variables disponibles. Cela entraînerait notamment des problèmes de colinéarité. Nous choisissons alors les variables entrant dans le calcul de la distance en fonction de l'indicateur. Ainsi avant tout calcul de distance, nous effectuons deux régressions : une régression sur les variables individuelles et une régression sur les variables de contexte. Nous gardons dans les deux cas uniquement les variables qui entrent

dans le meilleur modèle selon le critère AIC (fonction *step* de **R**). Le fait de séparer les variables individuelles et de contexte permet de ne pas noyer les variables individuelles au milieu des variables de contexte (16 variables individuelles et 45 variables de contexte) et ainsi de forcer la différenciation des exploitations au sein d'une commune qui auraient le même indicateur transféré s'il n'y avait que des variables de contexte. Ces régressions sont également faites par orientation et regroupement géographique, avant le calcul de la distance.

Nous recréons ainsi une liste spécifique de variables qui entreront dans le calcul de la distance composée obligatoirement à la fois de variables individuelles et de variables de contexte. Dans le cas où la régression sur variables individuelles donne moins de 3 variables, nous préférons intégrer à la distance toutes les variables individuelles. La liste complète des variables utilisées est disponible en annexe 5.3.

Nous pouvons faire varier le *caliper*, seuil de distance S^* à ne pas dépasser pour que les exploitations soient considérées comme voisines. On rejette ainsi les « couples » d'exploitations qui sont à une distance trop grandes l'une de l'autre. Nous avons testé plusieurs seuils de distance à partir duquel pour un individu du groupe de contrôle on rejettera une exploitation du groupe de traitement de ses plus proches voisins. Après étude nous choisissons le seuil de 50%, c'est-à-dire que seules les valeurs inférieures à la médiane seront conservées dans la matrice de distance (c.f. annexe 5.4).

2.4. Transfert

A chaque exploitation du groupe de contrôle, nous sommes désormais capables d'identifier ses plus proches voisins parmi le groupe de traitement, et nous connaissons pour chaque couple d'exploitations la distance associée. Nous déterminons des poids à affecter à chaque exploitation du groupe de traitement lors du transfert, égal à l'exponentielle inverse de la distance si celle-ci ne dépasse pas le seuil S^* choisit. Ces poids sont ensuite normalisés pour que la somme des poids par exploitation du groupe de contrôle soit égale à 1.

Soit $d(i, j)$ la distance entre l'exploitation i du groupe de contrôle et l'exploitation j du groupe de traitement, et S^* le seuil retenu. On définit ainsi $\omega(i, j)$ le poids qui sera affecté à l'exploitation j du groupe de traitement lors du transfert de l'indicateur sur l'exploitation i :

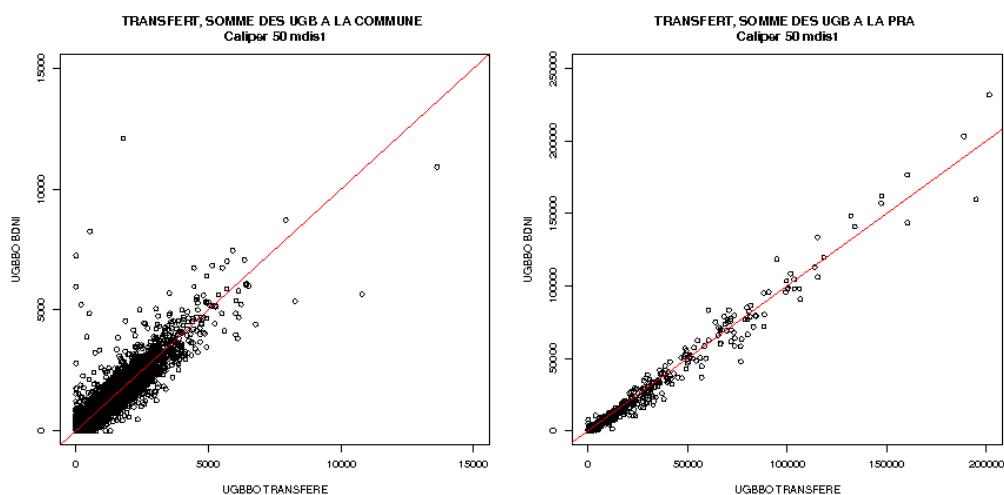
$$\omega(i, j) = \frac{\omega^*(i, j)}{\sum_j \omega^*(i, j)} \quad \text{avec} \quad \omega^*(i, j) = \begin{cases} e^{-d(i, j)} & \text{si } d(i, j) \leq S^* \\ 0 & \text{sinon} \end{cases}$$

La valeur de l'indicateur transféré sur l'exploitation i sera alors la moyenne pondérée des indicateurs des plus proches voisins de i , parmi les exploitations du groupe de traitement, affectés des poids $\omega(i, j)$.

$$\hat{y} = \sum_j \omega(i, j) \times y_j$$

3. Application transfert des UGB bovines

Via les données de la BDNI, nous disposons des UGB bovines par commune. Les UGB bovines sont également disponibles pour les exploitations RICA (UGBBO). Ainsi nous allons réaliser le transfert des UGB bovins du RICA sur les exploitations de la MSA. Nous pourrions valider la méthode en comparant les UGB bovins au niveau communal ou sommés à la PRA.



Plus la région est hétérogène, plus le transfert est difficile, de même les valeurs transférées sont moins proches de la réalité dans les régions où les exploitations appariées à des exploitations RICA sont peu nombreuses (cf annexe 5.5). Les p.values des tests des rangs appariées de Wilcoxon faits sur les régions, au niveau PRA sont présentés ci-dessous :

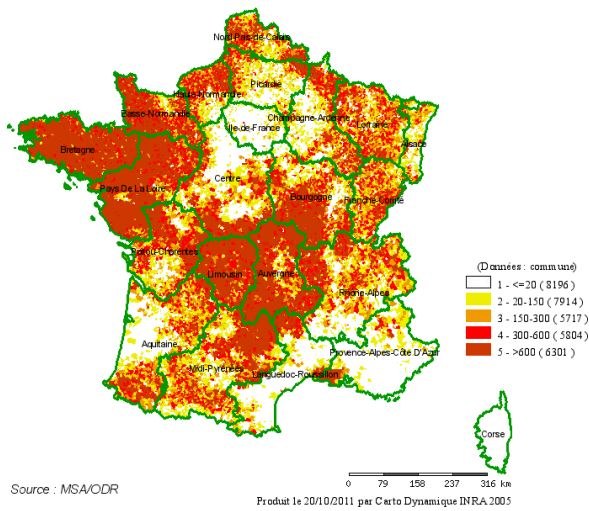
| Région | Code région | P.Value | Nombre de PRA |
|----------------------------|-------------|---------|---------------|
| Île-de-France | 11 | * | 37 |
| Champagne-Ardenne | 21 | . | 34 |
| Picardie | 22 | . | 19 |
| Haute-Normandie | 23 | * | 19 |
| Centre | 24 | . | 45 |
| Basse-Normandie | 25 | *** | 17 |
| Bourgogne | 26 | * | 35 |
| Nord-Pas-de-Calais | 31 | * | 20 |
| Lorraine | 41 | *** | 23 |
| Alsace | 42 | *** | 13 |
| Franche-Comté | 43 | . | 26 |
| Pays de la Loire | 52 | *** | 39 |
| Bretagne | 53 | * | 25 |
| Poitou-Charentes | 54 | . | 30 |
| Aquitaine | 72 | *** | 68 |
| Midi-Pyrénées | 73 | * | 61 |
| Limousin | 74 | *** | 17 |
| Rhône-Alpes | 82 | . | 76 |
| Auvergne | 83 | * | 36 |
| Languedoc-Roussillon | 91 | * | 32 |
| Provence-Alpes-Côte d'Azur | 93 | *** | 35 |

*** P.value>0.05
 * P.value>0.01
 . P.value<=0.01

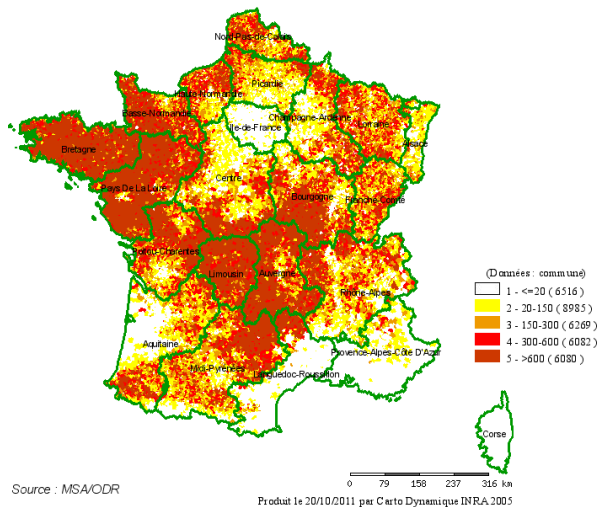
Pour 15 régions sur 21 nous ne rejetons pas l'hypothèse nulle du test des rangs de wilcoxon, ce qui laisse penser que pour ces régions les PRA sont ordonnées de la même façon en terme d'importance d'UGB, que ce soit transféré ou calculé sur la BDNI.

Les cartes ci-après confrontent la vraie valeur de l'indicateur UGB bovines de la BDNI avec la valeur transférée, par commune puis par PRA.

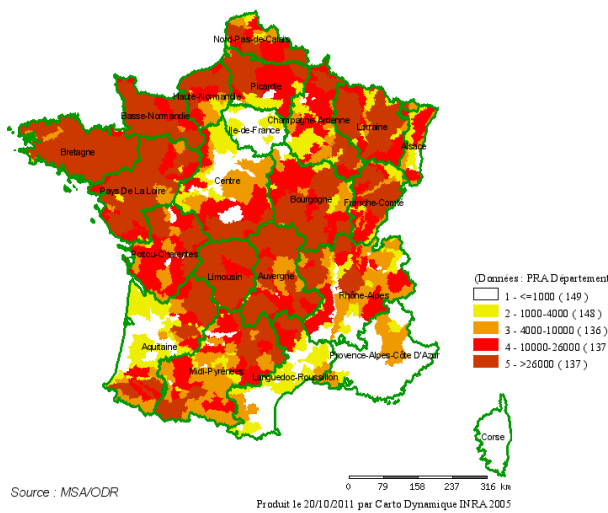
UGB Bovines (Têtes)



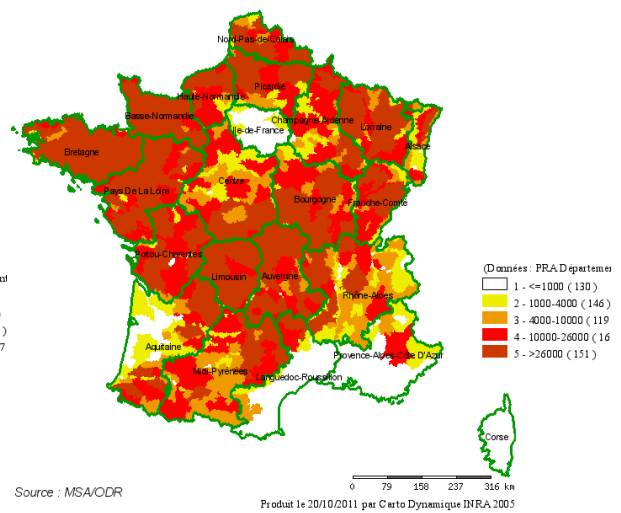
UGB Bovines Transférées (Têtes)



UGB Bovines (Têtes)



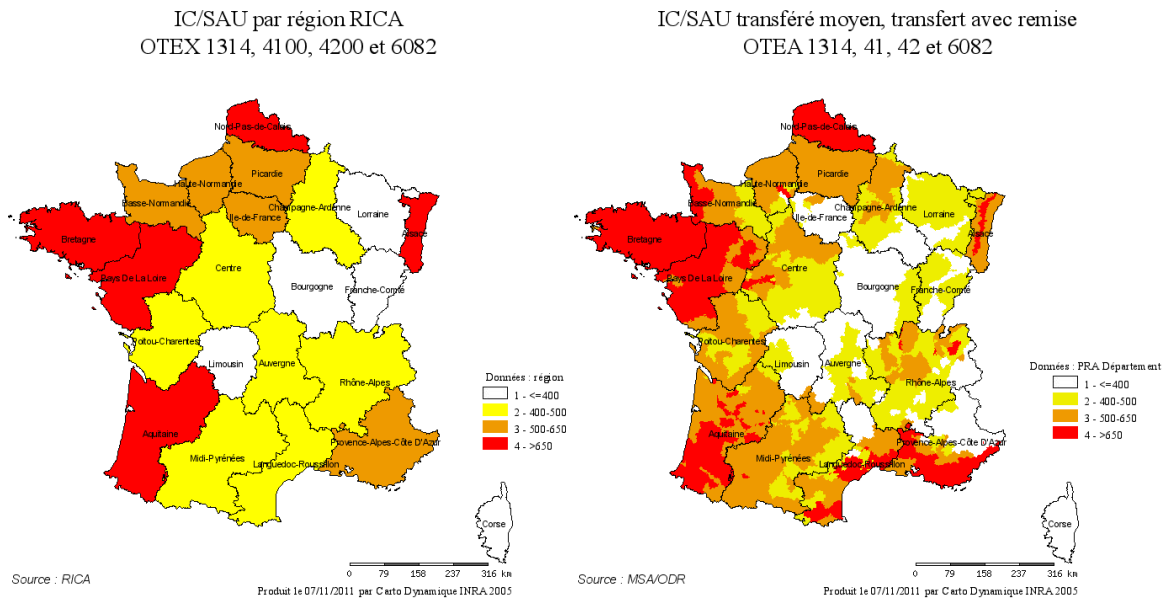
UGB Bovines Transférées (Têtes)



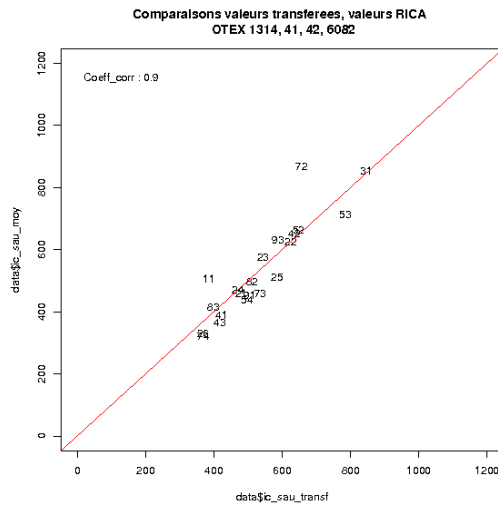
4. Transfert sur un indicateur d'intensité agricole

Ce transfert s'inscrit dans le projet de recherche FARBIRD sur les modèles de co-viabilité entre agriculture et biodiversité oiseaux (financé par l'ANR SYSTERRA). Ce projet a nécessité un changement d'échelle (de région à PRA) sur un indicateur d'intensité agricole calculé à partir du RICA (IC/SAU). Il est calculé comme le ratio des intrants de l'exploitation et de sa superficie (SAU). Les intrants sont : l'eau, le gaz, l'électricité, l'eau pour irrigation, les fournitures non stockées, les dépenses de transport sur achats et ventes, les charges réelles d'approvisionnement (semences, engrais, amendements, produits phytosanitaires, produits vétérinaires, aliments grossiers, aliments concentrés, carburant et lubrifiants, combustibles, fourniture stockées). Cet indicateur est calculé uniquement sur les OTEX grandes cultures, bovins lait, bovins viande et polyculture/polyélevage.

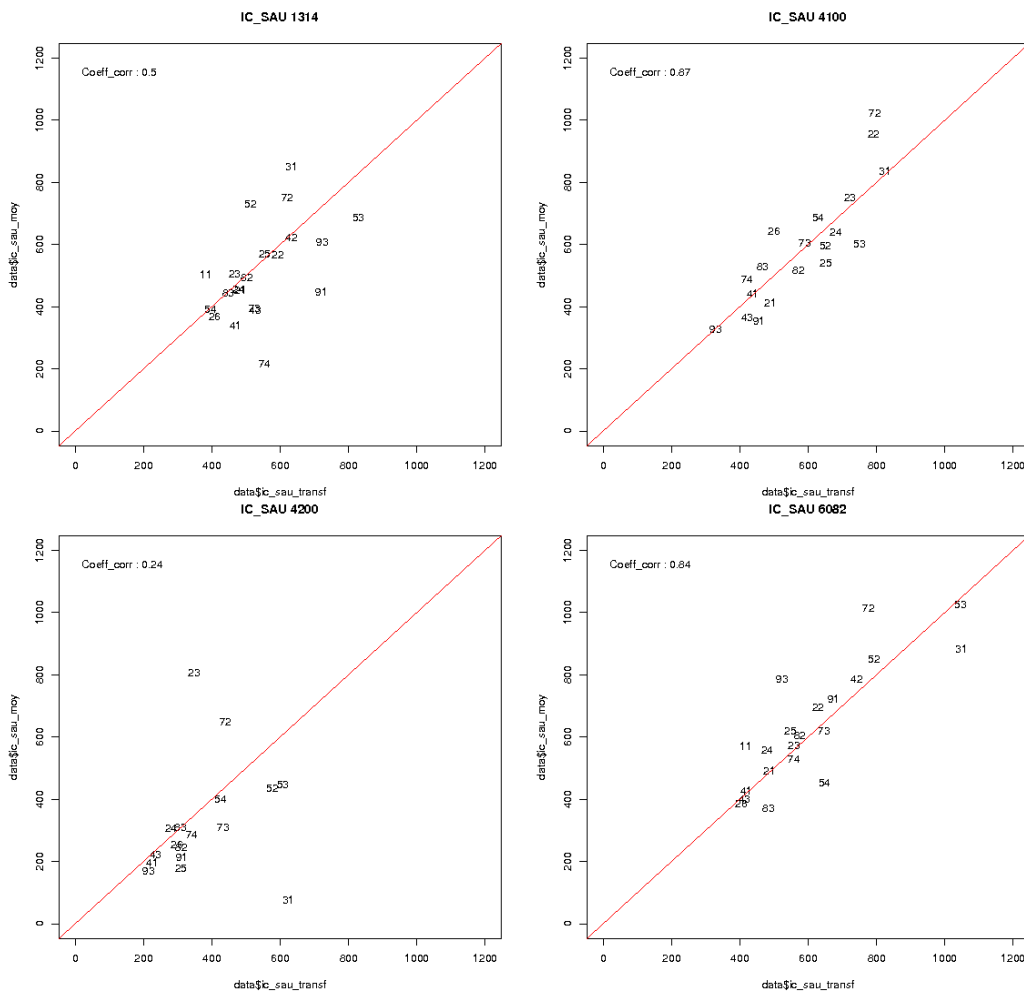
Les cartes suivantes montrent à gauche l'indicateur calculé par région à partir du RICA, et celle de gauche l'indicateur transféré au référentiel MSA 2006, par PRA.



Les graphiques ci-dessous présentent la valeur de l'indicateur transféré moyenné à la région, confrontée à la valeur régionale de l'indicateur calculée à partir du RICA. Cette confrontation est d'abord faite sur l'ensemble des orientations considérées (grandes cultures, bovins lait, bovins viandes et polyculture-élevage), puis par orientation.



On voit ainsi que les résultats régionaux sur les 4 orientations sont relativement proches des résultats régionaux du RICA (coefficient de corrélation de 0.9), excepté pour la région 72 (aquitaine) où la valeur de l'indicateur est sous-estimée par le transfert. Lorsque l'on observe les résultats par orientation, on peut constater que les résultats sont beaucoup plus variables notamment pour l'orientation bovins viande (OTEX 42) où le coefficient de corrélation est de 0.24. Ces faibles résultats peuvent s'expliquer en partie par la différence de définition de l'orientation des exploitations entre la MSA et le RICA. Les résultats paraissent ainsi s'équilibrer entre les orientations.



5. Bibliographie

Fougère, D. (2010), « Les méthodes économétriques d'évaluation », Revue française des affaires sociales 1-2, 105-128

Hansen, B.B. & Klopfer, S.O. (2006) Optimal full matching and related designs via network flows, JCGS 15 609-627.

Le Gallo, J. « Econométrie spatiale 1 - Autocorrélation spatiale », LATEC-Document de travail-Economie (1991-2003), 2000-05

Rosenbaum, P. & Rubin, D. (1983), « The central role of the propensity score in observational studies for causal effects », Biometrika 70, 41-55.

Sekhon, J.S. (2010). Matching : multivariate and propensity score matching with balance optimization. R package version 4.7-10.

Subervie, J. & Chabé-Ferret, S (2009), « Estimation des effets propres des mesures agroenvironnementales du plan de développement rural national 2000-2006 sur les pratiques des agriculteurs », rapport de recherche.

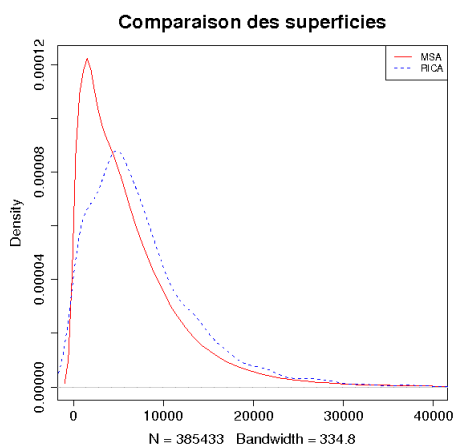
6. ANNEXES

6.1. Ajustement des variables

Sans restriction sur l'assiette et le revenu :

385 433 exploitations MSA vs 7363 exploitations RICA extrapolées. Le rapport des superficies des de 92.61% (Superficie MSA sur Superficie RICA).

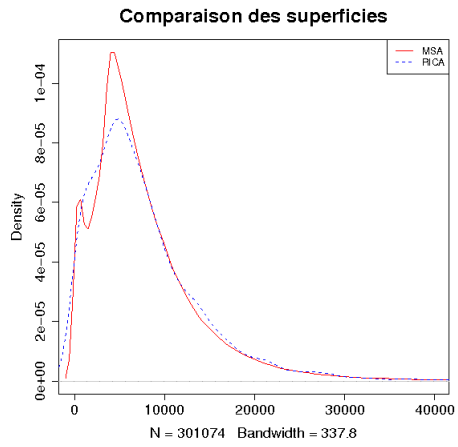
| OTEX | Code OTEX | Pourcentage superficie couvert | Superficie RICA (kha) |
|-------------------------------------|-----------|--------------------------------|-----------------------|
| Grandes cultures | 1314 | 78,13% | 8738,03 |
| Maraîchage, horticulture | 2829 | 307,82% | 81,43 |
| Viticulture | 3738 | 97,83% | 1113,47 |
| Fruits | 3900 | 82,89% | 306,96 |
| bovins lait | 4100 | 99,81% | 4124,33 |
| Bovins élevage et viande | 4200 | 102,03% | 3558,68 |
| Bovins lait, élevage et viande | 4300 | 162,18% | 890,26 |
| Ovins, caprins et autres herbivores | 4400 | 80,02% | 1694,99 |
| Porcins, volailles | 5072 | 92,82% | 632,04 |
| Polyculture, polyélevage | 6082 | 91,23% | 5445,46 |
| Total | | 92,24% | 26585,65 |



Restriction sur l'assiette et le revenu : >35ha ou assiette de cotisation >5000 euros

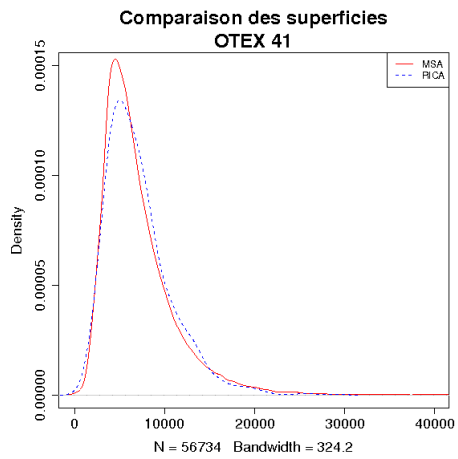
301 074 exploitations MSA vs 7363 exploitations RICA extrapolées. Le rapport des superficies des de 87.05% (Superficie MSA sur Superficie RICA).

| OTEX | Code OTEX | Pourcentage superficie couvert | Superficie RICA (kha) |
|-------------------------------------|-----------|--------------------------------|-----------------------|
| Grandes cultures | 1314 | 75,41% | 8738,03 |
| Maraîchage, horticulture | 2829 | 261,70% | 81,43 |
| Viticulture | 3738 | 71,78% | 1113,47 |
| Fruits | 3900 | 67,91% | 306,96 |
| bovins lait | 4100 | 96,87% | 4124,33 |
| Bovins élevage et viande | 4200 | 95,18% | 3558,68 |
| Bovins lait, élevage et viande | 4300 | 158,50% | 890,26 |
| Ovins, caprins et autres herbivores | 4400 | 71,56% | 1694,99 |
| Porcins, volailles | 5072 | 85,68% | 632,04 |
| Polyculture, polyélevage | 6082 | 87,88% | 5445,46 |
| Total | | 87,05% | 26585,65 |



Exemple orientation bovins lait : 56 734 exploitations MSA dans l’OTEA bovins lait vs 967 exploitations RICA OTEX bovins lait extrapolées.

Le rapport des superficies des de 96,87% (Superficie MSA sur Superficie RICA).



6.2. Représentativité

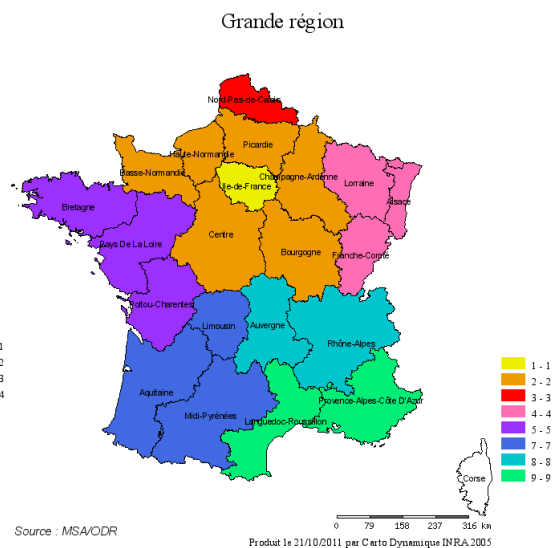
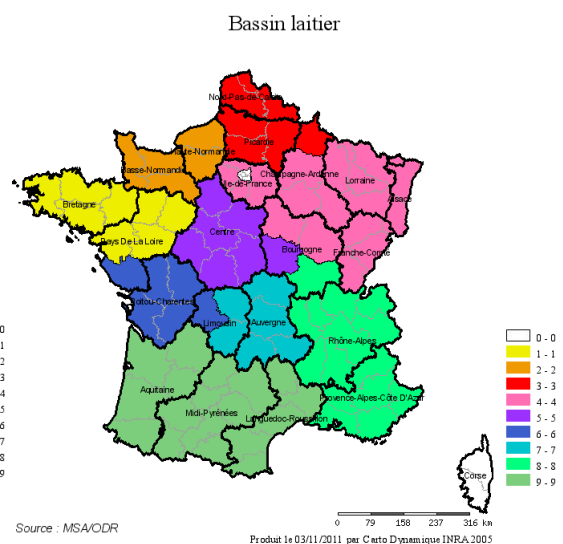
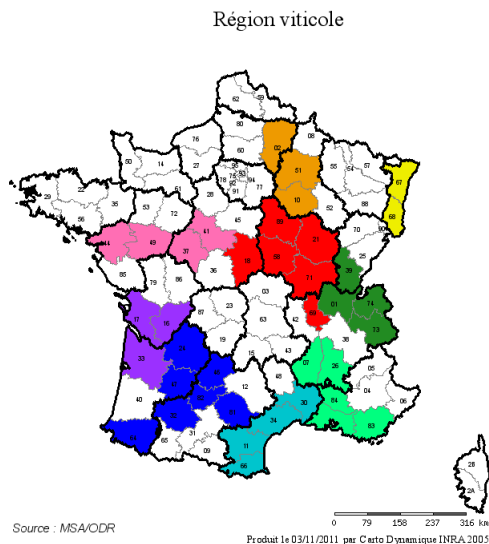
Le tableau suivant présente les effectifs pour chaque OTEX / Région pour les exploitations RICA et l'ensemble des exploitations MSA du champ agricole retenu (>35ha ou >5000 euros d'assiette).

| | | Région | | OTEX | | | | | | | | | |
|-----------------------|-------------------------------------|--------|-------|-------|-------|-------|-------|-------|-------|--------|-------|--|--|
| | | 1314 | 2829 | 3738 | 3900 | 4100 | 4200 | 4300 | 4400 | 5072 | 6082 | | |
| Nb appariés | | 134 | 16 | 0 | 6 | 1 | 0 | 0 | 0 | 2 | 6 | | |
| Nb non appariés | | 3116 | 306 | 12 | 76 | 24 | 23 | 5 | 21 | 19 | 248 | | |
| Appariés/Non appariés | | 4,30% | 5,23% | 0,00% | 7,89% | 4,17% | 0,00% | 0,00% | 0,00% | 10,53% | 2,42% | | |
| OTEX | Libellé des OTEX | | | | | | | | | | | | |
| 1314 | Grandes cultures | 161 | 2 | 109 | 0 | 35 | 4 | 18 | 0 | 1 | 57 | | |
| 2829 | Maraîchage, horticulture | 5748 | 246 | 5663 | 49 | 746 | 468 | 461 | 82 | 64 | 2437 | | |
| 3738 | Viticulture | 2,80% | 0,81% | 1,92% | 0,00% | 4,69% | 0,85% | 3,90% | 0,00% | 1,56% | 2,34% | | |
| 3900 | Fruits | 168 | 0 | 9 | 0 | 20 | 0 | 8 | 0 | 1 | 74 | | |
| 4100 | bovins lait | 5091 | 129 | 480 | 51 | 797 | 191 | 268 | 53 | 83 | 3153 | | |
| 4200 | Bovins élevage et viande | 3,30% | 0,00% | 1,88% | 0,00% | 2,51% | 0,00% | 2,99% | 0,00% | 1,20% | 2,35% | | |
| 4300 | Bovins lait, élevage et viande | 37 | 4 | 0 | 2 | 24 | 8 | 9 | 0 | 0 | 78 | | |
| 4400 | Ovins, caprins et autres herbivores | 1664 | 202 | 1 | 53 | 845 | 361 | 484 | 88 | 84 | 2838 | | |
| 5072 | Porcins, volailles | 2,22% | 1,98% | 0,00% | 3,77% | 2,84% | 2,22% | 1,86% | 0,00% | 0,00% | 2,75% | | |
| 6082 | Polyculture, polyélevage | 227 | 18 | 37 | 13 | 20 | 28 | 4 | 7 | 9 | 71 | | |
| | | 10283 | 342 | 964 | 274 | 656 | 1293 | 123 | 408 | 289 | 3479 | | |
| | | 2,21% | 5,26% | 3,84% | 4,74% | 3,05% | 2,17% | 3,25% | 1,72% | 3,11% | 2,04% | | |
| | | 30 | 14 | 0 | 7 | 103 | 17 | 25 | 7 | 6 | 26 | | |
| | | 1053 | 306 | 3 | 89 | 6812 | 1326 | 1739 | 444 | 246 | 1702 | | |
| | | 2,85% | 4,58% | 0,00% | 7,87% | 1,51% | 1,28% | 1,44% | 1,58% | 2,44% | 1,53% | | |
| | | 114 | 2 | 86 | 2 | 19 | 103 | 13 | 9 | 8 | 69 | | |
| | | 3334 | 277 | 2039 | 78 | 570 | 4487 | 417 | 265 | 195 | 2784 | | |
| | | 3,42% | 0,72% | 4,22% | 2,56% | 3,33% | 2,30% | 3,12% | 3,40% | 4,10% | 2,48% | | |
| | | 100 | 13 | 0 | 0 | 57 | 2 | 7 | 0 | 5 | 121 | | |
| | | 3130 | 407 | 4 | 71 | 1989 | 374 | 455 | 85 | 300 | 3790 | | |
| | | 3,19% | 3,19% | 0,00% | 0,00% | 2,87% | 0,53% | 1,54% | 0,00% | 1,67% | 3,19% | | |
| | | 50 | 1 | 0 | 2 | 38 | 8 | 31 | 6 | 0 | 97 | | |
| | | 1776 | 91 | 15 | 61 | 1306 | 474 | 1116 | 176 | 69 | 2584 | | |
| | | 2,82% | 1,10% | 0,00% | 3,28% | 2,91% | 1,69% | 2,78% | 3,41% | 0,00% | 3,75% | | |
| | | 107 | 3 | 58 | 3 | 0 | 0 | 12 | 0 | 0 | 4 | | |
| | | 3178 | 73 | 1481 | 112 | 0 | 0 | 338 | 57 | 52 | 81 | | |
| | | 3,37% | 4,11% | 3,92% | 2,68% | NaN | NaN | 3,55% | 0,00% | 0,00% | 4,94% | | |
| | | 18 | 4 | 13 | 0 | 114 | 9 | 15 | 2 | 3 | 42 | | |
| | | 473 | 99 | 145 | 29 | 3585 | 378 | 528 | 128 | 58 | 1080 | | |
| | | 3,81% | 4,04% | 8,97% | 0,00% | 3,18% | 2,38% | 2,84% | 1,56% | 5,17% | 3,89% | | |
| | | 41 | 51 | 32 | 14 | 103 | 75 | 43 | 16 | 38 | 66 | | |
| | | 1977 | 909 | 1132 | 406 | 7685 | 4681 | 2933 | 1056 | 2266 | 3438 | | |
| | | 2,07% | 5,61% | 2,83% | 3,45% | 1,34% | 1,60% | 1,47% | 1,52% | 1,68% | 1,92% | | |
| | | 34 | 38 | 0 | 3 | 271 | 26 | 20 | 18 | 125 | 51 | | |
| | | 1459 | 945 | 3 | 194 | 13774 | 1464 | 1420 | 729 | 4752 | 2012 | | |
| | | 2,33% | 4,02% | 0,00% | 1,55% | 1,97% | 1,78% | 1,41% | 2,47% | 2,63% | 2,53% | | |
| | | 88 | 12 | 73 | 3 | 35 | 39 | 9 | 27 | 11 | 67 | | |
| | | 5373 | 365 | 2499 | 122 | 1313 | 1912 | 360 | 1419 | 553 | 3260 | | |
| | | 1,64% | 3,29% | 2,92% | 2,46% | 2,67% | 2,04% | 2,50% | 1,90% | 1,99% | 2,06% | | |
| | | 100 | 22 | 154 | 14 | 44 | 42 | 2 | 41 | 24 | 90 | | |
| | | 3833 | 790 | 3448 | 853 | 2060 | 2765 | 252 | 1266 | 1277 | 3634 | | |
| | | 2,61% | 2,78% | 4,47% | 1,64% | 2,14% | 1,52% | 0,79% | 3,24% | 1,88% | 2,48% | | |
| | | 140 | 12 | 30 | 28 | 41 | 63 | 8 | 72 | 17 | 70 | | |
| | | 6456 | 545 | 1057 | 711 | 2936 | 6051 | 779 | 3793 | 771 | 4494 | | |
| | | 2,17% | 2,20% | 2,84% | 3,94% | 1,40% | 1,04% | 1,03% | 1,90% | 2,20% | 1,56% | | |
| | | 2 | 0 | 0 | 7 | 14 | 141 | 13 | 26 | 2 | 11 | | |
| | | 90 | 116 | 2 | 167 | 461 | 5966 | 588 | 782 | 101 | 621 | | |
| | | 2,22% | 0,00% | 0,00% | 4,19% | 3,04% | 2,36% | 2,21% | 3,32% | 1,98% | 1,77% | | |
| | | 49 | 20 | 93 | 53 | 105 | 34 | 15 | 27 | 18 | 44 | | |
| | | 2545 | 876 | 2349 | 1347 | 5384 | 2258 | 862 | 1268 | 581 | 2203 | | |
| | | 1,93% | 2,28% | 3,96% | 3,93% | 1,95% | 1,51% | 1,74% | 2,13% | 3,10% | 2,00% | | |
| | | 41 | 1 | 1 | 0 | 115 | 113 | 38 | 21 | 4 | 34 | | |
| | | 1227 | 176 | 56 | 47 | 4049 | 5255 | 2189 | 784 | 202 | 1627 | | |
| | | 3,34% | 0,57% | 1,79% | 0,00% | 2,84% | 2,15% | 1,74% | 2,68% | 1,98% | 2,09% | | |
| | | 33 | 29 | 219 | 46 | 9 | 22 | 4 | 25 | 2 | 7 | | |
| | | 949 | 801 | 4342 | 582 | 397 | 1253 | 297 | 978 | 98 | 312 | | |
| | | 3,48% | 3,62% | 5,04% | 7,90% | 2,27% | 1,76% | 1,35% | 2,56% | 2,04% | 2,24% | | |
| | | 42 | 136 | 134 | 47 | 4 | 3 | 2 | 21 | 1 | 13 | | |
| | | 1227 | 1764 | 1979 | 1044 | 175 | 224 | 77 | 961 | 116 | 705 | | |
| | | 3,42% | 7,71% | 6,77% | 4,50% | 2,29% | 1,34% | 2,60% | 2,19% | 0,86% | 1,84% | | |

On voit ainsi que pour un certain nombre de cellules, la représentativité est insuffisante pour réaliser un transfert. Nous regroupons ainsi les régions (de façon différente selon les OTEX) pour augmenter le nombre d'exploitations RICA desquelles se rapprocher pour effectuer un transfert.

Les regroupements sont les suivants :

- **OTEX Grandes cultures (1314)** : Création de régions de cultures : grandes régions si effectifs suffisants, regroupement de région si effectifs trop grands (modèle ne tourne pas).
- **OTEX Viticulture et vins d'appellation d'origine (3738)** : Régions viticoles. 13 bassins existent, regroupés en 9 ici pour des raisons d'effectifs. (Regroupements de départements).
- **OTEX Bovins lait (4100), bovins viande (4200) et bovins mixte (4300)** : Bassins laitiers en 9 classes.
- **OTEX Maraîchage (2829), Fruits (3900), Ovins caprins autres herbivores (4400), Porcins volailles (5072), Polyculture et polyélevage (6082)** : grandes régions



6.3. Variables utilisées

| | | |
|---|---|--|
| Variables individuelles (exploitation) | asb_aa | assiette de revenu |
| | sup | superficie de l'exploitation |
| | reg | région |
| | salaires | salaires versés |
| | inst_an_max | dernière année d'installation d'un chef d'exploitation |
| | age_moy | age moyen |
| | W_TRAVAIL | indicateur de travail dans l'exploitation |
| | w_conj | indicateur de travail des conjoints dans l'exploitation |
| | NB_EXPL | indicateur du nombre de chefs d'exploitation |
| | NB_EXPLH | indicateur de la présence de chefs d'exploitation homme |
| | NB_EXPLF | indicateur de la présence de chefs d'exploitation femme |
| | AF | indicateur de la présence d'aides familiaux |
| | nban_inst_max | Nombre d'années depuis la première installation (connue) sur l'exploitation |
| | nban_instmax | Nombre d'années depuis la dernière installation de chef sur l'exploitation |
| exo_ja | Au moins un des chefs est exonéré jeune agriculteur | |
| ecart_asb | ecart d'assiette de revenu entre n et n-1 | |
| Variables de contexte (commune) | altitude | altitude |
| | Altdelta | altitude du point le plus haut – altitude du point le plus bas |
| | Données premier pilier PAC | |
| | Somme_pmva | Somme touchées sur la Prime au Maintien du troupeau de Vaches Allaitantes |
| | Compte_pmva | Nombre de bénéficiaires à la Prime au Maintien du troupeau de Vaches Allaitantes |
| | Somme_pmsee | Somme touchées sur la Prime au Maintien des Systèmes Extensifs |
| | Compte_pmsee | Nombre de bénéficiaires à la Prime au Maintien des Systèmes Extensifs |
| | Somme_pab | Somme touchées sur la Prime à l'Abattage |
| | Compte_pab | Nombre de bénéficiaires à la Prime à l'Abattage |
| | Somme_psbm | Somme touchées sur la Prime Spéciale pour les Bovins Mâles |
| | Compte_psbm | Nombre de bénéficiaires à la Prime Spéciale pour les Bovins Mâles |
| | Somme_pbc | Somme touchées sur la Prime à la Brebis et/ou la Chèvre |
| | Compte_pbc | Nombre de bénéficiaires à la Prime à la Brebis et/ou la Chèvre |
| | Somme_pmr | Somme touchées sur la Prime au Monde Rural |
| | Compte_pmr | Nombre de bénéficiaires à la Prime au Monde Rural |
| | Données de surface (concernant toutes les exploitations de la commune) en 2005 | |
| | ZONES_URBANISEES | Surface en zones urbanisées |
| | terres_arables | Surfaces en terres arables |
| | forets | Surfaces en forêts |
| | zones_agricoles_heterogenes | Surfaces en zones agricoles hétérogènes |
| | prairies | Surfaces en prairies |
| | Données météo | |
| | preliq | précipitations liquides |
| | prenei | précipitations neigeuses |
| | nbpreliq | nombre de jour de pluie |
| | dli | rayonnement solaire |
| | vent | vent |
| | hr | humidité relative |
| | tn | température minimale |
| | tx | température maximale |
| | tmoy | temperature moyenne |
| | nbjgel | nmbre de jours de gel (moyenne entre 2000 et 2004) |
| | preliq_0004 | précipitations liquides (moyenne entre 2000 et 2004) |
| | prenei_0004 | précipitations neigeuses (moyenne entre 2000 et 2004) |
| | nbpreliq_0004 | nombre de jour de pluie (moyenne entre 2000 et 2004) |
| | dli_0004 | rayonnement solaire (moyenne entre 2000 et 2004) |
| | vent_0004 | vent (moyenne entre 2000 et 2004) |
| | hr_0004 | humidité relative (moyenne entre 2000 et 2004) |
| | tn_0004 | température minimale (moyenne entre 2000 et 2004) |
| | tx_0004 | température maximale (moyenne entre 2000 et 2004) |
| | tmoy_0004 | temperature moyenne (moyenne entre 2000 et 2004) |
| | Données MSA | |
| | nb_explsd | écart-type du nombre d'exploitants par exploitation |
| | supsd | écart-type de la superficie des exploitations |
| | AGE_moysd | écart type de l'âge moyen par exploitation |
| | salairesd | écart type des salaires versés par exploitation |
| | asb_aasd | écart type de l'assiette de revenus par exploitation |
| inst_an_maxsd | écart type de la dernière année d'installation par exploitation | |
| ecart_asbsd | écart type de l'écart de l'assiette de revenu entre n et n-1 par exploitation | |

6.4. Choix de la distance utilisée et du caliper

Nous comparons deux distances utilisées lors de l'étape du transfert d'indicateur : la distance de Mahalanobis et la distance calculée avec le score de propension.

Calcul de la distance de Mahalanobis :

Une exploitation i est représentée par le vecteur $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, Σ est la matrice de variance-covariance. La matrice de distance entre 2 individus i et j individus est alors :

$$D(X_i, X_j) = (X_i - X_j)^T \Sigma^{-1} (X_i - X_j)$$

Calcul de la distance basée sur le score de propension:

Le score de propension pour un individu dont les caractéristiques sont X_i est calculé par une modélisation linéaire :

$$P(X_i) = \Pr(R = 1 | X_i) = E(R | X_i)$$

où R est ici l'appariement à une exploitation RICA. La distance est alors l'écart entre les valeurs prédites du modèle linéaire.

Nous testons également plusieurs seuils de caliper :

10 : seuls les 10% inférieurs de la matrice de distance sont conservés.

20 : seuls les 20% inférieurs de la matrice de distance sont conservés.

25 : seuls les 25% inférieurs de la matrice de distance sont conservés.

50 : seuls les 50% inférieurs de la matrice de distance sont conservés.

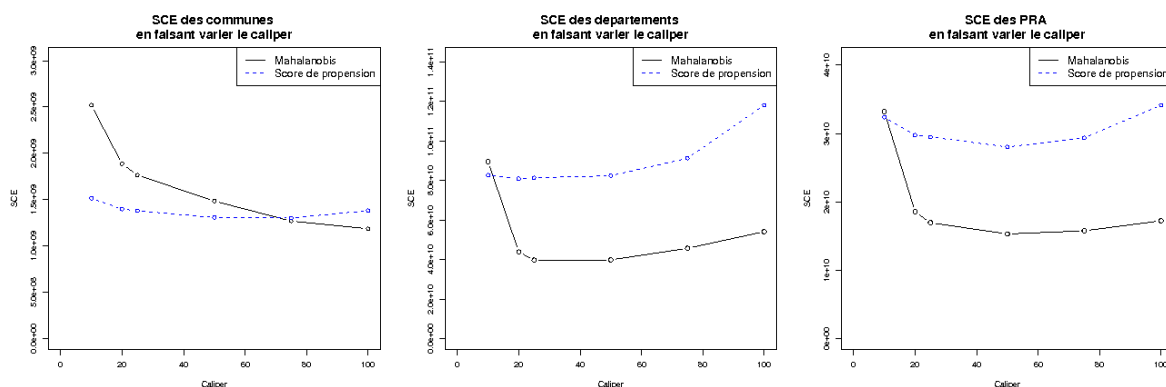
75 : seuls les 75% inférieurs de la matrice de distance sont conservés.

100 : toute la matrice de distance est conservée.

Les tests ci-dessous ont été faits sur l'indicateur UGB bovines.

Compte tenu de la variabilité au niveau communal, nous effectuons les choix de la méthode en tenant compte également des niveaux supérieurs : PRA et département.

Nous pouvons comparer la somme des carrés des erreurs des communes (ou PRA ou départements), indicées par j pour les deux distances et en faisant varier le caliper $SCE = \sum_j (UGBBO_{BDNI} - UGBBO_{transféré})^2$:



Le seuil de 10% est vraiment moins performant pour la distance de Mahalanobis que les autres valeurs de caliper, quelque soit le niveau géographique.

Au niveau communal, la distance de Mahalanobis est de plus en plus performante, au fur et à mesure qu'on diminue la restriction sur la matrice de distance. La distance avec le score de propension semble elle plus

stable, et plus performante que celle de Mahalanobis pour des seuils <75%. Le seuil de 50% semble optimal pour la distance calculée avec le score de propension au niveau communal, et pour les deux distances au niveau PRA.

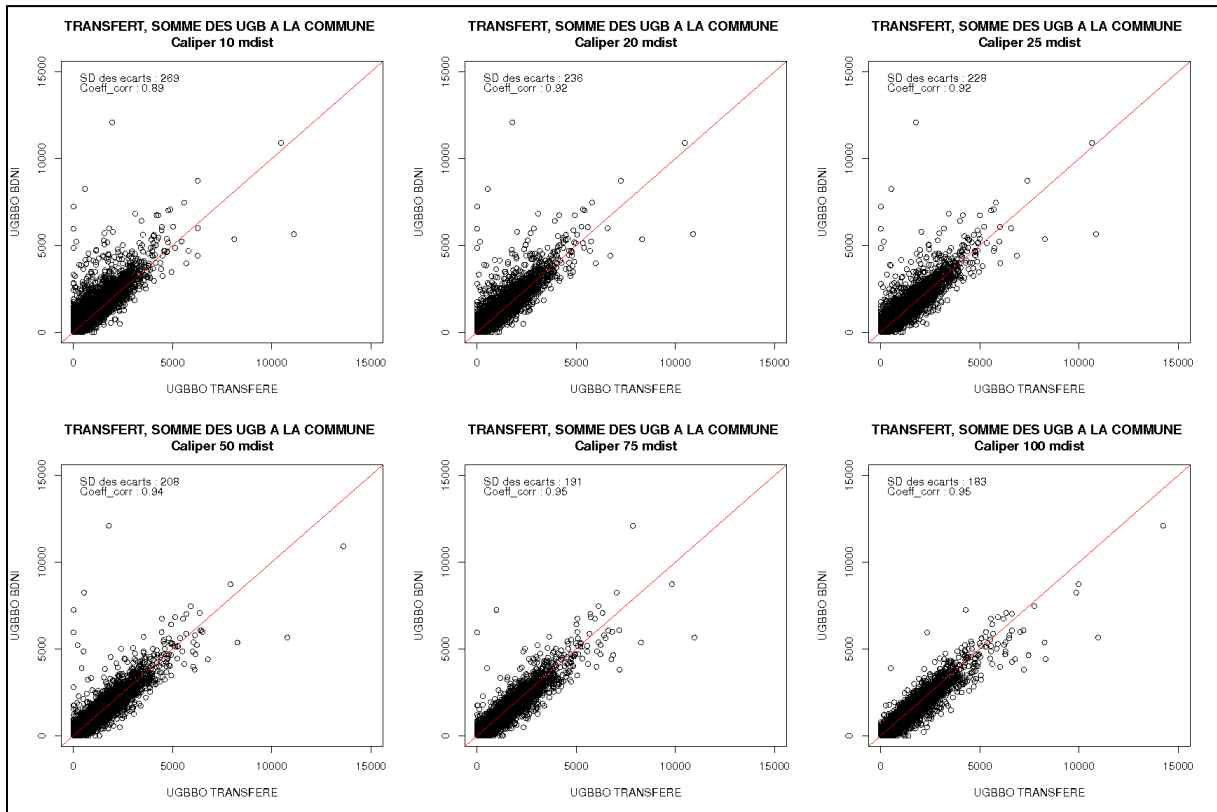
Au niveau PRA comme au niveau département, excepté le seuil de 10%, la distance de Mahalanobis est toujours plus performante pour le critère de la somme du carré des erreurs. En effet (hors seuil de 10%), le SCE pour la distance de Mahalanobis vaut entre 0.50 et 0.62% le SCE de la distance PS au niveau PRA, et entre 0.45 et 0.54% au niveau département.

Le critère du SCE fait préférer l'utilisation de la distance de Mahalanobis pour des niveaux supérieurs à la commune.

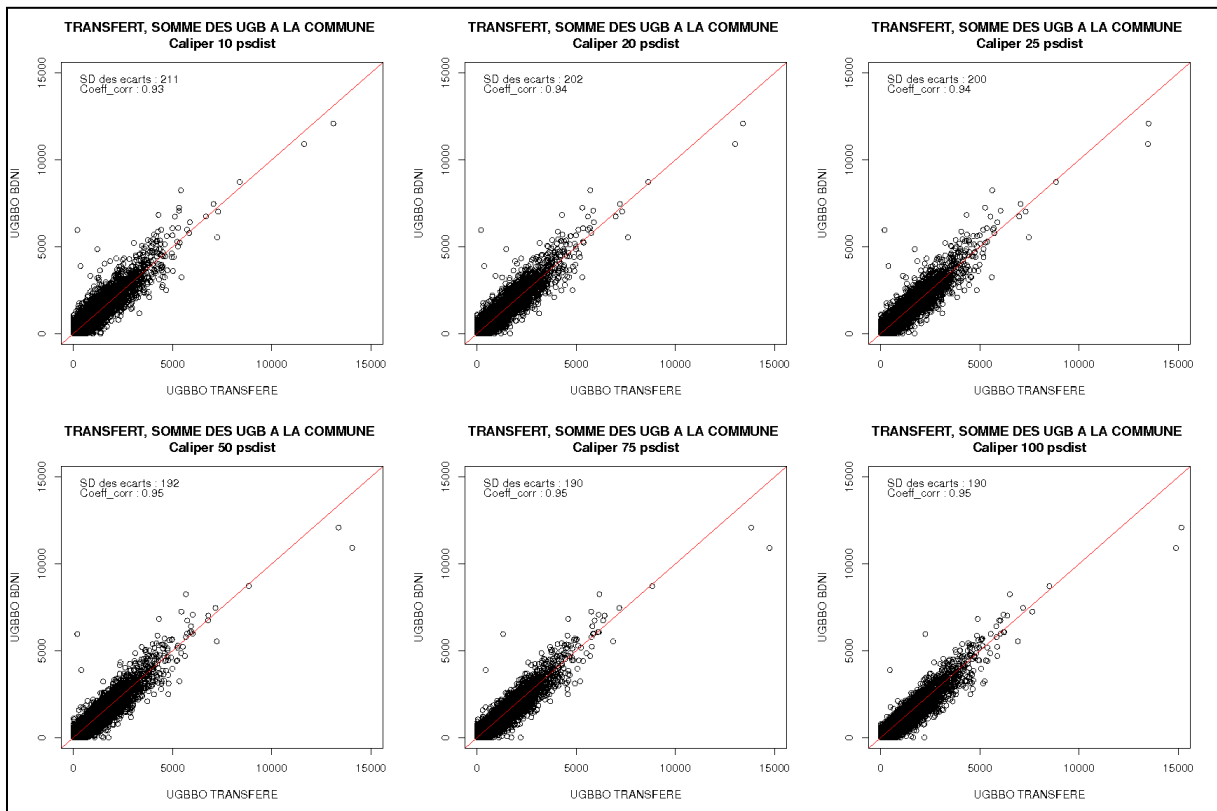
Nous regardons également l'ajustement pour chacun des seuils et les deux distances, aux « niveaux géographiques. Les graphiques ci-dessous représentent l'ajustement entre les valeurs de référence et les valeurs transférées.

Comparaison des valeurs transférées aux valeurs réelles, à la commune :

Distance de Mahalanobis, commune

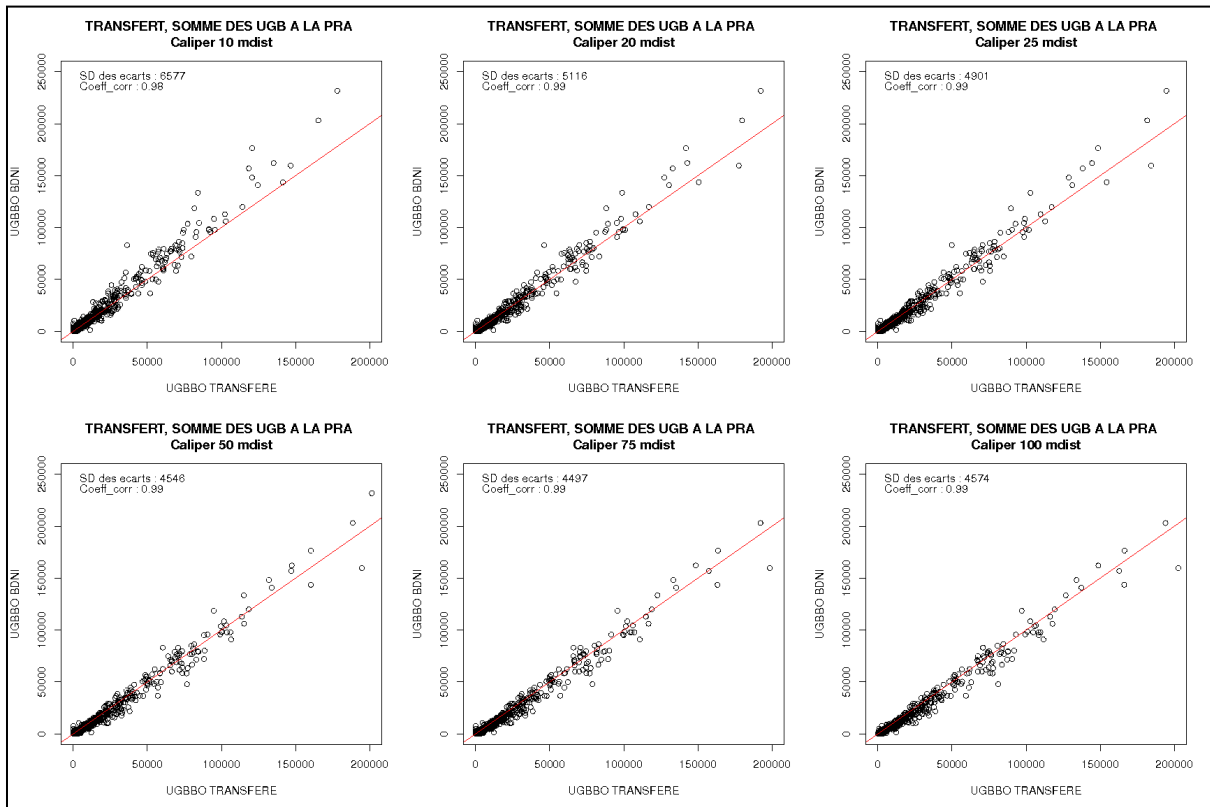


Distance score de propension, commune

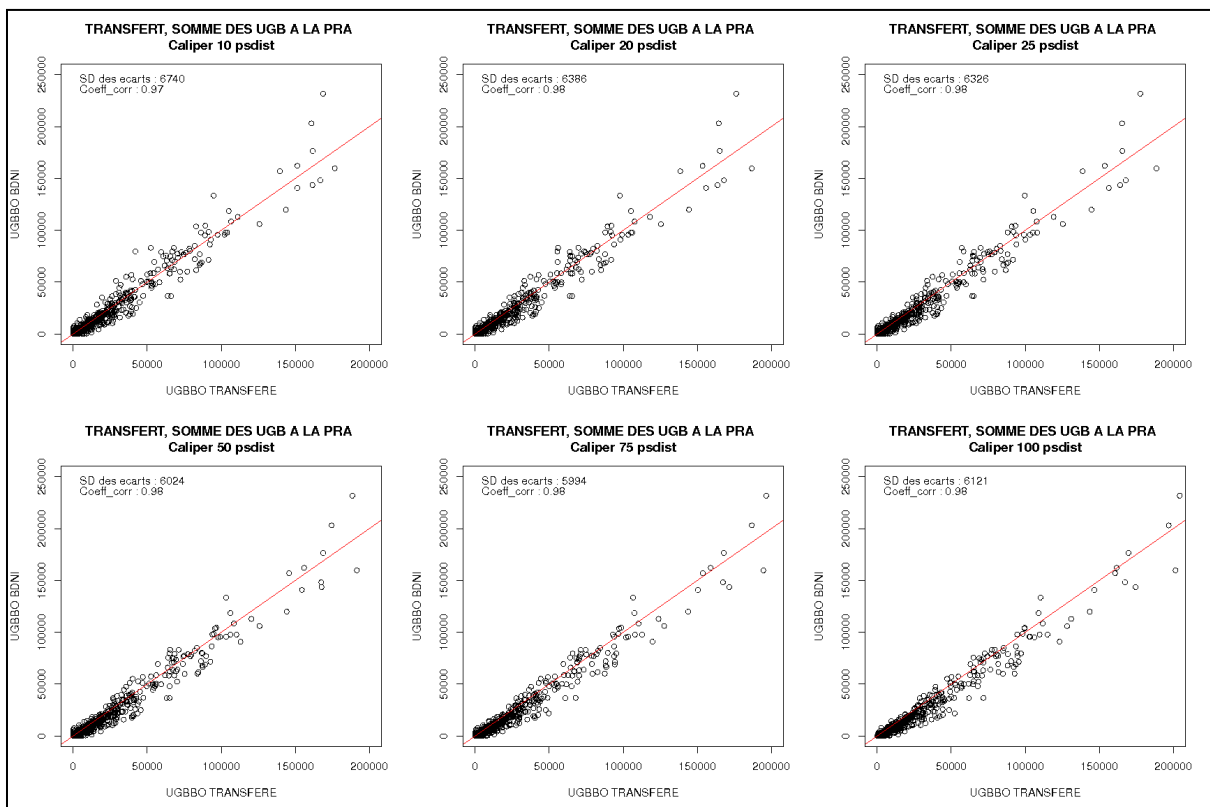


Comparaison des valeurs transférées aux valeurs réelles, à la PRA :

Distance de Mahalanobis, PRA

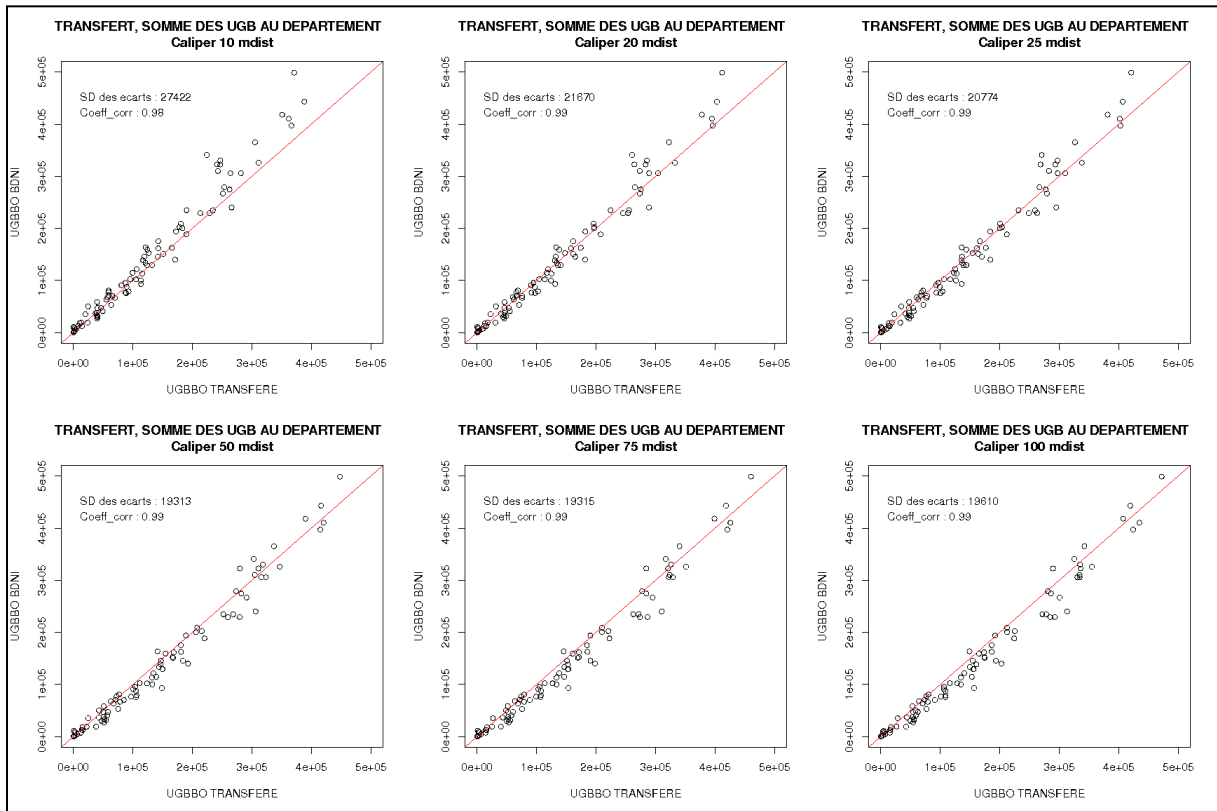


Distance score de propension, PRA

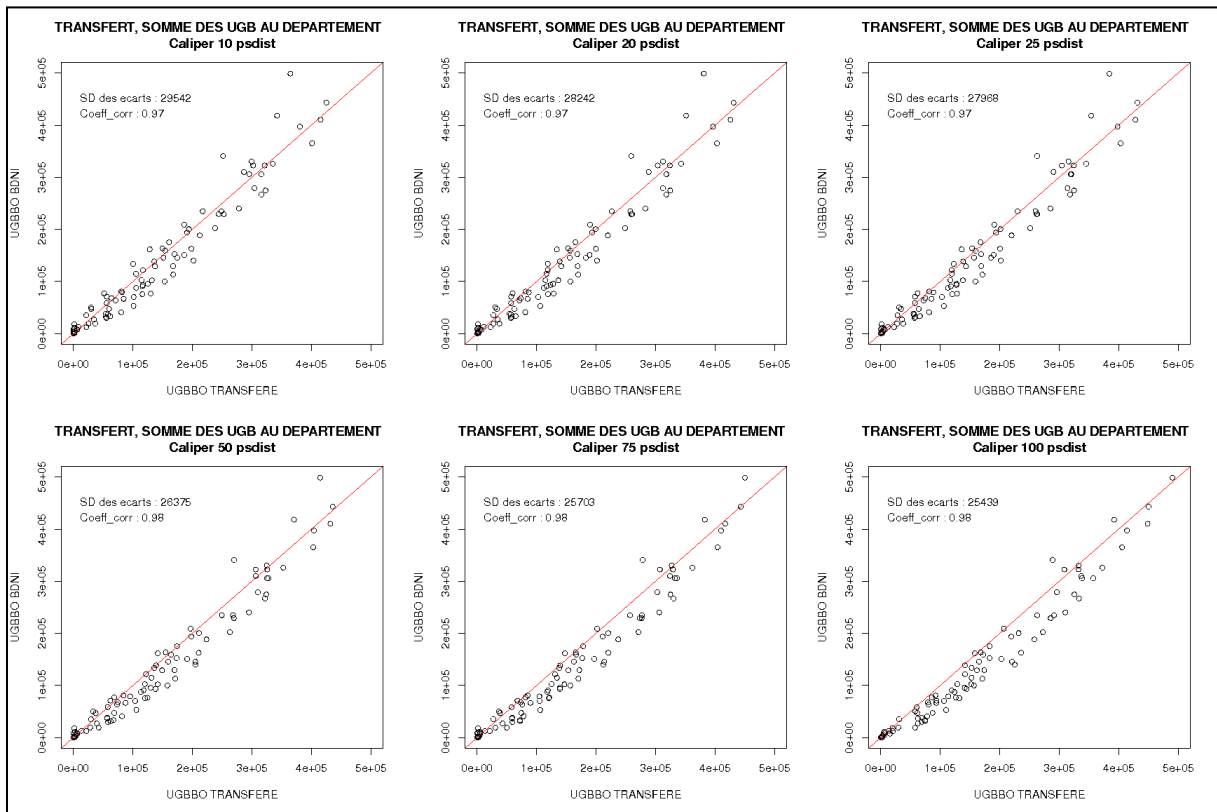


Comparaison des valeurs transférées aux valeurs réelles, au département :

Distance de Mahalanobis, département



Distance score de propension, département

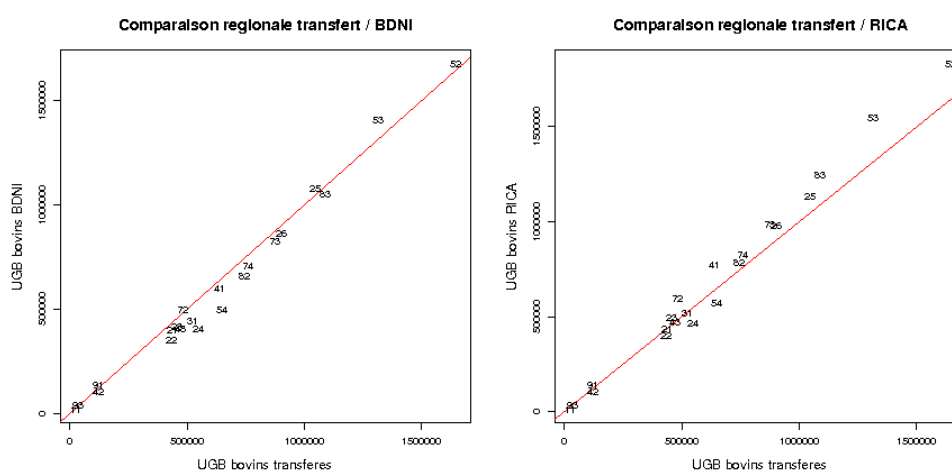


Ainsi au vu de ces résultats nous choisirons d'utiliser la distance de Mahalanobis pour le transfert des indicateurs, avec un caliper égal à la médiane de la matrice de distance. Ce choix est discutable et pourrait être différent selon l'indicateur transféré.

NB : ici, la matrice de poids est calculée comme l'exponentielle inverse de la matrice de distance. L'inverse de la distance a également été testé, et donnait au niveau communal des résultats comparables par rapport au coefficient de corrélation et des tests de rangs, mais moins satisfaisants aux niveaux PRA et département.

6.5. Performances par région

Nous pouvons observer les résultats régionaux comme précédemment au niveau PRA ou département avec la BDNI, mais nous pouvons également les comparer avec le nombre d'UGB bovines régionaux déterminées à partir du RICA. Les numéros des régions apparaissent sur le graphique.



On voit ainsi qu'il existe un décalage entre le nombre d'UGB bovines calculées à partir du RICA et celui de la BDNI. Le coefficient de corrélation est de 99.1 entre données transférées et BDNI, respectivement de 99.2 entre les données transférées et le RICA. L'ajustement avec la BDNI se fait mieux sur les grandes régions bovines que les régions d'importance moyenne en nombre d'UGB bovines. Ceci peut avoir plusieurs causes, nous regardons ainsi la représentativité des exploitations RICA dans les régions et l'homogénéité des exploitations.

Les graphiques ci-dessous représentent la somme des carrés des erreurs (niveau communal) par région, en fonction du pourcentage d'exploitations appariées RICA dans la région (niveau PRA) et en fonction de l'écart type des UGB bovines dans la BDNI.

