HAL
open science

Analyse spatio-temporelle des référentiels métiers de l'Infrastructure de Données Géographiques Géofoncier pour la description et la représentation de phénomènes statistiques inhérents à ces données

Emmanuel Labrousse

To cite this version:

Emmanuel Labrousse. Analyse spatio-temporelle des référentiels métiers de l'Infrastructure de Données Géographiques Géofoncier pour la description et la représentation de phénomènes statistiques inhérents à ces données. Sciences de l'ingénieur [physics]. 2015. dumas-01338160

HAL Id: dumas-01338160

https://dumas.ccsd.cnrs.fr/dumas-01338160v1

Submitted on 28 Jun 2016

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.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

ÉCOLE SUPÉRIEURE DES GÉOMÈTRES ET TOPOGRAPHES

MÉMOIRE

présenté en vue d'obtenir

le DIPLÔME D'INGÉNIEUR CNAM

Spécialité : Géomètre et Topographe

par

Emmanuel LABROUSSE

Analyse spatio-temporelle des référentiels métiers de l'Infrastructure de Données Géographiques Géofoncier pour la description et la représentation de phénomènes statistiques inhérents à ces données

Soutenu le 07 juillet 2015

JURY

PRESIDENT :	**M. Vincent HABCHI**	
MEMBRES :	**M. Jean-Michel FOLLIN**	**Professeur référent**
	M. Pierre TOUZARD	**Maître de stage**

Remerciements

Je tiens particulièrement à remercier les personnes m'ayant suivi lors de ce travail de recherche :

- Pierre TOUZARD, mon maître de stage, pour son ouverture, son implication et son aide précieuse dans la compréhension intrinsèque de l'IDG Géofoncier ;
- Mathilde de SULZERT WART, gérante de SOGEFI, pour ses conseils concernant l'analyse de données multidimensionnelles ;
- Pierre CLERGEOT, géomètre-expert *honoris causa*, auteur d'ouvrages sur l'évolution de la profession de géomètre-expert en France, pour ses conseils concernant la production cartographique et les analyses à mener ;
- Patrick BEZART-FALGAS, géomètre-expert et co-animateur national du portail Géofoncier au sein de l'OGE, pour son investissement dans la récupération de données financières sur la profession ;
- Jean-Michel FOLLIN, mon professeur référent, pour sa disponibilité, ses remarques qui m'ont conduit à réaliser des analyses concernant l'occupation du sol et ses conseils concernant la rédaction du rapport.

Enfin, je remercie pour son accueil chaleureux l'ensemble de l'équipe de SOGEFI[1] et SOGEXFO[2], sociétés voisines basées à Toulouse.

[1] Site web :http://www.sogefi-sig.com/
[2] Site web : http://www.sogexfo.com/

Liste des abréviations

ACP : analyse en composantes principales

AFAF : Aménagement Foncier Agricole et Forestier

APE : Activité Principale Exercée

API : Interface de programmation (issu de l'anglais : Application Programming Interface)

BRGM : Bureau de Recherches Géologiques et Minières

CA : Chiffre d'Affaires

CAH : Classification Ascendante Hiérarchique

CAO/DAO : Conception/Dessin Assisté par Ordinateur

CRAN : Comprehensive R Archive Network

BDD : Base De Données

DF : Dépendance Fonctionnelle

DGFiP : Direction Générale des Finances Publiques

DMPC : Document Modificatif du Parcellaire Cadastral

ECIR : Échanges et Cessions amiables d'Immeubles Ruraux

EBE : Excédent Brut d'Exploitation

ETL : Extract Transform and Load

GNSS : Système de positionnement par satellites (issu de l'anglais : Global Navigation Satellite System)

IDG : Infrastructure de Données Géographiques

IGN : Institut national de l'information géographique et forestière

INSEE : Institut National de la Statistique et des Études Économiques

INSPIRE : Infrastructure for Spatial Information in Europe

OGE : Ordre des Géomètres-Experts

ORES : observatoire régional économique et social

OSM : OpenStreetMap

PCI : Plan Cadastral Informatisé

RFU : Référentiel Foncier Unifié

RGF93 : Réseau Géodésique Français 1993

RPG : Registre Parcellaire Graphique

SGBD : Système de Gestion de Base de Données

SI : Système d'Information

SIG : Système d'Information Géographique

SQL : Structured Query Language

TMS : Tile Map Service

WKB : Well-Known Binary

WKT : Well-Known Text

WMS : Web Map Service

WMTS : Web Map Tile Service

Glossaire

Termes statistiques :

Dendrogramme : diagramme utilisé en classification hiérarchique représentant les regroupements successifs de classes d'individus.

Individu : élément d'une population statistique étudiée.

Modalité : valeur pouvant être prise par une variable qualitative.

R : logiciel libre et langage de programmation statistique.

Termes informatiques :

Application Programming Interface (API) : ensemble de fonctions, de protocoles permettant de communiquer, de transférer des données entre plusieurs logiciels.

Cluster : grappe de serveurs fonctionnant comme un seul système.

Serveur de recette : serveur dédié à l'expérimentation et la validation de nouvelles fonctionnalités.

Widget : élément d'une interface graphique d'une application (bouton ou menu déroulant par exemple)

Termes de comptabilité :

Capital : ensemble des biens propres représentés par les apports, en numéraire ou en nature, effectués par les associés lors de la création de la société.

Chiffre d'affaires : somme des ventes de marchandises, de produits fabriqués, des prestations de services et de travaux.

Excédent brut d'exploitation : solde du compte d'exploitation d'une société (excédent dégagé par les activités d'exploitation une fois la main-d'œuvre rémunérée). Contrairement au résultat net comptable, il ne prend pas en compte la gestion financière (intérêts d'un emprunt, dette…), les dotations aux amortissements (dépréciation, perte de valeur des biens) et les évènements exceptionnels de l'entreprise (prêt, caution…).

Résultat net comptable : différence entre les produits et les charges de l'entreprise sur la période correspondant à l'exercice comptable (généralement une année civile). On parle de bénéfice en cas de résultat positif et de perte en cas de résultat négatif.

Termes en lien avec la profession de géomètre-expert :

Aménagement foncier agricole et forestier (anciennement remembrement) : aménagement foncier visant à la redistribution de parcelles morcelées et dispersées sur des propriétés rurales non bâties dans l'objectif de constituer des exploitations rurales d'un seul tenant ou à parcelles groupées (article L123-1 du code rural).

Bornage : « opération qui a pour effet de définir juridiquement et de matérialiser sur le terrain les limites des propriétés privées contiguës, appartenant ou destinées à appartenir à des propriétaires différents. « (OGE, 2014, p. 1)

Délimitation de la propriété des personnes publiques : positionnement de « la limite de la propriété des personnes publiques et/ou la limite des différentes domanialités avec les propriétés riveraines. » (OGE, 2014, p. 18)

Document modificatif du parcellaire cadastral : document établi par le géomètre-expert et fourni au cadastre dans le cadre d'un changement de limite de propriété (division de propriété, réunion de parcelles…).

Échange et cession amiables d'immeubles ruraux : aménagement foncier réalisé dans l'objectif d'améliorer les conditions d'exploitation des immeubles ruraux, « d'assurer la mise en valeur des espaces naturels ruraux et de contribuer à l'aménagement du territoire communal ou intercommunal défini dans les documents d'urbanisme. » (OGE, 2014, p. 22).

État descriptif de division : document servant à l'identification des lots/volumes d'une copropriété/division en volumes pour la publicité foncière.

Publicité foncière : procédure visant à informer les tiers de tout changement concernant la situation juridique d'un immeuble.

Servitude : charge imposée sur un fonds (fonds servant) au profit d'un autre fonds (fonds dominant).

Autres termes :

Cartogramme : carte en anamorphose où la valeur d'une variable telle que la population est utilisée comme valeur de superficie des zones représentées (départements, pays…).

Trigramme : ensemble de trois caractères consécutifs.

Valeur : en sémiologie graphique, il s'agit du degré de clarté d'un élément pouvant aller du noir au blanc.

Sommaire

INTRODUCTION **8**

I STRUCTURATION ET DÉMARCHE D'ÉTUDE DE L'IDG **10**

I.1 PRÉSENTATION DE GÉOFONCIER: DES BASES DE DONNÉES AU PORTAIL GÉOGRAPHIQUE 10

I.1.1 Une infrastructure de données au service de la profession de géomètre-expert 10

I.1.2 Organisation des données accessibles sur Géofoncier (côté client) 10

I.1.2.1 Les données propres à l'IDG 10

I.1.2.2 Les données issues des partenaires 11

I.1.3 Les bases de données étudiées (côté serveur) 11

I.2 DÉMARCHE GÉNÉRALE ET OUTILS EMPLOYÉS 14

I.2.1 De l'IDG Géofoncier au serveur local dédié aux analyses statistiques 14

I.2.2 Traitement sous PostgreSQL et R 16

I.2.3 Production de graphes et rendus cartographiques 16

II ANALYSES STATISTIQUES MENÉES : STRUCTURE, ACTIVITÉ ET ÉVOLUTION DE LA PROFESSION **17**

II.1 APPROCHE DESCRIPTIVE DE LA PROFESSION ET DE SON ÉVOLUTION 17

II.1.1 Évolution quantitative de la profession 17

II.1.1.1 Méthodologie et fiabilité des données 17

II.1.1.2 Analyse des évolutions depuis 1999 : baisse du nombre de cabinets et rajeunissement de la profession 18

II.1.2 Évolution structurelle des cabinets 19

II.1.2.1 Méthodologie 19

II.1.2.2 Analyse des résultats : de fortes mutations structurelles depuis 2000 20

II.1.3 Situation financière des cabinets de géomètres-experts 21

II.1.3.1 Données et méthodologie 21

II.1.3.2 Analyse : une faible disparité des bénéfices par géomètre-expert 23

II.2 ÉTUDE DE L'ACTIVITÉ DES GÉOMÈTRES-EXPERTS À PARTIR DES TYPES DE DOSSIERS VERSÉS SUR GÉOFONCIER 24

II.2.1 Critères de l'étude 24

II.2.2 Analyse exploratoire des répartitions de dossiers versés 26

II.2.2.1 Objectif et variables de l'analyse 26

II.2.2.2 Principe et méthode 27

II.2.2.3 Déroulement de l'ACP 28

II.2.2.4 Analyse des résultats : distinction des domaines fonciers et immobiliers 30

II.2.3 Analyse complémentaire sur l'activité et l'économie d'un cabinet 31

II.2.4 Établissement d'une typologie des cabinets à partir du chiffre d'affaires 32

II.2.4.1 Sélection et fiabilité des données employées 32

II.2.4.2 Principes de la CAH 33

II.2.4.3 Déroulement de la classification sur le jeu de données 34

II.2.4.4 Analyse des classes 37

III ANALYSE SPATIALE ET CARTOGRAPHIE **39**

III.1 RÉPARTITION DES BUREAUX DE GÉOMÈTRES-EXPERTS 39

III.1.1 Méthodologie et données complémentaires 39

III.1.2 Une concentration des bureaux au cœur des aires urbaines 40

III.1.3 Une présence plus marquée dans le sud de la France 41

III.1.4 Étude complémentaire sur le secteur du bâtiment et les bureaux de géomètres-experts 42
III.2 CARTOGRAPHIE DES DOSSIERS PRÉSENTS SUR GÉOFONCIER 43
III.2.1 Répartition départementale des données du RFU 43
III.2.2 Carte de densité concernant les dossiers 43
III.2.3 Répartition départementale des dossiers fonciers et immobiliers 44
III.3 ANALYSE RÉGIONALE : LE CAS DE LA RÉGION MIDI-PYRÉNÉES 45
III.3.1 Sélection des données d'appui pour les analyses 45
III.3.2 Répartition de l'activité selon l'occupation du sol 46
III.3.3 Proximité géographique des cabinets et concurrence 47
IV GÉNÉRATION DYNAMIQUE DE RÉSULTATS STATISTIQUES À L'AIDE D'UN RENDU (CARTO-) GRAPHIQUE WEB 48
IV.1 SHINY, OUTIL DE VISUALISATION D'ANALYSES STATISTIQUES 48
IV.1.1 Principes de fonctionnement 48
IV.1.2 Gestion des interactions de l'utilisateur à l'aide de la programmation réactive . 49
IV.2 INTERFACE DE VISUALISATION GRAPHIQUE ET CARTOGRAPHIQUE 49
IV.2.1 Cartographie web à partir de leaflet 49
IV.2.2 Création de graphiques 51
CONCLUSION 52
Bibliographie 54
Packages employés sur R 56
Table des figures 57
Liste des tableaux 57
Table des annexes 59
Atlas cartographique 59

Introduction

Géofoncier est une infrastructure de données géographiques (IDG) également nommée infrastructure de données spatiales (IDS), terme traduit directement de l'anglais *Spatial Data Infrastructure*. Une IDG est une plateforme visant à « faciliter l'accès et le partage des ressources informationnelles géographiques détenues par un large éventail d'acteurs en vue de maximiser leur utilisation globale » (Masser, 2010).

Dans le cadre de Géofoncier, les « acteurs » principaux sont les géomètres-experts, seuls professionnels habilités par la loi à établir les documents et plans délimitant les propriétés foncières. Cette profession ordinale a su diversifier ses compétences depuis la création de l'Ordre des géomètres-experts en 1946 en intervenant désormais dans l'urbanisme, l'immobilier ou l'information géographique. Depuis 70 ans, les changements juridiques, économiques et politiques ont amené la profession à s'adapter au niveau de son organisation. En 1960, les cabinets étaient composés d'un unique géomètre-expert exerçant en individuel, sans aucun moyen de s'associer avec un confrère. Désormais, il est possible d'être à la fois salarié et géomètre-expert.

Les « ressources » principales dont dispose Géofoncier sont les versements réalisés par la profession en lien avec leurs activités, notamment foncières : l'IDG a pour objectif de centraliser sur une plateforme commune les opérations effectuées par les géomètres-experts de l'ensemble du territoire[3].

Géofoncier est officiellement opérationnel depuis le 1er juillet 2010. Il s'agit d'une mine d'informations jusqu'à présent inexploitée en ce qui concerne l'organisation et les activités de la profession. Il inclut les données issues de l'ancienne base Aurige qui rassemblait des informations concernant la localisation des dossiers des géomètres-experts. En 2015, l'IDG comprend plus de trois millions de dossiers versés avec plus de 30 000 versements mensuels.

Ce mémoire s'inscrit dans deux problématiques. En quoi l'infrastructure de données géographiques Géofoncier permet-elle de refléter l'état actuel de la profession et son évolution récente d'un point de vue géographique, économique et structurel ? Comment rendre compte de façon dynamique et dans un contexte web de l'évolution temporelle des données ?

L'exploitation de l'IDG pose *a priori* deux difficultés principales qui seront présentes tout le long de ce travail de recherche. La première est d'ordre quantitatif et concerne la gestion et le traitement d'un volume considérable de données en évolution permanente. Cela implique de chercher des méthodes peu coûteuses en termes de temps de calcul afin de réaliser certaines analyses complexes et parfois de réduire le nombre d'individus considérés, au sens statistique. Un travail d'optimisation doit donc être souvent effectué.

La seconde est d'ordre qualitatif : il s'agit de la fiabilité des informations pouvant être extraites. Ce second point soulève notamment le problème de l'exhaustivité des données présentes. En prospection de données, ces questions sont fréquentes et amènent à faire des choix au niveau des méthodes de traitement, de la sélection des données et des analyses pertinentes à mener.

L'aspect véritablement prospectif du projet conduit par ailleurs à utiliser des données externes en lien avec plusieurs thématiques. Ces sources complémentaires peuvent être confrontées ou mises en relation avec les données issues de l'IDG. Afin d'exploiter au mieux cet ensemble riche d'informations, des méthodes d'exploration de données (ou *data mining*) adaptées à l'étude d'un grand nombre de variables doivent alors être employées.

[3] Plus précisément, cela inclut les 101 départements français ainsi que les collectivités d'outre-mer de St-Barthélémy et St-Martin.

Géofoncier est le portail géographique de l'information foncière. Il est donc nécessaire de prendre en compte la composante spatiale présente dans les données. Cela implique l'utilisation de différents modes de représentation et de visualisation afin de conjuguer les statistiques, notamment exploratoires, et la cartographie. Cet aspect graphique est un élément prépondérant du mémoire.

La première partie présente Géofoncier, le portail géographique et la structuration des données hébergées ainsi que les outils, techniques et chaînes de traitement mises en œuvre lors de l'étude. Elle fixe les bases à partir desquelles les analyses ont ensuite été menées.

La seconde partie concerne les statistiques descriptives et exploratoires réalisées en précisant les données sélectionnées, la méthodologie spécifique à chaque analyse ainsi que les graphiques et résultats obtenus. La troisième partie décrit l'analyse spatiale de la profession et de ses activités au travers de la production de documents cartographiques.

La dernière partie s'intéresse à la mise en place d'une application web permettant la visualisation d'éléments statistiques et cartographiques dynamiques adaptés à l'évolution temporelle des données de l'IDG.

I Structuration et démarche d'étude de l'IDG

Cette première partie permet de présenter l'ensemble des données de Géofoncier accessibles pour l'étude, ainsi que les outils essentiels de ce travail de recherche.

I.1 Présentation de Géofoncier: des bases de données au portail géographique

I.1.1 Une infrastructure de données au service de la profession de géomètre-expert

Ouvert à la profession depuis le 1er juillet 2010, cette IDG vise en premier lieu à répondre aux exigences de la directive INSPIRE en matière de diffusion et d'interopérabilité des données géographiques produites par les acteurs publics ou assimilés de l'union européenne. L'OGE entre dans ce cadre de par sa délégation de service public sur l'organisation et la surveillance de la profession. L'IDG est aussi le support de l'élaboration du Référentiel Foncier Unifié (RFU) dont l'objectif est de pérenniser l'ensemble des limites foncières géoréférencées par les géomètres.

L'accessibilité du portail est élargie en avril 2011 au grand public. En 2012, une API Géofoncier est développée afin de permettre aux géomètres-experts d'effectuer des versements et de renseigner les objets RFU à partir de logiciels de CAO/DAO (Covadis, TopStation) et des logiciels de gestion/archivage (Cogeo, GX notamment).

À l'initiative du projet, l'OGE a mis en place dès l'automne 2010 la SAS Géofoncier qui est chargée de la gestion de l'IDG. Les serveurs sont hébergés par le BRGM sur son site d'Orléans, au quartier universitaire de La Source. L'administration technique ainsi que l'assistance aux utilisateurs sont pris en charge par la société SOGEFI. La maintenance applicative du portail est à la charge de la société NEOGEO TECHNOLOGIES.

Géofoncier est un outil rendu incontournable car au service de la déontologie ordinale formalisée dans le règlement intérieur de la profession. Les géomètres-experts ont l'obligation ordinale de géolocaliser certains types de dossiers concernant les opérations foncières et immobilières. Pour les opérations foncières, le versement d'objets RFU décrivant les limites géoéréférencées lors des travaux est également une obligation ordinale.

I.1.2 Organisation des données accessibles sur Géofoncier (côté client)

I.1.2.1 Les données propres à l'IDG

Cette partie présente brièvement les informations visibles sur le portail pour les géomètres-experts sur Géofoncier. Les données essentielles sont celles concernant les opérations réalisées et versées par la profession.

Pour chaque dossier visible, il est possible d'accéder aux informations suivantes en interrogeant la couche de données : la thématique dominante (voir figure 1), la nature de l'opération, le géomètre-expert responsable de l'intervention, le cabinet détenteur du dossier, sa date de création, les références cadastrales, la présence éventuelle d'objets RFU ou de documents matérialisés liés. Lorsqu'un utilisateur recherche des informations plus détaillées sur le détenteur d'un dossier ou désire le contacter, celui-ci est dirigé vers le site de la profession (*www.geometre-expert.fr*) à la page correspondant au cabinet ciblé.

En ce qui concerne le RFU, chaque géomètre-expert a accès aux limites et sommets RFU avec une information sur le type de sommet (borne, axe de cours d'eau…).

Des anciens documents cadastraux numérisés sont également disponibles localement (en 2015 : Haute-Garonne, Haute-Savoie, Lozère et Aveyron). Ils ont été mis en ligne à l'initiative des géomètres-experts locaux.

Figure 1 : données produites par les géomètres-experts visibles sur le portail Géofoncier

Ces données produites par la profession sont complétées par celles issues de l'OGE concernant la localisation des bureaux de géomètres-experts avec la précision du type de bureau (principal, secondaire, bureau de chantier ou permanence).

D'autres données utiles à la profession sont fournies ponctuellement concernant les aménagements fonciers avec leur période de réalisation et l'état de la vectorisation cadastrale notamment.

I.1.2.2 Les données issues des partenaires

Afin de compléter l'information fournie par les données issues de l'IDG, Géofoncier utilise des données provenant d'acteurs extérieurs :

- De nombreuses couches de données affichées proviennent directement du référentiel à grande échelle de l'IGN (BD Ortho, BD Topo et BD Parcellaire) avec lequel un partenariat a été établi. Certaines couches sont obtenues en utilisant l'API Géoportail. Cela permet d'afficher des données disponibles sur l'IDG Géoportail via des flux WMTS: les limites administratives, les cartes IGN, le registre parcellaire graphique, les cartes forestières, les photographies aériennes ou encore les cartes historiques. Les données de recensement des points géodésiques, des stations du réseau GNSS permanent ainsi que des repères de nivellement sont accessibles via le portail.
- Les données issues d'OpenStreetMap sont également diffusées avec le rendu Standard et Mapquest (flux TMS).
- Depuis février 2015, le PCI-Vecteur provenant de la DGFiP est aussi disponible à l'affichage, commune par commune (flux WMS).
- Des données sont également issues du ministère de l'écologie, du développement durable et de l'énergie concernant les risques et l'environnement (localisation des réserves naturelles, des zones Natura 2000 par exemple) et sont disponibles à l'aide de flux WMS[4].

I.1.3 Les bases de données étudiées (côté serveur)

Deux clusters de serveurs contiennent l'ensemble des données propres à l'IDG. Chaque cluster correspond à un type d'accès : le premier inclut l'ensemble des bases de données accessibles sur la version expert et sécurisée du portail, le second correspond aux informations accessibles sur

[4] Les flux WMS, WMTS et TMS permettent d'afficher des cartes sous forme d'images.

la version publique. Le premier rassemble donc toutes les données issues des versements réalisés par les géomètres-experts et c'est sur celui-ci que les études sont effectuées.

Nous n'allons pas ici présenter l'ensemble des BDD disponibles sur les serveurs de Géofoncier mais uniquement celles exploitées pour les besoins de l'étude.

Les données portent sur trois domaines, séparés au sein de différentes BDD :

- les géomètres-experts, au travers des données synchronisées quotidiennement avec le SI de l'OGE et contenues dans la BDD *gespro* (pour GEStion des PROfessionnels),
- les dossiers versés sur le portail,
- les objets RFU.

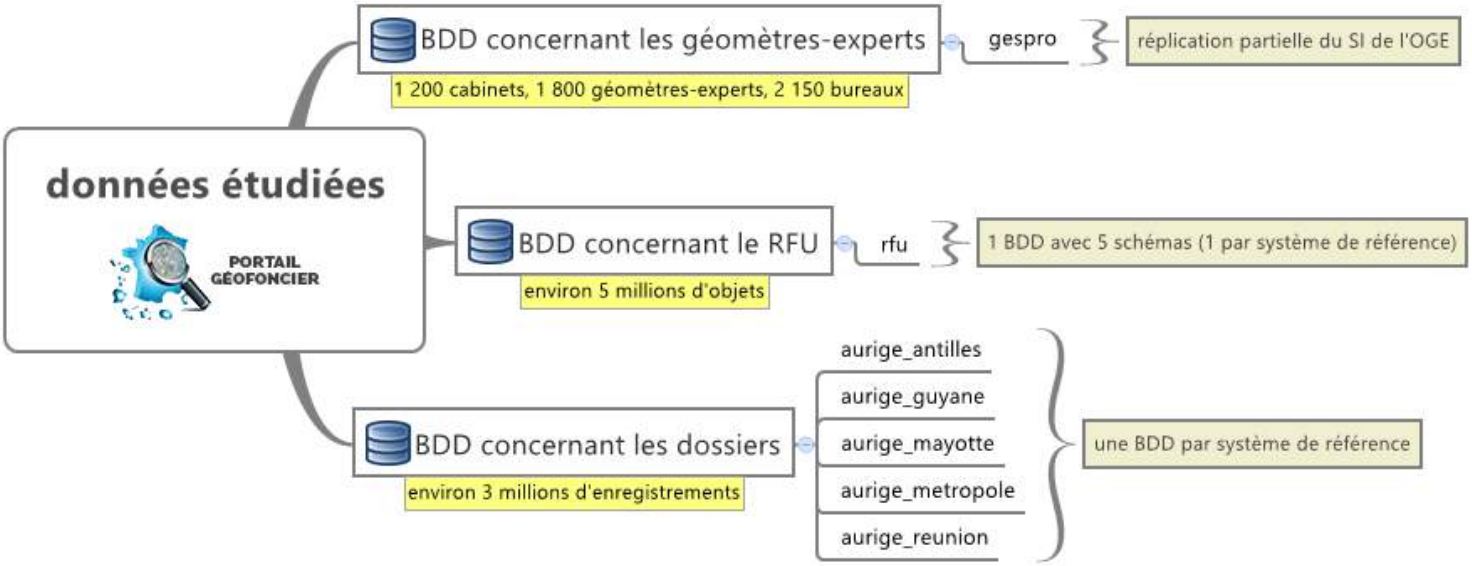


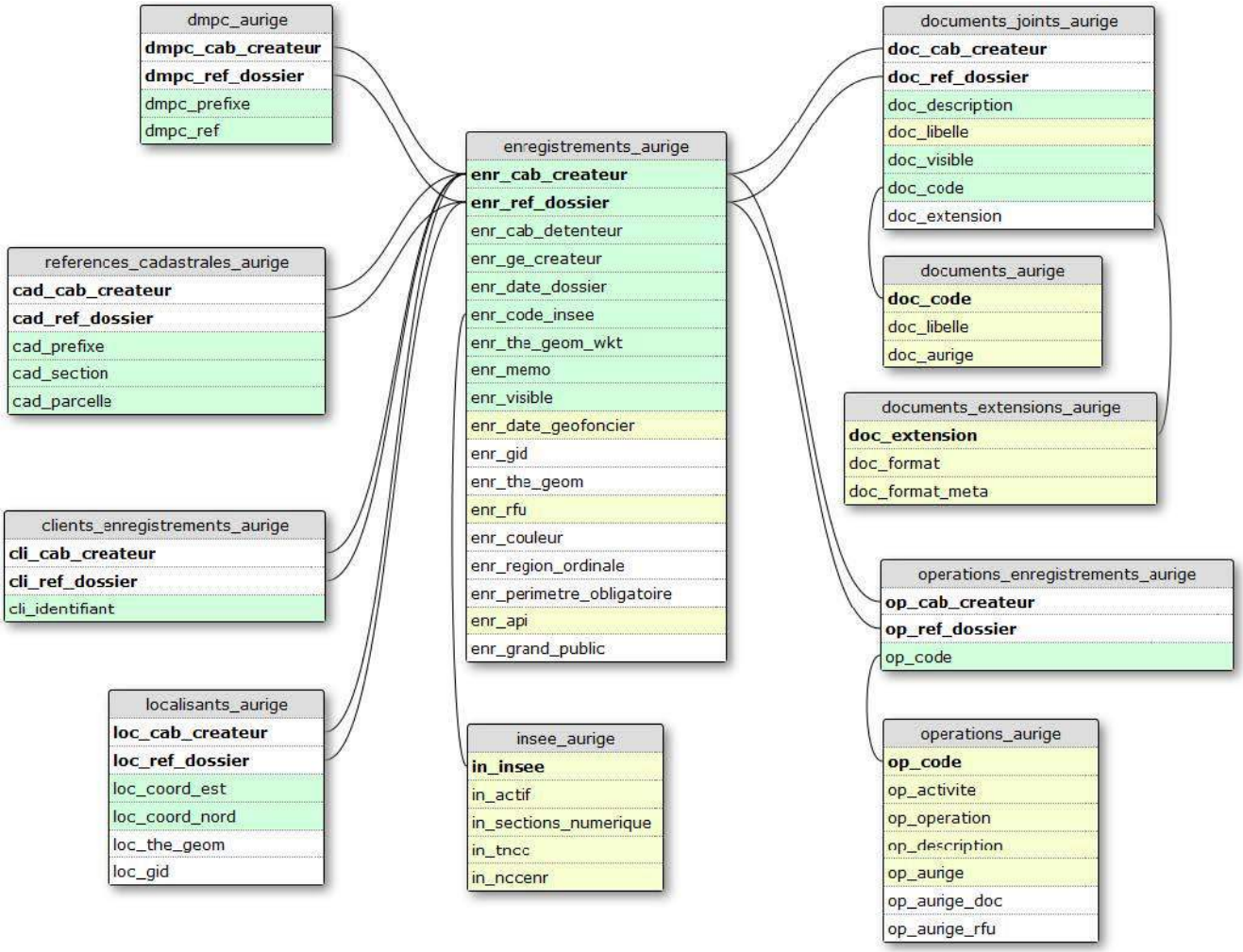
Figure 2 : schéma des bases de données employées

La mise à jour des dossiers et des objets RFU s'effectue en temps réel. Le volume de données varie considérablement selon la base de données considérée. Cela sera par la suite un facteur important au niveau des techniques de traitement utilisées et de la sélection des données traitées.

Les bases de données concernant les dossiers ont été structurées afin de pouvoir importer les données issues de la base Aurige, ancêtre de Géofoncier. Le modèle relationnel de ces BDD est présenté dans la figure 3 ci-dessous. Comportant une composante géographique, les dossiers sont répartis dans différentes bases de données correspondant chacune à un système de référence différent employé par la profession en France. En effet, le RGF93 est utilisé en métropole mais les géomètres travaillant dans des départements d'outre-mer s'appuient sur d'autres systèmes de référence adaptés à la localisation de leur territoire. C'est pourquoi chaque département d'outre-mer est visualisable séparément sur Géofoncier. Au niveau des géométries des objets, celles-ci sont systématiquement rattachées en coordonnées géographiques afin de se conformer à la directive INSPIRE.

Chaque BDD rassemblant les dossiers produits s'axe autour d'une table centrale *enregistrements_aurige* (voir figure 3) où chaque dossier est enregistré à l'aide d'une clé primaire composite comportant le numéro du cabinet créateur et la référence du dossier. Cette clé est utilisée dans d'autres tables liées à *enregistrements_aurige*. Ces tables connexes stockent des informations concernant les documents joints aux dossiers, les DMPC, les opérations réalisées, les

codes clients et les références cadastrales. La table *localisants_aurige* correspond à la localisation des dossiers et peut inclure une ou plusieurs positions pour un même dossier.



Source : Géofoncier SAS, 2015

Figure 3 : modèle relationnel type d'une des bases de données rassemblant les dossiers des géomètres-experts

en vert : champs accessibles en lecture et écriture
en jaune : champs accessibles en lecture
Les clés primaires sont en gras

Les serveurs comportent aussi une réplication partielle de la BDD de l'OGE, mise à jour quotidiennement, afin de fournir plusieurs informations concernant la profession : nom et prénom des géomètres-experts, nom des cabinets, localisation et type des bureaux notamment. Le modèle ci-dessous (figure 4) présente les tables essentielles de la BDD *gespro* qui seront par la suite exploitées dans l'analyse. La relation entre *t_gespro_ge* et *t_gespro_bureau* sert à identifier le géomètre-expert responsable du bureau s'il y en a un. La table *t_gespro_asso_cabinet_ge* permet d'établir dans quel cabinet travaille ou a travaillé un géomètre-expert. Enfin, la relation entre *t_gespro_cabinet* et *t_gespro_bureau* sert à rattacher chaque bureau à un cabinet.

En plus du volume parfois important de données à traiter, la seconde difficulté est la gestion de leur qualité et exhaustivité lors des analyses. En effet, même si les informations étudiées sont issues de versements obligatoires de la part de la profession, il est notable que des géomètres-experts n'utilisent pas encore suffisamment Géofoncier. De plus, la BDD *gespro* présente de nombreuses irrégularités, notamment pour des données anciennes comme nous le verrons par la suite.

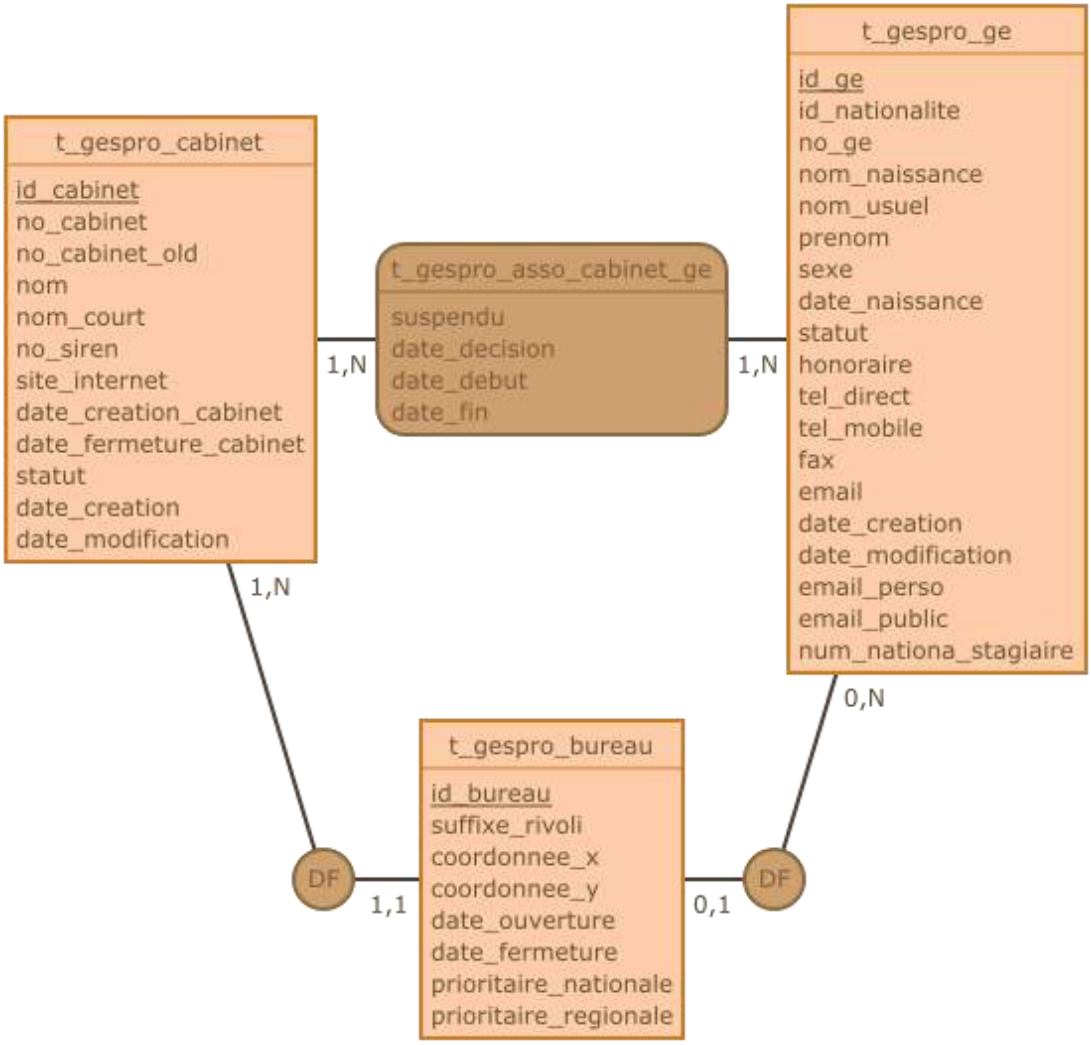


Figure 4 : modèle conceptuel simplifié de la BDD *gespro*

DF signifie ici dépendance fonctionnel

I.2 Démarche générale et outils employés

L'ensemble des étapes décrites ci-dessous sont résumées sous forme de graphe[5]. L'objectif de cette partie est de développer l'ensemble des outils techniques employés lors des analyses et d'expliquer le déroulement de l'extraction des données du serveur de Géofoncier jusqu'à la carte ou le graphique final.

I.2.1 De l'IDG Géofoncier au serveur local dédié aux analyses statistiques

L'IDG Géofoncier est basée sur des instances PostgreSQL 9.3.1 hébergées par le BRGM et accessibles en ligne via un accès sécurisé. La présence de géométries est gérée avec l'extension PostGIS 2.1.1. L'interface d'administration des serveurs employée est phpPgAdmin 5.1. Nous ne disposions que d'un accès en lecture seule aux serveurs de production et un accès en écriture aux serveurs de recette.

Les analyses s'appuient sur les données disponibles sur le serveur de production, afin de disposer des données actualisées. Comme aucune création de table ou de vue n'est possible, la seule méthode de travail à ce niveau est donc d'employer des requêtes de sélection afin d'afficher

[5] Voir annexe 1 : Démarche générale de travail : de la donnée brute à l'analyse

des sélections générées qu'il est ensuite possible de télécharger au format csv. Afin de conserver la géométrie au sein du format csv, celle-ci est exportée convertie en WKT. Elle est ainsi écrite de façon textuelle et non binaire.

L'étape suivante est l'import des fichiers csv sous un serveur en local. Le SGBD employé tout le long de cette analyse est PostgreSQL 9.4 combiné à l'extension PostGIS 2.1.5.

Les fichiers csv sont convertis en table au sein d'une base de données PostgreSQL (administrée avec PgAdmin III) à l'aide de GDAL, un outil d'ETL, permettant d'extraire, de transformer et de charger des données entre de nombreux formats.

Si le fichier dispose d'une géométrie écrite sous forme textuelle, il est nécessaire de créer un fichier virtuel (format vrt) en plus du fichier csv. Celui-ci va permettre de gérer la géométrie incluse sous forme de WKT dans le csv.

```
<OGRVRTDataSource>
        <OGRVRTLayer name="bureau">
                <SrcDataSource>bureau.csv</SrcDataSource>
                <GeometryField encoding="PointFromColumns"
reportSrcColumn="false" x="coordonnee_x" y="coordonnee_y">
                        <SRS>EPSG:4326</SRS>
                        <GeometryType>wkbPoint</GeometryType>
                        <ExtentXMin>-180</ExtentXMin>
                        <ExtentYMin>-90</ExtentYMin>
                        <ExtentXMax>180</ExtentXMax>
                        <ExtentYMax>90</ExtentYMax>
                </GeometryField>
        </OGRVRTLayer>
</OGRVRTDataSource>
```

L'exemple ci-dessus de fichier vrt permet de générer au format binaire WKB la géométrie incluse dans le fichier bureau.csv au travers de deux colonnes contenant la longitude et la latitude de chaque bureau. Le système de référence employé ici est le WGS84 (code EPSG 4326). On peut préciser également les limites de la zone couverte (ici en degrés comme il s'agit de coordonnées géographiques). L'attribut `reportSrcColumn= "false"` permet d'éviter l'ajout des colonnes contenant la géométrie sous forme textuelle.

On importe ensuite le fichier dans la base de données en utilisant la ligne de commande suivante dans la console GDAL :

```
ogr2ogr -f "PostgreSQL" -overwrite -a_srs "+init=EPSG:4326" -nlt
"POINT" PG:"dbname='etude_profession' host='localhost' port='5433'
user='postgres' password='postgres'" bureau.vrt -nln bureau
```

On précise notamment le format de sortie du fichier (sous forme de table dans une BDD PostgreSQL ici) et le fichier d'entrée. L'option `-overwrite` force l'écrasement de la table si celle-ci existe déjà. La commande `-nln` permet de nommer différemment le fichier ou la table de sortie. On remarque que certaines options sont redondantes avec les informations présentes dans le fichier vrt. Il est en effet possible de renseigner des informations comme le système de référence soit dans la commande GDAL, qui prévalent, soit dans le fichier vrt.

En ce qui concerne les fichiers issus de sources externes à Géofoncier, ceux-ci sont importés suivant le même principe comme il s'agit soit de fichiers shape soit de fichiers csv gérés par GDAL. Il est également possible d'importer les fichiers shape directement via un module d'extension sous l'interface de PgAdmin III. Les deux méthodes sont équivalentes. L'utilisation de GDAL est certes plus complexe au premier abord mais s'avère plus pertinente lorsqu'il s'agit d'importer de multiples fichiers dans une BDD du fait de l'automatisation de la procédure.

I.2.2 Traitement sous PostgreSQL et R

Une fois les données importées, celles-ci sont corrigées et traitées pour répondre aux besoins de l'analyse. De façon générale, l'utilisation du langage SQL va permettre de couvrir la majorité des besoins en termes de traitement sur les tables. Néanmoins, celui-ci reste un langage de manipulation de bases de données et atteint rapidement ses limites pour la réalisation de traitements statistiques.

C'est pourquoi le couplage du SQL avec un langage de programmation adapté à la réalisation d'études statistiques est nécessaire. R a été choisi pour répondre à ce rôle. Il s'agit d'un logiciel libre sous licence GPL, distribué par le CRAN et d'un langage de programmation dérivé du langage S. Celui-ci permet de réaliser des traitements statistiques et de visualiser des données. Il dispose aussi d'une communauté grandissante ayant étendu les fonctionnalités de R au travers de nombreux packages. Les versions 3.1 et 3.2 du logiciel ont été employées afin d'utiliser les derniers packages.

Il existe une extension pour PostgreSQL nommée PL/R permettant de concevoir des fonctions PostgreSQL en recourant au langage R. Celle-ci a été testée dans un premier temps, l'intérêt *a priori* étant de pouvoir travailler sur un seul et même logiciel en évitant d'ajouter R à la chaîne des traitements. Cependant, cela limite les possibilités qu'offre R notamment en termes de production de graphes (sorties graphiques). Pour cette raison, les analyses ont été réalisées directement avec le logiciel R et non l'extension PL/R. La connexion d'une session de R à une BDD PostgreSQL s'effectue à l'aide du package RPostgreSQL[6]. Cela permet de transférer des données sous environnement R voir d'utiliser R comme outil d'administration de la BDD.

Nous n'allons pas ici détailler les bases du langage pour lesquelles plusieurs ouvrages existent en français, y compris sous forme de livre « libre » (groupe ElementR, 2014). Une liste des packages essentiels à l'ensemble des recherches est disponible dans la bibliographie.

I.2.3 Production de graphes et rendus cartographiques

Au niveau des rendus graphiques, ceux-ci sont réalisés sous R à l'aide du package ggplot2 qui améliore les possibilités de création de graphiques. Un package en développement au moment de l'étude, rCharts, a également été testé afin de générer des graphes dynamiques. Celui-ci s'appuie sur des librairies JavaScript destinées à ce type de rendu. Néanmoins, les possibilités de personnalisation des graphiques obtenus sont très limitées par rapport à ggplot2. De plus, ce package n'est pas mis à jour pour la dernière version de R.

En ce qui concerne la production cartographique, le principal outil employé pour la mise en page est ArcGIS (version 10) complété par Inkscape (version 0.91) pour la finition du rendu. Celui-ci a été préféré à QGIS (version 2.8.1) dans la plupart des cas notamment en raison de la meilleure qualité de l'export au format svg, nécessaire pour travailler sous Inkscape, l'export de QGIS ayant parfois généré des erreurs de positionnement des éléments textuels.

ArcGIS dispose en outre de l'outil Cartogram, téléchargeable sur le site d'ESRI, qui permet de réaliser des cartogrammes. L'outil employé s'appuie sur le principe physique de diffusion des gaz (Gastner & Newman, 2004). Prenons l'exemple de la représentation de la population par département. La variable choisie (la population) est ramenée à une densité en utilisant la superficie du polygone représentant chaque département. L'algorithme utilisé va chercher à lisser les écarts de densité entre polygones : les départements avec une population faible vont voir leur surface diminuer au profit de ceux étant plus peuplés. On obtient ainsi une représentation en anamorphose exagérant les espaces les plus peuplés.

[6] Voir Annexe 2 : Connexion à une base de données PostgreSQL à partir de R à l'aide du package RPostgreSQL

II Analyses statistiques menées : structure, activité et évolution de la profession

Nous venons de présenter les outils et méthodes nécessaires pour l'étude des trois domaines considérés : les dossiers versés, les objets RFU et les données professionnelles concernant les géomètres-experts. Les premières analyses réalisées s'appuient sur des méthodes statistiques et visent à extraire des tendances concernant l'évolution de la profession.

II.1 Approche descriptive de la profession et de son évolution

II.1.1 Évolution quantitative de la profession

II.1.1.1 Méthodologie et fiabilité des données

Cette étude s'appuie exclusivement sur les données provenant de la BDD de l'OGE concernant les géomètres-experts. Les variables d'appui pour l'étude sont au nombre de deux :

- Le nombre de géomètres-experts actifs par an,
- Le nombre de cabinets et de bureaux en activité par an.

Il faut différencier la notion de cabinet qui renvoie à l'entreprise détenue par un ou plusieurs géomètres-experts et celle de bureau qui correspond au lieu de production pour la profession. Un cabinet peut rassembler un ou plusieurs bureaux.

Les collaborateurs libéraux sont également liés à un bureau au niveau de la BDD mais en pratique ces bureaux n'ont pas d'existence physique, c'est pourquoi ceux-ci sont exclus des calculs.

À partir des données disponibles, il est possible d'étendre l'étude de 1948 à 2015 soit la quasi-totalité de la période d'existence actuelle de l'OGE (figure 5). Jusqu'à la fin des années 70 et l'apparition des premières sociétés civiles professionnelles, les géomètres-experts exerçaient en individuel uniquement, c'est pourquoi les variables se confondent. Cependant, on remarque sur chacune des trois courbes obtenues une cassure au niveau de 1999. Cela correspond probablement à une mise à jour des tables et des méthodes de saisies afin de mieux prendre en compte les géomètres-experts retraités. En effet, les dates de fin d'activité au sein d'un cabinet n'étaient pas systématiquement renseignées sur la BDD auparavant.

Les données antérieures à 1999 n'ont pas été systématiquement mises à jour en ce qui concerne les fermetures de bureaux, cabinets et les géomètres quittant l'ordre. Il y a donc des valeurs anormalement élevées avant 1999. Il convient donc de se concentrer sur la période postérieure à cette année pour réaliser une étude en s'appuyant sur les données objectivement les plus fiables de la BDD.

Afin de compléter l'analyse sur l'évolution des effectifs de la profession, on s'appuie aussi sur la pyramide des âges, réalisée à différentes époques à partir des dates de naissance des géomètres-experts. Ces pyramides sont mises en forme sous R à l'aide du package *pyramid* spécifique aux graphes de type pyramide.

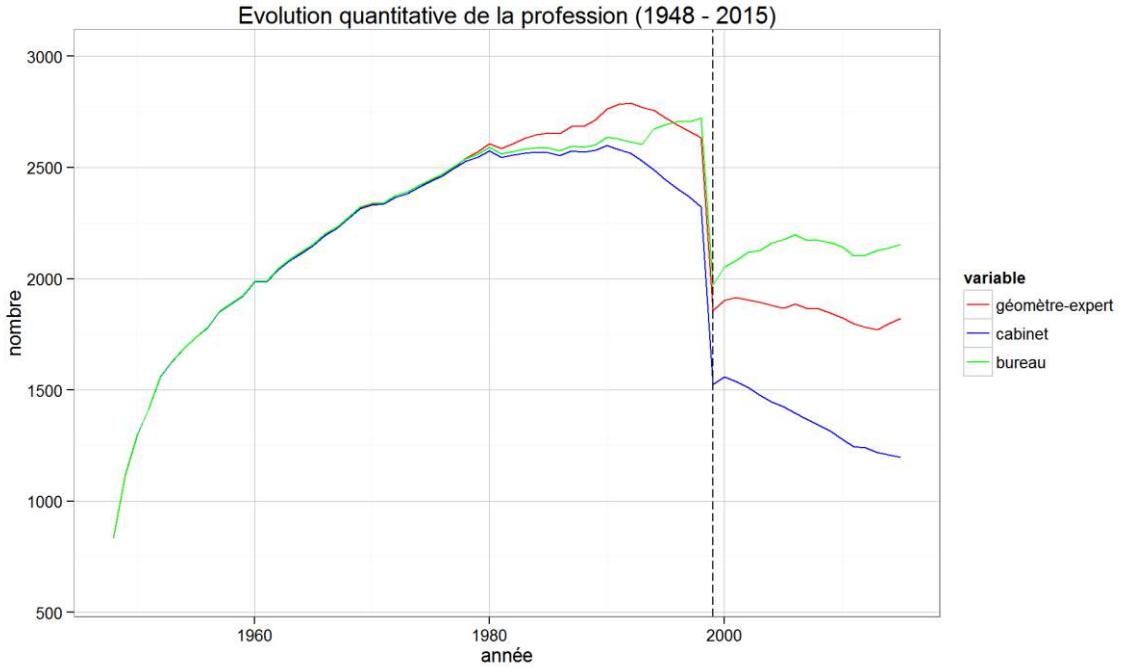


Figure 5 : évolution quantitative de la profession de 1948 à 2015 (cabinets, bureaux et géomètres-experts)

II.1.1.2 Analyse des évolutions depuis 1999 : baisse du nombre de cabinets et rajeunissement de la profession

À partir des résultats obtenus sur la période 1999 – 2015 (figure 6), on constate une baisse régulière et forte du nombre de cabinets, de plus de 20 % des effectifs. Le nombre de bureaux connait une légère augmentation.

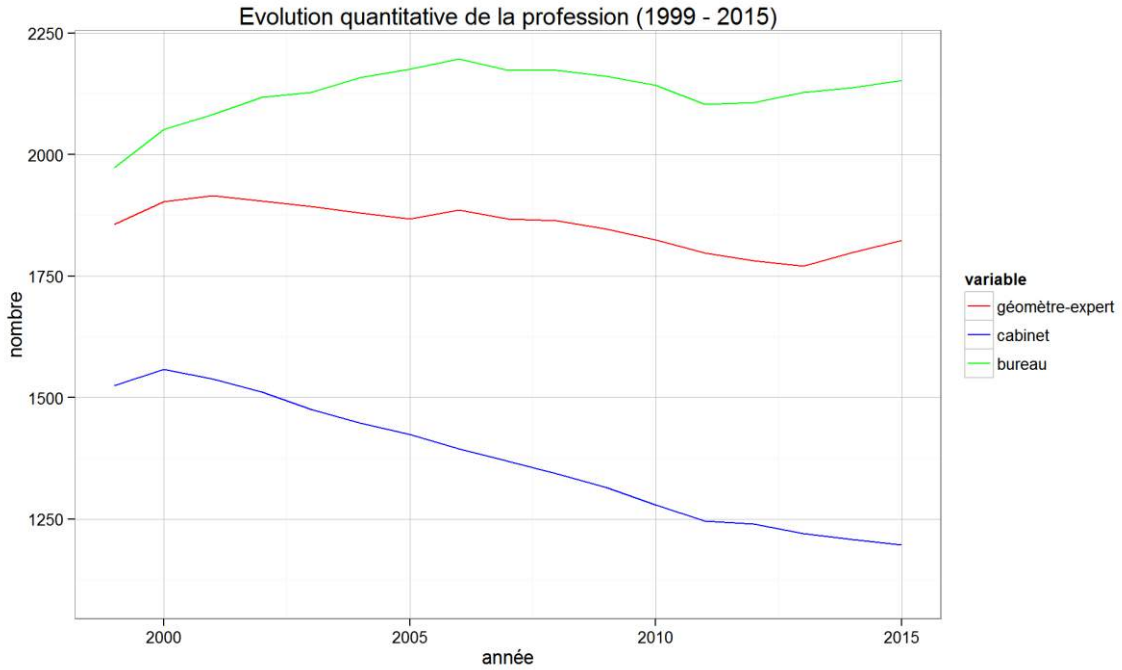


Figure 6 : évolution quantitative de la profession de 1999 à 2015 (cabinets, bureaux et géomètres-experts)

En parallèle, le nombre de géomètres-experts se maintient avec une légère baisse en lien avec le départ en retraite de la génération du baby-boom. Cela se traduit sur les pyramides des

âges (figure 7). Au niveau des pyramides de 2000 et 2005, on constate que la classe d'âge majoritaire est celle des 60-70 ans. À partir de 2005, un rajeunissement important de la profession s'opère pour aboutir en 2015 à une répartition plus homogène des classes d'âges entre 30 et 65 ans. La part des femmes dans la profession reste toujours fortement minoritaire bien qu'en progression depuis 2000. Au 1er janvier 2015, on dénombre 1 823 géomètres-experts (1 625 hommes et 198 femmes) pour 2 153 bureaux et 1 197 cabinets sur la totalité du territoire d'exercice.

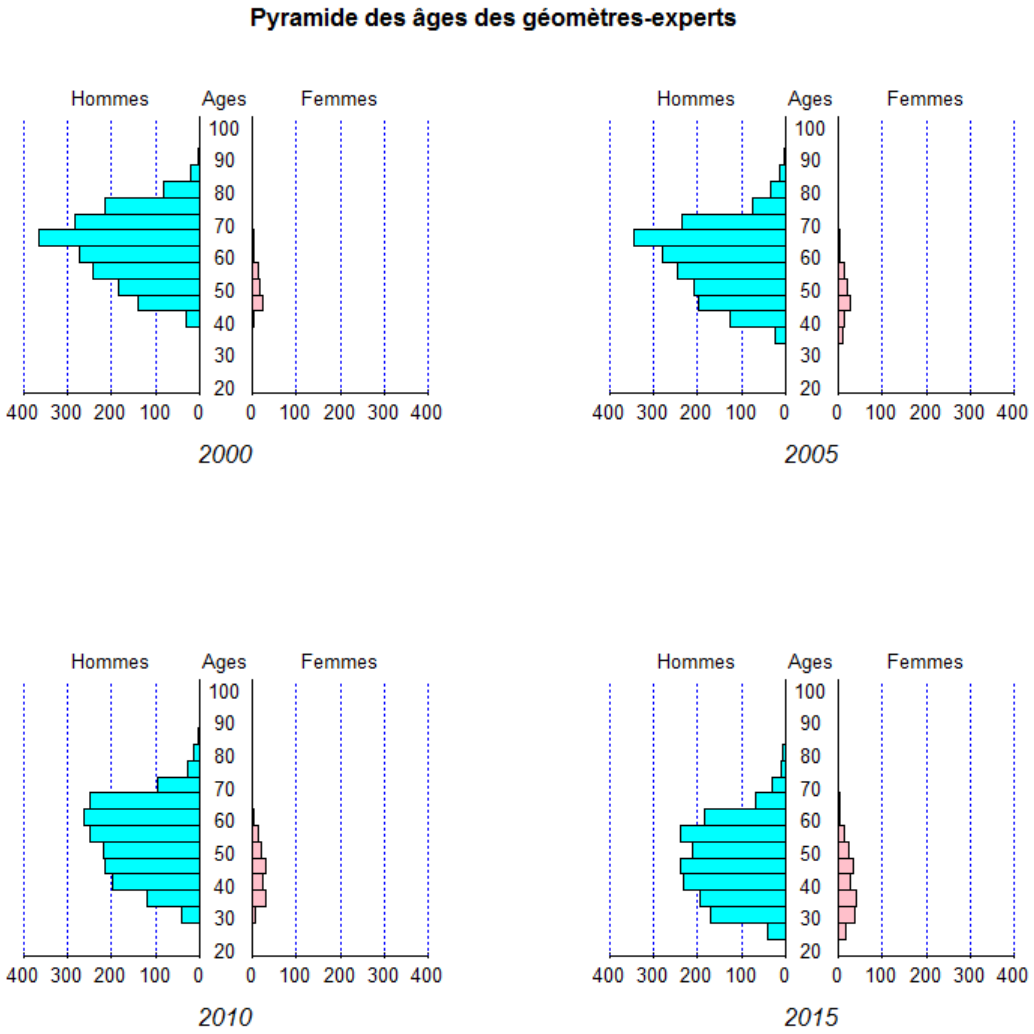


Figure 7 : pyramides des âges de la profession

Si l'on interprète conjointement ces trois variables, on conclut à une baisse du nombre de cabinets au profit d'une augmentation de leur taille pour un effectif sensiblement identique de la profession. Cette première analyse semble donc indiquer une évolution structurelle des cabinets entamée depuis au moins 2000 qu'il convient d'approfondir.

II.1.2 Évolution structurelle des cabinets

II.1.2.1 Méthodologie

L'analyse précédente a mis en lumière des évolutions concernant la structure des cabinets. Afin d'étudier cette évolution, on s'appuie sur trois variables prises à différentes périodes pour chaque cabinet :

- le nombre de géomètres-experts présent au sein de la structure,
- le nombre de bureaux détenus, incluant les bureaux principaux, les bureaux secondaires, les permanences et les bureaux de chantier,
- le type de société du cabinet.

Le choix des variables a été déterminé en fonction des informations disponibles au sein de l'IDG Géofoncier. Une autre variable pertinente aurait été par exemple l'effectif de chaque cabinet mais cette information n'est pas accessible sous forme de base de données aujourd'hui.

Comme nous l'avons constaté précédemment, les données antérieures à 1999 ne sont pas suffisamment fiables, c'est pourquoi la période d'étude s'étend de 2000 à 2015 suivant un palier de cinq ans.

II.1.2.2 Analyse des résultats : de fortes mutations structurelles depuis 2000

Chacune des variables permet d'obtenir la répartition en différentes catégories des individus de l'étude. Comme il s'agit d'une analyse univariée, il n'y a ici aucune étude des liaisons entre celles-ci. On peut visualiser l'évolution du nombre de cabinets au cours du temps conditionnée par les modalités de chacune des trois variables considérées. On obtient donc pour chaque variable, plusieurs panels de diagrammes.

Cette méthode de représentation est visuellement plus parlante : elle permet à la fois d'illustrer l'évolution au sein de chaque catégorie et de comparer les catégories entre elles.

Ces graphiques permettent de voir les évolutions d'ordre structurel des cabinets. On constate une forte réduction du nombre de cabinets de taille faible (caractérisés par un seul géomètre-expert exerçant en individuel et comprenant un bureau unique) au profit d'un accroissement des cabinets plus importants composés de plusieurs géomètres-experts et d'au moins deux bureaux.

Par ailleurs, la baisse régulière du nombre de cabinets sur cette même période constatée dans l'analyse précédente[7] concerne donc principalement les cabinets de taille faible. Cela traduit une tendance générale de la profession à se regrouper, à fusionner des cabinets de taille moindre amorcée depuis au moins quinze ans.

Cette tendance peut s'expliquer notamment par le coût élevé de l'équipement du géomètre-expert qui inclut aujourd'hui une station totale mais aussi un récepteur GNSS afin de géoréférencer efficacement les opérations foncières réalisées. De plus, de nombreux géomètres-experts correspondant à la génération du baby-boom et exerçant en individuel partent en retraite et ne trouvent pas de repreneur. Leur entreprise est souvent rachetée par des cabinets de plus grande ampleur.

En ce qui concerne les types de société, on remarque également une croissance du nombre de SARL et SELARL au dépend du nombre de SCP, première forme historique d'association des géomètres-experts autorisée pour la profession mais contraignante car à part égale pour chaque membre alors que les SELARL et SARL ont une approche de la répartition des bénéfices en terme de capital.

[7] Voir partie II.1.1.2

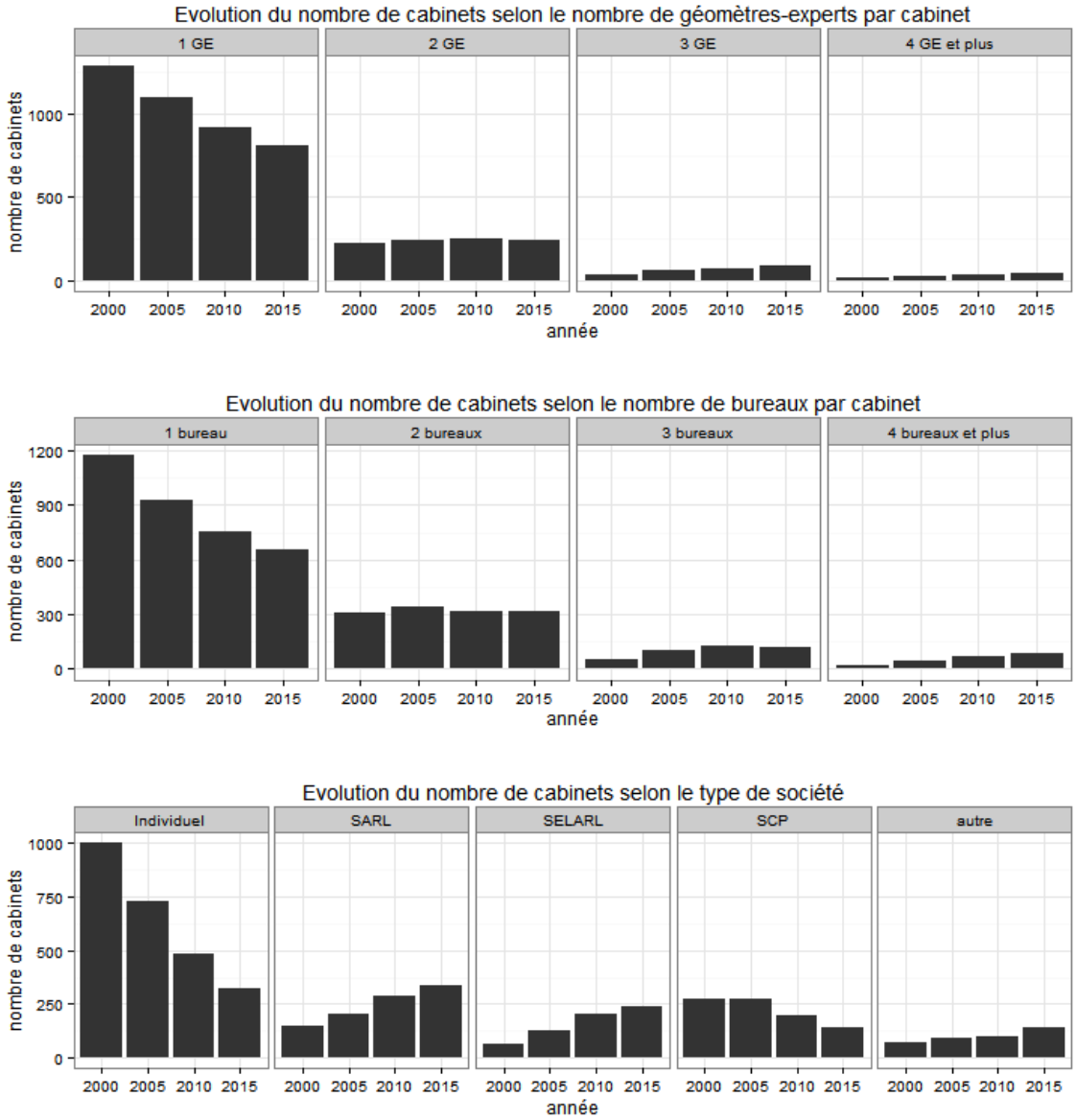


Figure 8 : diagrammes en barres sur l'évolution des cabinets de 2000 à 2015

II.1.3 Situation financière des cabinets de géomètres-experts

Nous venons de constater une évolution marquée de la structuration des cabinets de géomètres-experts. Celle-ci pouvant avoir des causes économiques, nous allons voir s'il existe un lien entre la taille du cabinet, ici traduite au travers du nombre de géomètres-experts, et son bilan financier.

II.1.3.1 Données et méthodologie

Malgré l'absence de données temporelles sur les finances des cabinets de géomètres-experts, nous disposons d'informations sur les résultats nets, l'excédent brut d'exploitation (EBE), le chiffre d'affaires, le capital et l'ordre de grandeur des effectifs à l'aide d'un fichier d'entreprise obtenu sur le site *societe.com*[8].

Le critère de sélection des entreprises s'appuie sur le secteur d'activité défini par le code APE. Le secteur d'activité des géomètres choisi ici (code APE 7112A) comprend les cabinets de

[8] Voir la liste des jeux de données employées en fin de bibliographie

géomètres-experts. Le fichier inclut également les sociétés composées de géomètres non-inscrits à l'OGE.

Parmi les 875 entreprises présentes dans les données, il faut conserver uniquement celles correspondant à des cabinets de géomètres-experts. Il faut donc réaliser une jointure avec les données disponibles sur Géofoncier afin de ne garder que les sociétés qui nous intéressent. La méthode de sélection s'appuie sur le numéro SIREN, le nom du dirigeant et le nom de l'entreprise[9] et permet d'identifier 354 cabinets sur le jeu de données financières. On ne travaille donc, dans le cas présent, que sur un échantillon de la profession (environ 30% des cabinets).

La variable sur laquelle les cabinets seront différenciées correspond au nombre de géomètres-experts présents au sein de celui-ci. Dans un premier temps, un calcul du coefficient de corrélation linéaire (aussi nommé coefficient de Bravais-Pearson) est effectué afin d'observer l'intensité de la liaison linéaire entre les variables.

Si l'on note $cov(x, y)$ la covariance des variables x et y, σ_x (respectivement σ_y) l'écart-type de la variable x (respectivement y) alors le coefficient de corrélation de x et y noté r_{xy} peut s'écrire de la façon suivante :

$$r_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

Si $r_{xy} = 0$ alors les variables sont indépendantes, si $|r_{xy}| = 1$ alors la liaison entre les variables est linéaire (Rakatomalala, 2015).

Dans un second temps, nous rassemblons comme précédemment les cabinets de quatre géomètres-experts et plus en une seule modalité, ceux-ci ne représentant qu'une minorité des cabinets de l'échantillon comme le montre le graphe ci-dessous (figure 9).

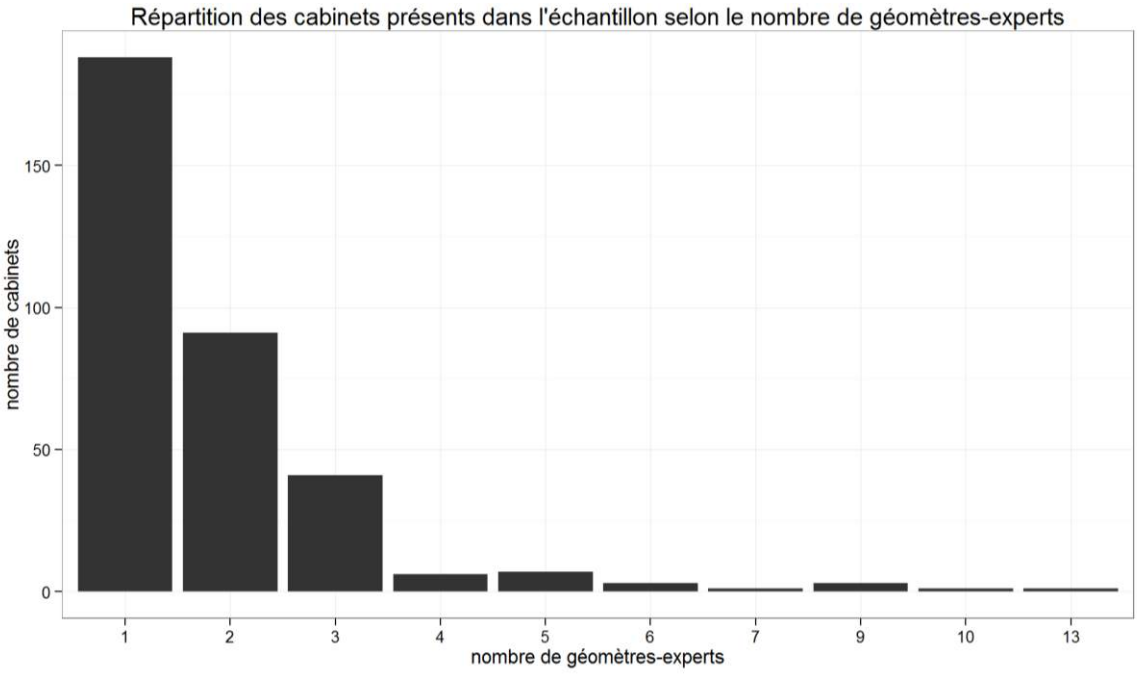


Figure 9 : diagramme en barres des cabinets selon le nombre de géomètres-experts

La méthode employée est très proche de l'analyse précédente[10] en utilisant les mêmes packages pour R. Afin de comparer les chiffres obtenus, ceux-ci sont divisés par le nombre de

[9] Voir annexe 3 : Méthode de jointure des données financières et des données issues de Géofoncier
[10] Voir partie II.1.2.2

géomètres-experts du cabinet. On effectue ensuite le calcul de la médiane pour chaque modalité afin de réduire l'influence des valeurs extrêmes.

II.1.3.2 Analyse : une faible disparité des bénéfices par géomètre-expert

On obtient une matrice de corrélation présentant les corrélations pour chaque paire de variable possible (tableau 1).

	capital	chiffre d'affaires	résultat	EBE	nombre de GE
capital	1	0,37	0,27	0,24	0,51
chiffre d'affaires	0,37	1	0,76	0,85	0,66
résultat	0,27	0,76	1	0,96	0,37
EBE	0,24	0,85	0,96	1	0,41
nombre de GE	0,51	0,66	0,37	0,41	1

Tableau 1 : matrice de corrélation des variables de l'étude

Comme la matrice est symétrique et que sa diagonale correspond à la corrélation d'une variable avec elle-même, on ne s'intéresse qu'au triangle inférieur strict de la matrice. Entre variables financières, on constate la liaison linéaire forte entre l'EBE, le résultat et le chiffre d'affaires d'un cabinet. Chacune de ces variables étant un indicateur de la santé financière d'une entreprise, cela est cohérent.

Le nombre de géomètres-experts apparait comme étant fortement corrélé avec le chiffre d'affaires et, dans une moindre mesure, avec le capital. En revanche, on tend davantage vers une relation d'indépendance entre le nombre de géomètres-experts et le résultat ou l'excédent brut d'exploitation. En d'autres termes, le lien entre les bénéfices d'un cabinet et son effectif de géomètres-experts est faible.

On s'intéresse ensuite aux valeurs prises par ces différentes variables en fonction du nombre de géomètre-experts (figure 10). L'excédent brut d'exploitation ainsi que le résultat net donnent des informations à propos de la rentabilité de l'entreprise. Sur l'échantillon étudié, l'EBE ne varie pas considérablement selon le nombre de géomètres-experts.

Cependant, les géomètres-experts seuls se démarquent par un EBE médian plus faible par rapport à la médiane de l'ensemble de l'échantillon (en rouge). La tendance est identique sur le résultat médian, on remarque aussi que ce sont les cabinets comportant seulement deux géomètres-experts qui obtiennent les meilleurs résultats nets.

Le chiffre d'affaires représente le total des ventes de l'entreprise, celui-ci s'accroit pour les cabinets les plus importants tout comme le capital, même si ces deux valeurs sont divisées par le nombre de géomètres-experts de l'entreprise.

Nous avons pu voir précédemment une décroissance du nombre de cabinets avec un seul géomètre-expert. On constate qu'actuellement, ce sont ceux qui dégagent les bénéfices les plus faibles par géomètre-expert. Les cabinets en comportant deux sont un cas à part : ils dégagent le plus de bénéfices par géomètre-expert tout en ayant un chiffre d'affaires plus faible. Ils dépensent donc moins que les autres. Enfin, les cabinets en comportant trois ou plus ont un chiffre d'affaires et des capitaux plus conséquents.

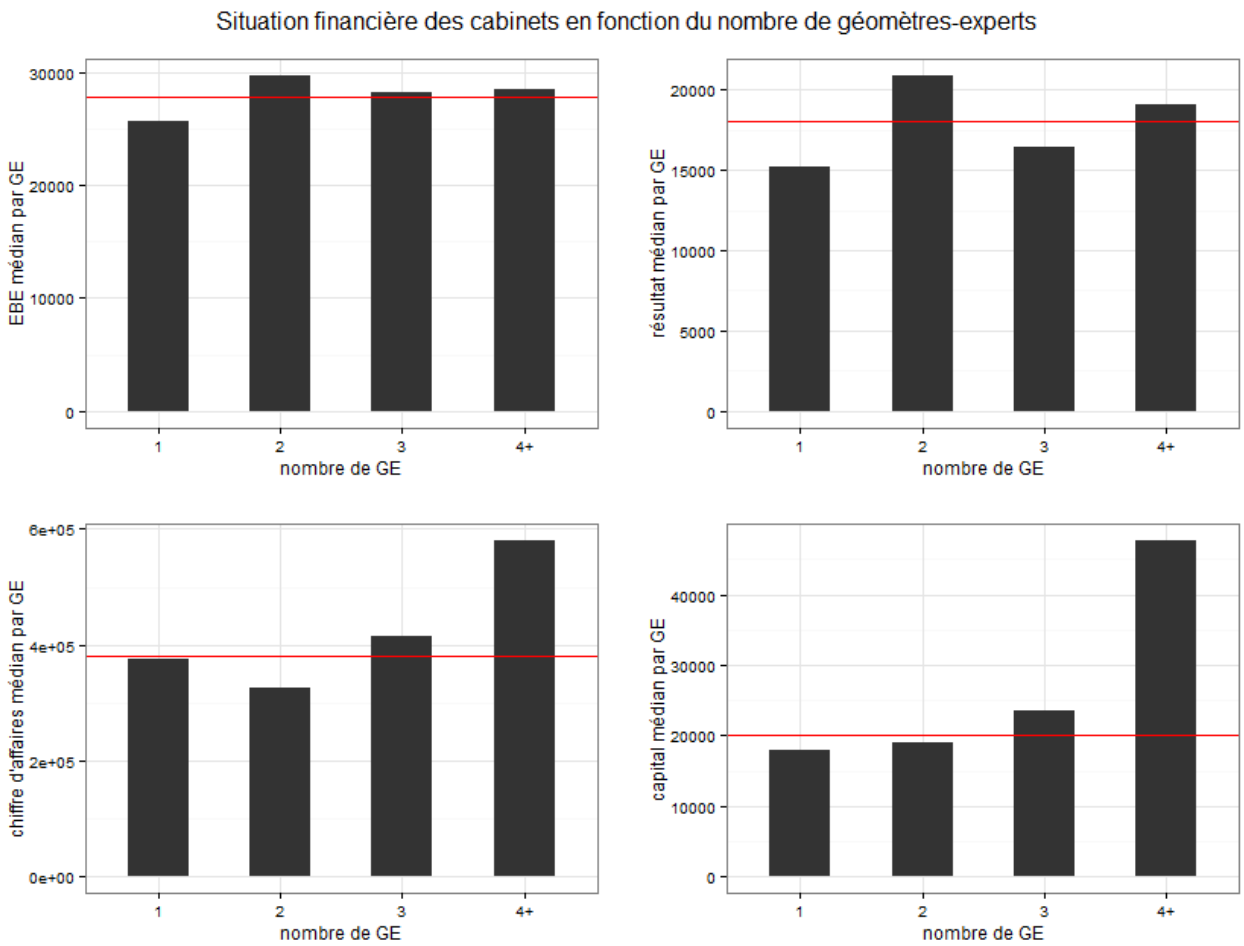


Figure 10 : diagrammes en barres sur la situation financière des cabinets

II.2 Étude de l'activité des géomètres-experts à partir des types de dossiers versés sur Géofoncier

En ayant la connaissance de l'activité exercée par chaque géomètre-expert sur ses dossiers (topographie, foncier, copropriété, aménagement…) il serait possible théoriquement d'établir plusieurs profils-types ou catégories de cabinets. C'est dans cet objectif que l'étude suivante a été menée sur les versements de dossiers effectués.

II.2.1 Critères de l'étude

Chaque dossier versé par Géofoncier est lié à au moins un type d'activité. Les activités proposées sont les suivantes, chaque activité regroupant plusieurs opérations :

- Études d'aménagement
- Urbanisme
- Aménagement foncier
- Ingénierie/maîtrise d'œuvre
- Foncier
- Copropriété/division en volumes
- Expertise en évaluation
- Topographie

- Immobilier
- Géomatique
- Prestations diverses

L'idée première était de s'appuyer sur l'ensemble des types de dossiers afin de calculer le nombre de dossiers versés par cabinet en fonction du type d'activité depuis l'ouverture de Géofoncier à la profession. L'objectif était alors d'en extraire une moyenne journalière de versement des cabinets afin de classifier les cabinets selon leurs activités.

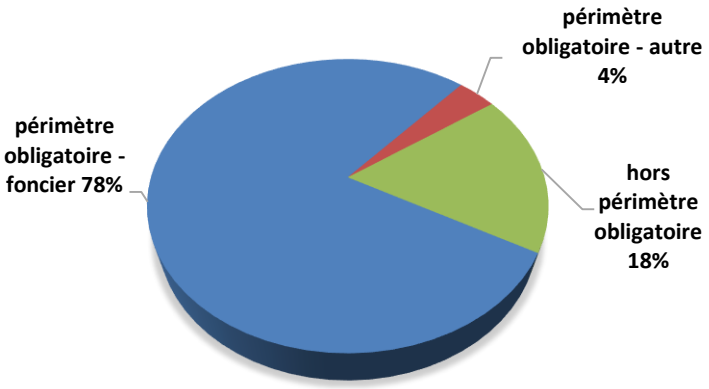


Figure 11 : répartition des dossiers versés sur Géofoncier depuis juillet 2010

Cela s'est avéré être une méthodologie ne pouvant être suivie si l'on tient compte de la fiabilité des données. En effet, seules certaines opérations dans le domaine du foncier, de la copropriété et de la division en volumes doivent être renseignées. Chaque géomètre a la liberté de renseigner ou non les dossiers ne concernant pas le périmètre obligatoire. Dans les faits, les dossiers fonciers soumis à versement obligatoire sont majoritaires à 78 %.

Afin d'obtenir des résultats suffisamment pertinents à partir des dossiers versés sur Géofoncier, l'analyse portera donc sur les dossiers appartenant au périmètre obligatoire et dont la date de création[11] a eu lieu entre 2012 et 2014. Ce choix des limites temporelles est justifié par le fait que les versements effectués sur ces trois années se sont régularisés. Depuis 2010, les géomètres ont pu se familiariser suffisamment avec le portail et l'utiliser de façon routinière pour ceux qui l'utilisent.

Pour expliciter ce choix, il faut tracer la courbe temporelle de l'évolution du nombre de dossiers de Géofoncier en fonction de leur date de création (voir figure 12). Pour rendre le graphique plus clair, on rassemble cette valeur par mois depuis juillet 2010, date d'ouverture du portail. On constate alors une périodicité qui s'établit dès 2012 au niveau de la courbe avec une baisse caractéristique au mois d'août, période de congés où les cabinets ont moins d'activité. Sur l'année 2014, on constate une baisse qui s'explique par le fait que tous les dossiers n'ont pas encore été complétés et versés par les géomètres : cela transparaît au niveau des courbes représentant respectivement les dossiers comportant un document joint et ceux comportant des objets RFU.

Par ailleurs, on peut constater que ceux faisant partie du périmètre obligatoire suivent la même tendance périodique que l'ensemble. Cela traduit bien l'importance des versements obligatoires (82 % de la totalité des dossiers) et portant sur des dossiers de type foncier.

On peut noter également que le versement des documents joints et des objets RFU, malgré l'obligation ordinale, sont encore loin d'être suivis par l'ensemble de la profession.

[11] Il s'agit théoriquement de la date de la commande reçue par le géomètre-expert et non la date à laquelle le dossier est versé sur Géofoncier.

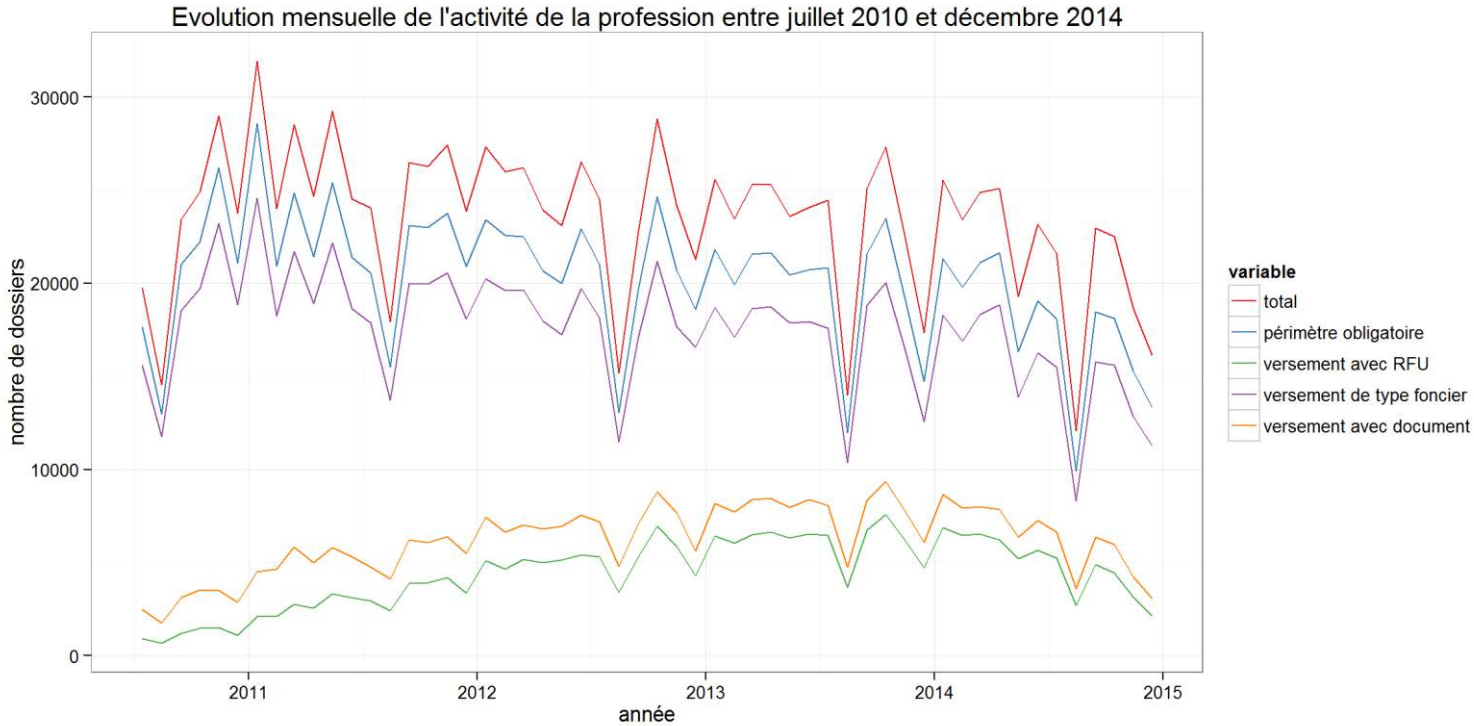


Figure 12 : évolution mensuelle de l'activité de la profession entre juillet 2010 et décembre 2014

II.2.2 Analyse exploratoire des répartitions de dossiers versés

II.2.2.1 Objectif et variables de l'analyse

À partir des critères énoncés précédemment, une extraction à partir de la base de données de Géofoncier est réalisée afin d'obtenir, pour chaque cabinet, le nombre de dossiers créés entre 2012 et 2014 inclus par type d'opération à versement obligatoire. On suppose donc que tous les dossiers créés sur cette période ont été versés. D'un point de vue statistique, chaque cabinet peut ainsi être considéré comme un individu pour lequel on dispose de plusieurs variables quantitatives. Les cabinets pris en compte dans l'analyse sont ceux en activités au 1er janvier 2015.

L'objectif est de voir s'il existe des corrélations entre les types d'opération versées par cabinet. La méthode employée est une analyse en composantes principales (ACP). Ce type d'analyse factorielle vise à « étudier les ressemblances entre individus du point de vue de l'ensemble des variables et dégage des profils d'individus » (Cornillon et al., 2012, p. 201). L'ACP est adaptée aux études portant sur des variables quantitatives.

Celle-ci est effectuée à l'aide du logiciel R qui dispose d'un package destiné à l'analyse factorielle : FactoMineR.

Parmi les opérations à versement obligatoire, certaines sont très peu présentes, notamment les AFAF et les ECIR. Ces deux types d'opération sont rassemblés sous une variable unique qui regroupe les aménagements fonciers afin de comptabiliser suffisamment de dossiers. On obtient alors la répartition présentée au tableau 2 sur les types d'opération.

Pour chaque cabinet, on ramène ensuite le nombre de dossiers versés à une moyenne journalière par géomètre-expert présent au sein du cabinet. Cela permet de considérer chaque cabinet indépendamment du nombre de géomètres-experts présents dans la structure.

opération	part de la totalité des dossiers
DMPC	57,19%
bornage	30,73%
état descriptif de division	4,55%
délimitation	2,60%
publicité foncière	2,32%
règlement de copropriété	1,15%
division en volumes	0,58%
voirie	0,44%
servitude	0,41%
aménagement foncier	0,04%

Tableau 2 : répartition des opérations obligatoires (période 2012-2014)

Au niveau de la création des cabinets, deux cas sont présents :

- si le cabinet a une date de création antérieure au 1er janvier 2012, on considère une moyenne journalière des versements sur toute la période,

- si le cabinet est créé au cours de la période d'étude, alors la moyenne est calculée à partir de la date de création de celui-ci.

II.2.2.2 Principe et méthode

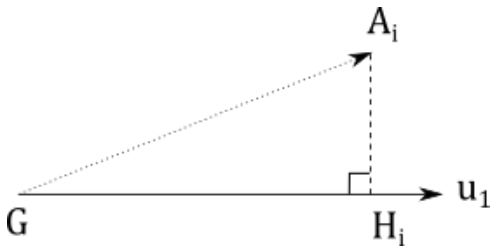


Figure 13 : projection sur l'axe factoriel du point A_i correspondant à l'individu i

Le principe de l'ACP est de considérer les *n* individus considérés comme étant un nuage de points dans un espace à *k* dimensions où *k* est le nombre de variables. L'ACP vise à trouver les axes les plus représentatifs du nuage de points, ceux pour lesquels l'inertie du nuage projeté sur ces axes est maximale.

Si l'on considère *G*, le centre de gravité du nuage de points et H_i, le projeté orthogonal de A_i sur l'axe u_1 alors on cherche à minimiser ${A_iH_i}^2$ ce qui revient à maximiser ${GH_i}^2$. On définit l'inertie I_1 du nuage projeté sur l'axe 1 de la façon suivante :

$$I_1 = \frac{1}{n}\sum_{i=1}^{n} {GH_i}^2$$

Le premier axe choisi lors du traitement est celui pour lequel l'inertie est maximale ce qui revient à maximiser $\sum_i {GH_i}^2$. Il est créé à partir d'une combinaison linéaire des variables résumant le mieux la totalité des variables. Le second axe est choisi de façon à maximiser cette même somme et à être orthogonal au premier axe. De même, on peut choisir un troisième axe orthogonal aux deux premiers en maximisant $\sum_i {GH_i}^2$, etc…

On peut faire l'analogie entre l'inertie et la variance. L'un comme l'autre mesurent la dispersion d'un jeu de données.

En termes d'analyse, l'ACP sert à synthétiser l'information en rassemblant les variables fortement corrélées. On distingue ainsi les variables qui permettent le plus de démarquer les individus. Cela fait apparaître les groupes d'individus présentant des caractéristiques communes.

II.2.2.3 Déroulement de l'ACP

Dans un premier temps, les variables sont centrées réduites afin de pouvoir les comparer. Ainsi, toutes les variables ont une espérance nulle et un écart-type égal à 1, le centre de gravité du nuage est confondu avec l'origine. Au vue de la répartition très inégale des dossiers, la réduction des variables, bien que facultative, s'avère ici nécessaire pour considérer avec la même pondération chaque type d'opération.

La réalisation de l'ACP aboutit à la visualisation des individus et des variables projetés sur le plan principal (ou plan factoriel) composé des deux premiers axes factoriels. Le pourcentage d'inertie associé à chaque axe est précisé sur les étiquettes (figure 14). Le nombre d'individus étant trop important, la représentation des individus ne permet pas d'interprétation. Ce graphe fait ressortir la présence d'individus atypiques qui contribuent fortement à la création des axes. Il est préférable d'exclure de l'ACP ce type d'individu qui peut à lui seul décider de l'axe factoriel créé. On remarque ici quatre individus extrêmes (en bleu) qui ont été exclus de l'analyse. Ils correspondent à des cabinets réalisant des versements hors normes.

Le graphe des variables est un cercle de corrélation (figure 15). Il permet de voir la qualité de la projection de chaque variable dans le plan en fonction de la longueur du vecteur. Comme les variables sont centrées, on a en effet l'égalité suivante concernant la corrélation des variables A et B :

$$cor\ (A, B) = \cos(A, B)$$

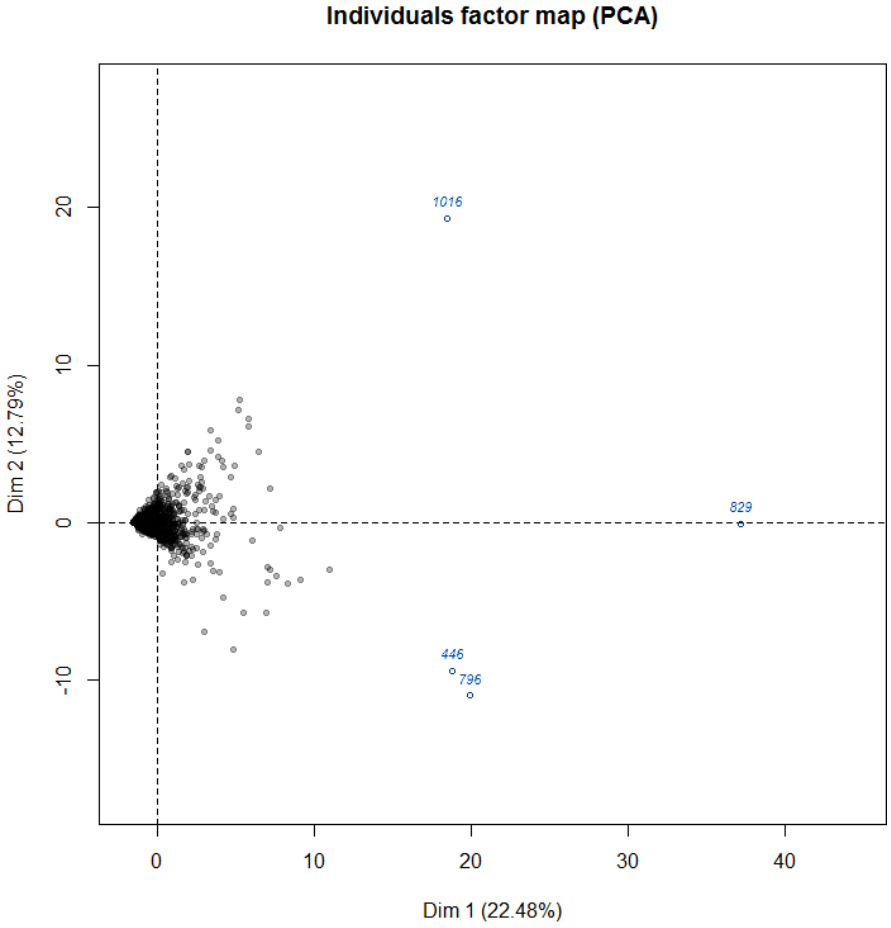


Figure 14 : graphe des individus (les cabinets de géomètres-experts) représentés dans le plan principal

Dossiers du périmètre obligatoire	nb_obli
Dossiers du périmètre obligatoire avec RFU	nb_rfu
Dossiers du périmètre obligatoire avec document joint	nb_doc
Délimitations de la propriété des personnes publiques	delim
Publicités foncières	pub_fonc
Aménagements fonciers	af
États descriptifs de division	copro_edd
Règlements de copropriété	regl_copro
Servitudes	servit
Divisions en volumes	div_vol
Documents modificatifs du parcellaire cadastral	dmpc

Tableau 3 : variables abrégées dans le cercle de corrélation

Plus deux variables sont proches ou diamétralement opposées, plus celles-ci sont corrélées (positivement ou négativement). Lorsque deux vecteurs forment un angle droit, celles-ci sont indépendantes.

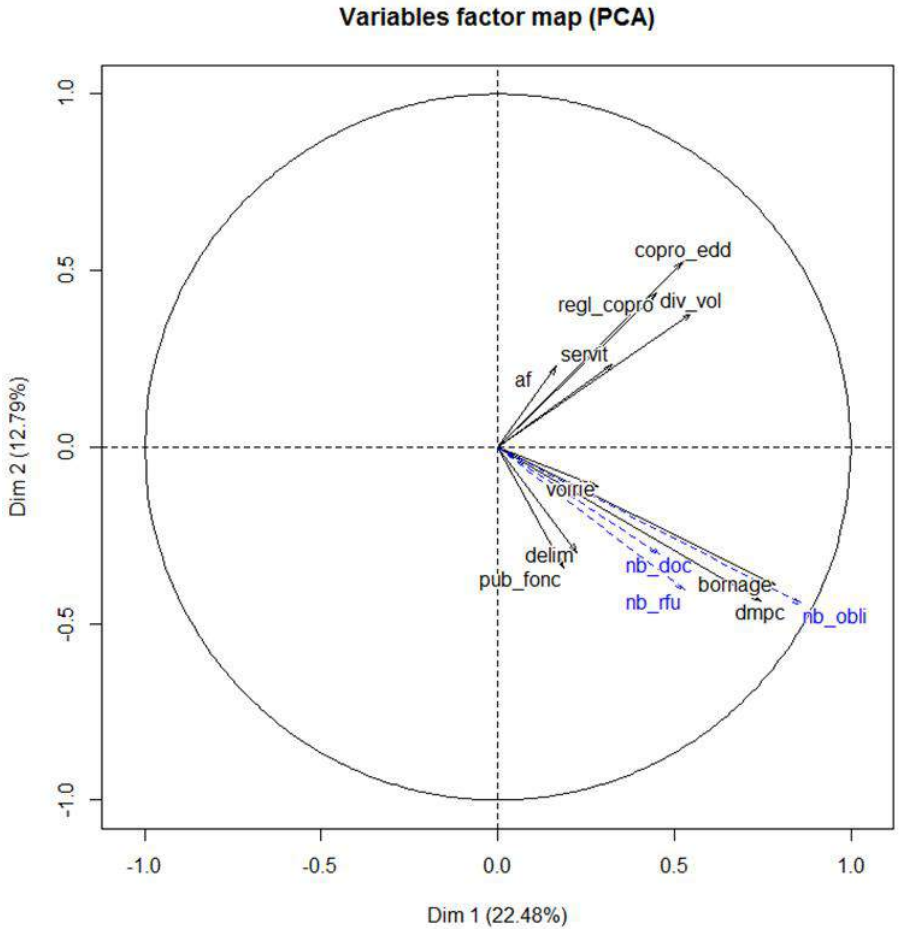


Figure 15 : graphe des variables (les types de versements) dans le plan principal

On distingue en bleu les variables supplémentaires.

Dans le cercle de corrélation et à condition que les variables soient bien projetées, on peut estimer visuellement la présence de liaisons entre celles-ci. Les variables représentées en bleu sont dites supplémentaires, elles ne participent pas au calcul des axes factoriels mais peuvent être lues sur le graphe. Contrairement aux autres variables, celles-ci ne correspondent pas à un type d'opération mais sont des indicateurs intéressants à confronter aux versements de chaque activité.

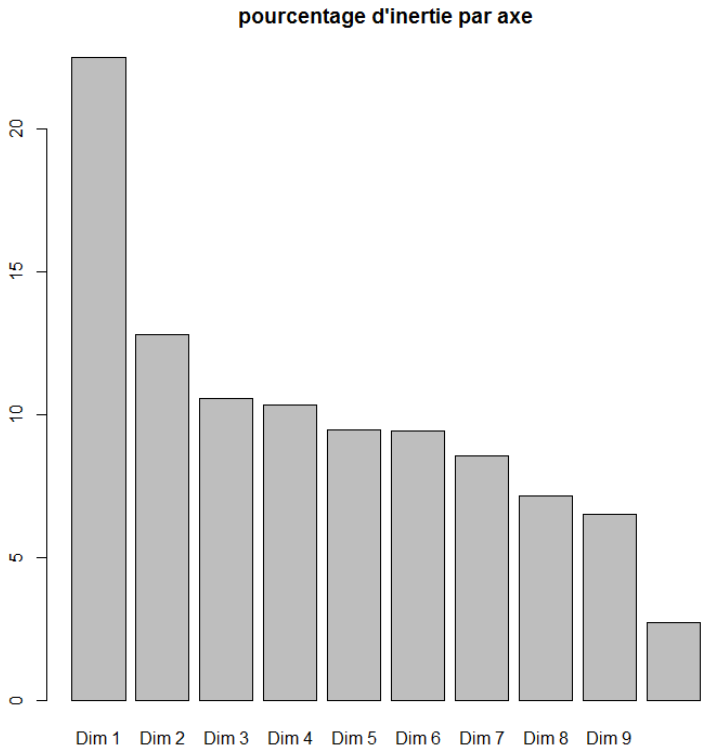


Figure 16 : diagramme en barres des pourcentages d'inerties des axes de l'ACP

Afin de sélectionner le nombre d'axes factoriels considérés pour l'étude, on s'appuie sur l'inertie représentée par chaque axe. Il est possible de sélectionner le nombre d'axes à interpréter en fonction de différentes méthodes. Selon Cornillon et al. (2012, p. 203), l'une d'elles consiste à s'appuyer sur le diagramme en barres des pourcentages d'inertie représentés par axe. Visuellement, on cherche une cassure sur ce graphe, on s'intéresse alors aux dimensions situées avant celle-ci. À partir de la figure 16, on constate cette cassure entre les dimensions deux et trois. On décide donc de ne conserver que les deux premiers axes représentés au travers des figures précédentes (figures 14 et 15), les autres axes étant difficilement interprétables du fait de leur trop faible inertie.

II.2.2.4 Analyse des résultats : distinction des domaines fonciers et immobiliers

Les résultats de l'ACP peuvent être interprétés à l'aide du graphe des variables (figure 15). On remarque tout d'abord les pourcentages d'inertie faibles des deux axes constituant le plan principal : 22 % et 13 %. En sommant, le premier plan ne représente donc que 35 % de l'inertie totale. Ce pourcentage indique quelle est la part d'information issue des données représentée dans le plan principal.

Cette portion est faible dans notre cas, cela s'explique en partie par la présence d'un grand nombre de variables (donc de dimensions), rendant d'autant plus difficile d'obtenir une bonne projection des données dans un plan. Le cercle de corrélation permet tout de même de souligner des liaisons existantes entre certaines variables.

On constate une corrélation forte entre les versements de DMPC et ceux de bornage. Ces deux variables étant bien projetées et proches dans le cercle, cette liaison peut être affirmée avec certitude. La variable supplémentaire correspondant au nombre de versements du périmètre obligatoire est également fortement liée aux versements de DMPC et de bornage. Il en est de même pour les variables du nombre de dossiers avec versement RFU et, dans une moindre mesure, du nombre de dossiers avec un document joint.

De façon moindre, les variables correspondant aux versements d'un état descriptif de division, d'une division en volumes ou d'un règlement de copropriété semblent également liées, bien que leur projection soit de moins bonne qualité. Les autres variables sont globalement mal projetées, il n'est donc pas possible de les interpréter.

Pour résumer, deux « blocs » d'opérations se détachent : les cabinets effectuant massivement des opérations foncières que sont le bornage et la division parcellaire d'une part et les cabinets réalisant des opérations immobilières d'autre part. Ces blocs sont indépendants, cela signifie qu'un cabinet réalisant beaucoup d'opérations foncières n'effectue pas nécessairement un grand nombre d'opérations immobilières et réciproquement. On remarque aussi que les cabinets versant le plus de dossiers du périmètre obligatoire et des objets RFU sont ceux tournés vers le

foncier, ce qui est cohérent. Les opérations concernant la copropriété ou la division en volumes ne nécessitent pas de versement RFU.

II.2.3 Analyse complémentaire sur l'activité et l'économie d'un cabinet

Dans la suite de l'analyse précédente, une ACP complémentaire visant à constater s'il existe des liens entre les revenus d'un cabinet et ses activités est réalisée.

Les données employées ici sont celles issues de l'échantillon de données financières déjà exploité précédemment[12] . À l'aide de l'excédent brut d'exploitation (EBE) et du chiffre d'affaires (CA), il est possible d'aboutir au taux ou marge brute d'exploitation (MBE) par la formule suivante :

$$MBE = 100 \times \frac{EBE}{CA}$$

Cet indicateur économique permet d'évaluer les revenus générés par l'activité d'une entreprise (ORES Pays de la Loire, s.d.). Comme il s'agit d'un taux, il présente l'intérêt d'être indépendant de l'envergure de la société.

Afin d'évaluer s'il existe une corrélation entre les revenus d'un cabinet et le type d'activité exercé, on réalise une seconde ACP en limitant les individus aux cabinets pour lesquels des données financières sont disponibles. Les variables quantitatives prises en compte lors du calcul de l'ACP restent identiques. On modifie seulement les variables supplémentaires en considérant cette fois-ci la marge brute d'exploitation notée « mbe » sur la figure 17.

On constate tout d'abord des résultats similaires à ceux obtenus précédemment : une corrélation forte entre les différents types de versements immobiliers d'une part et fonciers d'autre part. Les 354 entreprises sélectionnées ne diffèrent pas des autres en termes de versements obligatoires sur Géofoncier et sont donc assez représentatives de l'ensemble des cabinets. En ce qui concerne la marge brute d'exploitation, on constate que la variable est perpendiculaire aux deux premiers axes factoriels car la projection de celle-ci est quasi-nulle sur le cercle de corrélation. On peut donc en déduire que la marge brute d'exploitation est plutôt indépendante des variables bien projetées sur ces axes.

L'ACP montre que la marge brute d'exploitation ne dépend pas ou très peu des types de versements réalisés par les cabinets sur l'ensemble des individus considérés. Autrement dit, la rentabilité d'un cabinet n'est généralement pas liée à sa spécialisation dans les dossiers fonciers ou immobiliers.

Même si le recours à l'ACP ne permet pas d'établir un véritable profil-type des cabinets car l'information disponible de façon exhaustive se limite aux versements obligatoires, elle apporte tout de même des tendances sur la profession.

[12] Voir partie II.1.3 sur la situation financière des cabinets de géomètres-experts

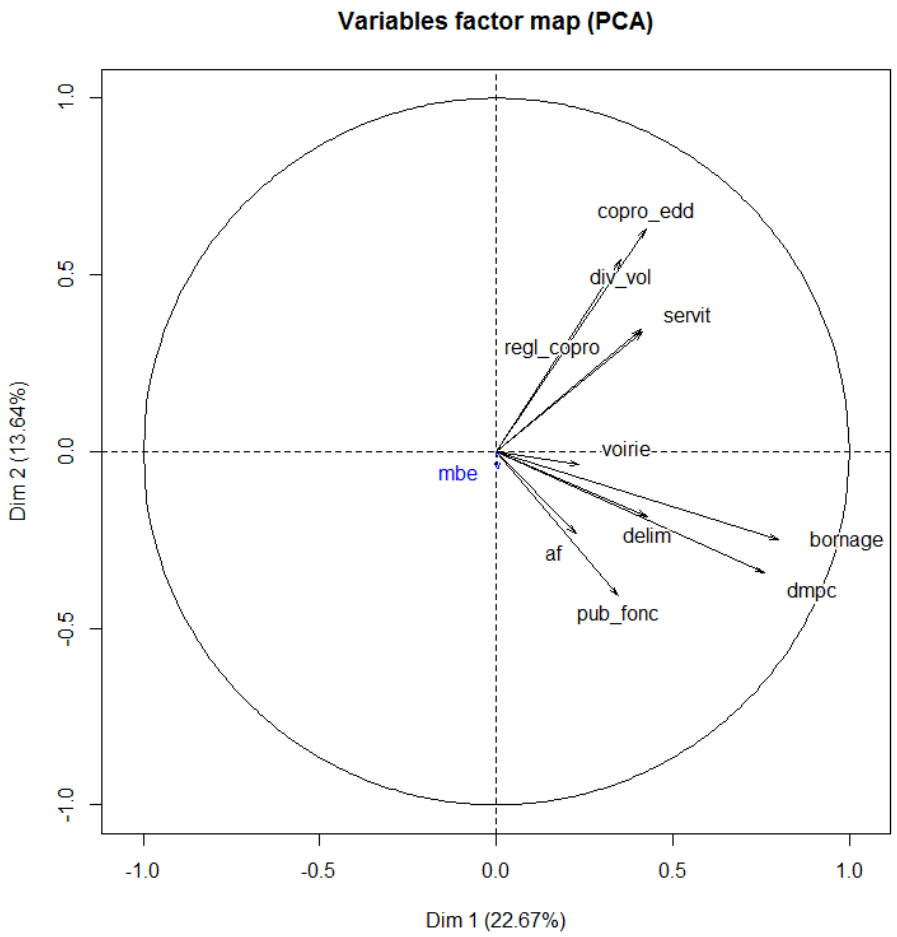


Figure 17 : cercle de corrélation de l'ACP

II.2.4 Établissement d'une typologie des cabinets à partir du chiffre d'affaires

II.2.4.1 Sélection et fiabilité des données employées

Jusqu'à présent, l'analyse des dossiers présents sur l'IDG Géofoncier nous a seulement permis de dégager deux « regroupements » concernant les travaux en lien avec le foncier d'une part et ceux en lien avec l'immobilier d'autre part. Mais aucune classification de la profession n'a été réalisée car toutes les activités ne sont pas soumises à versement obligatoire.

Pour cette raison, nous allons nous appuyer sur une source de données externe provenant de l'OGE : le chiffre d'affaires hors taxe par activité des cabinets de géomètres-experts en 2013, l'enquête permettant de recueillir ces données n'étant pas encore réalisée pour 2014.

Afin de ne garder que les résultats fiables des enquêtes, la somme des chiffres d'affaires par activité est comparée au chiffre d'affaires total obtenu à partir de la base de données des cotisations versées par les géomètres-experts à l'OGE. On ne conserve que les données pour lesquelles ces deux valeurs sont égales, ce qui correspond à environ cinq cents cabinets.

La répartition du chiffre d'affaires (figure 18) illustre la prépondérance de deux secteurs d'activité principaux : le foncier et la topographie. Trois secteurs secondaires se démarquent également : l'ingénierie et maîtrise d'œuvre, l'urbanisme et la copropriété, division en volume.

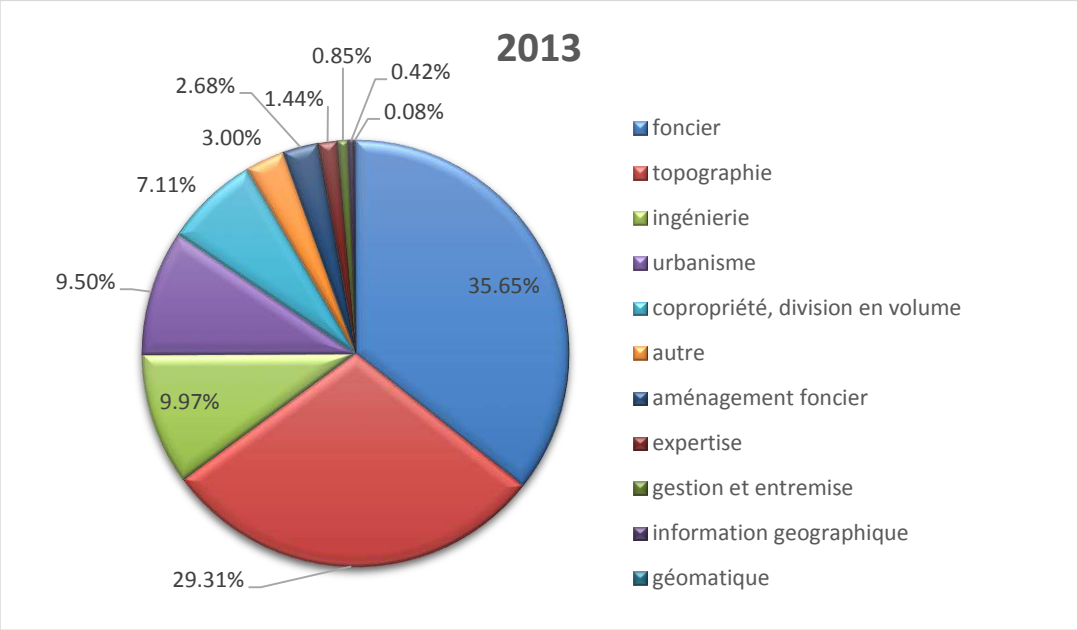


Figure 18: répartition du chiffre d'affaires par activité sur l'ensemble des cabinets étudiés en 2013

En plus des données sur les chiffres d'affaire par activité, d'autres variables issues des analyses précédentes sont prises en compte :

- le nombre de géomètres-experts et de bureaux par cabinet au 31 décembre de l'année considérée,
- le nombre de dossiers immobiliers (division en volumes, état descriptif de division et règlement de copropriété) et fonciers (DMPC et bornage) créés au cours de l'année et présents sur Géofoncier[13].

Toutes ces variables sont susceptibles de caractériser des cabinets et vont être employées au sein d'une classification ascendante hiérarchique (CAH).

II.2.4.2 Principes de la CAH

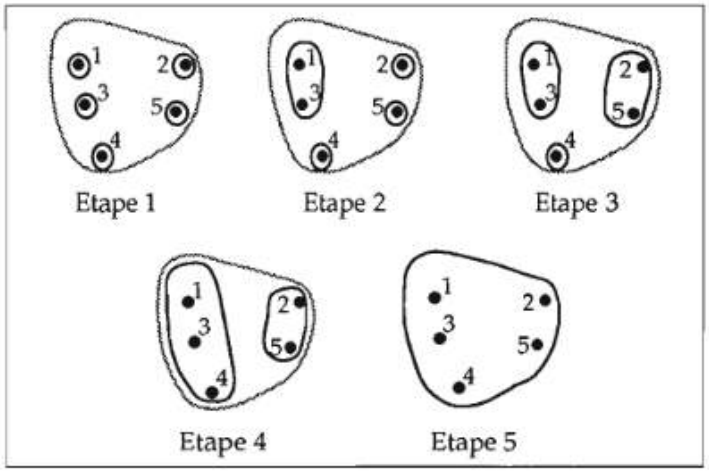


Figure 19 : agglomération progressive de cinq points
Source : Lebart & al. (1997)

La CAH permet de « regrouper des individus dans des classes les plus homogènes possibles à partir d'un jeu de données individus × variables » (Cornillon et al., 2012, p. 230). Les individus sont ici les cabinets de géomètres-experts et les variables sont celles présentées précédemment.

Cette méthode de classification fonctionne par regroupements successifs des individus ou groupes d'individus les plus proches (figure 19). On forme ainsi un dendrogramme ou arbre

[13] Ce choix de variables portant sur des opérations à versement obligatoire découle directement des conclusions de l'analyse exploratoire des répartitions de dossiers versés sur Géofoncier (voir partie II.2.2.4)

hiérarchique qui illustre la formation successive des classes d'individus.

De la même manière que pour une ACP, les *n* variables forment un espace à *n* dimensions au sein duquel la similarité ou proximité des individus correspond à la distance euclidienne les séparant. À chaque itération du processus de classification, le regroupement des individus s'effectue à partir du critère de Ward basé sur l'inertie[14].

L'algorithme employé pour réaliser la CAH est issu du package FactoMineR. Il présente l'intérêt d'effectuer au préalable une ACP sur les données ce qui va notamment permettre d'afficher le nuage de points projeté avec les classes créées sur les axes de l'ACP.

II.2.4.3 Déroulement de la classification sur le jeu de données

La première étape consiste à effectuer une ACP comportant ici quinze variables (tableau 4). L'ACP est réalisée à partir des chiffre d'affaires par activité. Les variables supplémentaires sont notées en rouge. Lors de l'ACP, trois individus atypiques sont exclus de l'analyse. Celle-ci va être employée pour visualiser les individus regroupés par classe à la suite de la CAH.

source d'information	type d'information	abbréviation employée
OGE - chiffre d'affaires par activité	CA total	*CA total*
	urbanisme	*urba*
	aménagement foncier	*af*
	ingénierie	*ingé*
	foncier	*foncier*
	copropriété/division en volumes	*copro div*
	expertise	*expertise*
	gestion et entremise immobilière	*gei*
	topographie	*topo*
	information géographique (SIG)	*info geo*
	géomatique (analyse et traitement de l'image)	*géomatique*
Géofoncier - données OGE	nombre de GE	*nb_ge*
	nombre de bureaux	*nb_bureau*
versements obligatoires	nombre de dossiers fonciers créés	*nb_foncier*
	nombre de dossiers immobiliers créés	*nb_immo*

Tableau 4 : variables employées lors de la CAH

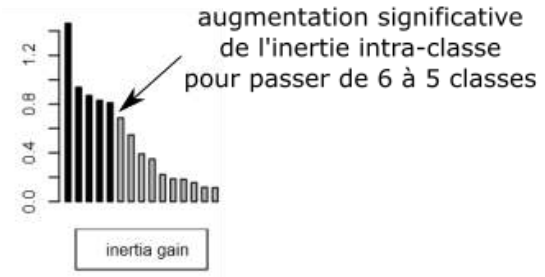


Figure 20 : gain d'inertie intra-classe selon le nombre de classes conservées pour la classification

Ensuite, l'algorithme de classification génère un dendrogramme représentant les étapes successives de regroupement d'individus jusqu'à ne former qu'une seule classe en haut du graphe. Un graphe secondaire (figure 20) présentant le gain d'inertie intra-classe pour passer de *n+1* à *n* classes vient compléter le dendrogramme (figure 21). La seconde barre correspond au gain d'inertie pour passer de trois à deux classes par exemple. Cela permet de choisir le nombre de classes à conserver. On remarque un saut important entre la cinquième et sixième barre, cela signifie qu'il y a une augmentation plus significative de l'inertie intra-classe pour passer de six à cinq classes. Comme on veut minimiser cette inertie, on décide de conserver six classes. Si on laisse ce choix à l'algorithme, celui-ci conserve de même six classes. Le nombre de

[14] Voir annexe 4 : Critère de Ward en CAH

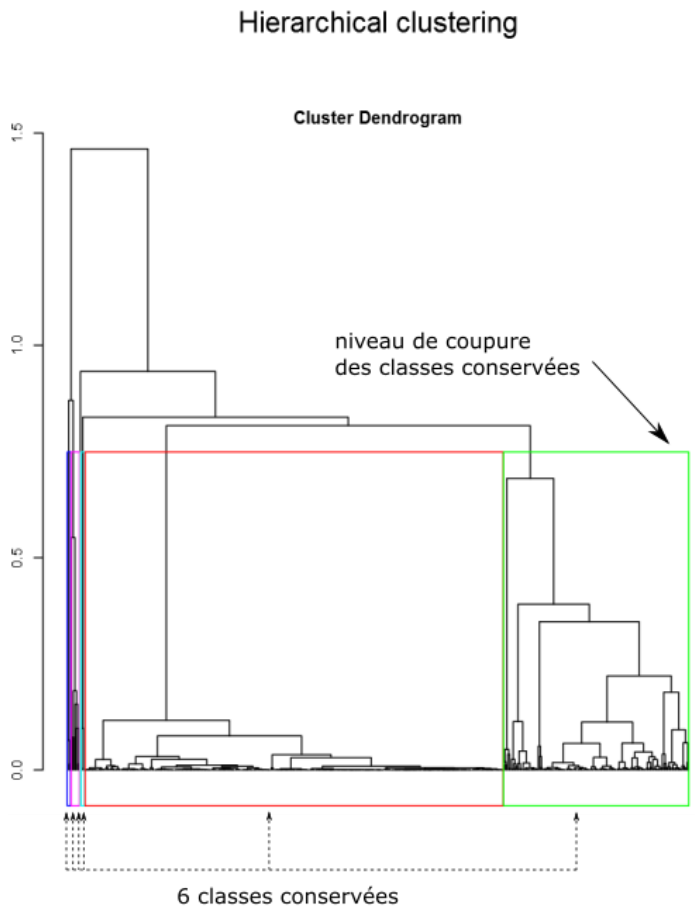


Figure 21 : dendrogramme commenté de la CAH et représentant les classes sélectionnées

classes conservées est spécifique à chaque étude. En outre, on constate qu'une classe est très majoritaire (rectangle rouge) alors que certaines classes sont à peine visibles sur le dendrogramme car elles comportent très peu d'individus.

Pour visualiser les individus et leur classe, on s'appuie sur les axes factoriels définis lors de l'ACP. On sélectionne les deux axes caractérisant et séparant au mieux les classes pour définir un plan sur lequel les individus sont projetés. Le plan composé de l'axe 1 et 5 choisi ici permet de différencier les classes (ou cluster) composées de peu d'individus. Les deux classes les plus conséquentes (en rouge et en vert) sont par contre peu séparées sur la figure 22.

Pour chaque classe, on obtient un descriptif des variables les plus représentatives de celle-ci. Chaque variable *k* est classée selon la valeur test suivante (Lebart & al., 1997) :

$$v_k = \frac{\bar{x}_{qk} - \bar{x}_k}{\sqrt{\frac{s_k^2}{N_q}\left(\frac{N - N_q}{N - 1}\right)}}$$

où s_k est l'écart type de la variable *k* pour l'ensemble des individus. Les notations employées sont identiques à celles de l'annexe détaillant le critère de Ward[15].

En valeur absolue, cette valeur test est équivalente à la valeur prise par une variable centrée réduite. Pour chaque classe, les variables sont ordonnées par ordre décroissant de valeur test en ne conservant que celles dont la valeur test est supérieure à 1,96 en valeur absolue. Ces variables valident alors avec un seuil de risque de 5%[16] l'hypothèse suivante : leur moyenne diffère significativement selon que l'on considère tous les individus ou uniquement ceux de la classe. On distingue ainsi une classe au travers des variables qui la caractérisent.

[15] Voir annexe 4 : Critère de Ward en CAH

[16] Si une variable suit une loi normale centrée réduite, la probabilité qu'un individu prenne une valeur supérieure à 1,96 en valeur absolue est de 5% (Lebart & al., 1997).

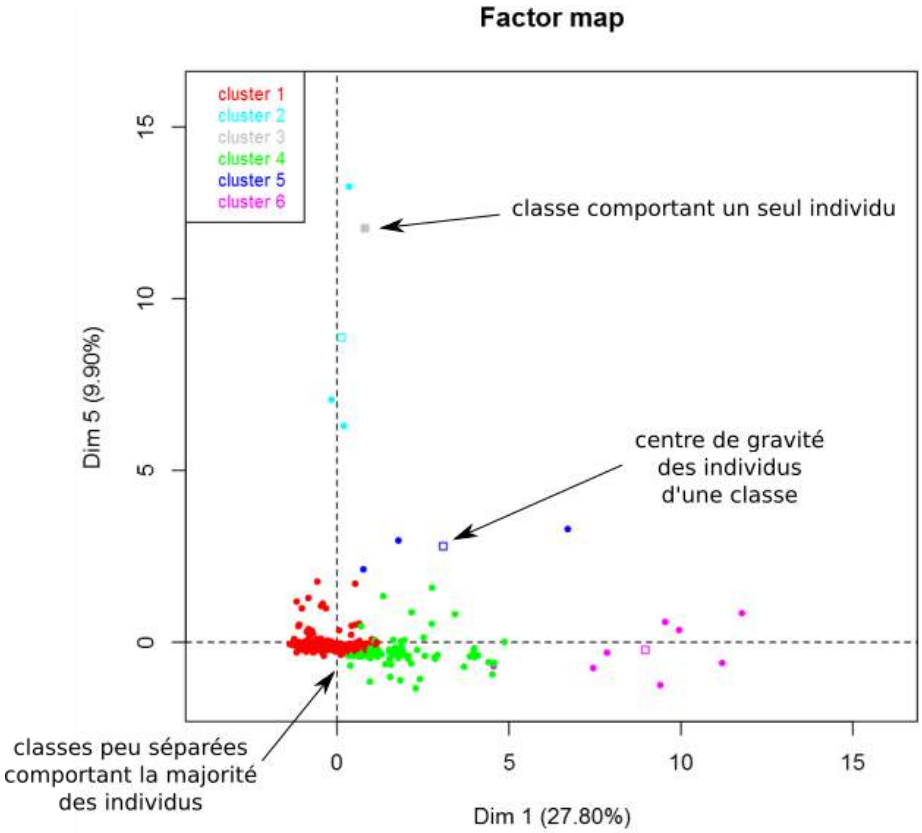


Figure 22 : graphe des individus représentés avec leur classe

Si l'on s'intéresse à la classe 1 représentée en rouge dans les figures précédentes, on constate que les quinze variables sont représentatives de la classe. Toutes les valeurs tests sont inférieures à -1,96, ce qui signifie que la moyenne de cette classe est inférieure à la moyenne générale pour toutes les variables.

Concrètement, le tableau 5 nous indique que la classe 1 correspond aux cabinets de petite taille : le nombre de géomètres-experts et de bureaux est inférieur en moyenne, tout comme le chiffre d'affaires total.

	valeur test	moyenne de la classe	moyenne globale	écart-type de la classe	écart-type global
géomatique	-2.6603	92.6	359.0	889.7	4 626.7
gei	-3.7328	2 020.1	5 786.9	10 496.4	46 619.0
info.geo	-4.0235	648.5	2 896.8	3 920.5	25 816.1
af	-5.9536	7 049.0	18 479.3	29 938.7	88 696.9
expertise	-6.082	4 972.8	8 442.8	11 305.5	26 357.9
nb_foncier	-7.3803	139.7	179.1	137.5	246.8
topo	-8.2873	106 607.3	189 565.4	129 151.7	462 462.0
nb_immo	-9.377	8.5	13.1	10.6	22.8
nb_bureau	-10.1397	1.5	1.8	0.8	1.4
nb_ge	-10.3463	1.3	1.5	0.5	1.0
ingé	-10.7941	28 078.5	68 382.8	63 885.3	172 503.3
foncier	-10.9175	175 911.8	235 061.6	128 615.2	250 300.1
urba	-12.3253	37 566.9	65 544.9	51 011.3	104 869.9
copro.div	-12.8596	27 880.5	48 661.0	33 022.9	74 654.7
CA.total	-13.776	405 489.8	662 838.5	285 295.8	863 036.4

Tableau 5 : variables représentatives de la classe 1 (cabinets de petite taille)

II.2.4.4 Analyse des classes

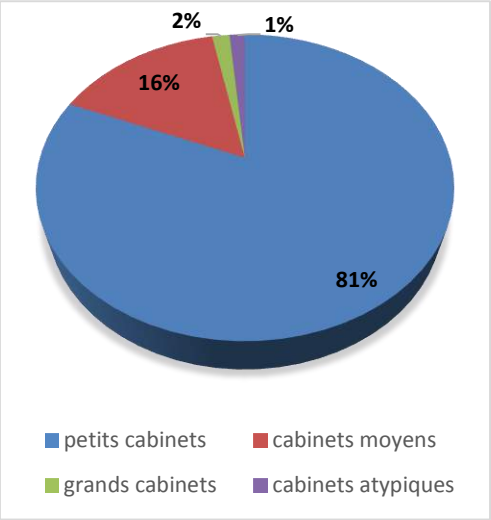


Figure 23 : répartition du nombre de cabinets par classe

La classification effectuée fait ressortir deux classes majoritaires comme l'illustre la figure 23 ci-contre. La première regroupe plus de 80% des cabinets, il s'agit des petits cabinets de géomètres-experts caractérisés par un chiffre d'affaires fortement inférieur à la moyenne quel que soit le type d'activité. Ceux-ci sont composés, pour la plupart, d'un seul géomètre-expert et d'un à deux bureaux. Leur chiffre d'affaires total moyen est inférieur de près de 40% par rapport à la moyenne générale.

La seconde classe majeure est représentative des cabinets de taille moyenne, comprenant deux géomètres-experts et deux ou trois bureaux en général. Au niveau des activités, ceux-ci se démarquent nettement sur deux domaines : la copropriété/division en volume d'une part et l'urbanisme d'autre part. Ce sont d'ailleurs les cabinets contribuant fortement au versement de dossiers obligatoires immobiliers sur Géofoncier. Ils se démarquent de façon plus nuancée sur l'ingénierie, le foncier et l'expertise judiciaire.

Les quatre autres classes obtenues sont mineures. On y distingue une classe correspondant aux cabinets les plus conséquents caractérisés principalement par un chiffre d'affaires total sept fois supérieure à la moyenne et huit bureaux ou plus en général. Les trois dernières classes correspondent à quelques cabinets atypiques se démarquant par une activité très spécifique : l'information géographique, la géomatique ou la gestion et entremise immobilière.

	petits cabinets	cabinets spécialisés dans l'urbanisme et l'immobilier	grands cabinets
effectif de la classe	398	76	8
CA total	405 490	1 447 142	5 231 842
nombre de GE	1 - 2	2 - 3	5 - 6
nombre de bureaux	1 - 2	2 - 3	8 - 9
aménagement foncier	- -	+	+++
topographie	- -	+	+++
foncier	- - -	++	+++
ingénierie	- - -	++	+++
copropriété/division en vol.	- - -	+++	++
urbanisme	- - -	+++	++
expertise judiciaire	- -	++	0
géomatique	-	0	0
information géographique	-	0	0
gestion et entremise immo.	-	0	0
versement foncier	- -	+	+++
versement immobilier	- -	++	+

Tableau 6 : caractéristiques des trois classes principales de cabinets

CA ou versement par rapport à la moyenne de l'ensemble des cabinets :	
+++	très supérieur
++	supérieur
+	légèrement supérieur
0	équivalent
-	légèrement inférieur
- -	inférieur
- - -	très inférieur

Figure 24 : légende du tableau comparatif

On a donc principalement trois classes qui correspondent aux différentes envergures de cabinets résumées dans le tableau comparatif (tableau 6). Les cabinets plus spécialisés dans l'immobilier et l'urbanisme sont de taille moyenne alors que les grands cabinets sont caractérisés par des activités plus diversifiées, notamment dans le foncier, l'ingénierie et la topographie. Cela s'illustre au travers de la répartition du chiffre d'affaires dans certaines activités, notamment le foncier et la copropriété et division en volumes (figure 25). Les activités d'information géographique, géomatique et gestion et entremise immobilière sont peu caractéristiques de ces classes car elles représentent une portion infime du chiffre d'affaires total (figure 18).

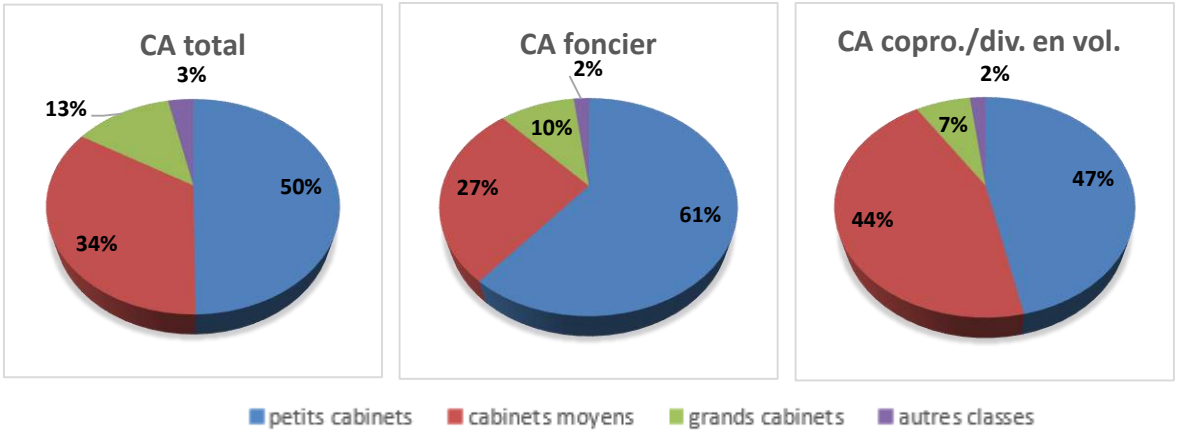


Figure 26: répartition du chiffre d'affaires (total ou pour une activité particulière) par classe sur l'ensemble des cabinets étudiés

Enfin, il est également intéressant de spatialiser cette classification par région ordinale afin d'avoir un aperçu de la répartition territoriale des trois classes principales (figure 26)[17]. Il faut cependant noter que les cabinets sont ici localisés à l'aide de leur bureau principal, les bureaux secondaires ne sont donc pas pris en compte.

On constate que les cabinets spécialisés dans la copropriété et la division en volumes se concentrent fortement en région parisienne où l'immobilier représente une part d'activité conséquente des géomètres-experts. À l'inverse, les grands cabinets ont leur bureau principal en province. Leurs activités phares (topographie, aménagement foncier, ingénierie) sont souvent connectées à des projets

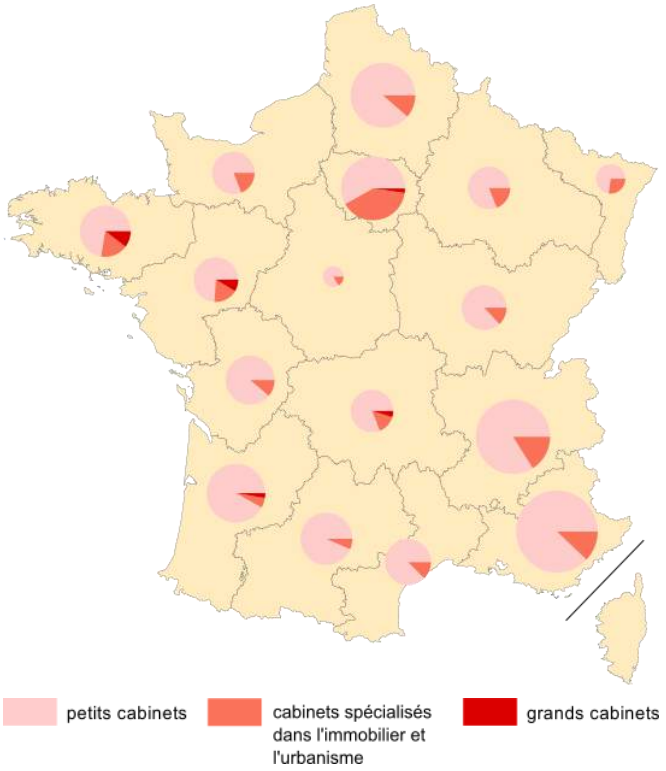


Figure 25 : répartition par région ordinale des classes de cabinets

[17] Voir également carte 1 dans l'atlas cartographique : Répartition des cabinets de géomètres-experts par classe et par région ordinale (année 2013)

d'envergure développés dans des espaces à aménager (création d'une ligne à grande vitesse par exemple). Leurs nombreux bureaux sont donc souvent dispersés sur un vaste territoire.

Cette première carte illustre l'importance d'une représentation spatiale d'un résultat, mettant en avant des spécificités géographiques et territoriales.

Cette partie d'analyse statistique a mis en avant le regroupement des cabinets en structures plus conséquentes. La classification établie présente une différenciation des activités selon l'envergure des cabinets avec notamment une classe spécialisée dans l'immobilier et l'urbanisme. Nous allons désormais nous tourner vers la composante géographique de l'étude.

III Analyse spatiale et cartographie

Avant de présenter de nouvelles cartes, il est important de préciser certaines règles de sémiologie graphiques suivies. En ce qui concerne les composantes quantitatives, celles-ci sont généralement réparties en classes. La méthode de Jenks a été choisie afin de définir les paliers des classes. Celle-ci vise à maximiser la variance entre les classes créées (variance inter-classe) tout en minimisant la variance à l'intérieur de chaque classe (variance intra-classe). Concrètement, cela permet d'avoir des classes homogènes et distantes au maximum les unes par rapport aux autres. Le nombre de classes choisis ne dépasse jamais cinq afin de pouvoir facilement les distinguer visuellement entre elles.

La sémiologie graphique employée ici s'appuie sur le classement des variables visuelles établi par Bertin (1999, p. 69). Les composantes ordonnées sont représentées à l'aide d'une variation de valeur ou de taille. Les composantes quantitatives sont symbolisées par une variation de taille et les composantes qualitatives par une variation de couleur.

III.1 Répartition des bureaux de géomètres-experts

III.1.1 Méthodologie et données complémentaires

Afin de poursuivre l'étude, il est nécessaire de comparer les données concernant les géomètres-experts à des données extérieures à la profession. L'individu étudié dans la suite de cette partie est le bureau de géomètre-expert. Plus précisément, il s'agit des bureaux où au moins un géomètre est censé être présent tous les jours de la semaine. Ce sont les bureaux qualifiés de « principaux » et « secondaires » au niveau de la base de données qui sont pris en compte. Cela exclut les bureaux de chantiers, qui ont vocation à être temporaire, ainsi que les permanences.

Le niveau administratif d'étude choisie est le département car de nombreuses données statistiques sont accessibles à cette échelle.

L'INSEE dispose de nombreuses études statistiques avec des résultats généralement recensés par commune. Dans l'étude suivante, deux jeux de données de l'INSEE seront employés :

- la démographie par département, résultats provisoires arrêtés à la fin 2014,
- la démographie des établissements du secteur de la construction par commune en 2013.

Les géométries des départements sont issues d'OSM[18]. Toutes ces données sont rassemblées à l'aide de requêtes SQL après import sur une base de données PostgreSQL. On y ajoute également les données concernant le nombre de bureaux de géomètres-experts par département. La table créée est alors exportée sous format shape afin d'être visualisée et mise en forme sous ArcGIS.

[18] Des liens sont disponibles en fin de bibliographie permettant d'accéder aux jeux de données libres employés.

Une donnée vectorielle issue du GéoCatalogue a également été employée, il s'agit du périmètre des aires urbaines établi en 2010. Celui-ci a été corrigé sur QGIS afin de correctement prendre en compte les villes de Paris, Lyon et Marseille car leur territoire communal n'était pas considéré comme faisant partie d'une aire urbaine dans le fichier.

III.1.2 Une concentration des bureaux au cœur des aires urbaines

Avant de réaliser une quelconque cartographie des bureaux par département, il est intéressant d'observer la localisation de ces bureaux sur le territoire[19]. On constate alors visuellement une certaine concentration des bureaux.

Si l'on met en parallèle la localisation des aires urbaines[20] françaises définies par l'INSEE, on constate que la concentration des bureaux se situe au cœur de celles-ci. En effet, elles recouvrent 45 % du territoire et rassemblent 89 % des bureaux de géomètres-experts au niveau de la métropole. Les dix aires urbaines principales en termes de population concentrent 28 % des bureaux sur 9 % du territoire métropolitain. Ce phénomène d'agrégation est cohérent : les aires urbaines concentrent les activités de façon générale, y compris celles exercées par le géomètre-expert, notamment en ce qui concerne l'immobilier.

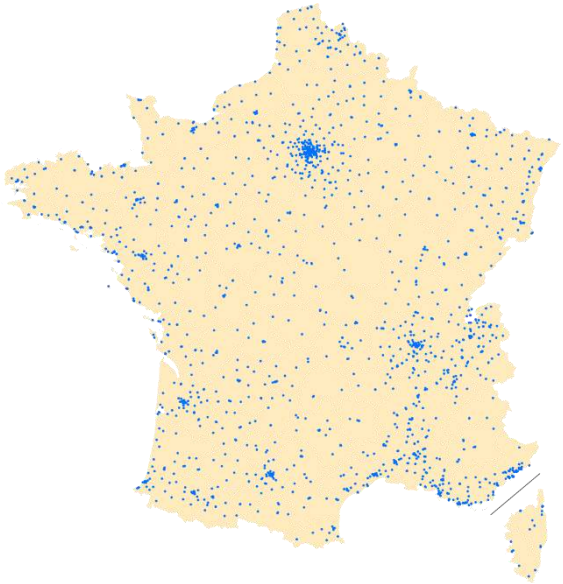


Figure 27 : localisation des bureaux de géomètres-experts en métropole

Le cartogramme représentant la répartition territoriale des bureaux de géomètres-experts en 2015[21] illustre également ce déséquilibre. Au nord, les départements de l'Ile-de-France sont énormément déformés, traduisant une forte densité de bureaux dans cette région contrastant avec celle des départements ruraux du centre de la France. Au sud, ce sont les départements de la vallée du Rhône, de l'Aquitaine et des bords de la Méditerranée qui se démarquent le plus.

Comme la partie précédente concernant l'évolution quantitative de la profession[22] l'a montré, l'analyse des données n'est pertinente qu'à partir de 1999. C'est pourquoi la période d'étude choisie est ici comprise entre 2000 et 2015. Les résultats obtenus sont systématiquement pris au 1er janvier de l'année.

La carte de la dynamique des bureaux de géomètres-experts est disponible en annexe[23]. Celle-ci présente deux variables : le nombre de bureaux par département en 2015 et l'évolution de ce nombre depuis 2000. Sur quinze ans, on constate que les départements présentant la dynamique positive la plus forte sont majoritairement des territoires comportant une métropole : la Loire-Atlantique (Nantes), le Bas-Rhin (Strasbourg), les Bouches-du-Rhône (Marseille) et la Haute-Garonne (Toulouse). D'autres départements connaissent un développement important concernant

[19] Voir carte 2 dans l'atlas cartographique : Localisation des bureaux de géomètres-experts en 2015 et influence des aires urbaines

[20] Dans le cas présent, une aire urbaine est un ensemble de communes, d'un seul tenant et sans enclave, constitué par un pôle d'au moins 1 500 emplois, et par des communes rurales ou unités urbaines dont au moins 40 % de la population résidente ayant un emploi travaille dans le pôle ou dans des communes attirées par celui-ci. Il s'agit donc d'une définition au sens large qui comprend les petites, moyennes et grandes aires urbaines définies par l'INSEE.

[21] Voir carte 3 : Répartition territoriale des bureaux de géomètres-experts en 2015

[22] Voir partie II.1.1.1

[23] Voir carte 4 : Dynamique des bureaux de géomètres-experts (2000-2015)

les bureaux de géomètres-experts, notamment en Corse et à la Réunion. À l'inverse, la Gironde et les Alpes-Maritimes connaissent la décroissance la plus marquée. Les causes de ces évolutions sont essentiellement spécifiques à chaque département.

	nombre de bureaux en 2015	évolution 2000 – 2015 (en valeur absolue)
Bouches-du-Rhône	56	8
Gironde	54	-8
Haute-Savoie	45	0
Var	43	-1
Rhône	40	-1
Seine-et-Marne	40	-1
Isère	39	0
Loire-Atlantique	39	6
Pyrénées-Atlantiques	38	6
Haute-Garonne	38	7

Tableau 7 : classement des dix départements comportant le plus de bureaux de géomètres-experts
en rouge sont représentés les départements comportant une aire urbaine majeure

Cela aboutit en 2015 à une répartition des bureaux par département en lien avec l'activité économique locale et les enjeux territoriaux. Sur les dix départements concentrant le plus de bureaux, neuf sont liés à une des quinze premières aires urbaines selon le classement établi par l'INSEE en 2012. On a donc une concentration forte de la profession à proximité des pôles principaux d'activité du territoire.

Le seul véritable cas particulier de ce classement est la place de la Haute-Savoie, département ne présentant pas de métropole majeure et comportant un nombre très important de bureaux. Une raison probable à cela est la forte attractivité du département qui bénéficie de la proximité de la Suisse limitrophe selon une étude de l'INSEE (Léger & Gilbert, 2014).

III.1.3 Une présence plus marquée dans le sud de la France

En s'appuyant sur la démographie départementale, on peut établir une variable relative du nombre de bureaux de géomètres-experts pour 1 000 habitants[24]. Bien entendu, les valeurs obtenues sont faibles. Le principal intérêt est l'étude relative des écarts que l'on peut constater entre départements.

Les départements où les bureaux de géomètres-experts sont nombreux par rapport à la population se situent sans exception dans la moitié sud de la France, dans des départements ruraux. On retrouve également la présence de la Haute-Savoie dans ce classement, seul département ayant la particularité d'avoir à la fois un nombre élevé de bureaux en absolu et relativement à la population départementale.

Afin d'avoir des résultats plus faciles à interpréter, le tableau 8 représente l'équivalent en nombre d'habitants d'un bureau de géomètre-expert. Le premier département en termes de présence de la profession relativement à la population est donc celui pour lequel le nombre d'habitants par bureau est le plus faible.

[24] Voir carte 5 : Répartition départementale des bureaux de géomètres-experts

	nombre d'habitants pour un bureau	nombre de bureaux en 2015
Haute-Corse	13 467	13
Gers	15 912	12
Alpes-de-Haute-Provence	16 244	10
Haute-Savoie	17 275	45
Pyrénées-Atlantiques	17 545	38
Landes	18 248	22
Aveyron	18 338	15
Corrèze	18 427	13
Corse-du-Sud	18 503	8
Savoie	18 579	23

Tableau 8 : classement des dix premiers départements en termes de bureaux par habitant

On peut trouver notamment deux explications à cela :

- l'attractivité des départements du sud et de l'ouest, particulièrement prononcée au début des années 2000 comme l'illustre une étude publiée par l'INSEE sur le recensement de la population de 2006 (Baccaïni & Levy, 2009),
- la faible densité et la dispersion de la population en milieu rural.

Le géomètre-expert est un professionnel travaillant avant tout sur le foncier, que celui-ci se situe en milieu urbain ou rural. Il y a donc un besoin, même dans les espaces peu peuplés, de pouvoir disposer de cette profession.

III.1.4 Étude complémentaire sur le secteur du bâtiment et les bureaux de géomètres-experts

Pour compléter les analyses précédentes, une étude a été menée afin de mettre en parallèle la répartition des établissements du secteur de la construction et les bureaux de géomètres-experts. Ce sont a priori des domaines liés, le géomètre-expert étant souvent présent avant les travaux de construction (création des lots d'un lotissement, implantation de bâtiment par exemple). Une carte mettant en avant le nombre d'établissements de la construction avec le nombre de bureaux de géomètres-experts par département a ainsi été produite[25].

Parmi les dix départements où se concentre le plus le secteur de la construction (tableau 9), on retrouve la moitié des départements comportant le plus de bureaux de géomètres-experts. Le secteur de la construction se focalise là où l'activité économique est la plus importante : dans les grandes métropoles, au bord de la Méditerranée notamment. Les deux activités ont donc une répartition assez similaire sur le territoire bien que Paris présente une concentration très importante d'établissements du secteur de la construction qui est plus nuancée au sein de la profession de GE.

En effet, le domaine de la construction travaille majoritairement en milieu urbain alors que le géomètre-expert est fréquemment amené à travailler en milieu rural (division de parcelles agricoles par exemple).

[25] Voir carte 6 : Les géomètres-experts et le secteur de la construction

	établissements de la construction	bureaux de géomètres-experts
Paris	49 416	32
Bouches-du-Rhône	23 762	56
Seine-Saint-Denis	16 697	19
Alpes-Maritimes	16 478	35
Rhône	16 059	40
Var	13 769	43
Gironde	13 445	54
Hérault	12 143	33
Nord	11 768	36
Haute-Garonne	11 094	38

Tableau 9 : classement des dix départements comportant le plus d'établissements de la construction
en rouge sont représentés les départements classés parmi les dix comprenant le plus de bureaux
de géomètres-experts

III.2 Cartographie des dossiers présents sur Géofoncier

L'étude spatiale des bureaux de géomètres-experts a mis en avant une représentation plus forte de la profession au sud de la France et à proximité des principales métropoles. On peut alors se demander si la répartition des objets RFU et des dossiers, notamment en ce qui concerne le foncier et l'immobilier, présente des similarités avec celle des géomètres-experts.

III.2.1 Répartition départementale des données du RFU

Au contraire des bureaux de géomètres-experts qui représentent un nombre limité d'individus (environ 2 000 entités), l'étude des dossiers ou des objets RFU est plus complexe. En effet, Géofoncier comporte plus de trois millions de dossiers et cinq millions d'objets RFU, ce qui limite l'ampleur des opérations réalisables sur ces éléments à l'échelle du pays[26].

Une cartographie départementale illustrant la répartition des objets RFU versés depuis l'ouverture du portail a pu tout de même être réalisée[27]. Il en ressort une répartition fortement inégale des versements entre des départements au Nord-Ouest comportant plus de 100 000 objets (notamment la Loire-Atlantique et le Morbihan) et d'autres au Nord-Est en comprenant moins de 5 000. Il convient également de mentionner le cas à part de l'Alsace-Moselle disposant d'un cadastre spécifique expliquant en partie le très faible versement.

Ce déséquilibre interdépartemental est principalement lié à la volonté variable des géomètres-experts locaux d'effectuer leurs versements. Même s'il s'agit d'une obligation ordinale, le RFU n'est pas systématiquement renseigné[28]. À l'échelle nationale, il est donc délicat pour le moment de mener des analyses illustrant l'activité de la profession ou l'évolution du foncier à partir de ces données. C'est pourquoi nous allons nous concentrer davantage sur la BDD concernant les dossiers.

III.2.2 Carte de densité concernant les dossiers

Afin d'avoir une représentation claire et lisible des dossiers sur l'ensemble du territoire, une carte de densité a été réalisée en s'appuyant sur le style d'une carte de chaleur[29] et en utilisant les localisations des dossiers, chaque dossier pouvant avoir une ou plusieurs localisations. À l'aide

[26] Voir annexe 5 : Problématique du volume des données : exemple de méthode de récupération des objets RFU par département sur l'ensemble du territoire
[27] Voir carte 7 : Versement d'objets RFU par département
[28] Ce constat rejoint celui présenté au travers de la figure 12 sur l'évolution de l'activité, partie II.2.1
[29] Méthode de représentation de la densité similaire à celle employée pour les températures : les espaces les plus denses sont en rouge (couleur chaude) alors que les moins denses sont en bleu (couleur froide).

d'ArcGIS, il est possible de générer un fichier raster où l'on affecte à chaque pixel une valeur correspondant à une densité de dossiers. Pour cela, on s'appuie sur l'outil *densité de points* disponible dans l'extension *Spatial Analyst*.

On réalise la carte de densité en s'appuyant sur une grille avec des mailles de un kilomètre environ et en configurant un masque correspondant à la France métropolitaine. Le calcul de la valeur attribuée à chaque pixel est généré en prenant en compte les pixels voisins dans un rayon de huit unités. La même méthode est appliquée aux DOM ensuite. Une fois les fichiers rasters générés au format TIFF, on applique sur la mise en page d'ArcGIS une égalisation de l'histogramme afin d'améliorer le rendu visuel de la carte. Le reste de la mise en forme du rendu est achevée sous Inkscape.

Cette opération est réalisée sur l'ensemble des dossiers de 2014. À travers la carte de densité produite[30]., on constate une concentration des dossiers au cœur de l'Ile-de-France, le long de la vallée du Rhône et des Alpes, à proximité du littoral méditerranéen et atlantique. On retrouve également les principales aires urbaines (Lille, Strasbourg, Bordeaux, Nantes ou Toulouse), celles-ci comportant une forte densité de dossiers.

Les espaces où l'activité est la plus éparse sont essentiellement ruraux, forestiers ou montagneux avec une densité de population faible. On notera le cas particulier de la Guyane où la densité de dossiers est quasi-nulle en lien avec la forêt tropicale. Cette carte de densité des dossiers rejoint celle présentée précédemment concernant la répartition départementale des bureaux de géomètres-experts[31]

On remarque également un effet de tâche rouge et ronde sur la carte. Ces tâches sont de taille similaire au voisinage considéré lors du calcul, à savoir un cercle de rayon huit pixels. Elles sont générées par une concentration locale forte des dossiers au milieu d'un espace très peu dense.

III.2.3 Répartition départementale des dossiers fonciers et immobiliers

Les dossiers de type foncier et immobilier représentent la quasi-totalité des versements obligatoires réalisés par les géomètres-experts sur l'IDG Géofoncier. Nous avons aussi pu constater que ces deux activités ont tendance à être exercées par des cabinets différents au travers de l'ACP présentée sur les types de dossiers versés[32]. C'est pourquoi celles-ci sont représentées séparément avec d'une part, les dossiers de DMPC et bornages et d'autre part, les dossiers de type : état descriptif de division, règlement de copropriété et division en volumes. Il convient de mentionner qu'un dossier peut ici être comptabilisé plusieurs fois s'il comporte différentes opérations (un dossier comportant un bornage et un état descriptif de division est comptabilisé en tant qu'opération d'ordre foncière et immobilière par exemple).

Contrairement à la carte de densité précédente, une répartition départementale des opérations ne nécessite pas de s'appuyer sur la géométrie des objets mais seulement sur l'attribut correspondant au code INSEE de la commune où se situe le dossier. En effet, celui-ci inclut le numéro du département. La carte concernant l'activité foncière et immobilière sur la période 2012-2014[33] met en parallèle la répartition des bureaux avec des diagrammes proportionnels au nombre d'opérations foncières et immobilières.

On constate la forte prédominance de l'activité foncière sur l'ensemble des départements à l'exception de Paris et de la petite couronne. L'immobilier se concentrant en milieu urbain, la copropriété et les divisions en volumes sont des activités absentes des espaces ruraux.

[30] Voir carte 8 : Densité de dossiers créés en 2014
[31] Voir carte 4 : Dynamique des bureaux de géomètres-experts (2000-2015)
[32] Voir partie II.2.2.4
[33] Voir carte 9 : Activité foncière et immobilière - répartition départementale (2012-2014)

Les variables concernant le total des opérations foncières et immobilières à caractère obligatoire d'une part et le nombre de bureaux par département d'autre part présentent une corrélation égale à 0,6. Il y a donc un lien significatif entre ces deux variables : les départements concentrant le plus d'opérations obligatoires ont tendance à être ceux où l'on trouve le plus de bureaux de géomètres-experts.

Au-delà de ce lien global, de nombreux départements font exceptions notamment ceux situés sur la façade méditerranéenne où les opérations obligatoires sont faibles par rapport au nombre de bureaux. Les départements où les celles-ci sont les plus nombreuses se situent sur la façade atlantique, en Rhône-Alpes et dans le Nord-Pas-de-Calais.

	opérations		total
	foncières	immobilières	
Loire-Atlantique	16 439	1 136	17 575
Charente-Maritime	13 898	608	14 506
Manche	14 063	190	14 253
Gironde	12 661	990	13 651
Nord	12 484	1 144	13 628
Vendée	11 911	229	12 140
Haute-Savoie	10 629	1 237	11 866
Pas-de-Calais	10 954	630	11 584
Ain	10 592	511	11 103
Loiret	10 826	259	11 085

Tableau 10 : classement des dix départements comportant le plus d'opérations foncières et immobilières du périmètre obligatoire sur la période 2012-2014

Outre l'incertitude concernant l'exhaustivité des versements obligatoires, cela illustre une répartition de la profession sur le territoire en partie inégale par rapport à l'activité de chaque bureau de géomètre-expert.

III.3 Analyse régionale : le cas de la région Midi-Pyrénées

Dans cette partie sont détaillées des analyses dans la continuité des précédentes mais à une échelle plus grande : la zone d'étude devient la région ordinale. Une approche plus fine est en effet nécessaire en ce qui concerne les analyses menées ici : l'impact de l'occupation du sol sur l'activité et la dimension géographique de la concurrence au sein des géomètres-experts. Ces études nécessitent d'exploiter un nombre d'entités géométriques d'autant plus important que la zone étudiée est étendue.

III.3.1 Sélection des données d'appui pour les analyses

La composante essentielle à exploiter est ici la localisation des dossiers car il s'agit de la donnée représentant le mieux la répartition géographique du travail de la profession. La sélection des travaux s'appuie sur les mêmes critères que ceux énoncés lors de l'étude de l'activité du géomètre-expert en fonction des versements de dossiers[34]. On ne s'intéresse donc qu'à ceux créés entre 2012 et 2014, appartenant au périmètre obligatoire et se situant en Midi-Pyrénées.

En ce qui concerne l'occupation du sol, deux sources de données sont à disposition sur la France entière : Corine Land Cover, datant de 2006, et OpenStreetMap. La seconde a l'avantage de s'appuyer sur les données de la première et d'être complétée par les contributeurs. Elle est donc plus complète et actualisée mais aussi moins homogène sur le territoire et plus lourde à traiter.

[34] Voir partie II.2.1

Cependant, la précision de Corine Land Cover est limitée[35] et l'apport des contributeurs à OSM est d'une meilleure précision. Dans le cas présent, OSM a été privilégié car le volume de données à traiter ici est restreint à une région.

Le site Geofabrik fournit gratuitement sous différents formats les données d'OSM, par région, actualisées régulièrement. L'ensemble des polygones utilisant la clé *landuse* correspondant à l'occupation du sol sont accessibles au sein d'un fichier unique sous format shape. Chaque contributeur étant libre sur le choix des « tags »[36] qu'il emploie, seuls les plus significatifs doivent être gardés pour l'étude.

occupation du sol	valeurs de la clé *landuse* employées	superficie (km²)
cultures et pâtures	*meadow, farm, farmland*	9144
espace urbain résidentiel	*residential*	993
vignes	*vineyard*	423
vergers	*orchard*	150
espace urbain industriel et commercial	*industrial, retail, commercial*	126

Tableau 11 : ensemble des occupations du sol sélectionnées sur la région Midi-Pyrénées

De plus, le RPG est utilisé pour compléter l'information concernant l'espace agricole. Il est composé de l'ensemble des îlots de culture déclarés par les exploitants agricoles dans le cadre de la politique agricole commune. Datant de 2012, il a l'avantage d'être plus complet et plus précis que l'occupation du sol issue d'OSM ou de Corinne Land Cover. L'ensemble des données nécessaires est importée sous une base de données locale PostgreSQL afin d'être traitée pour chaque analyse.

III.3.2 Répartition de l'activité selon l'occupation du sol

L'occupation du sol est un indicateur de la valeur d'un terrain et des enjeux fonciers autour de celui-ci. L'objectif de cette partie est d'observer dans quelle mesure l'activité des géomètres-experts est en lien avec l'occupation du sol. Pour cela, on s'intéresse au type de terrain sur lesquels se situent les dossiers.

À l'échelle de la région, les résultats obtenus sont présentés dans le tableau 12.Ce qui ressort le plus des valeurs obtenues est la prédominance des dossiers localisés en milieu urbain et tout particulièrement en milieu résidentiel. La densité des dossiers est seize fois supérieure en milieu résidentiel par rapport aux espaces agricoles traditionnels. Les dossiers à dominante immobilière sont présents exclusivement en milieu urbain.

occupation du sol	nombre de dossiers	densité de dossiers au km²		
		foncier	immobilier	ensemble des dossiers
cultures et pâturages	5983	0,65	0,00	0,65
espace urbain résidentiel	11002	10,48	0,58	11,08
vignes	391	0,92	0,00	0,92
vergers	254	1,70	0,00	1,70
espace urbain industriel et commercial	778	5,92	0,23	6,16

Tableau 12 : densité de dossiers en fonction de l'occupation du sol

[35] La précision des données issues de Corine Land Cover 2006 est inférieure à 100 m (Copernicus, 2015).
[36] Au sein d'OSM, les attributs d'une entité sont présents sous la forme d'un « tag » composé d'une clé et d'une valeur. Par exemple, *highway = secondary* est le tag indiquant une route secondaire.

Ce constat rejoint celui présenté précédemment sur les aires urbaines[37]. Les bureaux de géomètres-experts sont présents à proximité des métropoles, là où se situent l'espace urbain et la majorité de l'activité foncière et immobilière.

En espace rural, la concentration des dossiers fonciers est variable selon le type d'exploitation. Les vergers et les vignes dans une moindre mesure sont des espaces où l'on trouve davantage de dossiers que les espaces cultivés et les pâturages. Cela se constate autant à l'aide du RGP que des données d'OSM. Les estives et les landes sont des espaces vastes où l'on ne trouve que très peu de dossiers.

Même si les ces deux jeux de données conduisent aux mêmes conclusions concernant l'espace agricole, la densité de dossiers obtenue diffère fortement. Le RGP, du fait de son homogénéité et de son exhaustivité, aboutit vraisemblablement à des résultats plus réalistes. OSM présente en outre une hétérogénéité entre les espaces urbains très bien renseignés et les espaces ruraux où la qualité des données est proche de celle de Corinne Land Cover.

destination agricole	superficie (km²)	nombre de dossiers fonciers	densité de dossiers au km²
prairies permanentes et temporaires	10181	3556	0,35
cultures[38]	10141	3116	0,31
estives et landes	3827	235	0,06
vignes	339	153	0,45
vergers et fruits à coque	146	104	0,71

Tableau 13 : densité de dossiers fonciers selon la destination agricole

Sur l'ensemble des résultats, on constate l'influence de la valeur du parcellaire sur l'activité foncière du géomètre-expert. Le besoin de sécurisation foncière apportée par le bornage est plus présent sur des terrains à fort enjeu et les mutations parcellaires sont plus fréquentes au niveau des zones constructibles généralement situées à la frontière de l'espace urbain et rural.

III.3.3 Proximité géographique des cabinets et concurrence

L'objectif est d'observer les zones où l'activité se concentre et ainsi avoir un aperçu visuel de la pression concurrentielle entre cabinets. Pour cela, on estime la zone de chalandise de chaque cabinet à l'aide de la localisation des dossiers et des bureaux de géomètres-experts situés dans la région. Elle représente dans notre cas l'espace géographique où l'on trouve la moitié de la clientèle la plus proche de l'entreprise[39].

Ce type de traitement est plus pertinent sur un espace à l'échelle régionale même s'il exclut des calculs les versements extrarégionaux. La multitude des polygones générés au niveau national rendrait tout rendu cartographique illisible.

Afin de visualiser la superposition des zones de chalandises produites, celles-ci sont représentées à l'aide d'une fusion de type « produit ». Pour deux objets superposés, les valeurs des pixels des deux couches (comprises entre 0 et 255) sont multipliées entre elles et divisées par 255, le nombre de valeurs possibles. La couleur obtenue sera donc plus foncée lorsque plusieurs objets se superposent. Le rendu cartographique obtenu est disponible en annexe[40]. On constate tout d'abord que l'espace recouvert par les zones de chalandises considérées est loin de recouvrir l'ensemble du territoire. Globalement, on constate généralement des superpositions de plusieurs zones au cœur des principales villes de la région : Toulouse, Montauban, Rodez, Albi, Cahors

[37] Voir partie III.1.2
[38] Cette catégorie regroupe les cultures céréalières, le tournesol et le colza.
[39] Voir annexe 6 : Méthode de création des zones de chalandise
[40] Voir carte 10 : Zones de chalandise des bureaux de géomètres-experts en Midi-Pyrénées

notamment, l'agglomération toulousaine étant l'espace le plus concurrentiel suivi par Rodez. Le Gers fait exception : il y a une seule zone où des zones de chalandises se superposent à l'est du département, en lien avec la proximité de la métropole toulousaine.

En dehors des villes, le chevauchement des zones de chalandise n'existe pratiquement pas. Dans la région, la distance médiane entre le bureau du géomètre-expert et la localisation d'un dossier est de 10,4 kilomètres à vol d'oiseau. Neuf dossiers sur dix sont situés à moins de 26 kilomètres du bureau.

Cette cartographie n'est cependant pas complète car seuls les dossiers à versement obligatoire sont considérés. Elle s'appuie donc principalement sur les travaux fonciers et immobiliers et exclut notamment les travaux topographiques.

À travers l'atlas cartographique et les analyses réalisées dans cette partie, la répartition des géomètres-experts sur le territoire est inégale. La profession est davantage implantée dans les aires urbaines et le sud de la métropole. Les espaces ruraux, aux enjeux fonciers variables selon l'occupation du sol, ne concentrent qu'une faible partie de l'activité.

IV Génération dynamique de résultats statistiques à l'aide d'un rendu (carto-) graphique web

L'ensemble des études menées, des résultats numériques et graphiques obtenus s'appuient sur des données prises à un instant donné. Ils ont une fixité temporelle. Cette partie qui fait suite à l'analyse vise à mettre en place des outils facilitant la visualisation de rendus graphiques et cartographiques dynamiques s'adaptant à l'évolution des données sources.

De nombreux outils comme MapServer ou GeoServer permettent d'afficher des données cartographiques sur un site web. Nous allons cependant employer ici R et ses packages. Ce choix s'inscrit dans la continuité des travaux de recherche présentés au cours desquels ce logiciel a été fréquemment utilisé afin de gérer l'aspect graphique et statistique.

IV.1 Shiny, outil de visualisation d'analyses statistiques

IV.1.1 Principes de fonctionnement

Le package essentiel de R dans cette partie est Shiny (version 0.12). Celui-ci est destiné à la mise en place d'applications web de visualisation de graphes statistiques et de cartes. Très concis, il permet de créer des interfaces web à partir du langage R en s'appuyant sur des bibliothèques existantes adaptées à la création de sites web telles que Bootstrap pour la mise en place de widgets et le style de rendu ou DataTables pour la gestion de tables affichées sur une page web. Le package Shiny et ses extensions intègrent des fonctions R faisant appel directement aux fonctionnalités offertes par ces bibliothèques. Les fichiers CSS, HTML et JavaScript nécessaires sont préconfigurés et peuvent être paramétrés avec R.

La mise en place d'une application de partage d'une étude statistique nécessite l'écriture généralement de deux fichiers : *server.R* contenant les traitements exécutés côté serveur et *ui.R* dans lequel est configurée l'interface utilisateur, en plus des données éventuelles auxquelles l'application fait appel[41]. Ces deux scripts fonctionnent ensemble et peuvent être d'ailleurs assemblés dans un fichier unique. Cependant, on génère souvent deux fichiers afin de différencier les deux parties de l'application.

[41] Voir annexe 7 : Fonctionnement de Shiny pour la création d'applications web

Exécuté au moment du chargement de la page web, le fichier *ui.R* sert à définir les blocs qui constitueront l'interface (graphes, zones de texte, tables, panneaux d'options…) en utilisant soit du code R, soit la syntaxe des langages web (HTML, CSS et JavaScript). Des widgets permettant à l'utilisateur d'interagir avec la page web sont également paramétrables. Il est ainsi possible de créer un outil d'analyse statistique avancé disponible à partir d'un simple navigateur. Le package FactoShiny par exemple permet de réaliser les mêmes traitements que ceux effectués avec FactoMineR portant sur l'analyse de données multidimensionnelles[42] mais sur un navigateur avec une interface adaptée.

IV.1.2 Gestion des interactions de l'utilisateur à l'aide de la programmation réactive

On peut paramétrer avec Shiny des variables correspondant à une action de l'utilisateur au sein du navigateur web. Ces flux de données sont stockés dans l'objet appelé `input`. Il peut s'agir du fait de cocher une case, choisir une valeur dans une liste déroulante ou cliquer sur un élément par exemple. Ces variables « entrantes » sont généralement utilisées au sein d'expressions définissant des variables « sortantes » stockées dans l'objet `output` qui correspond à la réponse du serveur comme le calcul d'une valeur demandée ou la génération d'un rendu cartographique à afficher sur la page web.

Le style de programmation employé dans Shiny facilite la mise à jour au sein du code des variables dépendantes d'un flux de données (mouvement du curseur entraînant l'affichage d'une information par exemple).

Supposons que l'on affecte la valeur de la variable entrante A à la variable sortante B. En programmation réactive, les variables sont alors dépendantes : si *A* change de valeur, la variable *B* sera systématiquement mise à jour en conséquence. Il n'est pas nécessaire de préciser explicitement que *B* doit être recalculé, ce qui simplifie l'écriture du code.

Cette réactivité n'est cependant possible que dans certaines fonctions spécifiques à Shiny capables de gérer ces variables dites « réactives ». Ce sont principalement les fonctions produisant un rendu sur le navigateur qui ont cette capacité.

IV.2 Interface de visualisation graphique et cartographique

IV.2.1 Cartographie web à partir de leaflet

Leaflet est une bibliothèque JavaScript destinée à la création de cartes interactives sur site web. Un package du même nom a été développé afin de faciliter la création de rendus cartographiques sur des applications créées à l'aide de Shiny. Celui-ci fournit des fonctions issues de la bibliothèque éponyme permettant de configurer les cartes générées directement dans R. Bien qu'utilisable, il est encore en développement.

Il s'appuie sur la version 0.8 de Leaflet et permet de créer des cartes à partir de multiples sources de données : la latitude et la longitude de points renseignés dans un tableau de données, des géométries au format GeoJSON ou des classes d'objets spatiaux spécifiques au package sp.

[42] Voir l'ACP réalisée sur les types de dossiers de Géofoncier (partie II.2.2.3)

Cette dernière méthode est celle que nous avons privilégiée car il s'agit du format auquel sont généralement converties les données spatiales chargées dans un environnement R. C'est notamment ce que fait rgdal, le package utilisant GDAL et permettant la conversion d'une donnée spatiale en objet R et inversement.

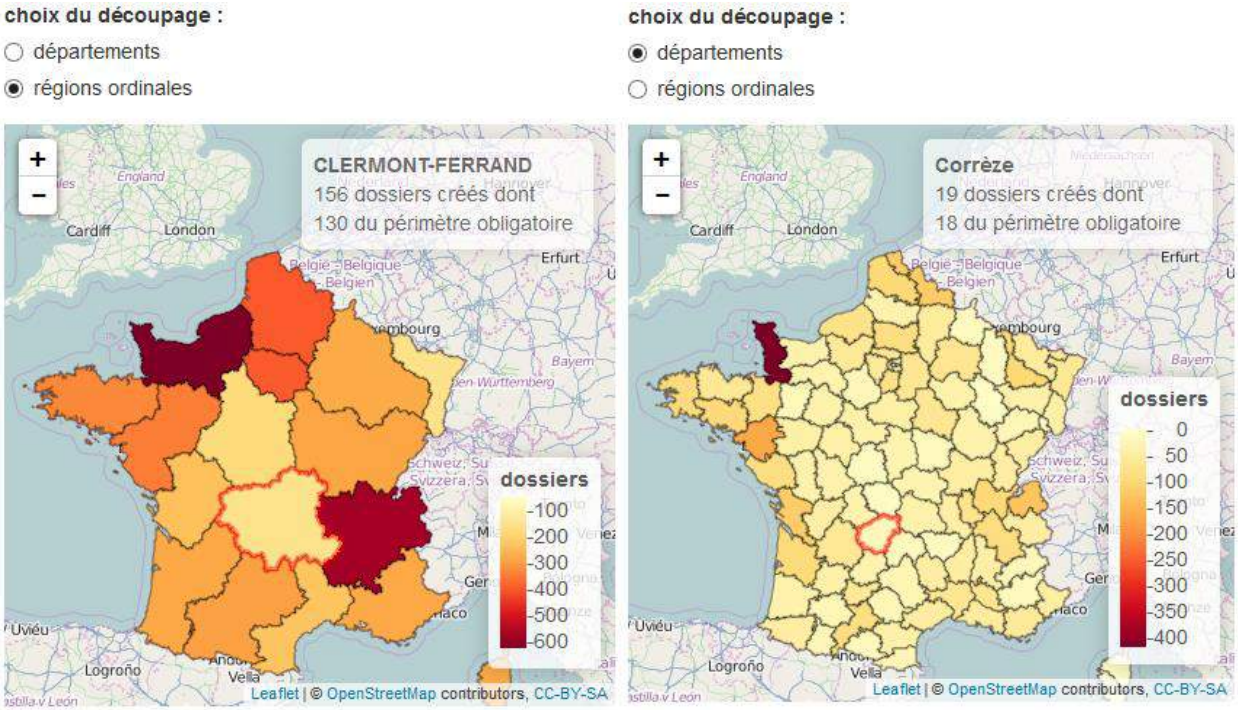


Figure 28 : visualisation sur navigateur de la création de dossiers sur Géofoncier selon un découpage par régions ordinales et départements

La figure 27 présente un exemple d'interface créé avec les outils précédemment cités[43]. Il s'agit d'une cartographie des dossiers présents sur Géofoncier et créés sur une période donnée. L'utilisateur peut sélectionner le découpage territorial à l'aide d'un widget composé de deux boutons radio au-dessus de la carte. Il est possible d'afficher des informations concernant une région ou un département en cliquant dessus (territoire affiché avec une bordure rouge).

[43] L'application est disponible à l'adresse suivante : http://91.206.198.152/datadep-app2/

IV.2.2 Création de graphiques

L'étude que nous avons menée a conduit à la création de cartes mais aussi à la production

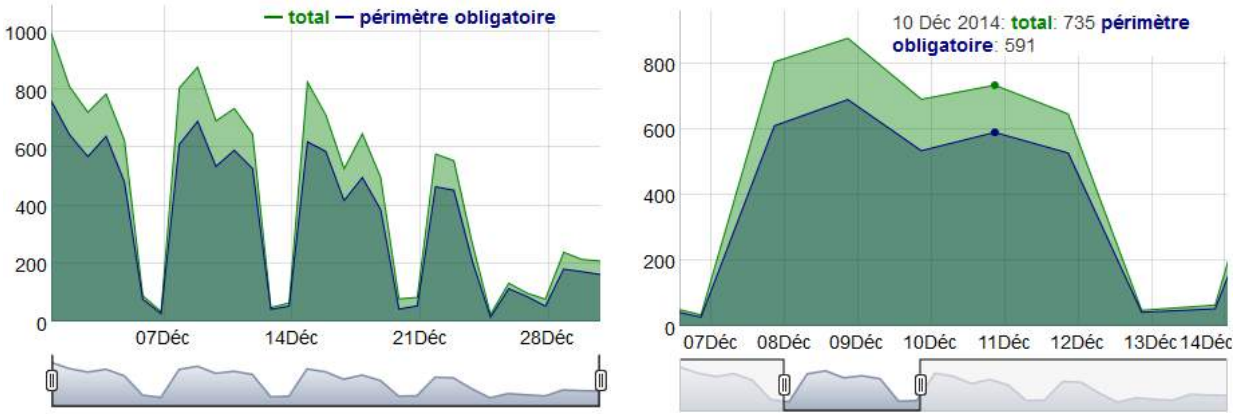


Figure 29 : visualisation sur navigateur du nombre de dossiers créés sur Géofoncier sur une période donnée

de graphes, notamment temporels. La visualisation de données au travers de graphiques est plus développée que la cartographie sous Shiny. Nous avons décidé d'utiliser le package dygraphs destinée à la création de graphiques dynamiques sous R. Plusieurs raisons justifient ce choix : celui-ci est bien documenté, les fonctions proposées suivent la même logique de programmation que celles du package leaflet et il est pensé pour être utilisé avec Shiny.

La figure 28 montre un exemple d'utilisation de ce package avec le tracé en fonction du temps du nombre de dossiers créés au total et appartenant au périmètre obligatoire. Il s'agit d'un graphique temporel sur lequel l'affichage évolue en fonction du positionnement du curseur sur le tracé. La barre horizontale en dessous du graphique permet de délimiter la période représentée. Le code source des éléments présentés sur les figures 27 et 28 est détaillé en annexe[44].

[44] Voir annexe 8 : Interface cartographique réalisée avec Shiny : exemple sur la visualisation des dossiers créés sur Géofoncier

Conclusion

L'ensemble des analyses mises en œuvre au cours de cette étude et détaillées dans ce mémoire ne permettent pas de dégager de conclusions catégoriques mais des tendances sur la profession de géomètre-expert. La fiabilité variable des données rend également plus complexe les interprétations.

Néanmoins, certains éléments sont significatifs de l'ensemble des études tant statistiques que cartographiques. Tout d'abord, la structuration de la profession est en mutation avec un fort rajeunissement constaté depuis vingt ans ainsi qu'un regroupement des géomètres-experts au sein de cabinets moins nombreux mais plus importants. Le nombre de cabinets individuels se réduit. On constate par ailleurs que les activités exercées par les cabinets sont liées à leur envergure. Certaines activités telles l'aménagement foncier, la topographie et l'ingénierie semblent caractériser ceux disposant d'importants effectifs alors que les plus modestes se tournent vers les activités en lien avec l'immobilier et l'urbanisme.

Au niveau géographique, l'implantation de la profession est concentrée dans les aires urbaines dynamiques avec une présence plus marquée dans la partie sud du territoire et le littoral maritime. Elle intervient très majoritairement sur les territoires à fort enjeu économiques et fonciers. C'est d'ailleurs ce que tend à confirmer l'exemple midi pyrénéen avec une superposition des zones d'activité des cabinets auprès des principales agglomérations régionales.

Pour compléter le travail de recherche effectué notamment en ce qui concerne l'évolution au cours du temps de la profession, il est nécessaire d'avoir des données disponibles avec un recul plus important et une fiabilité homogène sur la période étudiée. Reprendre l'étude dans quelques années permettrait d'avoir une visibilité plus détaillée sur les changements en cours. Il serait alors intéressant de disposer d'informations concernant l'effectif de chaque cabinet, indicateur plus pertinent que le nombre de géomètres-experts, et de simplifier la chaîne de traitement des données, souvent complexe à cause de leur volume. En outre, une autre perspective d'étude tournée vers les mutations foncières à l'échelle nationale serait envisageable, sous réserve d'une meilleure homogénéité spatiale concernant le versement d'objets RFU.

Cependant, toute analyse menée nécessite la mise en place de traitements qui doivent être reproduits à chaque fois. L'élaboration d'une solution web paramétrable accédant aux données et s'actualisant avec l'accroissement de la base permettrait un suivi continu. Cet aspect a été initié en dernière partie des travaux effectués mais peut être considérablement développé. Cela ouvre également la possibilité de véritablement valoriser les données auprès des utilisateurs de Géofoncier à commencer par les géomètres-experts. Leurs propres versements pourraient être utilisés afin de fournir des analyses concernant leur périmètre d'action ou l'évolution de l'activité du cabinet dans un espace personnalisé accessible uniquement par identifiant. Il y aurait alors un intérêt pour le professionnel à fournir des données plus exhaustives afin d'améliorer la qualité des analyses.

D'un point de vue technique, certaines solutions employées en lien avec R sont extrêmement récentes et encore en développement, notamment en ce qui concerne la création d'une interface cartographique. Elles sont vraisemblablement amenées à progresser et s'enrichir de nouvelles fonctionnalités. Shiny est un outil prometteur en termes d'application web à des fins de visualisation et d'analyses de données. Il répond parfaitement au besoin de mettre en place des éléments graphiques voir cartographiques en ligne sans avoir une forte connaissance des langages web traditionnels.

Enfin, je considère ce travail de fin d'études comme particulièrement formateur au niveau des méthodes et outils destinées à l'analyse de données, y compris géographiques. Cela a aussi été un projet de recherche où j'ai pu être autonome, expérimenter sur l'analyse, le traitement de gros

volumes de données et la cartographie. Ces compétences s'inscrivent dans l'objectif de m'investir professionnellement dans la géomatique. J'ai pu apprendre des méthodes statistiques d'analyse de données multidimensionnelles pouvant être employées en *data mining* et acquérir de l'expérience qui me permettra d'apprendre de façon autonome de nouveaux langages informatiques à l'avenir.

Bibliographie

Ouvrages imprimés

BERTIN J., 1999. Sémiologie graphique. Les diagrammes – Les réseaux – Les cartes. Éditions de l'École des Hautes Études en Sciences Sociales, Paris, 432 p.

CORNILLON P.-A., GUYARDER A., HUSSON F., JEGOU N., JOSSE J., KLOAREG M., MATZNER-LØBER E., 2012. Statistiques avec R. 3e édition revue et augmentée. Presses Universitaires de Rennes, Rennes, 296 p.

Groupe ElementR, 2014. R et espace. Traitement de l'information géographique. Framasoft, 244 p. Publié sous licence CC By-SA 3.0. Disponible également en ligne sur http://framabook.org/16-r-et-espace/ (consulté le 04/05/15).

LEBART L., MORINEAU A., PIRON M., 1997. Statistique exploratoire multidimensionnelle. Dunod, Paris, 439 p.

MASSER I., 2010. Building European Spatial Data Infrastructures. ESRI Press, Redlands (États-Unis), 95 p.

Publications en ligne

BACCAÏNI B., LEVY D. (juillet 2009). Recensement de la population de 2006 – Les migrations entre départements : le Sud et l'Ouest toujours très attractifs. *Insee Première* [en ligne]. n°1248. Disponible sur : http://www.insee.fr/fr/themes/document.asp?reg_id=0&ref_id=ip1248 (consulté le 08/04/15).

GASTNER M. T., NEWMAN M. E. J. (mai 2004). Diffusion-based method for producing density-equalizing maps. *PNAS* [en ligne]. **vol. 101** n°20. Disponible sur : http://www.pnas.org/content/101/20/7499 (consulté le 01/06/15).

LÉGER F., GILBERT A. (janvier 2014). Haute-Savoie : un des départements les plus dynamiques de France. *La Lettre - Analyses* [en ligne]. n°220. Disponible sur : http://insee.fr/fr/insee_regions/rhone-alpes/themes/syntheses/lettre_analyses/02220/02220_lettre_RP_Haute_Savoie.pdf (consulté le 22/05/15).

Pages, sites web et documents dématérialisés

Boston Geographic Information Systems. PostGIS Concave Hull, [en ligne]. Disponible sur : http://www.bostongis.com/postgis_concavehull.snippet (consulté le 24/04/15).

Copernicus. Copernicus Land Monitoring Services, [en ligne]. Disponible sur : http://land.copernicus.eu/pan-european/corine-land-cover (consulté le 28/04/15).

Géofoncier SAS, Documentation technique API Géofoncier v1.3, 2015.

INSEE. Les 60 premières aires urbaines en 2012, [en ligne]. Disponible sur : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=nattef01203&id=205 (consulté le 07/04/15).

INSEE. Définitions [en ligne]. Disponible sur : http://www.insee.fr/fr/methodes/default.asp?page=definitions/liste-definitions.htm (consulté le 06/05/15).

OGE, 2014. Recueil des prestations [en ligne]. Disponible sur : http://www.geometre-expert.fr/upload/docs/application/pdf/2014-01/recueilprestations2012_2.pdf (consulté le 18/05/15).

ORES Pays de la Loire. Marge brute d'exploitation, [en ligne]. Disponible sur : http://ores.paysdelaloire.fr/1056-marge-brute-d-exploitation.htm (consulté le 21/04/15).

RAKOTOMALALA R., 2015. Analyse de corrélation. Étude des dépendances - variables quantitatives [en ligne]. Disponible sur : http://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse_de_Correlation.pdf (consulté le 19/05/15).

Jeux de données employés accessibles en ligne

Agence de services et de paiement (ASP), 2014. Registre Parcellaire Graphique 2012: contours des îlots culturaux et leur groupe de cultures majoritaire des exploitations. Disponible sur https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-2012-contours-des-ilots-culturaux-et-leur-groupe-de-cultures-majorita/ (consulté le 28/04/15)

Conseil Régional de Basse-Normandie, 2013. Périmètre des aires urbaines. Disponible sur : http://www.geocatalogue.fr/Detail.do?id=157759 (consulté le 08/04/15)

Geofabrik, 2015. Données OSM sur la France. Disponible sur http://download.geofabrik.de/europe/france.html (consulté le 16/04/15)

INSEE, 2014. Démographie des entreprises 2006-2013. Disponible sur : http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=base-cc-demo-entreprises (consulté le 08/04/15)

Ministère de l'Écologie, du Développement Durable et de l'Énergie, 2012. Corine Land Cover 2006. Disponible sur http://www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/1825.html (consulté le 16/04/15)

OpenStreetMap, 2012. Limites administratives départementales. Disponible sur : http://openstreetmap.fr/contours-departements-et-regions (consulté le 08/04/15)

Societe.com, 2015. Données financières concernant le secteur d'activité des géomètres. Disponible sur http://fichier.societe.com/pages_html/fichier-entreprises.html (consulté le 10/04/15), données payantes.

Packages employés sur R

dygraphs : package complémentaire de Shiny destiné à la création de graphiques interactifs et basé sur la bibliothèque JavaScript éponyme (source : http://cran.r-project.org/web/packages/dygraphs/index.html et documentation : https://rstudio.github.io/dygraphs/).

FactoMineR : permet de réaliser des analyses factorielles sur R (source : http://cran.r-project.org/web/packages/FactoMineR/index.html, site dédié au package : http://cran.r-project.org/web/packages/FactoMineR/index.html).

Factoshiny : package complémentaire de FactoMineR qui permet d'utiliser une interface web pour la mise en œuvre des analyses factorielles (source : ftp://cran.r-project.org/pub/R/web/packages/Factoshiny/index.html).

ggplot2 : package offrant des options pour des graphiques avancés, notamment sous forme de grille (source : http://cran.r-project.org/web/packages/ggplot2/index.html, documentation : http://docs.ggplot2.org/current/).

leaflet : package complémentaire de Shiny, celui-ci permet de mettre en place une interface cartographique en s'appuyant sur la bibliothèque JavaScript éponyme (source : https://github.com/rstudio/leaflet et documentation : https://rstudio.github.io/leaflet/).

pyramid : fournit des fonctions adaptées à la réalisation de pyramides des âges (source : http://cran.r-project.org/web/packages/pyramid/index.html).

rCharts : package destiné à la création de graphes dynamiques en s'appuyant sur des bibliothèques JavaScript existantes (source : http://ramnathv.github.io/rCharts/).

reshape2 : package simplifiant la restructuration de tableaux de données (source : http://cran.r-project.org/web/packages/reshape2/index.html).

RPostgreSQL : permet de se connecter à des bases de données PostgreSQL et d'effectuer des requêtes SQL sur la base (source : http://cran.r-project.org/web/packages/RPostgreSQL/index.html).

Shiny : package assurant le développement d'applications web destinées à la visualisation de résultats d'analyses statistiques (source : http://cran.r-project.org/web/packages/shiny/index.html et documentation : http://shiny.rstudio.com/).

sp : package définissant des classes d'objets spatiaux permettant de gérer des géométries avec des données attributaires au sein d'un environnement R (source : http://cran.r-project.org/web/packages/sp/index.html).

Table des figures

Figure 1 : données produites par les géomètres-experts visibles sur le portail Géofoncier 11
Figure 2 : schéma des bases de données employées 12
Figure 3 : modèle relationnel type d'une des bases de données rassemblant les dossiers des géomètres-experts 13
Figure 4 : modèle conceptuel simplifié de la BDD *gespro* 14
Figure 5 : évolution quantitative de la profession de 1948 à 2015 (cabinets, bureaux et géomètres-experts) 18
Figure 6 : évolution quantitative de la profession de 1999 à 2015 (cabinets, bureaux et géomètres-experts) 18
Figure 7 : pyramides des âges de la profession 19
Figure 8 : diagrammes en barres sur l'évolution des cabinets de 2000 à 2015 21
Figure 9 : diagramme en barres des cabinets selon le nombre de géomètres-experts 22
Figure 10 : diagrammes en barres sur la situation financière des cabinets 24
Figure 11 : répartition des dossiers versés sur Géofoncier depuis juillet 2010 25
Figure 12 : évolution mensuelle de l'activité de la profession entre juillet 2010 et décembre 2014 26
Figure 13 : projection sur l'axe factoriel du point A_i correspondant à l'individu i 27
Figure 14 : graphe des individus (les cabinets de géomètres-experts) représentés dans le plan principal 28
Figure 15 : graphe des variables (les types de versements) dans le plan principal 29
Figure 16 : diagramme en barres des pourcentages d'inerties des axes de l'ACP 30
Figure 17 : cercle de corrélation de l'ACP 32
Figure 18: répartition du chiffre d'affaires par activité sur l'ensemble des cabinets étudiés en 2013 33
Figure 19 : agglomération progressive de cinq points 33
Figure 20 : gain d'inertie intra-classe selon le nombre de classes conservées pour la classification 34
Figure 21 : dendrogramme commenté de la CAH et représentant les classes sélectionnées 35
Figure 22 : graphe des individus représentés avec leur classe 36
Figure 23 : répartition du nombre de cabinets par classe 37
Figure 24 : légende du tableau comparatif 38
Figure 26 : répartition par région ordinale des classes de cabinets 38
Figure 25: répartition du chiffre d'affaires (total ou pour une activité particulière) par classe sur l'ensemble des cabinets étudiés 38
Figure 27 : localisation des bureaux de géomètres-experts en métropole 40
Figure 28 : visualisation sur navigateur de la création de dossiers sur Géofoncier selon un découpage par régions ordinales et départements 50
Figure 29 : visualisation sur navigateur du nombre de dossiers créés sur Géofoncier sur une période donnée 51

Liste des tableaux

Tableau 1 : matrice de corrélation des variables de l'étude 23
Tableau 2 : répartition des opérations obligatoires (période 2012-2014) 27
Tableau 3 : variables abrégées dans le cercle de corrélation 29
Tableau 4 : variables employées lors de la CAH 34
Tableau 5 : variables représentatives de la classe 1 (cabinets de petite taille) 36
Tableau 6 : caractéristiques des trois classes principales de cabinets 37
Tableau 7 : classement des dix départements comportant le plus de bureaux de géomètres-experts 41
Tableau 8 : classement des dix premiers départements en termes de bureaux par habitant 42
Tableau 9 : classement des dix départements comportant le plus d'établissements de la construction 43

Tableau 10 : classement des dix départements comportant le plus d'opérations foncières et immobilières du périmètre obligatoire sur la période 2012-2014..... 45
Tableau 11 : ensemble des occupations du sol sélectionnées sur la région Midi-Pyrénées 46
Tableau 12 : densité de dossiers en fonction de l'occupation du sol 46
Tableau 13 : densité de dossiers fonciers selon la destination agricole 47

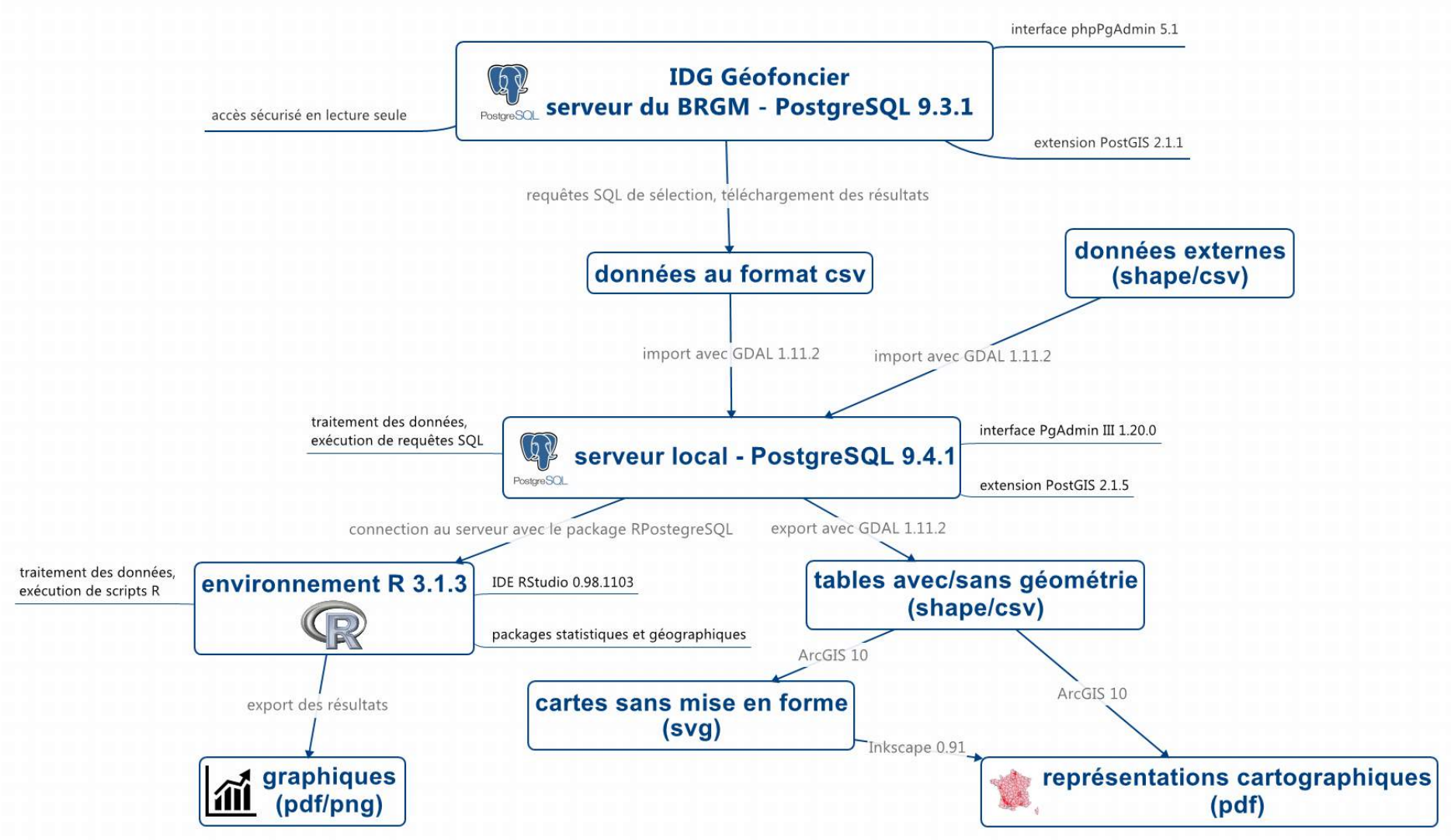
Table des annexes

Annexe 1 : Démarche générale de travail : de la donnée brute à l'analyse I
Annexe 2 : Connexion à une base de données PostgreSQL à partir de R à l'aide du package RPostgreSQL II
Annexe 3 : Méthode de jointure des données financières et des données issues de Géofoncier..... III
Annexe 4 : Critère de Ward en CAH VI
Annexe 5 : Problématique du volume des données : exemple de méthode de récupération des objets RFU par département sur l'ensemble du territoire VIII
Annexe 6 : Méthode de création des zones de chalandise..... XI
Annexe 7 : Fonctionnement de Shiny pour la création d'applications web..... XIII
Annexe 8 : Interface cartographique réalisée avec Shiny : exemple sur la visualisation des dossiers créés sur Géofoncier XIV
Annexe 9 : Résumé du mémoire XXII
Annexe 10 : Poster scientifique..... XXVI

Atlas cartographique

Carte 1 : Répartition des cabinets de géomètres-experts par classe et par région ordinale (année 2013)
Carte 2 : Localisation des bureaux de géomètres-experts en 2015 et influence des aires urbaines
Carte 3 : Répartition territoriale des bureaux de géomètres-experts en 2015
Carte 4 : Dynamique des bureaux de géomètres-experts (2000-2015)
Carte 5 : Répartition départementale des bureaux de géomètres-experts
Carte 6 : Les géomètres-experts et le secteur de la construction
Carte 7 : Versement d'objets RFU par département
Carte 8 : Densité de dossiers créés en 2014
Carte 9 : Activité foncière et immobilière - répartition départementale (2012-2014)
Carte 10 : Zones de chalandise des bureaux de géomètres-experts en Midi-Pyrénées

Annexe 1
Démarche générale de travail : de la donnée brute à l'analyse



Annexe 2
Connexion à une base de données PostgreSQL à partir de R à l'aide du package RPostgreSQL

Après avoir chargé le package RPostgreSQL, la connexion s'effectue en précisant les paramètres de celle-ci via la fonction `dbConnect` :

```
drv <- dbDriver("PostgreSQL")
con <- dbConnect(drv, host="localhost", port="5433",
dbname="etude_dossier", user="postgres",
                        password="postgres")
```

Il est ensuite possible d'exécuter des requêtes SQL via les commandes `dbGetQuery` ou `dbSendQuery`. La première fonction permet en plus de récupérer le résultat de la requête dans une variable sous forme de *data frame*, qui est l'équivalent sous R de la table d'une BDD.

```
requete2 <- "
SELECT no_cabinet, count(ogc_fid) AS nb_bureau
FROM bureau_corrige
GROUP BY no_cabinet;
"
nb_bur <- dbGetQuery(con, requete2)
```

Une fois les données récupérées, on peut se déconnecter avec la commande suivante ci-dessous :

```
dbDisconnect(con)
```

Annexe 3
Méthode de jointure des données financières et des données issues de Géofoncier

Plusieurs champs au niveau des données financières récupérées permettent une jointure avec les données de Géofoncier afin d'identifier les cabinets de géomètres-experts :

- le numéro SIREN, identifiant unique d'une entreprise, est présent dans les données au travers du numéro SIRET (identifiant unique d'un établissement) mais souvent non renseigné au sein des données de Géofoncier,
- la dénomination de l'entreprise correspond au nom du cabinet mais l'écriture au sein des deux bases est souvent différente,
- l'adresse est également disponible mais présente le même problème que le nom de l'entreprise (variations d'écriture entre les données)
- le nom, prénom et date de naissance du dirigeant, dont l'écriture ne varie pas quelles que soient les données.

Parmi ces possibilités de jointure, le numéro SIREN a été retenu en premier lieu. Il s'agit de la seule méthode permettant d'effectuer une jointure sans aucune erreur. Seulement, comme les données de Géofoncier concernant le numéro SIREN sont peu ou mal renseignées (on trouve en fait des numéros SIRET et SIREN mélangés dans un même attribut). Il est nécessaire de "nettoyer" cette donnée après l'avoir importée sous PostgreSQL. Le numéro SIRET est composé de quatorze chiffres, on ne conserve que les neuf premiers correspondant au numéro SIREN :

```
UPDATE cabinet
SET no_siren = REPLACE(no_siren, ' ', '');
UPDATE cabinet
SET no_siren = SUBSTRING(no_siren FROM 1 FOR 9);
```

Cette méthode de jointure limite le nombre de cabinets identifiés. Afin d'obtenir un maximum d'individus pour l'étude statistique, une seconde méthode de jointure est employée en utilisant les noms, prénoms et dates de naissance des dirigeants d'entreprise. On joint ces données avec celles correspondant aux géomètres-experts. Ce type de jointure présente un risque d'erreur. Afin de le réduire, on double la jointure d'un "contrôle" au niveau des dénominations des entreprises. Pour cela on utilise deux extensions de PostgreSQL :

```
CREATE EXTENSION unaccent;
CREATE EXTENSION pg_trgm;
```

La première extension permet de supprimer les accents d'un champ de texte et la seconde fournit une fonction `similarity` permettant d'évaluer la similarité de deux textes en utilisant les trigrammes ou ensemble consécutifs de trois caractères. Deux textes identiques renverront une valeur de 1, deux textes sans trigramme commun renverront une valeur de 0.

Prenons l'exemple des deux abréviations IGN et SIG. Les trigrammes correspondants sont les suivants :

- IGN : " I", " IG", "IGN" et "GN ",
- SIG : " S", " SI", "SIG" et "IG ".

La fonction prend en compte deux espaces en préfixe et un espace en suffixe du texte traité. Il n'y a ici aucun trigramme commun entre les deux termes, la fonction renverra donc une valeur de 0.

Celle-ci n'est pas sensible à la casse mais tient compte des accents, c'est pourquoi ceux-ci sont supprimés avec la fonction `unaccent`. On assure ainsi une identification correcte de "GEOMETRE" avec "géomètre" pour lesquels on obtient une valeur égale à 1 avec la fonction `similarity`.

```
--création de la table joignant données financières et données issues de
Géofoncier
CREATE TABLE donnees_financieres_ge AS (
--sélection rassemblant les GE en activité
WITH ge_actuel AS (
	SELECT
		a.id_ge,
		nom_naissance,
		prenom,
		date_naissance,
		c.id_cabinet,
		no_cabinet,
		nom,
		type_cabinet
	FROM
		ge AS a
		INNER JOIN
		cabinet_ge AS b
		ON a.id_ge = b.id_ge
		INNER JOIN
		cabinet AS c
		ON b.id_cabinet = c.id_cabinet
	WHERE
		date_fin IS NULL --date_fin correspond à la date de fin
d'exercice d'un GE dans un cabinet
)
--jointure sur le numéro siren présent sur Géofoncier et celui issu des
données financières (méthode 1)
SELECT
	id_cabinet,
	no_cabinet,
	nom,
	type_cabinet,
	capital,
	"chiffre d_affaire"::integer AS ca,
	"résultat net"::integer AS resultat,
	"excédent brut d_exploitation"::integer AS ebe,
	"tranche effectif entreprise" AS eff
FROM
	cabinet,
	donnees_financieres
WHERE no_siren = siren
UNION
--jointure entre la liste des GE et les dirigeants d'entreprises sur le
noms, prénoms et dates de naissances (méthode 2)
SELECT
	id_cabinet,
	no_cabinet,
	nom,
	type_cabinet,
	capital,
	"chiffre d_affaire"::integer AS ca,
	"résultat net"::integer AS resultat,
```

```
        "excédent brut d_exploitation"::integer AS ebe,
        "tranche effectif entreprise" AS eff
FROM
        ge_actuel AS a
        INNER JOIN donnees_financieres
        ON
                (nom_naissance, prenom, date_naissance::date) =
                ("nom dirigeant", "prénom dirigeant", "date naissance
dirigeant"::date)
--restriction des données sélectionnées sur la similarité entre
--le nom de l'entreprise sur Géofoncier et sa dénomination dans les données
financières,
--choix du seuil de façon empirique
WHERE similarity(unaccent(nom)::text, "dénomination"::text) > 0.1
);
```

Le seuil fixé sur la similarité du nom de l'entreprise est fixé à 0,1 afin d'éviter la jointure d'entreprises avec des noms vraiment différents.

Annexe 4
Critère de Ward en CAH

À chaque itération de la classification, le critère de Ward permet de sélectionner le regroupement pour lequel on maximise l'inertie inter-classe tout en minimisant l'inertie intra-classe. Celui-ci s'appuie sur le fait que l'inertie totale est égale à la somme de l'inertie inter-classe et de l'inertie intra-classe. La démonstration de cette égalité s'appuie sur le théorème de Huygens énoncé ci-dessous.

Pour un ensemble de points $A_1, \dots, A_N$ de pondérations respectives $p_1, \dots, p_N$ et de centre de gravité G, on a la relation suivante pour tout point A :

$$\sum_{i=1}^{N} p_i {AA_i}^2 = \sum_{i=1}^{N} p_i {GA_i}^2 + \left(\sum_{i=1}^{N} p_i\right) GA^2$$

On définit tout d'abord les différentes inerties. On considère N, le nombre total d'individus, Q le nombre de classes et K le nombre de variables. N_q est le nombre d'individus de la classe q.

Soit $A_{iq} = \begin{pmatrix} x_{iq1} \\ \vdots \\ x_{iqk} \\ \vdots \\ x_{iqK} \end{pmatrix}$, le i-ème individu de la classe q. x_{iqk} est la valeur prise par l'individu i de la classe q pour la variable k.

Soit $G = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_k \\ \vdots \\ \bar{x}_K \end{pmatrix}$, le centre de gravité du nuage de points. $\bar{x}_k$ est la moyenne des valeurs prises par l'ensemble des individus pour la variable k :

$$\bar{x}_k = \frac{1}{N} \sum_{q=1}^{Q} \sum_{i=1}^{N_q} x_{iqk}$$

$G_q = \begin{pmatrix} \bar{x}_{q1} \\ \vdots \\ \bar{x}_{qk} \\ \vdots \\ \bar{x}_{qK} \end{pmatrix}$, le centre de gravité des individus de la classe q. $\bar{x}_{qk}$ est la moyenne des valeurs prises par les individus de la classe q pour la variable k :

$$\bar{x}_{qk} = \frac{1}{N_q} \sum_{i=1}^{N_q} x_{iqk}$$

On définit l'inertie totale par :

$$I_t = \frac{1}{N} \sum_{q=1}^{Q} \sum_{i=1}^{N_q} {GA_{iq}}^2 = \frac{1}{N} \sum_{k=1}^{K} \sum_{q=1}^{Q} \sum_{i=1}^{N_q} (x_{iqk} - \bar{x}_k)^2$$

De même, l'inertie intra-classe se définit par :

$$I_a = \frac{1}{N}\sum_{q=1}^{Q}\sum_{i=1}^{N_q} G_q {A_{iq}}^2 = \frac{1}{N}\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{N_q}(x_{iqk} - \bar{x}_{qk})^2$$

Enfin, l'inertie inter-classe est définie comme suit :

$$I_e = \frac{1}{N}\sum_{q=1}^{Q} N_q \times G_q G^2 = \frac{1}{N}\sum_{k=1}^{K}\sum_{q=1}^{Q} N_q(\bar{x}_{qk} - \bar{x}_k)^2$$

D'après le théorème de Huygens, on a la relation suivante pour les individus d'une classe q :

$$\sum_{i=1}^{N_q} {GA_{iq}}^2 = \sum_{i=1}^{N_q} G_q {A_{iq}}^2 + N_q \times G_q G^2$$

En sommant sur l'ensemble des classes, on obtient donc :

$$\sum_{q=1}^{Q}\sum_{i=1}^{N_q} {GA_{iq}}^2 = \sum_{q=1}^{Q}\sum_{i=1}^{N_q} G_q {A_{iq}}^2 + \sum_{q=1}^{Q} N_q \times G_q G^2$$

D'où :

$$I_t = I_a + I_e$$

C'est pourquoi maximiser l'inertie inter-classe revient à minimiser l'inertie intra-classe.

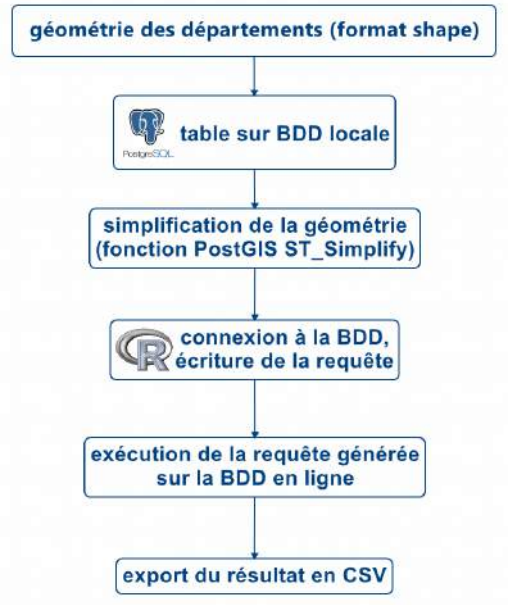
Annexe 5
Problématique du volume des données : exemple de méthode de récupération des objets RFU par département sur l'ensemble du territoire

Les objets RFU sont stockés au sein d'une base de données qui leur est consacrée sur le serveur de Géofoncier. Cette BDD ne comporte aucune donnée d'ordre attributaire (un code INSEE par exemple) permettant d'identifier simplement le département où se situe un objet RFU. Il est donc nécessaire d'utiliser la géométrie des objets pour comptabiliser le nombre d'entités RFU présentes dans un département. Or, cette BDD ne comporte aucune géométrie des départements et il n'est pas possible de créer ou modifier une table sur le serveur en ligne.

La première solution, presque instinctive, est l'extraction de l'ensemble des objets RFU de la BDD pour réaliser des traitements en local. Seulement, la BDD comporte plus de 5 millions d'entités. Cela ne rend pas impossible l'utilisation de cette méthode mais implique un temps d'import sur la base de données locale et un temps de traitement des géométries extrêmement long, d'autant plus sur un ordinateur de bureau.

Une approche différente est de réaliser le traitement directement sur les données du serveur en ligne lorsque cela s'avère possible. Cela présente plusieurs avantages :

- aucun import volumineux de données sur un disque local,
- gain de temps en exploitant directement les capacités du serveur, réduction de la chaîne de traitements,
- itération du processus facilitée.



Démarche générale employée

L'inconvénient principal est d'écrire une requête souvent plus complexe et plus longue. Dans le cas présent, il est nécessaire d'avoir dans la requête envoyée au serveur la donnée concernant la géométrie de chaque département français. Cette donnée doit être écrite en WKT afin d'être correctement interprétée.

La méthode employée a donc été de s'appuyer sur une base de données locale comportant la géométrie des départements, celle-ci pouvant être importée soit par l'utilitaire GDAL, soit par le plugin *PostGIS Shapefile and DBF loader* de PgAdmin directement.

1) requêtes SQL, simplification des géométries départementales :

```
--création d'une table comportant la géométrie simplifiée des départements
après conversion en coordonnées géographiques du système de référence RGF93
CREATE TABLE france_dep_simp AS (
       SELECT
              nom,
              numero,
              ST_SimplifyPreserveTopology(ST_Transform(geom, 4171), 0.01)
AS geom
       FROM france_departements
);

--création d'un index sur la nouvelle table
CREATE INDEX france_dep_simp_geom_idx
  ON france_dep_simp
  USING gist
        (geom);
```

2) script R, écriture de la requête à envoyer au serveur (ici, exemple concernant les sommets RFU) :

```
#script de génération d'une requête SQL comptabilisant, pour chaque
département :
#- le nombre de sommets RFU
#- le nombre de sommets créés avant 2012
#- le nombre de sommets créés en 2012, 2013, 2014

#suppression des variables enregistrées
rm(list=ls())

#nettoyage de la console (ou ctrl + L)
cat("\014")

#fermeture des graphiques
dev.off()

#encodage des caractères
options(encoding = "UTF-8")

#import package
library(RPostgreSQL)

#connection à la base de données
drv <- dbDriver("PostgreSQL")
con <- dbConnect(drv, host="localhost", port="5433", dbname="etude_rfu",
user="postgres",
                 password="postgres")

db <- dbGetQuery(con, "SELECT numero, nom, ST_AsText(geom) AS geom
                 FROM france_dep_simp
                 WHERE char_length(numero) = 2
                 ORDER BY numero;")

#début de la requête
sql <- "WITH commune AS ("

#génération de la table temporaire comportant les départements avec leur
géométrie
```

```
for (i in 1:(nrow(db)-1)) {
  sql <- paste0(
    sql, "
    SELECT '", db[i, 1], "' as insee, st_geomfromtext('", db[i, 3], "',
4171) as comgeom
    UNION")
}

sql <- paste0(
  sql, "
    SELECT '", db[nrow(db), 1], "' as insee, st_geomfromtext('",
db[nrow(db), 3], "', 4171) as comgeom
)
SELECT
  insee,
  count(node_id) AS nb_sommet,
  sum(CASE WHEN date_part('year', som_date_creation::date) < 2012 THEN 1
ELSE 0 END) AS nb_sommet_av_2012,
  sum(CASE WHEN date_part('year', som_date_creation::date) = 2012 THEN 1
ELSE 0 END) AS nb_sommet_2012,
  sum(CASE WHEN date_part('year', som_date_creation::date) = 2013 THEN 1
ELSE 0 END) AS nb_sommet_2013,
  sum(CASE WHEN date_part('year', som_date_creation::date) = 2014 THEN 1
ELSE 0 END) AS nb_sommet_2014
FROM
  topology_metropole.sommets_fonciers_courants,
       commune
WHERE
  ST_Intersects(node_geom, comgeom)
GROUP BY insee
ORDER BY insee
;")

#export de la requête
cat(sql, file = "export_de_R/stats_RFU_sommet.sql")

#déconnection de la BDD
      dbDisconnect(con)
```

On obtient alors un fichier stats_RFU_sommet.sql qu'il suffit de copier dans l'espace dédié à l'écriture de requêtes sur le serveur de Géofoncier. Le script permet d'automatiser la requête et notamment la partie correspondant à l'ensemble des géométries des départements qui sont écrites une à une en WKT en s'appuyant sur la table *france_dep_simp* de la BDD locale.

La seule limitation de cette méthode est la longueur de la requête. Lors de son exécution, elle permet d'afficher normalement le résultat mais ne permet pas de télécharger celui-ci. Ceci est une limitation du langage php sur lequel s'appuie phpPgAdmin en termes de longueur de nombre de caractères d'une requête. Dans le cas présent, il faut donc copier les résultats affichés dans le navigateur directement sur un fichier excel. Il suffit ensuite d'utiliser GDAL pour intégrer ceux-ci dans une BDD et les exploiter pour réaliser des traitements statistiques ou des rendus cartographiques.

Annexe 6
Méthode de création des zones de chalandise

Les zones de chalandise sont générées à partir d'une table comportant la localisation des dossiers sur Géofoncier (nommée ici *localisants*), table différente de celle des enregistrements. Au vue de la structure des données sur Géofoncier, un dossier peut correspondre à plusieurs localisations (dans le cadre d'un aménagement foncier par exemple).

En s'appuyant sur les données de positionnement des bureaux, on estime dans un premier temps la distance du dossier au bureau le plus proche appartenant au cabinet ayant réalisé les travaux. Les bureaux principaux et secondaires sont donc pris en compte.

```
--ajout d'une colonne indiquant la distance du dossier au bureau
ALTER TABLE localisants
ADD COLUMN distance integer;

WITH t1 AS (
SELECT
       enr_cab_createur,
       enr_ref_dossier,
       min(ST_Distance(a.wkb_geometry::geography,
b.wkb_geometry::geography, true)) AS distance
FROM
       localisants AS a
       INNER JOIN
       bureau AS b
       ON enr_cab_createur = no_cabinet
GROUP BY
       enr_cab_createur,
       enr_ref_dossier
)
UPDATE localisants AS a
SET distance = b.distance
FROM t1 AS b
WHERE   (a.enr_cab_createur,   a.enr_ref_dossier)   =   (b.enr_cab_createur,
b.enr_ref_dossier);
```

Dans l'analyse menée, un calcul de médiane est réalisé sur les distances. PostgreSQL ne comporte aucune fonction d'agrégation permettant d'avoir facilement la médiane, même si le calcul de celle-ci est souvent plus pertinent que celui de la moyenne pour de grands jeux de données comme c'est le cas ici. La médiane est en effet insensible aux valeurs extrêmes, au contraire de la moyenne.

Son calcul s'appuie sur les fonctions de fenêtrage permettant d'ordonner des lignes d'une table et de les positionner relativement les unes par rapport aux autres. Dans le cas suivant, la commande `ntile(10) OVER (ORDER BY distance ASC)` renvoie le décile de la ligne selon un classement basé sur la distance. Il faut ensuite générer la distance minimale et maximale de chaque décile. Pour avoir la médiane notamment, il suffit de calculer la moyenne du maximum du cinquième décile et du minimum du sixième décile.

```
--calcul des déciles sur la distance bureau/dossier
WITH t1 AS (
SELECT
       id_bureau,
       distance,
       ntile(10) OVER (ORDER BY distance ASC) AS decile
```

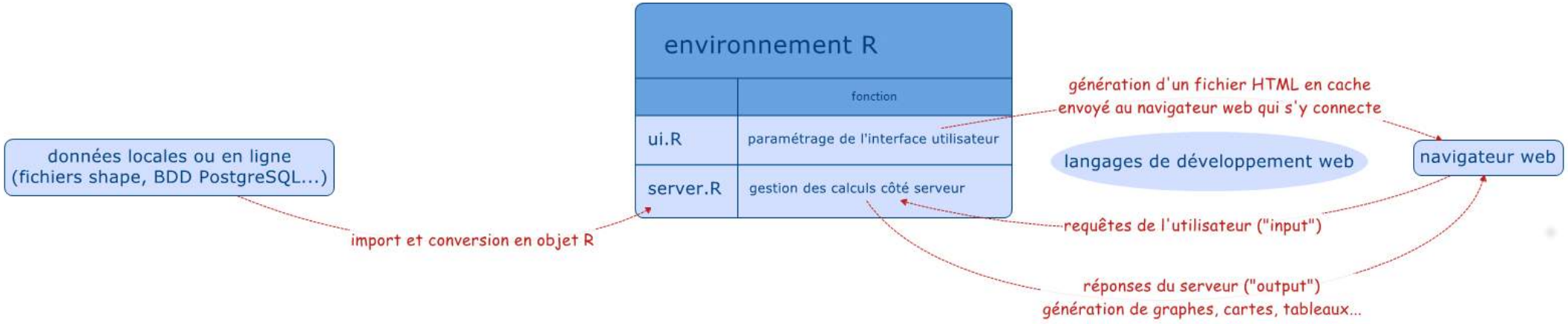
```
FROM localisants
WHERE
        distance IS NOT NULL
        AND id_bureau IS NOT NULL
)
SELECT
        part,
        min(distance),
        max(distance)
FROM t1
GROUP BY decile
ORDER BY decile;
```

Les mêmes fonctions sont employées pour ne conserver que la moitié des dossiers les plus proches des bureaux lorsqu'il s'agit de générer le polygone de la zone de chalandise. Celui-ci est obtenu à l'aide de la fonction PostGIS `ST_ConcaveHull` générant un polygone concave englobant l'ensemble des géométries rassemblées. Celles-ci étant initialement en WGS 84 et créant des polygones aberrants, elles ont été transformées en RGF 93 (projection Lambert 93) au préalable avec la fonction `ST_Transform`.

Pour plus de détails sur les mécanismes de cette fonction, un lien est disponible en bibliographie vers le site *Boston Geographic Information Systems* où des exemples d'utilisation illustrés sont présentés.

```
--création des zones de chalandise : polygone concave incluant 50% des
localisants les plus proches
DROP TABLE IF EXISTS zone_chalandise;
CREATE TABLE zone_chalandise AS (
WITH t1 AS (
SELECT
        id_bureau,
        distance,
        ntile(2) OVER (PARTITION BY id_bureau ORDER BY distance ASC) AS
part,
        ST_Transform(wkb_geometry, 2154) AS wkb_geometry
FROM localisants
WHERE
        distance IS NOT NULL
        AND id_bureau IS NOT NULL
),
t2 AS (
SELECT
        id_bureau,
        count(*) AS nb_dossier,
        ST_Transform(ST_ConcaveHull(ST_Collect(t1.wkb_geometry),0.99, true),
4326) AS geom
FROM  t1
WHERE part = 1
GROUP BY id_bureau
)
SELECT
        a.*,
        no_cabinet,
        nom
FROM
        t2 AS a
        INNER JOIN
        bureau AS b
        ON a.id_bureau = b.id_bureau
);
```

Annexe 7
Fonctionnement de Shiny pour la création d'applications web



Annexe 8
Interface cartographique réalisée avec Shiny : exemple sur la visualisation des dossiers créés sur Géofoncier

Les deux fichiers essentiels pour la mise en place de l'interface (disponible à l'adresse : http://91.206.198.152/datadep-app2/) sont présentés et décrits ici. Voici tout d'abord le fichier de configuration du serveur : server.R.

La première partie du fichier correspond à ce qui est chargé au moment du lancement de l'application. Ce sont des packages nécessaires, des fonctions créées pour simplifier le code par la suite et des données importées.

```
#server.R
#chargement des packages nécessaires
require(shiny)
require(rgdal)
require(leaflet)
require(RPostgreSQL)
require(dygraphs)

#import des géométries des départements et des régions ordinales
#encodage adapté pour Windows
fds <- readOGR("data", layer = "france_dep_simp", encoding = "ISO8859-1")
oge <- readOGR("data", layer = "reg_oge_wgs84")

#vecteur contenant les noms des schémas à partir desquels les données des
DOM sont extraites
domList <- c("reunion", "guyane", "antilles", "mayotte")

#fonction d'import des données pour le graphe temporel et formatage de
celles-ci
getTimeSerie <- function(start = "2014-12-01", end = "2014-12-31", dom =
domList) {

  #connection à la BDD
  drv <- dbDriver("PostgreSQL")
  con <- dbConnect(drv, host="localhost", port="5432", dbname="dossiers",
user="postgres",
                   password="postgres")

  #écriture de la requête
  req <- paste0("
  WITH t1 AS (
  SELECT enr_date_dossier::date, COUNT(*) AS nb_dossier,
    sum(CASE WHEN enr_perimetre_obligatoire = 't' THEN 1 ELSE 0 END) AS
nb_peri_obli
  FROM metropole.enregistrements
  WHERE enr_date_dossier::date BETWEEN '", start, "' AND '", end, "'
  GROUP BY 1
  ")

  #boucle for pour intégrer les DOM à la requête
  for (i in dom) {
    req <- paste0(req,
    " UNION
```

```
    SELECT enr_date_dossier::date, COUNT(*) AS nb_dossier,
      sum(CASE WHEN enr_perimetre_obligatoire = 't' THEN 1 ELSE 0 END) AS
nb_peri_obli
    FROM ", i, ".enregistrements
    WHERE enr_date_dossier::date BETWEEN '", start, "' AND '", end, "'
    GROUP BY 1")
  }

  req <- paste0(req, ")
    SELECT enr_date_dossier, sum(nb_dossier) AS nb_dossier,
      sum(nb_peri_obli) AS nb_peri_obli
    FROM t1
    GROUP BY 1;")

  #envoi de la requête et récupération du résultat
  result <- dbGetQuery(con, req)

  #déconnexion de la BDD
  dbDisconnect(con)

  #date en nom de ligne nécessaire pour être compatible au format xts
utilisé ensuite
  row.names(result) <- result$enr_date_dossier

  #renvoie le résultat en excluant le champ des dates, désormais renseigné
dans le nom de la ligne
  return(result[, -1])

}


#fonction d'import des données attributaires pour les départements et les
régions ordinales
getData <- function(lim = "dep", start = "2014-12-01", end = "2014-12-31",
dom = domList) {

  #connexion
  drv <- dbDriver("PostgreSQL")
  con <- dbConnect(drv, host="localhost", port="5432", dbname="dossiers",
user="postgres",
                   password="postgres")

  #choix entre département et région ordinale
  col <- ifelse(lim == "dep",
                "SUBSTRING(enr_code_insee FROM 1 FOR 2)",
                "enr_region_ordinale")

  #choix entre département et région ordinale pour les DOM
  colDom <- ifelse(lim == "dep",
                   "SUBSTRING(enr_code_insee FROM 1 FOR 3)",
                   "enr_region_ordinale")

  #écriture de la requête
  req <- paste0("
  WITH t1 AS (
    SELECT ", col, " AS num, COUNT(*) AS nb_dossier,
      sum(CASE WHEN enr_perimetre_obligatoire = 't' THEN 1 ELSE 0 END) AS
nb_peri_obli
    FROM metropole.enregistrements
    WHERE enr_date_dossier::date BETWEEN '", start, "' AND '", end, "'
```

```
    GROUP BY 1")

  #boucle for pour intégrer les DOM à la requête
  for (i in dom) {
    req <- paste0(req,
      " UNION
      SELECT ", colDom, " AS num, COUNT(*) AS nb_dossier,
        sum(CASE WHEN enr_perimetre_obligatoire = 't' THEN 1 ELSE 0 END) AS
nb_peri_obli
      FROM ", i, ".enregistrements
      WHERE enr_date_dossier::date BETWEEN '", start, "' AND '", end, "'
      GROUP BY 1")
  }

  req <- paste0(req, ")
    SELECT num, sum(nb_dossier) AS nb_dossier,
      sum(nb_peri_obli) AS nb_peri_obli
    FROM t1
    GROUP BY 1;")

  #envoi de la requête et récupération du résultat
  result <- dbGetQuery(con, req)

  #déconnexion
  dbDisconnect(con)

  return(result)
}
```

La partie suivante correspond à ce qui est lancé à chaque nouvelle session côté serveur, c'est-à-dire chaque fois qu'un nouvel utilisateur ouvre la page web. Cette partie est définie avec la fonction `shinyServer`. On y trouve ici les éléments intégrés dans la variable `output` et des évènements s'exécutant lorsque certaines variables en entrée prennent une nouvelle valeur.

L'utilisation du symbole `%>%` est une méthode d'écriture plus synthétique du code et issue à l'origine du package *magrittr*. Il permet d'enchaîner les fonctions avec une syntaxe lisible sans créer des variables intermédiaires. Le symbole sert à transférer le résultat d'une fonction à la suivante qui va l'utiliser comme premier argument : `addTiles` utilise en premier argument `leaflet()`, `setView` utilise en premier argument `addTiles()`. Les deux écritures suivantes sont équivalentes pour la production du fond de carte par exemple :

```
  output$map <- renderLeaflet({
    leaflet() %>%
      addTiles() %>%
      setView(0, 46, zoom = 5)
  })


  output$map <- renderLeaflet({
        myMap <- leaflet()
        myMapOSM <- addTiles(myMap)
        myMapConfigured <- setView(myMapOSM, 0, 46, zoom = 5)
        myMapConfigured
            })
```

Définition du serveur commentée:

```
shinyServer(function(input, output, session) {

  #définition de la carte leaflet
  #seul le fond OSM est ici défini (addTiles) avec des options de
positionnement de la vue (setView)
  output$map <- renderLeaflet({
    leaflet() %>%
      addTiles() %>%
      setView(0, 46, zoom = 5)
  })

  #définition du graphe
  output$graph <- renderDygraph({
    dygraph(getTimeSerie()) %>%
      dySeries("nb_dossier", label = "total") %>%
      dySeries("nb_peri_obli", label = "périmètre obligatoire") %>%
      dyOptions(fillGraph = TRUE, fillAlpha = 0.4) %>%
      dyLegend(show = "auto") %>%
      dyRangeSelector(dateWindow = c(), keepMouseZoom = FALSE)

  })

  #création de variables réactives :
  #attention : LastId est ici un environnement adapté aux variables
réctives
  #LastId$x est la variable définie à l'intérieur de cet environnement
  #elle servira à stocker par la suite le numéro du département/de la
région cliquée
  LastId <- reactiveValues(x = list())

  #compteur d'evènement
  count <- reactiveValues(x = 0)

  #variable correspondant à l'heure du système
  timer <- reactiveValues(x = Sys.Date())


  #évènement s'activant lorsque le choix du découpage (input$lim) ou de la
période (input$graph_date_window) évolue
  #il permet de redessiner la carte lorsque nécessaire
  #l'event s'active deux fois juste après le chargement de la page (une
fois pour chacune des variables qui le déclenchent)
  observeEvent(c(input$lim, input$graph_date_window), {

    #arrête l'event si celui-ci a été exécuté il y a moins de 4 sec
    #évite de recharger trop souvent la carte lorsqu'on bouge le slider
    #la deuxième condition "count$x > 2" assure que le chargement initial
se fasse correctement,
    #car celui-ci déclenche 2 fois l'event au départ
    if (difftime(Sys.time(), timer$x, units = "secs") < 4 & count$x > 2)
      return()

    #incrémentation du nombre d'events déclenchés
    count$x <- count$x + 1

    #reset du timer
    timer$x <- Sys.time()
```

```
    #arrêt si NULL, évite les plantages
    if (is.null(input$graph_date_window))
      return()

    #arrêt si période inférieure à 1 jour, évite certains plantages apparus
lorsqu'on zoom à l'excès
    #sur une période trop courte avec le graphe
    if (difftime(input$graph_date_window[2],
as.Date(input$graph_date_window[1]), units = "days") < 1)
      return()

    #si l'on choisit les régions ordinales:
    if (input$lim == "reg") {

      #récupération des géométries
      geom <<- oge

      #récupération des données
      dbData <<- getData("reg", start = input$graph_date_window[1], end =
input$graph_date_window[2])

    #sinon, on choisit nécessairement les départements :
    } else {

      #récupération des géométries
      geom <<- fds

      #récupération des données
      dbData <<- getData("dep", start = input$graph_date_window[1], end =
input$graph_date_window[2])
    }

    #assemblage données et géométries
    geom@data <<- data.frame(geom@data, dbData[match(geom@data[, 'numero'],
dbData[, 'num']), ])

    #conversion des NA en 0
    geom@data[is.na(geom@data)] <<- 0

    #ajustement de la palette de couleurs
    pal <<- colorNumeric("YlOrRd", geom$nb_dossier)

       #reset de la variable LastId$x
    LastId$x <- list()

    #tracé des polygones après purge des précédents (clearShapes)
    leafletProxy("map", data = geom) %>%
      clearShapes() %>%
      addPolygons(layerId = geom$numero,
                  stroke = TRUE, color = "Black", smoothFactor = 0.2,
                  fillOpacity = 1, fillColor = ~pal(nb_dossier),
                  weight = 1) %>%
      addLegend(layerId = "legend","bottomright", pal = pal, values =
~nb_dossier,
                opacity = 1, title = "dossiers",
                labFormat = labelFormat(suffix = ""))
  })


  #évènement s'activant lorsque l'on clique sur une géométrie de la carte
```

```
  #"map" renvoie ici à l'élément sortant output$map
  observeEvent(input$map_shape_click, priority = -1, {

  #arrête l'event le temps que la carte des régions/départements se
recharge
  #délai de 4 s en principe suffisant pour laisser le navigateur charger la
carte
  #cette "sécurité" évite un plantage pouvant apparaître lorsque l'on
clique "excessivement" au niveau de la carte
  #avant la fin du chargement complet
    if (difftime(Sys.time(), timer$x, units = "secs") < 4 & count$x > 2)
      return()

    #retrace l'objet dans sa forme initale une fois un autre objet cliqué
    if (length(LastId$x)>0) {
      if (LastId$x == input$map_shape_click$id) {
        return()
      } else {
      leafletProxy("map") %>%
        addPolygons(data = geom[geom$numero == LastId$x, ],
                    layerId = LastId$x,
                    stroke = TRUE, color = "Black", smoothFactor = 0.2,
                    fillOpacity = 1, fillColor = ~pal(nb_dossier),
                    weight = 1)
      }
    }

    #récupération de l'identifiant de l'objet cliqué
    LastId$x <- input$map_shape_click$id

    #trace l'objet cliqué avec un rendu spécifique
    leafletProxy("map") %>%
      addPolygons(data = geom[geom$numero == LastId$x, ],
                  layerId = LastId$x,
                  stroke = TRUE, smoothFactor = 0.2, color = "red",
                  fillOpacity = 1, fillColor = ~pal(nb_dossier),
                  weight = 3)
  })

  #mise en place du texte dynamique en haut à droite
  #apparaissant lorsqu'on clique sur une géométrie
  output$info <- renderUI({
    if (length(input$map_shape_click$id) == 0) {
      return(tags$div("Cliquez sur un territoire pour avoir des
informations"))
    } else {
      nb <- geom$nb_dossier[geom$numero == input$map_shape_click$id]
      nom <- geom$nom[geom$numero == input$map_shape_click$id]
      nb_obli <- geom$nb_peri_obli[geom$numero == input$map_shape_click$id]
      return(tags$div(
        tags$strong(nom),
        tags$div(nb, HTML("dossiers créés dont")),
        tags$div(nb_obli, HTML("du périmètre obligatoire"))
      ))
    }
  })

})
```

La partie correspondant à l'interface est définie dans le fichier ui.R présenté ci-dessous. Celle-ci est globalement organisée sous forme d'une grille fluide, c'est-à-dire s'adaptant à la taille de l'écran. Ce fonctionnement est issu du système de grille employé par Bootstrap. On place les éléments ligne après ligne, répartis sur douze colonnes.

La seule portion écrite explicitement en HTML est la barre de navigation supérieure avec la fonction HTML mais on retrouve des éléments de ce langage stockés dans l'objet `tags`. Les éléments stockés dans cet objet sont convertis en langage HTML. Il est donc possible de créer l'interface en combinant les deux méthodes.

```
# ui.R
#chargement des packages nécessaires
require(shiny)
require(leaflet)
require(dygraphs)

#définition de l'interface
shinyUI(fluidPage(
  #titre de la page
  title = "GéoAnalyse - Dynamique territoriale",

  tags$head(
    # CSS pour l'encadré en haut à droite de la carte et le texte en bas à
gauche
    tags$style("
        .floater { background-color: white; padding: 8px; opacity: 0.7;
border-radius: 6px; box-shadow: 0 0 15px rgba(0,0,0,0.2); }
        .indication { position: relative; }
        .indication span{ position: absolute; bottom: 0; right: 0; }
      "),
    # scripts de dygraph modifiés pour écrire les dates en français
    tags$script(src = "dygraphs-fr.js"),
    tags$script(src = "dygraph-combined-fr.js")
  ),

  # barre de navigation
  HTML('
    <nav class="navbar navbar-default navbar-static-top" role="navigation">
      <div class="container">
      <div class="navbar-header">
      <span class="navbar-brand">Géofoncier - GéoAnalyse</span>
      </div>
      <ul class="nav navbar-nav">
      <li class="active">
      <a href="http://91.206.198.152/heatmap-app2/">Carte de chaleur</a>
      </li>
      <li class="active">
      <a data-toggle="tab" href="http://91.206.198.152/datadep-
app2/">Dynamique territoriale</a>
      </li>
      </ul>
      </div>
      </nav>
  '),

  #1ère ligne : titre central
  fluidRow(
```

```
    h2("Dynamique territoriale : création de dossiers sur Géofoncier",
align = "center")
  ),

  #2ème ligne : boutons radios sur 2 colonnes et carte sur 8
  fluidRow(
    column(2,
           radioButtons("lim", label = "choix du découpage :", selected =
"reg",
                        choices = list("départements" = "dep", "régions
ordinales" = "reg"))
    ),

    column(8,
           leafletOutput("map", height = 400, width = "100%"),
           absolutePanel(
             right = 30, top = 10, width = 200, class = "floater",
             draggable = TRUE,
             uiOutput("info")
           )
    )
  ),

#   fluidRow(
#     column(1, offset = 1,
#       uiOutput("text")
#     )
#
#   ),
  #retour à la ligne pour séparer la carte du graphique
  br(),

  #3ème ligne : graphique
  fluidRow(
    column(8, offset = 2,
           dygraphOutput("graph", height = 300))
  ),

  #4ème ligne : texte informatif
  fluidRow(
    column(2,
           div(span(h4('sélection de la période :'))), class =
"indication")
  )
))
```

Résumé du mémoire

Introduction

Lancé en 2010, Géofoncier est une Infrastructure de Données Géographiques (IDG) permettant de rassembler sur une unique plateforme les travaux réalisés par les géomètres-experts. Une partie des versements de la profession fait l'objet d'une obligation ordinale, notamment dans les domaines du foncier et de l'immobilier. Cette IDG est également le support de l'élaboration du Référentiel Foncier Unifié (RFU) dont l'objectif est de pérenniser l'ensemble des limites foncières géoréférencées par la profession.

Il s'agit d'une mine d'informations jusqu'à présent inexploitée sur le plan statistique concernant les géomètres-experts. L'exploration des données de Géofoncier pose deux difficultés principales. La première est celle liée au traitement d'un volume considérable de données. L'IDG comporte en effet plus de trois millions de dossiers et cinq millions d'objets RFU. La seconde concerne la fiabilité, l'exhaustivité des informations présentes dans l'IDG.

L'étude réalisée s'inscrit alors dans deux problématiques. En quoi l'infrastructure de données géographiques Géofoncier permet-elle de refléter l'état actuel de la profession et son évolution récente d'un point de vue géographique, économique et structurel ? Comment rendre compte de façon dynamique et dans un contexte web de l'évolution temporelle des données ?

Ce résumé présente tout d'abord les données employées provenant de Géofoncier ainsi que les méthodes utilisées lors de ce travail. Ensuite, les principaux résultats des analyses menées et des cartes réalisées sont brièvement détaillés. Enfin, la dernière partie évoque la mise en place d'une interface web interactive.

I. Données présentes sur Géofoncier

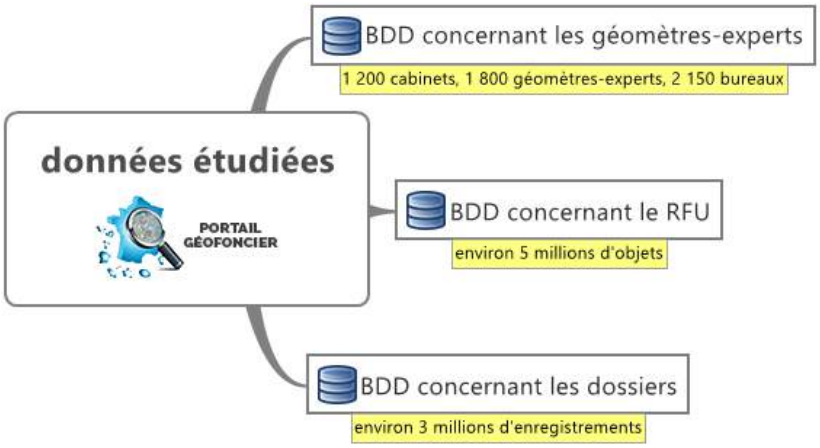


Figure 1 : schéma des Bases De Données (BDD) employées

Les données considérées lors de ce travail portent sur les trois domaines détaillés ci-contre (figure 1). Cependant, la plupart des analyses ont été réalisées avec les BDD fournissant des informations sur les géomètres-experts et sur leurs dossiers enregistrés. Lorsqu'il est nécessaire de disposer de données exhaustives, seuls les enregistrements faisant partie du périmètre obligatoire sont généralement pris en compte. Cela inclut les principales opérations foncières et immobilières.

Le RFU, bien que comportant cinq millions d'objets, est trop hétérogène sur l'ensemble du territoire, certains géomètres-experts n'effectuant pas les versements d'objets RFU pourtant obligatoires.

II. Méthodes employées

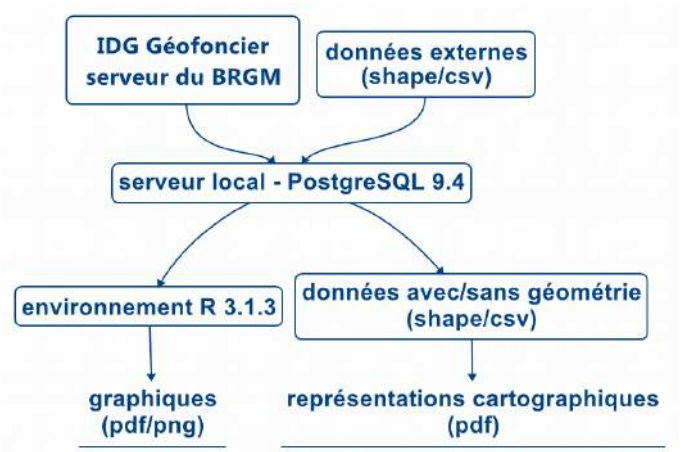


Figure 2 : démarche générale de travail

La démarche générale employée lors des analyses est détaillée ci-contre (figure 2). Celle-ci peut être globalement scindée en trois étapes au niveau du schéma.

La première consiste à extraire du serveur de production hébergeant Géofoncier les données nécessaires aux analyses. Bien que théoriquement possible, l'export massif des données brutes a été le plus fréquemment évité à cause de leur volume trop important. Les données extraites sont obtenues au format CSV et sont ensuite transférées sur une base de données locale à l'aide de GDAL, un convertisseur de fichiers comportant des données géographiques entre différents formats. De même, on importe les données externes venant compléter les analyses.

Les traitements sont ensuite réalisés sous PostgreSQL, un système de gestion de bases de données, avec l'extension PostGIS afin de travailler sur les géométries des objets. Les analyses statistiques sont ensuite effectuées à l'aide de R, un langage et logiciel libre dédié à la statistique, en se connectant directement au serveur local depuis cet environnement.

Par ailleurs, des cartes sont également produites à partir des données exportées du serveur local en utilisant GDAL. Leur mise en page est effectuée avec les logiciels ArcGIS et/ou Inkscape généralement.

III. De fortes évolutions structurelles de la profession

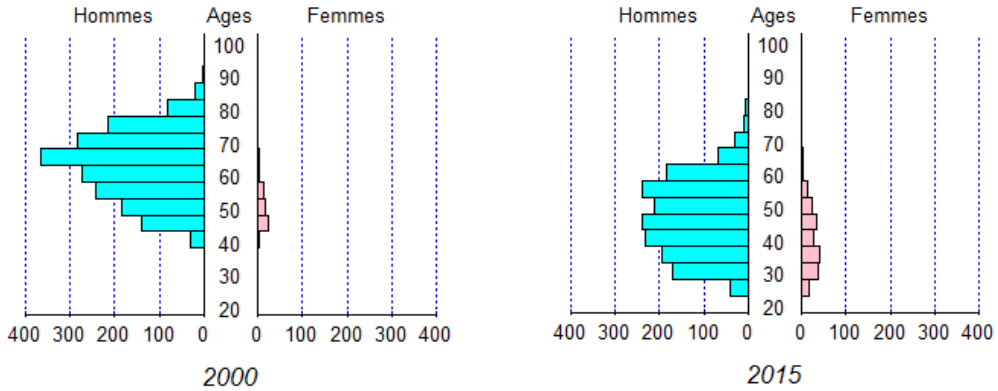


Figure 3 : pyramides des âges de la profession en 2000 (à gauche) et 2015 (à droite)

La première partie des analyses réalisées à partir des données de Géofoncier s'intéresse aux mutations récentes de la profession. Au travers de l'évolution de la pyramide des depuis quinze ans (figure 3), nous avons constaté un fort rajeunissement des géomètres-experts en activité. La part des femmes s'accroit également mais reste très minoritaire.

Par ailleurs, nous avons observé un changement structurel important au niveau de la taille des cabinets. Ceux de faible envergure, constitués généralement d'un unique géomètre-expert exerçant en libéral, sont de moins en moins nombreux au profit des cabinets plus importants qui, eux se développent. Les effectifs de la profession sont relativement constants depuis quinze ans mais le

nombre de cabinets est en baisse significative, passant de plus de 1 500 en 2000 à moins de 1 200 en 2015.

IV. La classification des cabinets de géomètre-expert

À l'aide de données économiques sur la profession et des informations sur les versements réalisés sur Géofoncier, une typologie des cabinets de géomètres-experts est définie en utilisant des méthodes d'analyse de données multidimensionnelles.

Une première analyse a mis en lumière deux types de cabinets différents au niveau des versements obligatoires réalisés sur Géofoncier. D'une part, on distingue des structures réalisant des versements spécifiques au domaine foncier (bornage, document modificatif du parcellaire cadastral). D'autre part, on trouve des structures effectuant des versements spécifiques au domaine immobilier (règlement de copropriété, division en volumes notamment).

Par la suite, ces statistiques exploratoires permettent de dégager des tendances à partir d'un nombre plus important de variables. Il en ressort une différenciation des cabinets de géomètres-experts selon trois classes :

- les petites structures, avec des effectifs et une activité de faible envergure, représentent la majorité des cabinets,
- les cabinets de taille moyenne, fortement présents en Ile-de-France, comportent plusieurs géomètres-experts et bureaux ; ils se focalisent sur l'immobilier (copropriété, division en volumes) et l'urbanisme principalement,
- les grandes structures, travaillant sur un large territoire et spécialisées dans les travaux d'envergure en topographie, foncier et ingénierie.

V. Une dispersion géographique de la profession au profit des aires urbaines et du sud du territoire

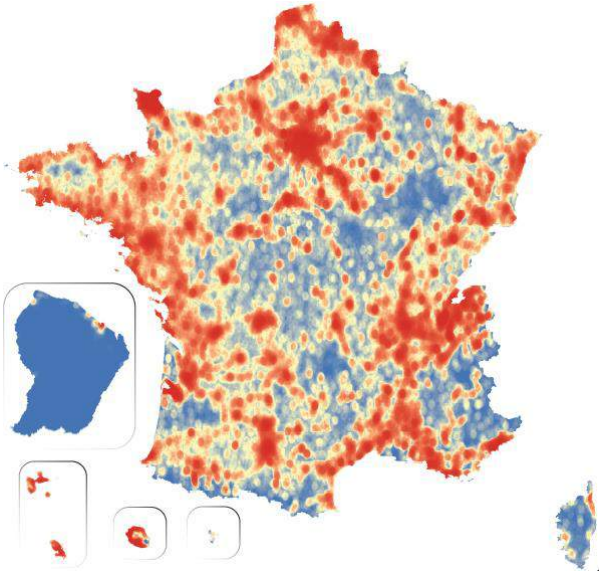


Figure 4 : carte de densité des dossiers créés en 2014
les couleurs chaudes traduisent une forte densité

Le caractère géographique des données à disposition a été étudié au niveau de la répartition territoriale des bureaux de géomètres-experts sur l'ensemble du territoire. Les aires urbaines concentrent 89 % des bureaux sur 45 % du territoire. Les géomètres-experts ont également une présence nettement plus marquée dans la partie sud du pays.

Cette concentration des implantations de la profession transparaît au niveau des dossiers (figure 4). On remarque également une forte activité le long de la façade atlantique et du littoral méditerranéen.

Les géomètres sont donc présents en milieu urbain et dans les espaces à fort enjeu foncier et économique.

VI. La mise en place d'une interface web de visualisation des résultats

L'utilisation de R a permis d'expérimenter l'extension Shiny. Celle-ci étend les fonctionnalités de ce langage à la réalisation d'une application web apportant de manière interactive la visualisation

de cartes et graphiques. L'outil permet de créer une interface valorisant les données issues de Géofoncier.

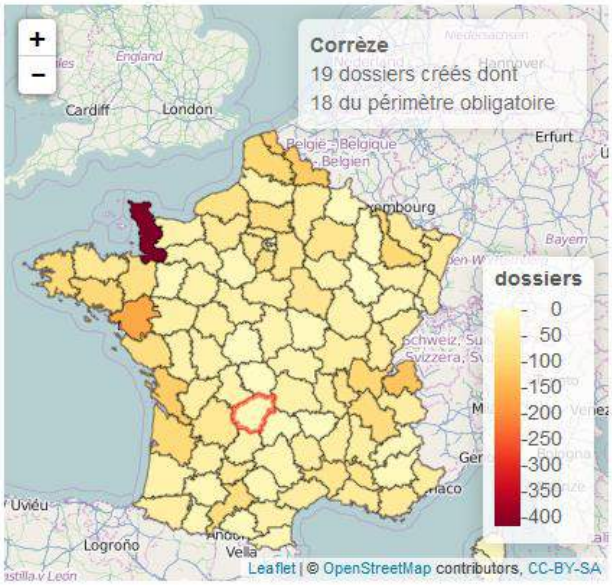


Figure 5 : exemple de visualisation sur navigateur de la création de dossiers sur Géofoncier selon un découpage départemental

L'application est créée à partir de deux fichiers R : l'un gérant le serveur et les requêtes de calcul, l'autre paramétrant l'interface utilisateur. Ceux-ci génèrent au moment du lancement de l'application sur un navigateur le contenu en langage web nécessaire à son exécution. Elle s'appuie sur des bibliothèques web comme Bootstrap pour les éléments interactifs, Leaflet pour l'aspect cartographique ou Dygraphs pour les graphiques temporels.

Elle permet de visualiser, de façon continue et interactive, l'évolution de Géofoncier en utilisant directement les données. On peut notamment observer le nombre de dossiers créés selon un découpage territorial (figure 5) et une période donnée.

Conclusion

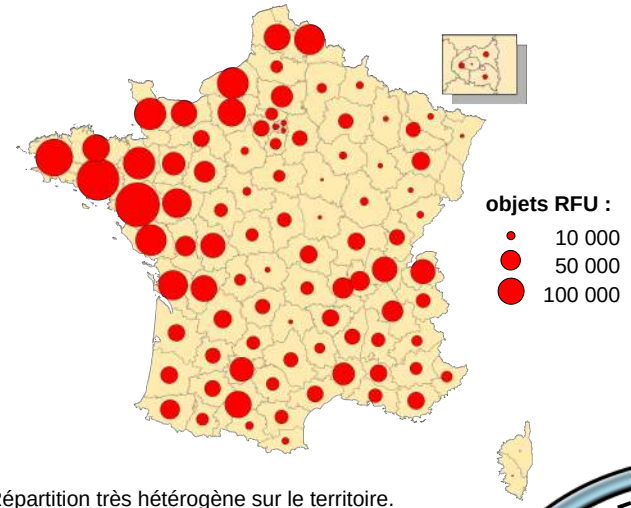
En tenant compte de la fiabilité des données, les analyses tant statistiques que cartographiques mises en œuvre font ressortir des tendances sur l'ensemble des géomètres-experts. Ce travail de *data mining* permet finalement d'avoir un regard présent sur la profession au travers de l'IDG Géofoncier et de son activité en prenant en compte la gestion d'un grand volume de données et la fiabilité de celles-ci.

D'un point de vue technique, les expérimentations réalisées sur R et ses extensions ont montré qu'il s'agit d'un langage destiné en premier lieu aux statistiques mais aussi performant sur d'autres utilisations, notamment la visualisation de données ainsi que l'intégration dans des applications web.

Analyse spatio-temporelle des référentiels métiers de l'Infrastructure de Données Géographiques Géofoncier pour la description et la représentation de phénomènes statistiques inhérents à ces données

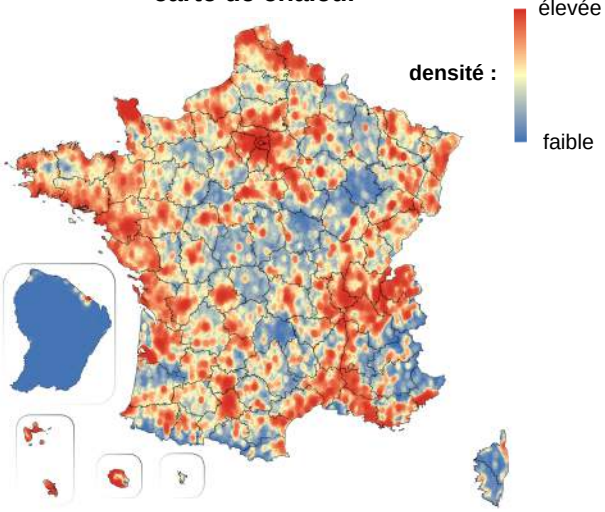
Présentation et objectif :
Ouvert depuis 2010, Géofoncier est une Infrastructure de Données Géographiques (IDG) visant à rassembler sur une seule plateforme les travaux réalisés par les géomètres-experts. Cette IDG est également le support de l'élaboration du Référentiel Foncier Unifié (RFU) dont l'objectif est de collecter l'ensemble des limites foncières géoréférencées par la profession. Le travail réalisé vise à obtenir une vision de la profession et de son évolution au travers de l'analyse statistique et cartographique des données de Géofoncier.

Nombre d'objets RFU versés par département (juillet 2010 - mars 2015)



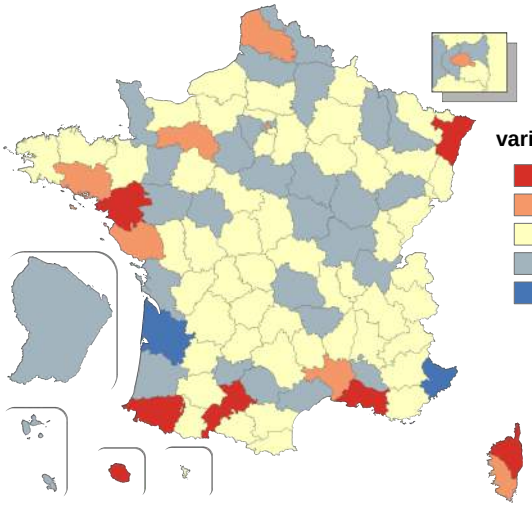
- Répartition très hétérogène sur le territoire.
- Versement variable selon les géomètres-experts malgré l'obligation ordinale de compléter le RFU.

Création de dossiers en 2014 - carte de chaleur -



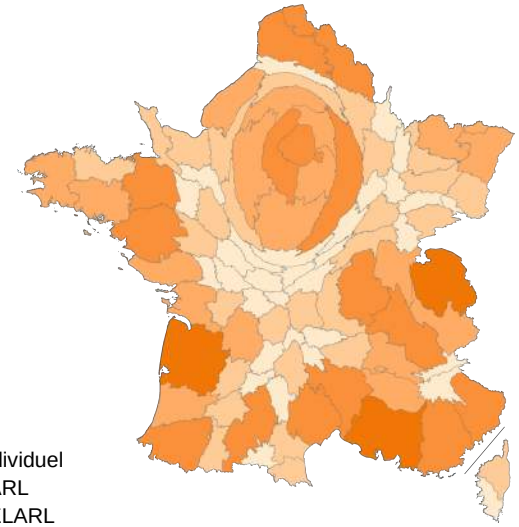
- Concentration des dossiers dans les zones urbaines et littorales.
- Densité élevée de l'activité dans la vallée du Rhône, sur la façade atlantique et méditerranéenne.

Evolution du nombre de bureaux de géomètres-experts de 2000 à 2015



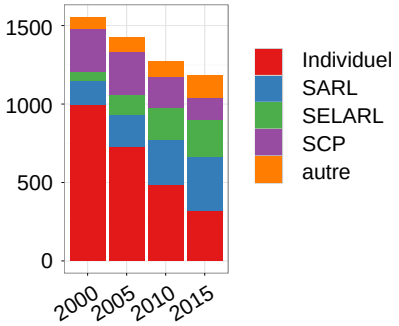
- Les fortes variations du nombre de bureaux depuis 2000 se situent sur les départements comportant une métropole.
- Evolution quantitive sur 15 ans :
  - ↗ +8 % de bureaux sur le territoire
  - ↘ -23 % de cabinets
  - ↘ -4 % de géomètres-experts en activité.
- Dépréciation des formes historiques d'exercice libéral (entreprise individuelle, SCP) au profit des sociétés de capitaux (SARL, SELARL).

Répartition des bureaux de géomètres-experts en 2015 - cartogramme -



- Les aires urbaines concentrent 89 % des bureaux sur 45 % du territoire.
- La profession est peu présente sur les départements du centre et du nord-est, essentiellement ruraux.

Evolution du nombre de cabinets par type de société

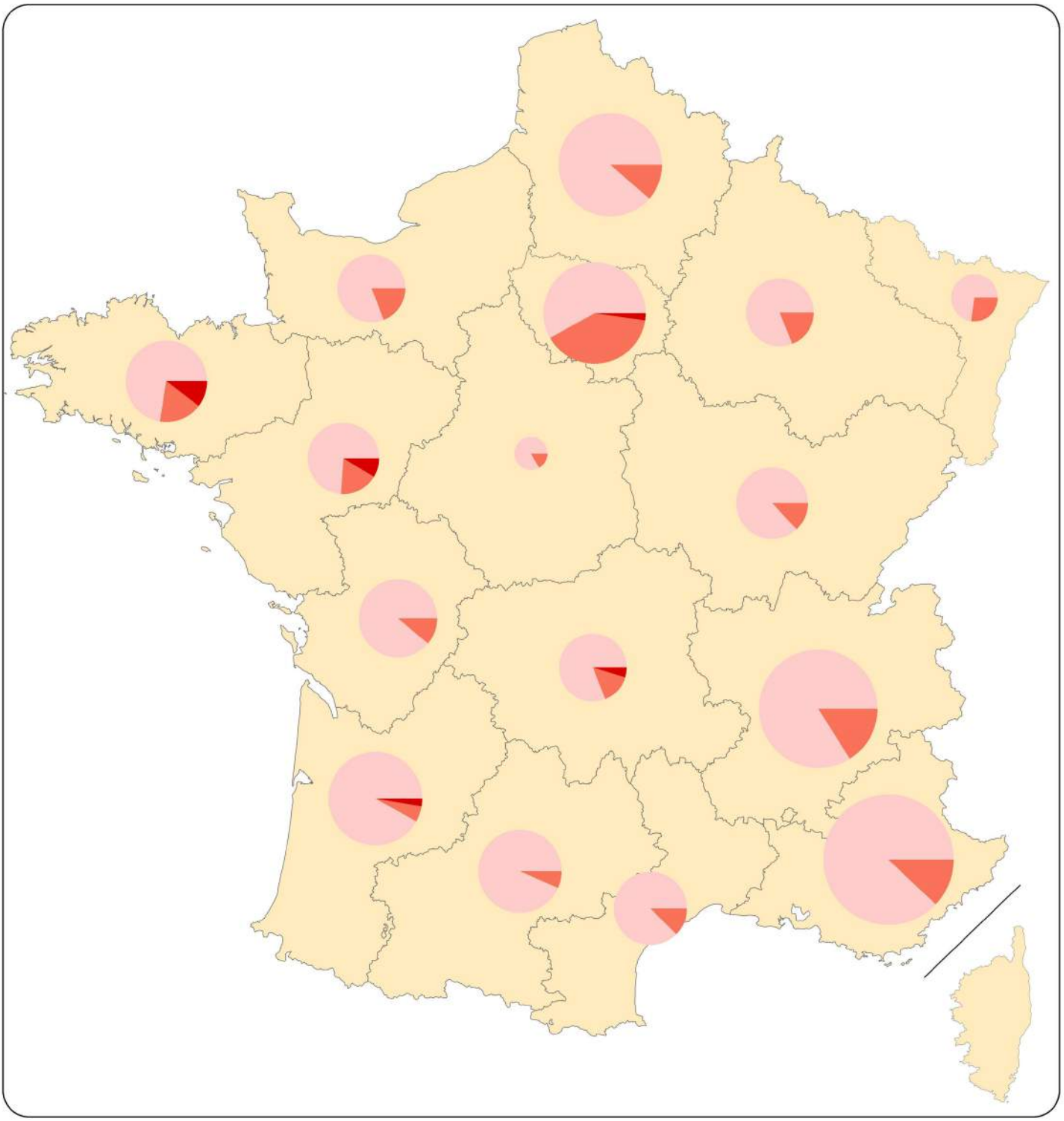


Conclusion :
L'analyse des données de Géofoncier révèle une tendance d'association des géomètres-experts au sein de cabinets plus importants structurés en sociétés commerciales. L'exercice à titre individuel, forme historique d'installation de la profession, est désormais minoritaire. L'activité et l'implantation des cabinets se concentrent dans les espaces présentant un fort enjeu foncier et économique, principalement les zones urbaines et littorales.

source des données : Géofoncier
auteur : Emmanuel Labrousse
date : juin 2015

- carte 1 -

Répartition des cabinets de géomètres-experts par classe et par région ordinale (année 2013)



Légende

classification des cabinets

9,8 cabinets

petits cabinets

cabinets spécialisés dans l'immobilier et l'urbanisme

grands cabinets

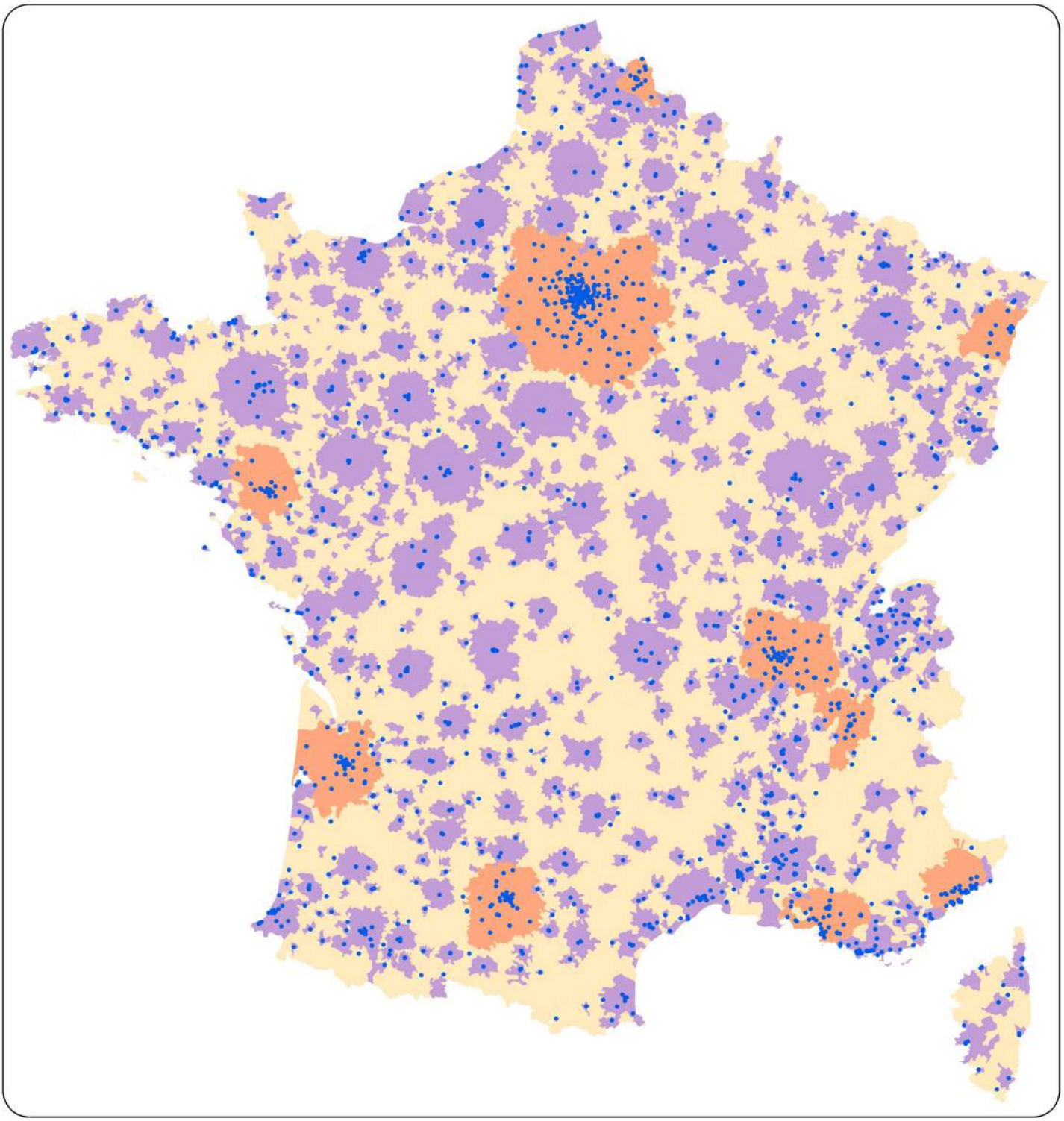
Sogefi®

Source : IDG Géofoncier

E. Labrousse, le 08/06/15

- carte 2 -

Localisation des bureaux de géomètres-experts en 2015 et influence des aires urbaines



Légende

- les 10 premières aires urbaines en 2010
- autre aire urbaine
- bureau de géomètre-expert

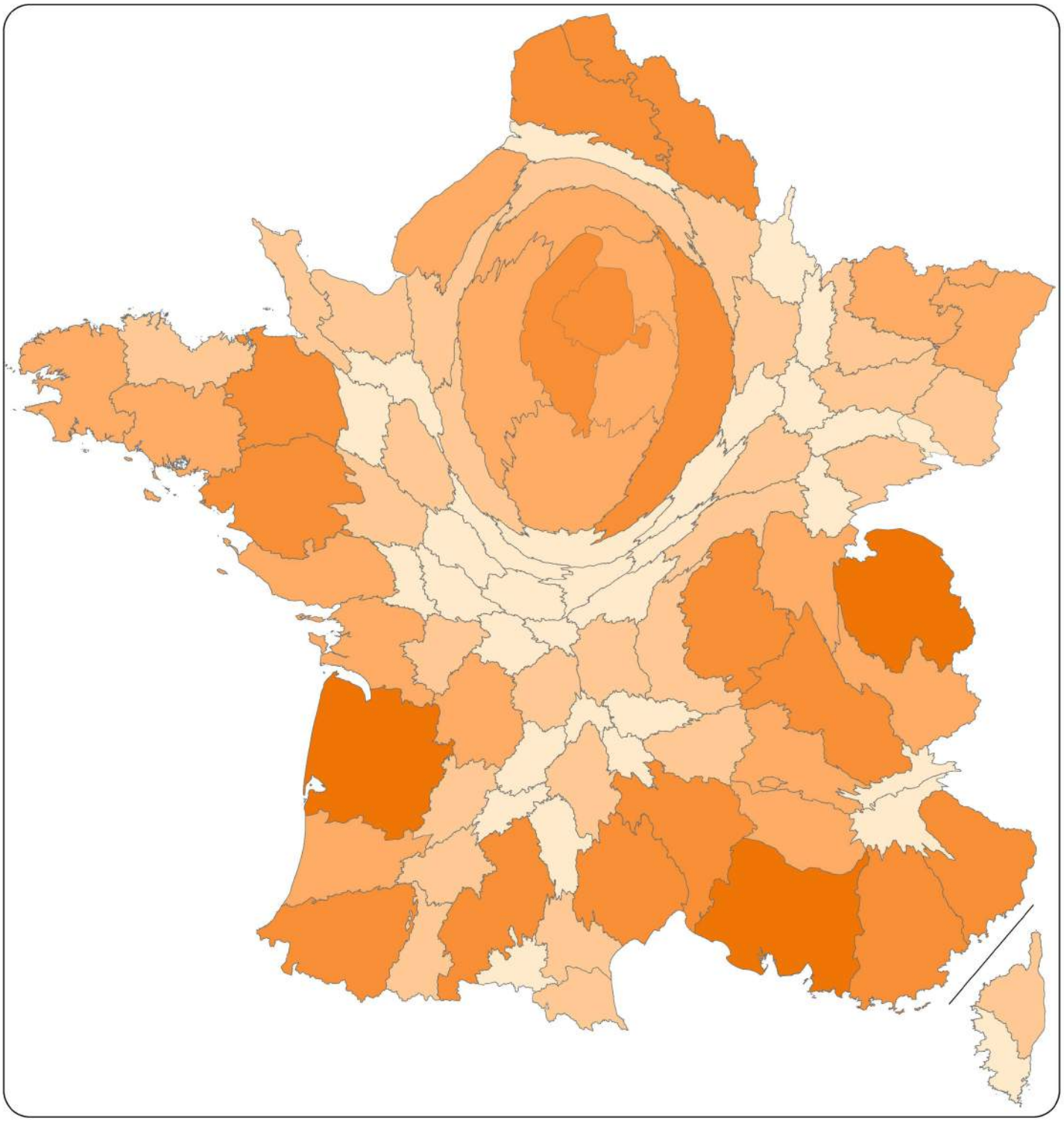
Sogefi®

Source : IDG Géofoncier, Conseil Régional de Basse-Normandie

E. Labrousse, le 31/03/15

- carte 3 -

Répartition territoriale des bureaux de géomètres-experts en 2015



Légende
nombre de bureaux

3 - 10 | 11 - 18 | 19 - 27 | 28 - 43 | 44 - 56

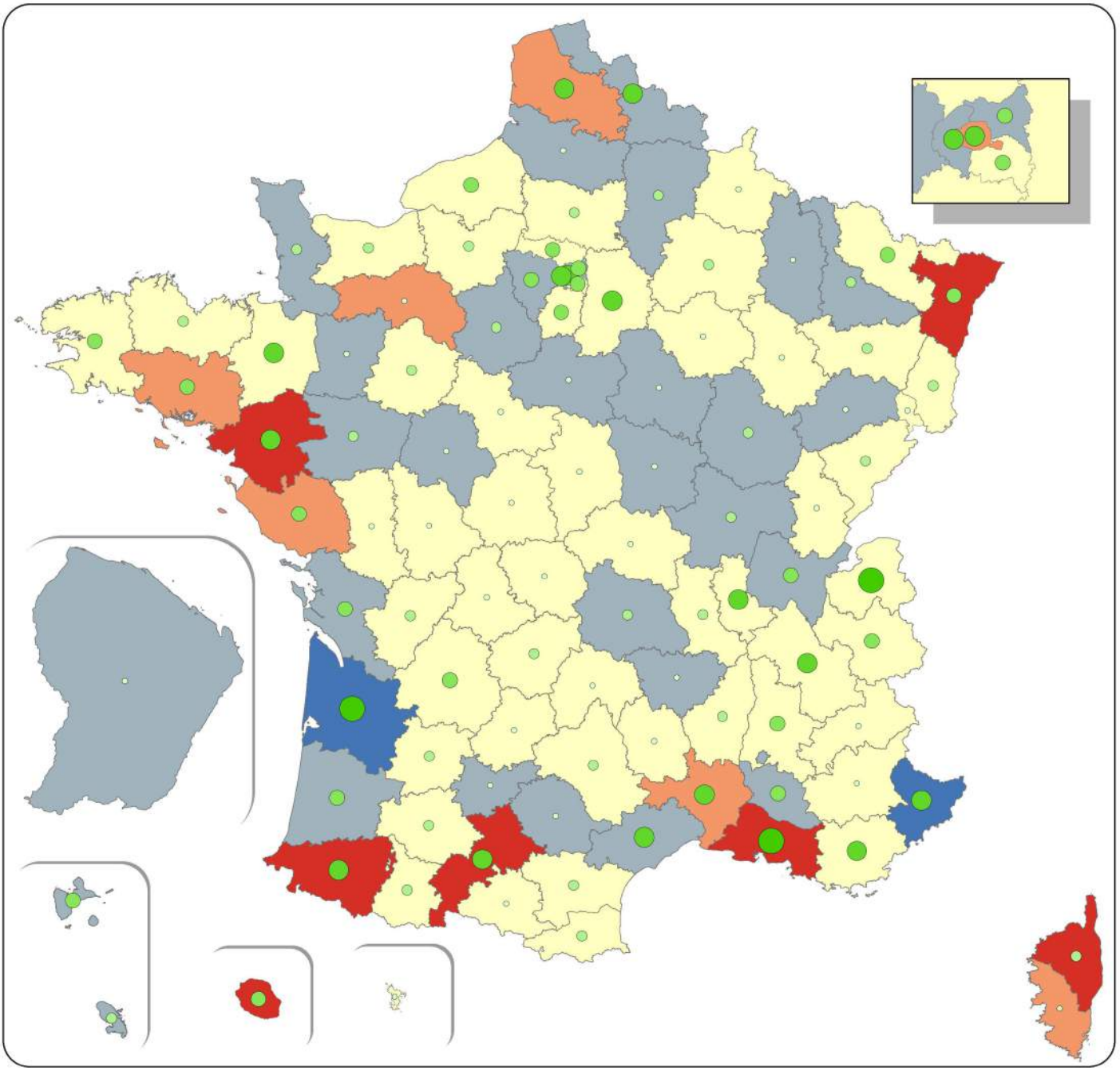
Sogefi®

Source : IDG Géofoncier

E. Labrousse, le 11/05/15

- carte 4 -

Dynamique des bureaux de géomètres-experts (2000 - 2015)



Légende

bureaux en 2015

- 3 - 10
- 11 - 18
- 19 - 27
- 28 - 43
- 44 - 56

évolution du nombre de bureaux de 2000 à 2015

- 5 à 8
- 2 à 4
- -1 à 1
- -4 à -2
- -8 à -5

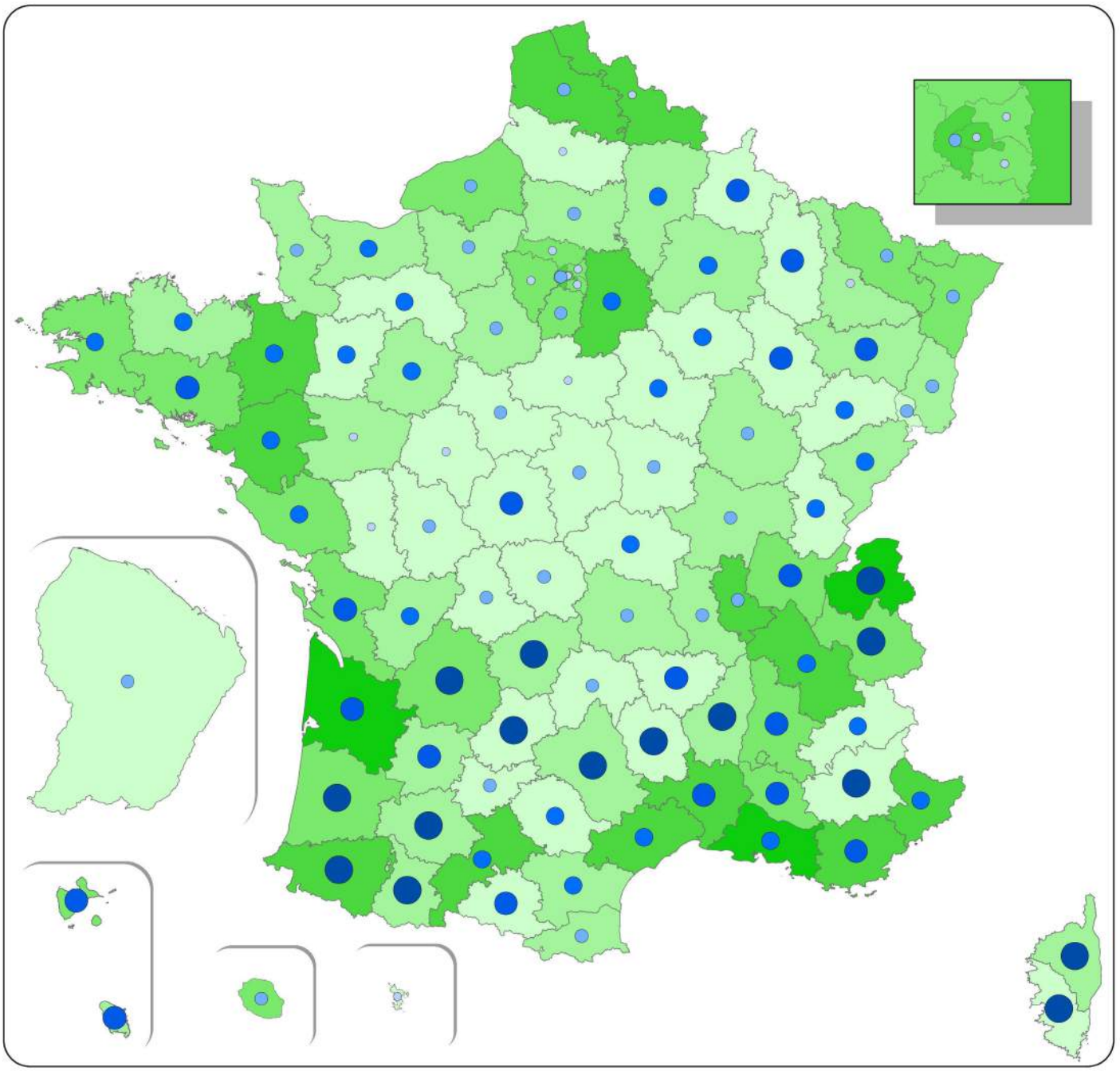
Sogefi®

Source : IDG Géofoncier

E. Labrousse, le 16/06/15

- carte 5 -

Répartition départementale des bureaux de géomètres-experts



Légende

bureaux pour 1 000 hab.

- 0 - 0,018
- 0,018 - 0,025
- 0,025 - 0,033
- 0,033 - 0,047
- 0,047 - 0,074

nombre de bureaux

- 3 - 10
- 11 - 18
- 19 - 27
- 28 - 43
- 44 - 56

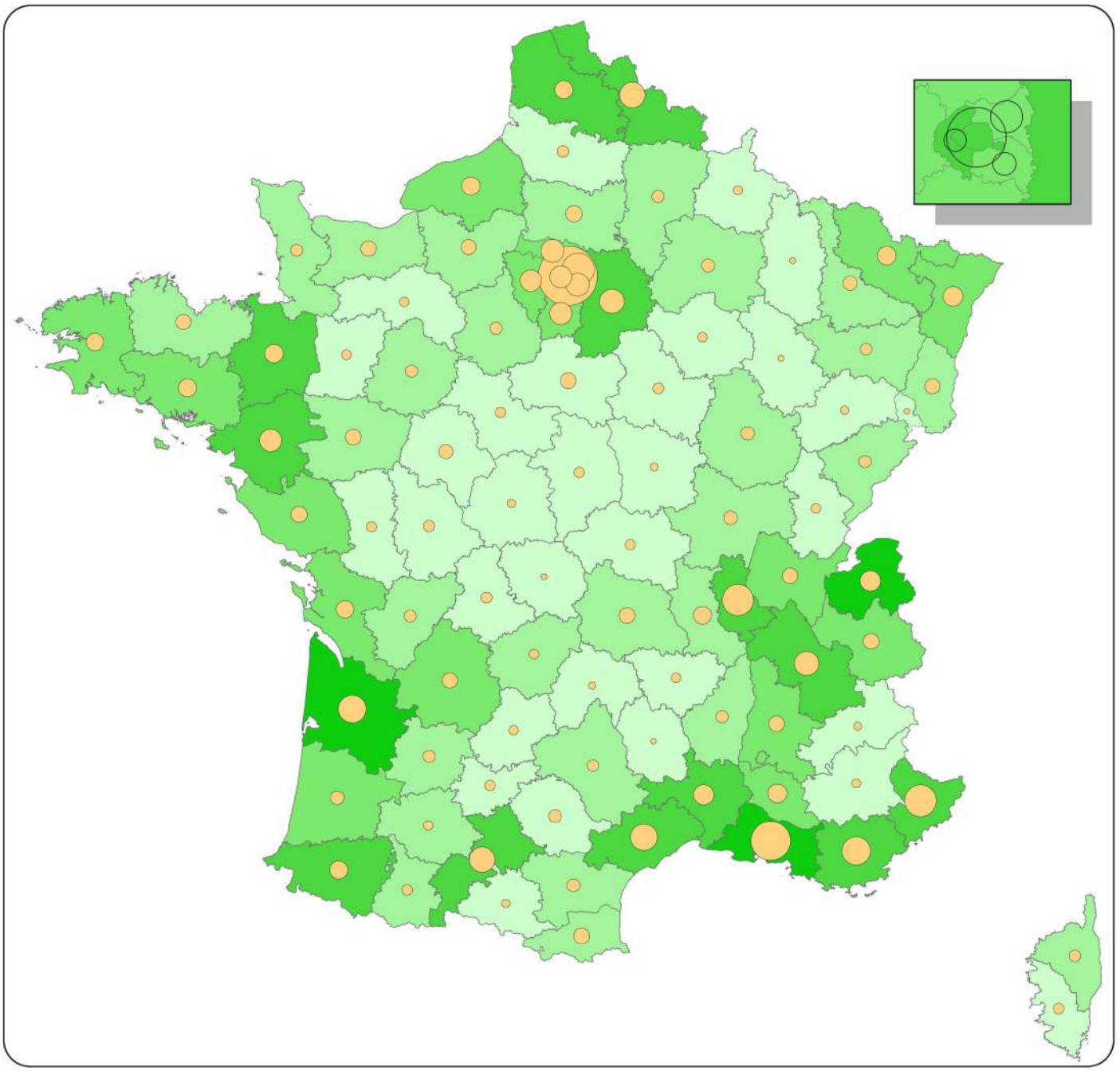
Sogefi®

Source : IDG Géofoncier

E. Labrousse, le 22/05/15

- carte 6 -

Les géomètres-experts et le secteur de la construction



Légende

établissements de la construction

- 1 000
- 5 000
- 10 000

bureaux de géomètres-experts

- 3 - 10
- 11 - 18
- 19 - 27
- 28 - 43
- 44 - 56

Sogefi®

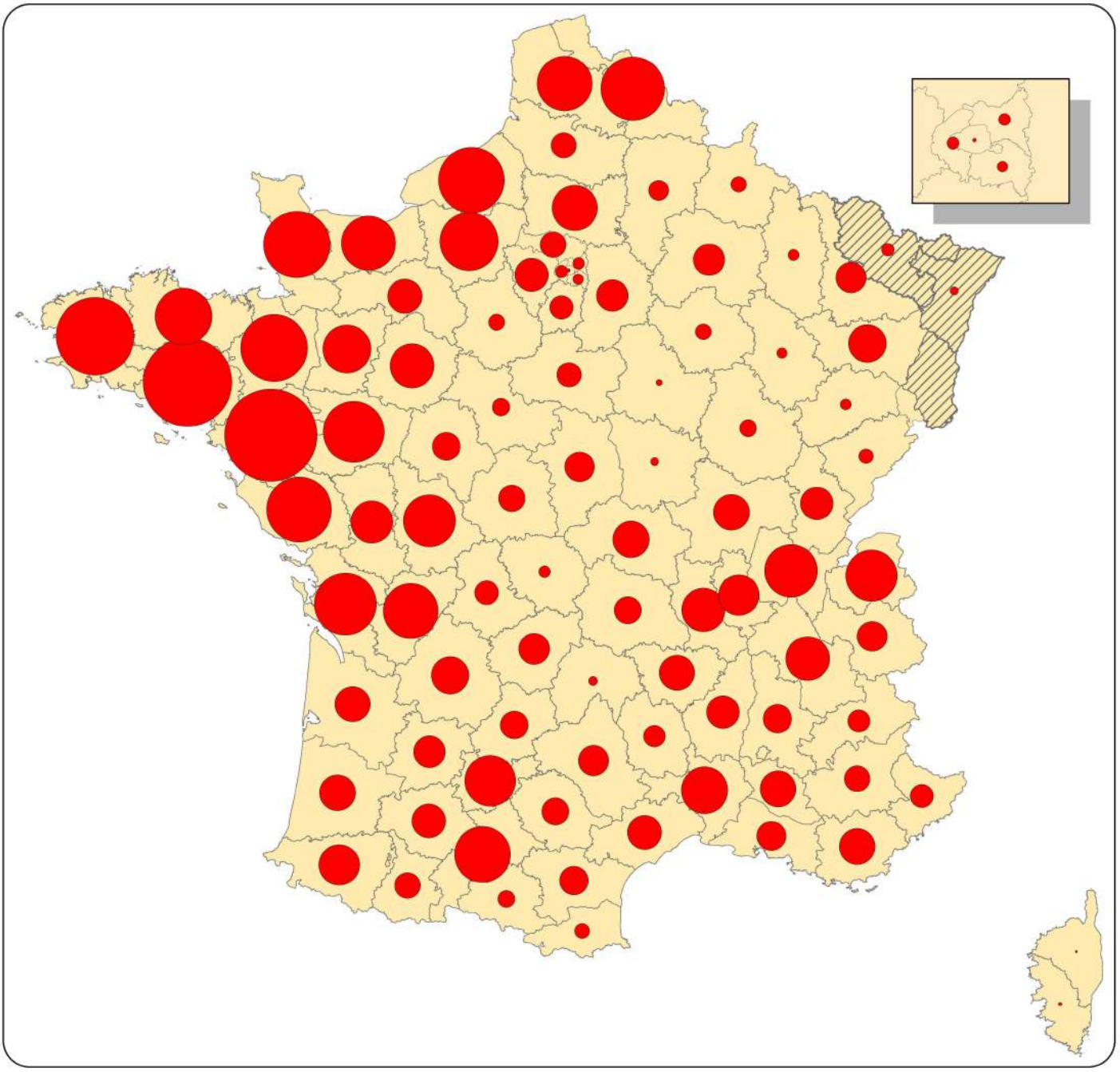
Sources : IDG Géofoncier; Insee, Répertoire des entreprises et des établissements (Sirene)

E. Labrousse, 30/03/15

- carte 7 -

Versement d'objets RFU par département

situation au 25/03/2015

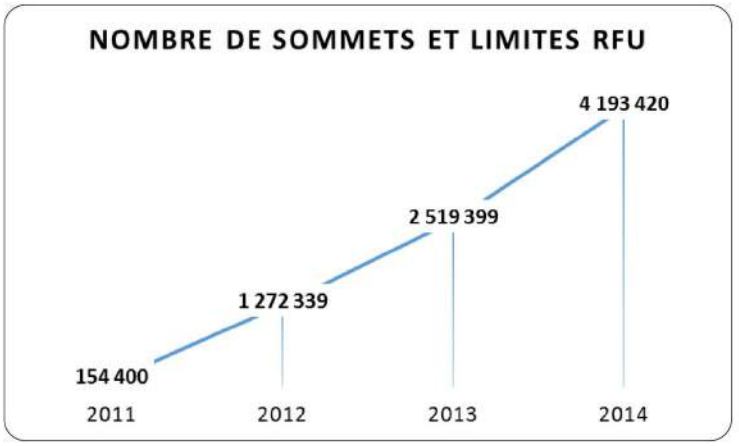


Légende

cadastre spécifique (Alsace-Moselle)

nombre d'objets RFU

- 100
- 1 000
- 5 000
- 10 000
- 100 000

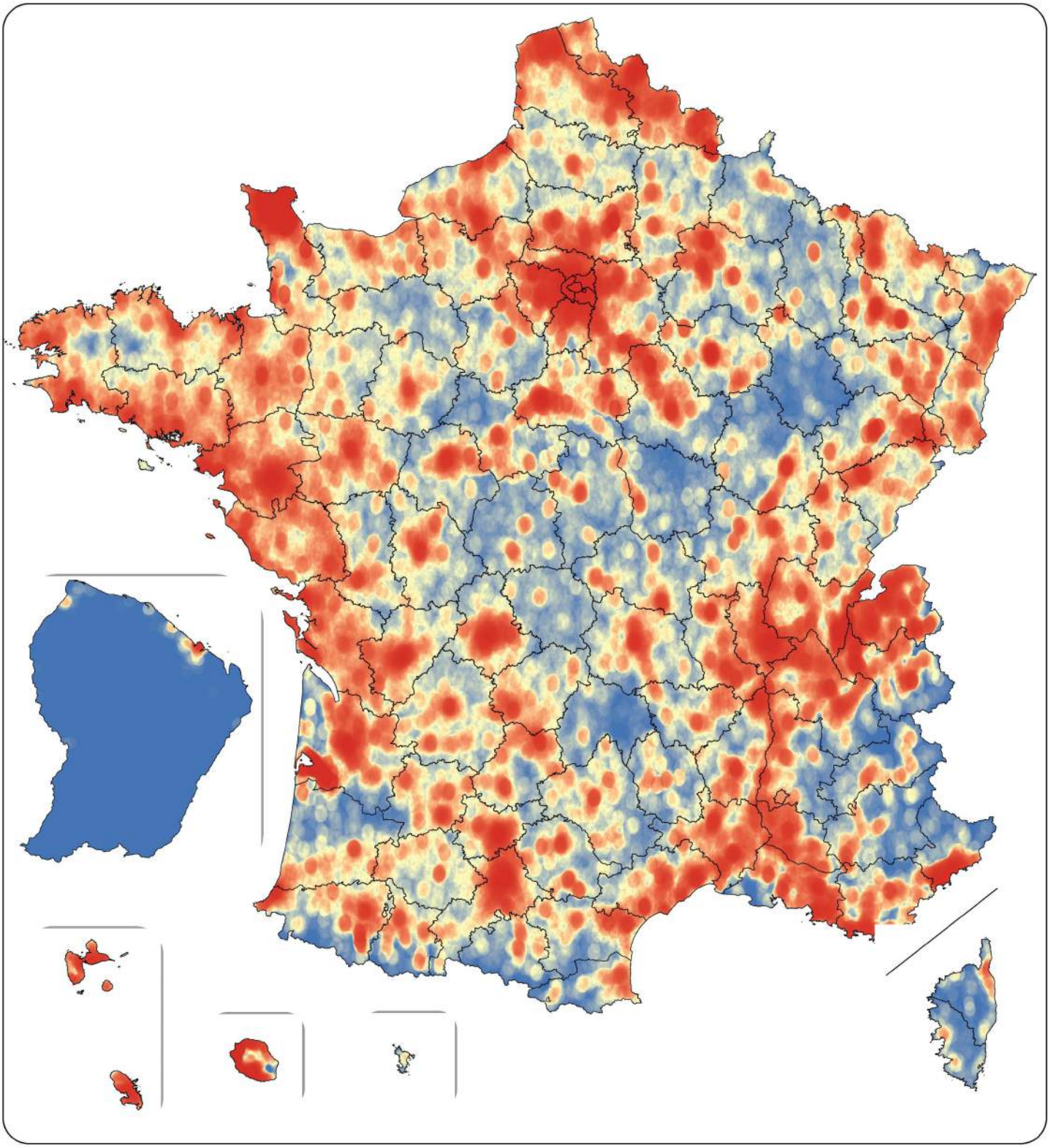


Source : IDG Géofoncier

E. Labrousse, le 30/03/15

- carte 8 -

Densité de dossiers créés en 2014



densité

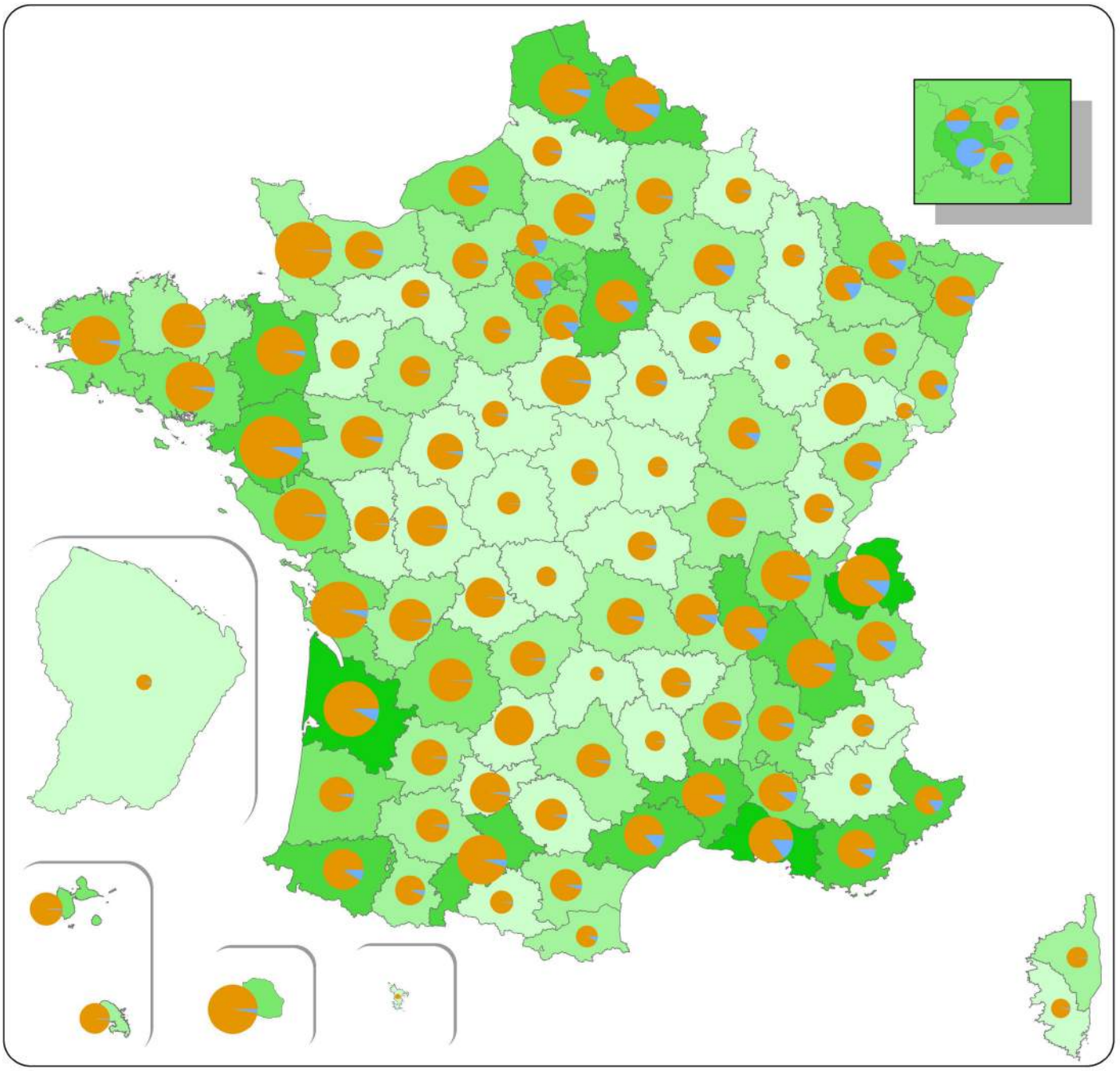
élevée — faible

Source : IDG Géofoncier

E. Labrousse, 20/04/15

- carte 9 -

Activité foncière et immobilière répartition départementale (2012-2014)



Légende

nombre de dossiers par département

9 000 dossiers

foncier

copropriété/division en volumes

nombre de bureaux de géomètres-experts

3 - 10

11 - 18

19 - 27

28 - 43

44 - 56

Sogefi®

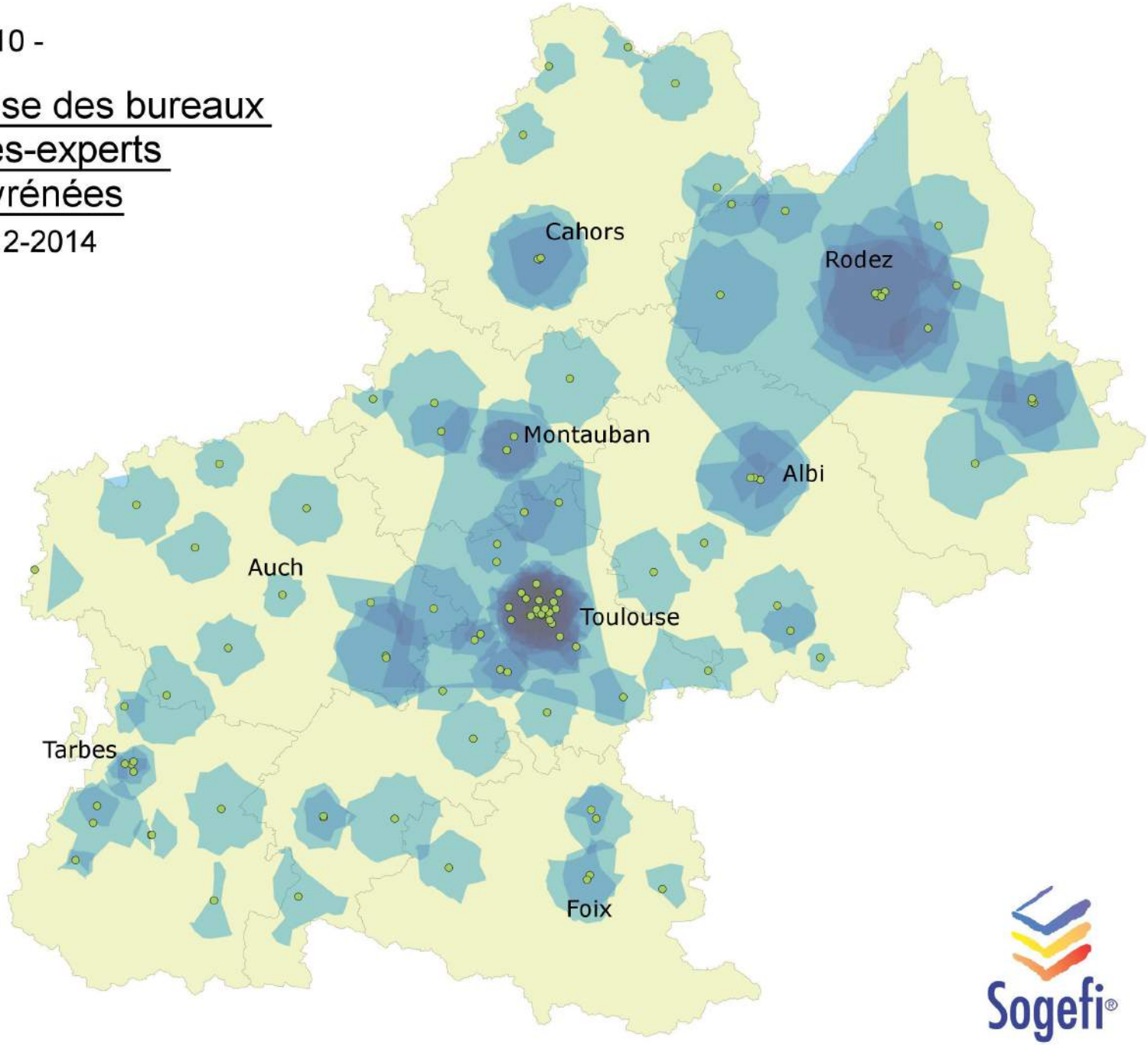
Source : IDG Géofoncier

E. Labrousse, 07/04/15

- carte 10 -

Zones de chalandise des bureaux de géomètres-experts en Midi-Pyrénées

période 2012-2014



Légende

- bureau
- zone de chalandise

Source : IDG Géofoncier

E. Labrousse, 23/04/15

Analyse spatio-temporelle des référentiels métiers de l'Infrastructure de Données Géographiques Géofoncier pour la description et la représentation de phénomènes statistiques inhérents à ces données

---

Mémoire en vue de l'obtention du titre d'ingénieur CNAM

Soutenu par Emmanuel LABROUSSE

RÉSUMÉ

Lancé depuis 2010, Géofoncier est une infrastructure de données géographiques sur laquelle les géomètres-experts versent des informations concernant leurs travaux. Ce portail géographique tourné vers le foncier centralise une quantité considérable de données jamais exploitées jusqu'à présent.

L'étude contient un ensemble d'analyses concernant la profession de géomètre-expert à partir des données présentes sur Géofoncier. Les questions de fiabilité des données sont abordées dans un premier temps. Des analyses basées sur des statistiques descriptives et exploratoires réalisées à l'aide du logiciel R sont ensuite détaillées. Plusieurs documents cartographiques produits fournissent également un regard sur la profession et sur sa répartition en France.

Enfin, la mise en place d'une application web interactive utilisant les données de Géofoncier afin de produire un rendu cartographique dynamique est détaillée à la fin du mémoire.

SUMMARY

Launched since 2010, Géofoncier is a spatial data infrastructure where land surveyors register information about their tasks. This geoportal focused on land ownership centralizes a substantial amount of data which has never been exploited so far.

This study includes a set of analysis about the surveying profession from Géofoncier databases. The issues of data reliability are addressed then several analysis are submitted based on descriptive and exploratory statistics with R software. Several generated cartographical materials provide a vision of the profession and its distribution in France.

Finally, the implementation of an interactive web mapping application using Géofoncier database in order to generate a map with dynamic rendering is detailed at the end of the study.

Keywords : Géofoncier, GIS, data mining, R, web mapping, data analysis, surveying.