Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems"
(UNIGIS MSc) am Interfakultären Fachbereich für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Geographical Equity of the EU's Agricultural Subsidies in Belgium"

vorgelegt von

Mag. Christian Lagona
103839, UNIGIS MSc Jahrgang 2014

Zur Erlangung des Grades
„Master of Science (Geographical Information Science & Systems) – MSc(GIS)"

Gutachter:
Ao. Univ. Prof. Dr. Josef Strobl

Lichtenwörth, 19.07.2017

Declaration of Honesty

I hereby declare that the work submitted is my own and that all passages and ideas that are not mine have been fully and properly acknowledged.

Lichtenwörth, July 19, 2017
(Place, Date)

(Signature)

Abstract

The agricultural subsidies paid under the scope of the Common Agricultural Policy of the European Union amounted to around € 55 billion in 2014 and € 57 billion in 2015. Of this amount, nearly € 748 million or € 790 million were paid to Belgian farmers and agricultural enterprises in 2014 and 2015 respectively.

This work attempts to answer the question whether these subsidies were geographically evenly distributed across Belgium. The subsidies for the budget years 2014 and 2015 were examined. The analysis was carried out at the municipal level as well as at the arrondissement level, the latter mainly to control the Modifiable Area Unit Problem (MAUP). The subsidies were compared with agricultural data about different livestock species and cultivation areas of different crops. In addition, an exclusive analysis of Flanders only was attempted with very detailed data from a different source, containing just agricultural areas of the most important cultivated plants. The study of the whole country of Belgium was carried out at three different hierarchical levels, at which the predictors were combined increasingly. Overall, the analysis was carried out at 16 different levels.

Firstly, global linear regression models were calculated and checked for outliers. Subsequently, the geographically weighted regression (GWR) was used, three different weighting functions being tested. Repeatedly, some predictors proved to be stationary while others were spatially variable. Therefore, the mixed geographically weighted regression (MGWR) was also employed. Finally, the results of OLS, GWR and MGWR were compared using the corrected Akaike information criterion (AICc) to determine which of the models described the data best.

In seven cases, the global linear regression was the best model, two times the GWR had the lowest AICc and three times the MGWR. In four cases, including the exclusive examination of Flanders, no suitable model was found at all. The overall results turned out to be too different to recognize a clear pattern. GWR and MGWR found exceptionally high values only in those municipalities and areas, which had previously been identified as outliers. These were mostly Leuven and Sint Truiden, two cities, in which large agricultural cooperatives with superregional importance are based, that receive very high subsidy payments. However, since there are many farmers and agricultural enterprises organized in those cooperatives, who often don't have their seat in one of these two cities, one can't claim those

municipalities are being favored in terms of agricultural subsidies. In the end, this work did not find any evidence for significant differences in the geographical distribution of subsidies within Belgium.

Zusammenfassung

Die im Rahmen der Gemeinsamen Agrarpolitik der Europäischen Union ausbezahlten Agrarsubventionen betrugen in den Jahren 2014 und 2015 ca. 55 bzw. 57 Mrd. €. Davon wurden 2014 fast 748 Mio. € bzw. 2015 fast 790 Mio. € an belgische Landwirte und Agrarbetriebe ausbezahlt.

Diese Arbeit versucht die Frage zu beantworten, ob diese Subventionen innerhalb Belgiens auch geographisch gleichmäßig verteilt wurden. Untersucht wurden die Subventionen der Budgetjahre 2014 und 2015. Die Analyse erfolgte dabei auf Gemeindeebene sowie auf Ebene der Arrondissements, letzteres vor allem, um das Modifiable Areal Unit Problem (MAUP) zu kontrollieren. Den Subventionen wurden landwirtschaftliche Daten zu verschiedenen Nutztierarten und Anbauflächen verschiedener Kulturpflanzen gegenübergestellt. Zudem wurde mit aus anderer Quelle stammendem, sehr detailliertem, jedoch lediglich Anbauflächen umfassendem Datenmaterial eine gesonderte, ausschließlich auf Flandern beschränkte Analyse versucht. Die Untersuchung ganz Belgiens erfolgte auf drei verschiedenen Hierarchieebenen, auf denen die Prädiktoren in zunehmendem Maße aggregiert wurden. Insgesamt erfolgte die Analyse somit auf 16 verschiedenen Ebenen.

Zunächst wurden globale lineare Regressionsmodelle berechnet und auf Ausreißer überprüft. Im Anschluss wurde die geographisch gewichtete Regression (GWR) eingesetzt, wobei drei verschiedene Gewichtungsfunktionen getestet wurden. Immer wieder erwiesen sich manche Prädiktoren als stationär, während andere räumlich variabel waren. Deshalb kam auch die gemischte geographisch gewichtete Regression (MGWR) zur Anwendung. Die Ergebnisse von OLS, GWR und MGWR wurden schließlich mittels korrigiertem Akaike-Informationskriterium (AICc) einem Vergleich unterzogen, um festzustellen, welches der Modelle die Daten am besten beschrieb.

In sieben Fällen war die einfache lineare Regression das beste Modell, zwei Mal hatte die GWR den niedrigsten AICc und drei Mal die MGWR. In vier Fällen konnte überhaupt kein geeignetes Modell gefunden werden. Die Ergebnisse waren insgesamt zu unterschiedlich, um ein klares Muster erkennen zu können. GWR und MGWR fanden meist nur in jenen Gemeinden und Gegenden außergewöhnlich hohe Werte, die zuvor bereits als Ausreißer identifiziert worden waren. Dabei handelte es sich meist um Leuven und Sint Truiden, zwei Städte, in denen große landwirtschaftliche Kooperativen mit überregionaler Bedeutung ihren Sitz haben, die sehr hohe Subventionszahlungen erhielten. Da in

diesen aber sehr viele Landwirte und Agrarbetriebe organisiert sind, die ihren Sitz oftmals gar nicht in einer dieser beiden Städte haben, kann man hier nicht von Bevorzugung dieser Gemeinden sprechen. Diese Arbeit konnte somit letztlich keine bedeutsamen Unterschiede bei der geographischen Verteilung der Subventionen innerhalb Belgiens finden.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Symbols

α   Significance Level

β   Standardized β-Coefficient

AIC   Akaike Information Criterion

AICc   Corrected Akaike Information Criterion

CAP   Common Agricultural Policy

CN   Condition Number

CSV   Comma Separated Values

EAFRD   European Agricultural Fund for Rural Development

EAGF   European Agricultural Guarantee Fund

EU   European Union

FTP   File Transfer Protocol

GWR   Geographically Weighted Regression

M   Regression Model

MAUP   Modifiable Areal Unit Problem

MGWR   Mixed Geographically Weighted Regression

OLS   Ordinary Least Squares

Q-Q   Quantile-Quantile

r   Pearson's Correlation Coefficient

p   probability value

t   Student's t-value

R²   Coeffcient of Determination

s   Standard Deviation

s²   Variance

SD   Standard Deviation (only in legends)

VIF   Variance Inflation Factor

1 Introduction

1.1 Motivation

In 1962, the Common Agricultural Policy (CAP) of the European Union was introduced. It is financed through the European Agricultural Guarantee Fund (EAGF) and the European Agricultural Fund for Rural Development (EAFRD). The share of the budget for the Common Agricultural Policy in the total budget of the European Union for the period 2014 - 2020 is 37.8 percent[1], in 2014 (2015) the sum of the agricultural subsidies paid out to farmers in the European Union amounted to approximately € 55 billion (€ 57 billion)[2]. Initially, the aim of the Common Agricultural Policy was to stimulate agricultural production, which had declined after the Second World War in most European countries, in order to guarantee food security of the population and ensure a secure income for the agricultural population[3]. The objectives have shifted with the changing conditions of the world economy and the Common Agricultural Policy has been and still is being reformed ever since the 1980s. Nevertheless, there has been criticism of the Common Agricultural Policy for a long time. For example, it is criticized that the agricultural sector disproportionately benefits from the CAP. Although, the share of the Common Agricultural Policy in the EU's total budget has declined for several years, it is still the largest item despite the fact that the percentage of people working in agriculture is declining throughout the EU. In 2010, the agricultural sector employed only around 10 million people (about 5 percent of the working population[4]). It is also being criticized that larger farms and agricultural businesses, which already have larger resources anyway, are receiving far more subsidies than small farmers, since they are also the larger producers. As a result, a large number of family-run businesses are squeezed out of the market, which distorts it. There are also debates on equity regarding the distribution of agricultural subsidies among the member states, as some countries have a much larger agricultural sector than

---

[1] http://www.europarl.europa.eu/atyourservice/en/displayFtu.html?ftuId=FTU_5.2.2.html (retrieved May 2, 2017)
[2] http://ec.europa.eu/budget/library/biblio/documents/2015/internet-tables-2000-2015.xls (retrieved May 5, 2017)
[3] http://ec.europa.eu/agriculture/cap-overview/history_de (retrieved May 8, 2017)
[4] https://ec.europa.eu/agriculture/sites/agriculture/files/rural-area-economics/briefs/pdf/08_en.pdf (retrieved May 2, 2017)

others and thus benefit disproportionately. Numerous other criticisms exist. In order to defuse the increasingly louder criticism, the EU Regulation No. 1306/2013 of the European Parliament and of the Council of the European Union as well as the EU Commission Implementing Regulation No. 908/2014 of the European Commission have been enacted which oblige the EU member states to publish all beneficiaries of subsidies paid under the scope of the Common Agricultural Policy.

This work does not deal with any of the criticisms cited above, but rather addresses the question how equal the distribution of agricultural subsidies at the domestic level is, i.e., whether there is at least fairness in the distribution of payments to the municipalities or higher administrative units within a state. As it would not have been possible to scrutinize the European Union in its entirety and all of its member states in the course of this study, I decided to select a single country as an example. Since I am living in Austria, I originally wanted to investigate the situation of Austria. However, the acquisition of necessary data would have involved considerable costs and the data on agricultural subsidies were only available for reading, but not in downloadable form. After a long web search, the choice finally fell on Belgium as it was the country with the best data and which offered them all free of charge and in an uncomplicated manner for use and download.

With a national territory of 30.528 km² and about 11.3 million inhabitants (2016)[5], Belgium is one of the small to medium-sized countries in the European Union. Its share of the EU's total area is 0.7 percent, and of its total population 2.2 percent (2016)[6]. Its share of the total expenditure of the European Union of around € 142.5 billion (2014) and € 145.2 billion (2015) was around 5 percent (€ 7 billion)[2] for both years. The share of the population working in the agricultural sector was less than 1 percent in 2013[7].

Belgium's administrative structure is a peculiarity within the European Union. The country is, so to say, divided: The Dutch-speaking region of Flanders lies to the north and the French-speaking region of Wallonia to the south. On the other hand, the Brussels-Capital Region is a mixed-language entity. For a long time, there has been a political conflict between the Flemings in the north and the Walloons in the south, which also has economic backgrounds. Thus, some parties and politicians of Flanders demand the independence of their region from Belgium. They argue that, as an economically stronger region, they would finance Wallonia, which has been in recession since the decline of the coal and steel industry.

---

[5] http://statbel.fgov.be/nl/binaries/2_WEB_NL_kerncijfers_2016_tcm325-280618.pdf (retrieved May 2, 2017)
[6] http://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table&plugin=1&pcode=tps00001&language=en (retrieved May 3, 2017)
[7] http://statbel.fgov.be/nl/binaries/NL_Kerncijfers%20Landbouw_2015k_tcm325-270401.pdf (retrieved May 5, 2017)

1.2 Research Questions

The purpose of this master thesis is to clarify whether the agricultural subsidies paid out to farmers and agricultural businesses in Belgium within the framework of the EU's Common Agricultural Policy in 2014 and 2015 are distributed geographically evenly among the administrative entities. Firstly, the investigation takes place at municipal level and, secondly, at the arrondissement level. The subsidies of all beneficiaries of a political municipality and arrondissement, respectively, are summed up regardless of their purpose or origin (EAGF or EAFRD). The sum of subsidies of each municipality or arrondissement functions as response variable and is compared with various explanatory variables from the agricultural sector. These are the number of farms, the total area used for agricultural purposes, the cultivation areas of different crops and the number of different livestock animals. The study is carried out at three levels of aggregation, which are based on the classification of Statistics Belgium, the agency for statistics of the Belgian Federal Government: At the highest level, all variables about cultivation areas and livestock animals are grouped into one each, at the second level there is a simple disaggregation of the variables, and at the third level an enhanced one.

Of course, the Flemish-Walloon conflict is also of interest to the present work as the question arises whether one of the two parts of the country is benefiting excessively from the EU's Common Agricultural Policy. While this conflict is not explicitly the subject of this work and only an incidental aspect resulting from the selection of Belgium as target country, a potential unequal treatment of the two regions can also be identified on the basis of the results obtained.

When searching for data material for my evaluations, I came across geodata of Flanders on the website of the Flemish Office for Geoinformation with very detailed information on the cultivation areas of many crops, but without information on livestock animals. Unfortunately, I didn't succeed in finding a Walloon counterpart. Nonetheless, I absolutely wanted to include these data in my analyses, because, in contrast to the geodata used in the investigation of the whole Belgian territory, they contain information for all individual cultivation areas. Therefore, in addition to the investigation of the whole country of Belgium, I tried to conduct an analysis limited to Flanders based solely on data on crops and cultivated plants by making use of these geodata to see if it would produce similar results for the region of Flanders.

An important concern for me, alone from financial reasons, was the exclusive use of open source software. Since it is accessible to everyone and its acquisition does not incur any costs, it can also be easily used by everyone. Open source solutions exist for a great number of applications, which are very

often in no way inferior to their proprietary counterparts, indeed often surpassing their functionality. I am an absolute advocate of open source solutions and wanted to proof that no expensive software is required to carry out geostatistical analyses for a master thesis.

Also for reasons of cost, I have decided only to employ freely accessible and usable data. At this point, however, I must add that there is still a lot of catching up to do, concerning this matter. Data, whose collection or acquisition is funded by public means anyway should also be freely available to all interested parties, which, despite Open Data, is still a great exception.

2 Methods

2.1 Data

2.1.1 Data Acquisition

The data on all EU agricultural subsidies paid out to beneficiaries for the budget years 2014 and 2015 are provided by the Belgian Paying Agencies and have been downloaded free of charge as CSV files on their website[8]. The payout periods run from October 16 to December 15 of the respective budget year. The statistical data on agriculture originate from the website of the Belgian Office for Statistics[9] and are available for download in tabular form. The information contained in these tables is based on statistical updating and therefore, to a degree, subject to uncertainties. The reference period for the survey of cultivated plants in 2014 and 2015 was April / May and October / November for the survey of livestock. The agricultural data exclusively used for the investigation of Flanders are derived from geodata, which were obtained from the Geoportal of the Flemish government[10]. The reference period used starts on April 21 of the preceding year and ends on the same day of the following year.

Geodata with the municipal boundaries of Flanders (as of May 22, 2003) with a scale of 1:5,000 were also downloaded from the Flemish government's geoportal. The Walloon counterpart with a scale of 1:10,000 (as of November 14, 2013) comes from the FTP server[11] of the geoportal of the Walloon Region. Geodata with a scale of 1:2,000, which contain the agricultural areas of Flanders broken down by crop species, have already been mentioned above. As no geodata could be found on the Walloon side, I had to resort to the sectoral plan of Wallonia with a scale of 1:10,000 (as of November 14, 2013), provided on the FTP server of the geoportal of the Walloon Region.

---

[8] `http://www.belpa.be/belpa/pub/EN/data.html#download` (retrieved November 29, 2016)

[9] `http://statbel.fgov.be/nl/statistieken/cijfers/economie/landbouw/bedrijven/` (retrieved December 14, 2016)

[10] `http://www.geopunt.be/` (retrieved October 6, 2016)

[11] `ftp://docum1.wallonie.be/DONNEES/Amenagement/Plans/Plans_secteurs/Plans_Secteur_Vectoriel/` (retrieved October 10, 2016)

2.1.2 Data Preparation

Since the geodata with the administrative boundaries of the municipalities of Wallonia and Flanders differ in their scales, they were initially processed with QGIS version 2.18 (QGIS Development Team 2016). The adjoining regions of Wallonia and Flanders were aligned with one another. Municipalities located at sea had their marine areas removed to avoid statistical distortions. For the geodata, which contains the sectoral plan of Wallonia, all areas not classified "agriculture" or "specific agro-economic activity" were removed. The two remaining classes were combined into a common one and intersected with the municipal boundaries of Wallonia. The geodata containing the agricultural surfaces of Flanders were used in two ways: For the analysis of all of Belgium unclassified areas and all those areas containing the class "water" were removed. Once again, all the remaining classes were merged into one. For the exclusive analysis of Flanders, the geodata adjusted for the class "water" and the unclassified areas were used, with the remaining classes left in their condition. Both the generated geodata with the agricultural areas for the analysis of the whole of Belgium as well as for the analysis of Flanders were then intersected with the municipal boundaries of Flanders. The agricultural areas of Wallonia and Flanders, which had been aggregated to form a common class and intersected with the municipal borders were combined into a single file.

Due to data privacy concerns, recipients who received up to € 1,250 in a budget year are published in anonymous form[12] (European Parliament & Council of the European Union 2013; European Commission 2014). Since there is no information on the postal code and the municipality for anonymous recipients, they were removed from the data on the beneficiaries. Likewise, those recipients whose residence or business seat is not in Belgium were removed as well. Also, all beneficiaries from those 44 municipalities for which no agricultural data are available, including all 19 from Brussels-Capital Region (Table 2.1; Fig. 2.1), were dropped. In addition, this was the case for the municipality of Martelange in 2014 and Vilvoorde in 2015. Of course, these areas were removed from the geodata too. Since the statistical data on agriculture also contain a row with confidential municipalities and, therefore, no allocation is possible, they were also deleted. The generated geodata were joined with the agricultural and subsidy data using QGIS and finally saved as new files.

---

[12] `https://ec.europa.eu/agriculture/sites/agriculture/files/cap-funding/beneficiaries/shared/`
`threshold_en.pdf` (retrieved April 12, 2017)
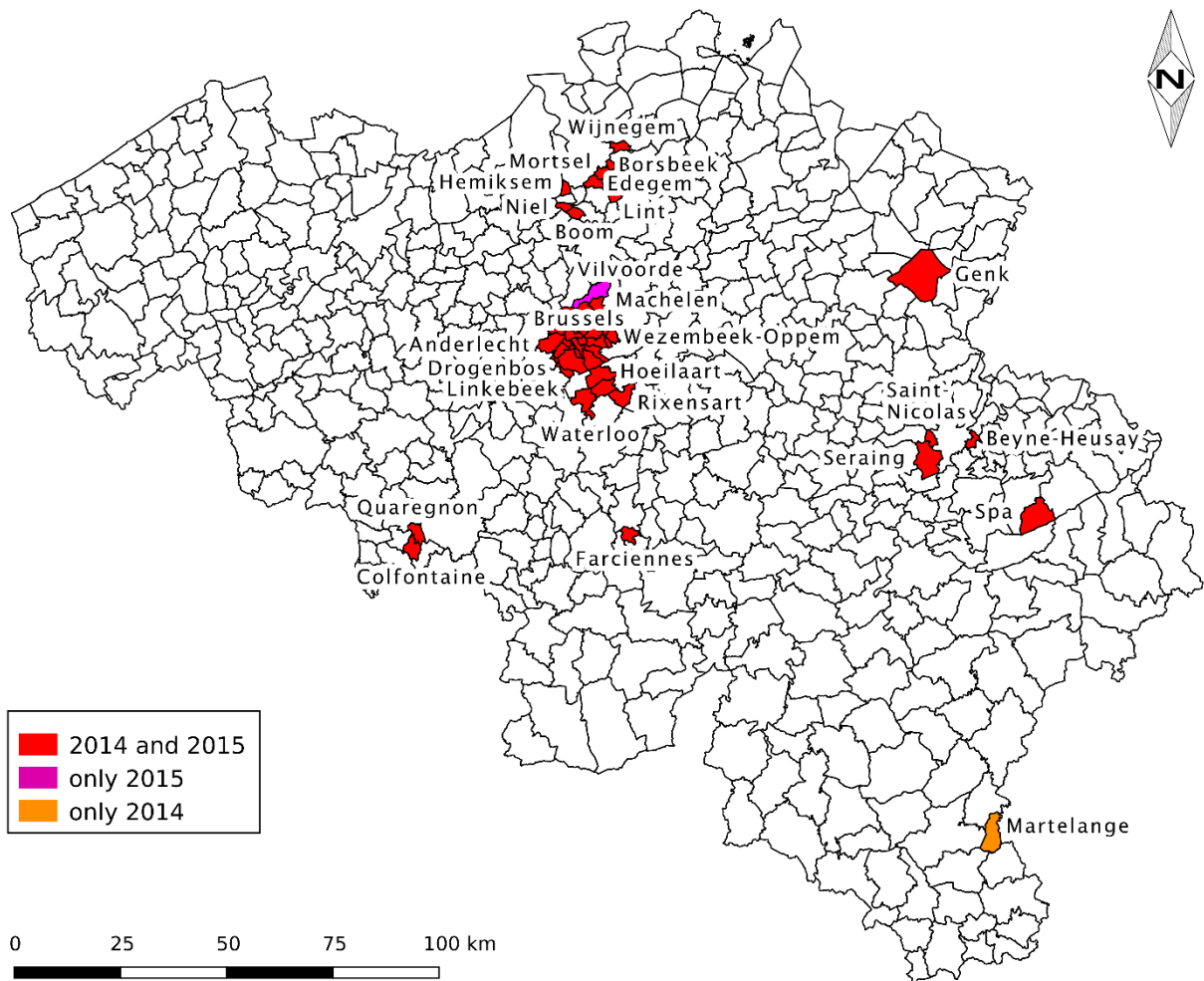
Fig. 2.1: Excluded municipalities.

It is important to mention that the areas of the Walloon sectoral plan do not indicate the actual use of the areas it contains, but merely the intended use, which of course leads to uncertainties. However, here I see no significant problem since the extent of the uncertainty is probably the same for all municipalities in Wallonia and thus, ensures a balance. Only in the comparison of Flemish and Walloon municipalities could there be considerable uncertainties. Unfortunately, in the absence of alternatives, there was no other choice. This circumstance is also responsible for differences between the size of the polygons of the Walloon municipalities and the areas given in the tables from the Belgian Office for Statistics, which is why a compromise had to be found here. Therefore, the area sizes provided by the Belgian Office for Statistics in combination with the polygons of the Walloon sectoral plan were used for the geostatistical analyses. This problem is also present in the Flemish geodata, which is why the overall analysis of Belgium was carried out according to the same pattern. The Walloon sectoral plan does not provide detailed information on the cultivated plants but has all areas generally designated as agricultural areas. In an analysis of the whole country of Belgium all regions have to be

treated statistically alike, so despite all the detailed information on the Flemish side all areas were aggregated regardless of their use. Thus, when analyzing the whole state of Belgium, each polygon was treated as if within its boundaries every crop category that had entered the statistics was cultivated. Since no detailed geodata were available for farm animals at all, they were treated in the same manner. In the analysis limited to Flanders, the detailed geodata could be used, but for the reasons mentioned above, without livestock. Unlike in the case of the geodata of the whole country of Belgium, in the analysis limited to Flanders, each area has only a single form of use, which, unfortunately, makes it impossible to carry out a regression involving all forms of use. In order to solve this, the table data had to be designed in such a way that only one field per row had a subsidy amount given, i.e., the field the form of usage applied. That is, a table was produced in which each area faced all types of use, but all the fields except the one whose use applied for the respective area had the subsidy amount of € 0.

Table 2.1: Overview of analyzed and excluded subsidies.

	2014	% 2014	2015	% 2015
Total subsidies	€ 747,650,156.45	100.00	€ 789,532,360.83	100.00
Foreign and anonymized recipients	€ 4,356,259.94	0.58	€ 3,557,907.96	0.45
Excluded municipalities	€ 8,922,342.31	1.19	€ 4,561,219.14	0.58
Not analyzed subsidies total	€ 13,278,602.25	1.78	€ 8,119,127.10	1.03
Analyzed subsidies Wallonia	€ 381,888,546.27	51.08	€ 386,168,171.53	48.91
Analyzed subsides Flanders	€ 352,483,007.93	47.15	€ 395,245,062.20	50.06
Analyzed subsidies total	€ 734,371,554.20	98.22	€ 781,413,233.73	98.97

2.1.3 Data Summary

In 2014 the sum of all agricultural subsidies paid by the Belgian Paying Agencies under the EU's common agricultural policy amounted to € 747,650,156.45, which went to 41,792 recipients. The subsidies not included in the analysis due to the above reasons are € 13,278,602.25. That is, the sum analyzed for the year 2014 is € 734,371,554.20, which is 98.2 percent of the total subsidy sum. In 2015, a total of € 789,532,360.83 was paid out to 43,150 beneficiaries and the sum not taken into account was € 8,119,127.10. Thus, for 2015 € 781,413,233.73 (99 percent) are analyzed, of which in 2014 (2015) € 352,483,007.93 (€ 395,245,062.20) are attributable to Flanders and € 381,888,546.27 (€ 386,168,171.53) to Wallonia (Table 2.1). The analysis limited to the region of Flanders required an allocation of the subsidies to the individual areas of the cultivated plants. It should be noted that small polygons with an area of less than one square meter had been removed from the data prior to the analysis. The sum of the agricultural areas of each municipality before removing these small polygons differs by a few square meters from the area sums afterwards. Since the second value was used as the

total area for the factoring of the subsidies, but the area sums of the individual crops of each municipality contain these small polygons, there are marginal differences in the total sum analyzed, which amounts to € 352,483,302.88 (difference: € 294.95) in 2014 and € 395,246,380.39 (difference: € 1,318.19) in 2015. This difference has absolutely no effect on the statistical analysis whatsoever, it is merely an insignificant procedural error, but has to be mentioned anyway. The highest amount received by a beneficiary in 2014 (2015) was € 28,812,795 (€ 36,445,938.94), followed by ten (nine) recipients, each with at least € 1 million. Six (eight) recipients received between € 1 million and more than € 0.5 million, 503 (564) recipients between € 0.5 million and more than € 100,000, 2,477 (2,689) recipients between € 100,000 and more than € 50,000 and another 16,262 (16,045) between € 50,000 and more than € 10,000 in subsidies. The remainder is distributed among the remaining 22,532 (23,836) recipients, many of whom also had negative sums.
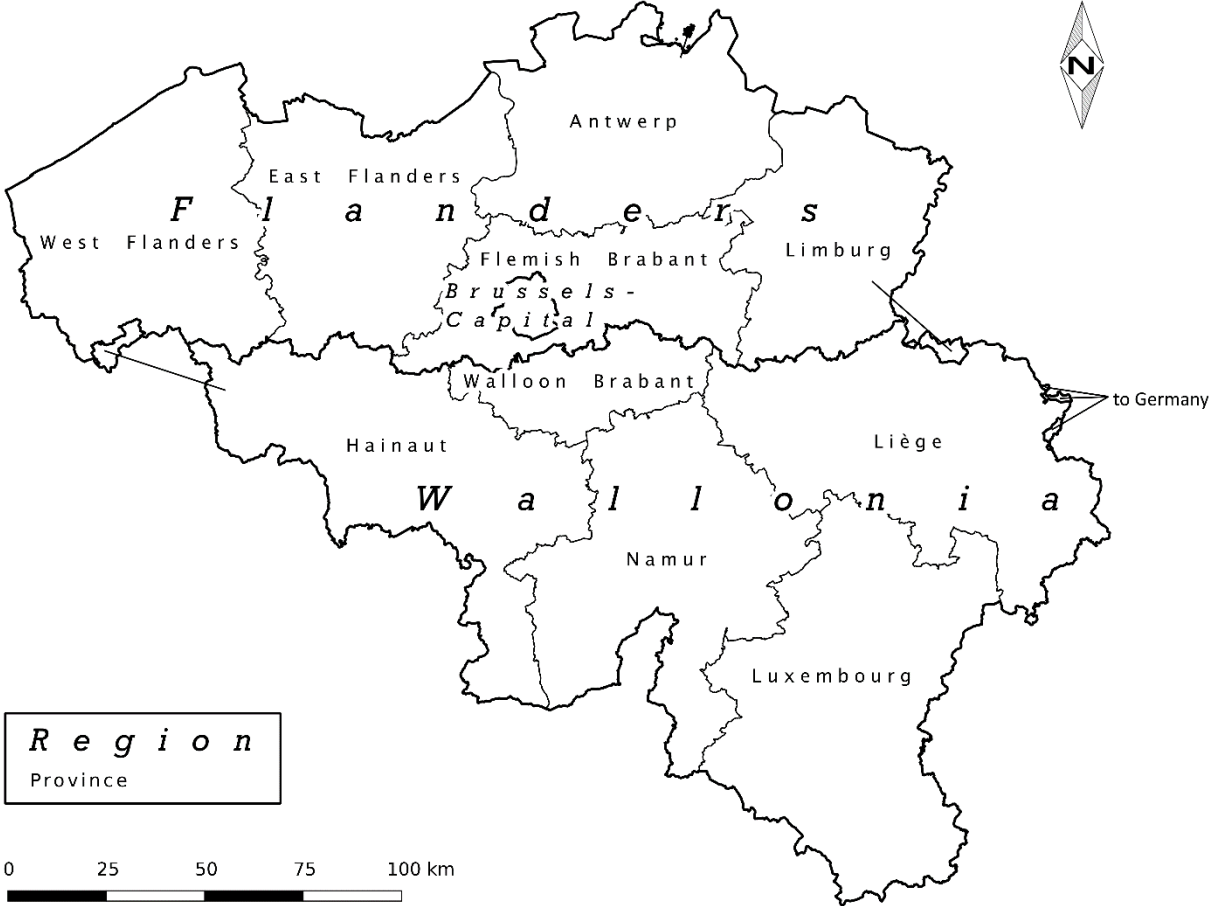
Fig. 2.2: Regions and provinces of Belgium.

Fig. 2.3: Arrondissements of Belgium.

As mentioned above, Belgium is divided into the three regions of Flanders with an area of 13,522 km² and a population of approximately 6.5 million inhabitants (479 inhabitants per km²), Wallonia with an area of 16,844 km² and a population of about 3.6 million inhabitants (214 inhabitants per km²) and Brussels-Capital with an area of 161 km² and a population of ca. 1.1 million inhabitants (7,074 inhabitants per km²). The regions of Flanders and Wallonia, for their part, are divided into five provinces each, each of which is composed of several arrondissements (Fig. 2.2 and 2.3) (a total of 43 including the arrondissement Brussels-Capital). Thereof, Flanders has 22 and Wallonia 20 arrondissements. The number of Belgian political municipalities (Fig. 2.4 and 2.5) is 589, of which 308 are in Flanders, 262 in Wallonia, and 19 in the region of Brussels-Capital. Of the 589 municipalities, all 19 of the Brussels-Capital Region were not analyzed because they had no agricultural data available. This was the case with a further 15 municipalities in Flanders and 10 municipalities in Wallonia. For the same reason, the Walloon municipality of Martelange in 2014 and the Flemish municipality of Vilvoorde in 2015 were excluded from the analysis so that a total of 544 municipalities were examined

each year. Broken down by region, there were 293 (292) Flemish and 251 (252) Walloon municipalities in 2014 (2015).

The data from the Belgian Office for Statistics show that the number of farms in the municipalities of northern and northeastern Flanders is generally higher than in Wallonia and central Flanders. Pig and poultry farming is concentrated mainly on West Flanders and the north of the province of Antwerp. In these areas, in northern East Flanders and Hainaut as well as in the east of the province of Luxembourg, cattle keeping is very well established. In the central part of Belgium, on the other hand, there are significantly fewer cattle. The agricultural area is larger in Wallonia and in the north and north-east Flanders than in the rest of Belgium. A quite similar pattern is also evident in the subsidies, especially the municipalities of Leuven 2014 (2015) with almost € 31.5 million (€ 39.3 million) and Sint Truiden with over € 13.3 million (€ 23.4 million) are clearly visible. The cultivation area of perennial plants is largely distributed uniformly, but there is a focus around Sint Truiden. The perennial grassland area is highest in southeastern Wallonia, whereas greenhouse crops are found especially in northern Flanders. The cultivation of industrial crops, potatoes, vegetables and cereals is very common along the Flemish-Walloon border, especially in the Flemish part, as well as in West Flanders. The cultivation of fodder plants correlates geographically with cattle farming. The largest area of cultivation of legumes is in southern Limburg, Luxembourg, Liège and Namur. Fallows are found in Hainaut, Namur and Walloon Brabant.
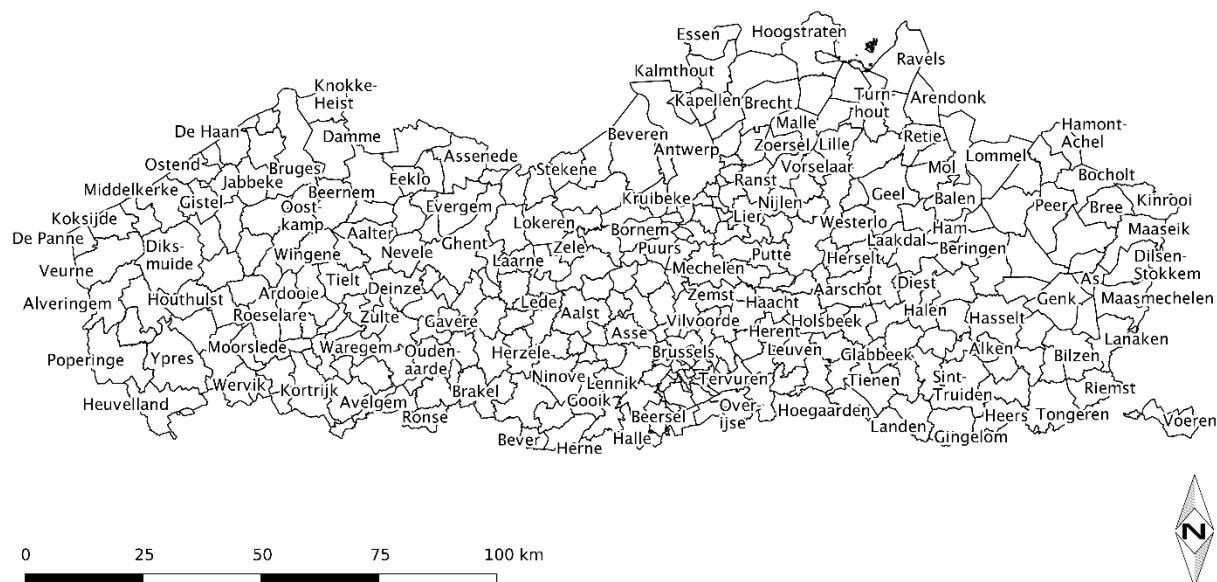
Fig. 2.4: Municipalities of the regions Flanders and Brussels-Capital.

The data about Flanders suggest that potatoes, vegetables, herbs and horticultural plants are cultivated mainly in West Flanders, but fruits and nuts are found especially around Sint Truiden. Cereals, seeds and legumes have their main focus in the north-eastern part of West Flanders and in the southern part of Limburg, grassland and maize in the northern and western part of Flanders.
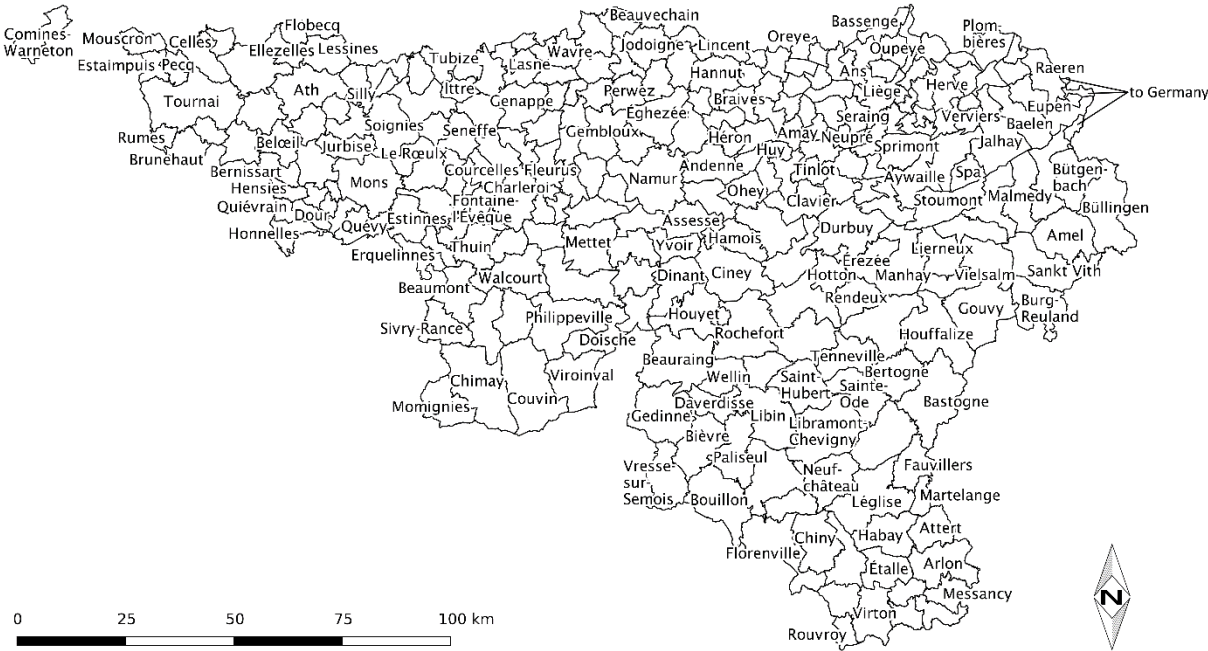
Fig. 2.5: Municipalities of the region Wallonia.

2.2 (Geo-)Statistical Analysis

2.2.1 Preliminary Notes

All statistical analyses were carried out using the free statistics package R, version 3.3.1 (64-bit) (R Core Team 2016) and with the help of the also freely available integrated development environment RStudio, version 1.044 (RStudio Team 2016). An R script was created that, except for a few points, automatically performs the statistical analyses after selecting a desired data set. The code of this R script, together with the necessary auxiliary scripts, is available in the Appendix. All analyses were carried out for two budget years. In this way, contingencies or outliers can be mitigated. It was not possible to analyze any further years since only the last two subsidy periods are available on the web. Unless otherwise stated, the common significance level of α < 0.05 was used for all significance tests.

The modifiable areal unit problem (MAUP) explained by Openshaw (1984) describes the dependence of the results of spatial analyses on the form of the aggregation of the initial data. Though using the very same data for an analysis, depending on the subdivision of the investigated area, completely different results can be obtained. In order to deal with this problem, all analyses were carried out at the municipal level as well as at the arrondissement level. Originally, it was also planned to carry out the analyses at provincial and regional level. However, this plan was abandoned because the number of units (ten provinces or two regions) simply would have been too small for a meaningful analysis.

Table 2.2: **A**ggregation levels for the analysis of the entire country of Belgium (abbrevations in brackets).

Aggregation level 1	Aggregation level 2	Aggregation level 3
	Businesses (bus)	
Agricultural area (aar)	Arable land (arl)	Cereals (cer)
		Industrial crops (ind)
		Potatoes (pot)
		Legumes (leg)
		Fodder plants (fod)
		Vegetables (veg)
		Horticultural plants (hor)
		Fallows (fal)
	Perennial plants (ppl)	Tree nurseries (nur)
		Orchards (orc)
		Small fruits (sfr)
	Perennial grassland (pgr)	
	Greenhouse crops (gcr)	Greenhouse vegetables (gvg)
		Horticultural plants in greenhouses (hrg)
Cattle (cat)		Calves < 1 yr. (cal)
		Cattle 1 - < 2 yr. (c12)
		Cattle ≥ 2 yr. (c>2)
Pigs (pig)		Piglets (pgl)
		Breeding sows > 50 kg (sow)
		Pigs 20 - 50 kg and porkers (por)
Poultry (pou)		Laying hens (lay)
		Broilers (bro)

Note: All variables about plants and crops are given in ares whereas "Businesses" and variables about livestock are count data.

Table 2.3: Predictors used in the exclusive analysis of Flanders (abbrevations in brackets).

Potatoes (pot)	Woody plants (woo)	Flax and hemp (flh)
Fruits and nuts (frn)	Agricultural infrastructure (inf)	Fodder plants (fod)
Cereals, seeds and legumes (csl)	Maize (mai)	No information (nin)
Grassland (gra)	Other plants (oth)	
Vegetables, herbs and horticultural plants (vhh)	Sugar beets (sug)	

Note: All variables are given in ares.

As already mentioned, the investigation is carried out at three levels of aggregation. At the highest level all variables on cultivated areas or livestock are combined into one variable each, at the second level a simple disaggregation of these variables takes place and at the third level an enhanced one (Table 2.2). Experimentally, at the highest aggregation level, the analysis was also carried out using a hexagonal grid (Fig. 2.6) whose parallel edges have a distance of five kilometers. The grid was intersected with the agricultural areas fitted to the municipal boundaries and the areas obtained were then analyzed. For this analysis, the variable sums were divided proportionally between the individual areas. Parallel to this, an analysis limited to the region of Flanders was performed, in which, as already mentioned, other data and predictors were used (Table 2.3). Overall, the analysis was conducted at 16 different levels. For the sake of clarity, each of these levels has been given a code (Table 2.4.) where the letter "m" denotes municipalities, "a" arrondissements, "g" grid, "f" Flanders, the number after the letter "a" or "m" the aggregation level, and the numbers "14" and "15" the years 2014 and 2015 respectively.

Table 2.4: Analysis levels.

m1 14	Municipalities of Belgium at aggregation level 1 2014
m1 15	Municipalities of Belgium at aggregation level 1 2015
m2 14	Municipalities of Belgium at aggregation level 2 2014
m2 15	Municipalities of Belgium at aggregation level 2 2015
m3 14	Municipalities of Belgium at aggregation level 3 2014
m3 15	Municipalities of Belgium at aggregation level 3 2015
a1 14	Arrondissements of Belgium at aggregation level 1 2014
a1 15	Arrondissements of Belgium at aggregation level 1 2015
a2 14	Arrondissements of Belgium at aggregation level 2 2014
a2 15	Arrondissements of Belgium at aggregation level 2 2015
a3 14	Arrondissements of Belgium at aggregation level 3 2014
a3 15	Arrondissements of Belgium at aggregation level 3 2015
g14	Municipalities of Belgium rasterd by a hexagon grid at aggregation level 1 2014
g15	Municipalities of Belgium rastered by a hexagon grid at aggregation level 1 2015
f14	Municipalities of Flanders 2014
f15	Municipalities of Flanders 2015

Note: m, municipality; a, arrondissement; g, grid; f, Flanders; 1, aggregation level 1; 2, aggregation level 2; 3, aggregation level 3; 14, 2014; 15, 2015.

2.2.2 Global Linear Regression

In order to numerically identify correlations between predictors and to remove redundant variables prior to the analysis, correlation matrices were created. There is no universally valid limit from which variables are regarded as actually correlated. However, it can be assumed that a significant correlation exists from a correlation coefficient after Pearson (1895) of about $r \geq 0.8$. Combinations of variables

whose correlation coefficient exceeded this value were declared inadmissible and removed from the potential regression models.

The beginning of the actual analysis was the calculation of a (multiple) global linear regression model. The model selection for the global linear regression was carried out using the R package glmulti, version 1.0.7 (Calcagno & de Mazancourt 2010), which calculates all possible models and their respective corrected Akaike information criterion (AICc) (Akaike 1973; Sugiura 1978). The general view is that models whose AICc deviate less than the value of 2 of each other (Raftery 1995; Burnham & Anderson 2004) bear the same information content and thus are equal.

All regression models, whose AICc were up to a maximum of the value 2 above the model with the lowest AICc (M), were subjected to a more detailed analysis. Models, in which not all predictors were significant, were deleted, with the intercept perceived not of any importance. The model, which had the lowest AICc value and at the same time had all predictors significant, was finally selected. If there was no model within the interval M + 2 in which all predictors were significant, non-significant predictors were taken from the model with the lowest AICc value, and the calculation was performed again until finally the best possible model was found, subjected to a deeper analysis. For the purpose of verifying the goodness-of-fit of the regression model, the adjusted measure of determination R^2 was used. β-coefficients were calculated to study the relative magnitude of the influence of the respective predictors. The variables were standardized so that their variances s^2 have the value 1. The standardized β-coefficients thus give information about how many standard deviations s the dependent variable changes with the respective predictor.

The regression models selected for each analysis run were subsequently subjected to regression diagnostics. For this purpose, the variance inflation factors (VIF) of the predictors were calculated first. The VIF is a measure of the multicollinearity of the variables. A somewhat arbitrary, but generally accepted approach is that, from a VIF of 10 multicollinearity is given (Pardoe 2006) and therefore the redundant variables must be removed.

Diagnostic plots (Fox 1991) were prepared for visual inspection. The plot "Residuals vs Fitted" indicates whether the relationship between the predictors and the dependent variable is linear. The diagram, which plots the square root of the standardized residuals against the fitted values, indicates whether heteroscedasticity is given. In case of a completely random pattern, or when the y-values have the same spread at each point of the x-axis, the residuals are homoscedastic. The quantile-quantile plot indicates whether the residuals are normally distributed. The plot "Residuals vs Leverage" indicates influential points or outliers that could distort the result. Points that lie outside the lines marked as "Cook's Distance" are such leverage points. Outliers can be detected with all four plots. In

addition, the studentized residuals were examined for outliers too. Values larger than 3 and less than -3 were displayed in the map of Belgium or Flanders explicitly in red.

The residuals were then examined for first order autocorrelation by conducting a Durbin-Watson test (Durbin & Watson 1950; 1951). The global spatial autocorrelation was analyzed with the Moran's I index (Moran 1950). For visual analysis, Moran scatter plots (Anselin 1996) were generated which also provide evidence of outliers. The studentized Breusch-Pagan test proposed by Koenker (Breusch & Pagan 1979; Koenker 1981) was employed to numerically test the residuals for heteroscedasticity. The Jarque-Bera test (Bera & Jarque 1980) and the Shapiro-Wilk test (Shapiro & Wilk 1965) served as numerical verification of the normal distribution of the residuals. With the standard deviation being used as scaling, the residuals were mapped to determine in which regions they take on high values.

2.2.3 Geographically Weighted Regression

The global linear regression implies a constant linear relationship for the entire study area. Frequently, however, this assumption does not apply at all. For geographical analyses the geographically weighted regression (GWR) (Brunsdon et al. 1996) is therefore appropriate. For the GWR, a separate, local regression model is calculated for each point of the investigation area. In order to accomplish this, a so-called "moving window" approach is followed in which only a portion of the examination area around the respective point of interest is included in the calculation. The points surrounding the respective point of interest are weighted in dependence of their distance from the point of interest. Nearer points gain greater weight, with increasing distance the weighting factor decreases. The weighting of the points depends on the selected weighting function. The range that is taken into account for each local regression is limited by the bandwidth, the larger it is, the larger the considered space (kernel (Fig. 2.7)) becomes. The bandwidth can be defined either subjectively by the user or by objective criteria such as the cross-validation or the AICc. The bandwidth can also be chosen in such a way that the kernel always has the same size (fixed distance or fixed kernel (Fig. 2.8)) or a fixed number of points are used for the calculation (adaptive distance or adaptive kernel (Fig. 2.9)).

Table 2.5: Weighting functions for spatial kernels used in GWR analysis.

Weighting function	Formula	continuous/disontinuous	used
Gaussian (= normal)	$w_{ij} = \exp(-0.5 * d_{ij} / h)^2$	continuous	yes
Bisquare	if $d_{ij} < h$: $w_{ij} = (1 - (d_{ij} / h)^2)^2$, otherwise: $w_{ij} = 0$	discontinuous	yes
Exponential	$w_{ij} = \exp(-d_{ij} / h)$	continuous	no
Tricube	if $d_{ij} < h$: $w_{ij} = (1 - (d_{ij} / h)^3)^3$, otherwise: $w_{ij} = 0$	discontinuous	no
Boxcar	if $d_{ij} < h$: $w_{ij} = 1$, otherwise: $w_{ij} = 0$	discontinuous	yes

Note: w_{ij}, weight of point j around point i; d_{ij}, Euclidian distance between location i and location j ; h, bandwith;
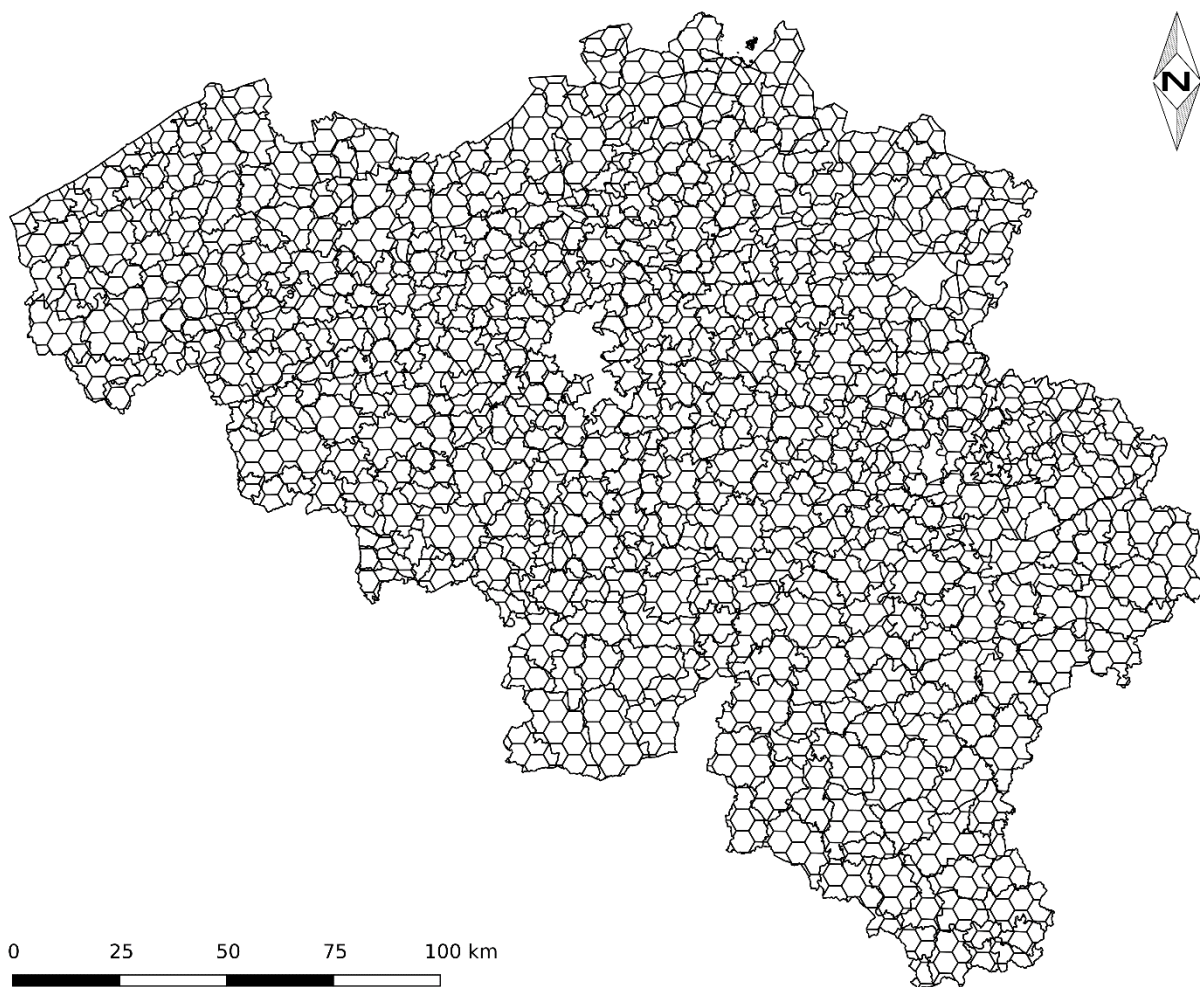
Fig. 2.6: Municipalities intersected with a hexagon raster whose parallel edges have a distance of 5 km.

Since the selection of a weighting function by the user always contains a subjective element, which is what I wanted to avoid, no specific weighting function was defined for this work, but three of the five functions of the R package GWmodel version 1.2.5 (Gollini et al. 2015) used for the implementation of geographically weighted regressions were used both for determining the optimal bandwidth as well as for the actual GWR itself. These are the weighting functions Gaussian, bisquare, boxcar, tricube and exponential. The Gaussian and exponential function are continuous which means that despite prior determination of bandwidth all points are included in the calculation, but points which are outside the specified bandwidth are extremely downweighted. On the other hand, the functions bisquare, boxcar and tricube are discontinuous, points which are outside the bandwidth are completely excluded from the calculation. Due to pretests, I could see that the Gaussian and exponential functions yielded nearly the same bandwidths as well as GWR results which is why I only considered the Gaussian function and eliminated the exponential function. This was also the case with the functions bisquare and tricube, so

I discarded the second one. In the end, the three functions Gaussian, bisquare and boxcar were used. Although, the use of several weighting functions has the disadvantage that one does not receive one but several results, at the same time it has the advantage that they can be matched with one another which results in a higher degree of safety. The Euclidean distance pre-set by the program was employed as distance measure, also being used for all subsequent steps.

In order to determine the bandwidths to be used, the global regression models were applied, using the AICc approach and choosing an adaptive kernel because the regression points to be examined are not regularly distributed across space. The finally selected bandwidths determined with the respective weighting function were then used together with the corresponding weighting function for the GWR model selection.

X regression point w_{ij} is the weight of data point *j* at regression point *i*
● data point d_{ij} is the distance between regression point *i* and data point *j*
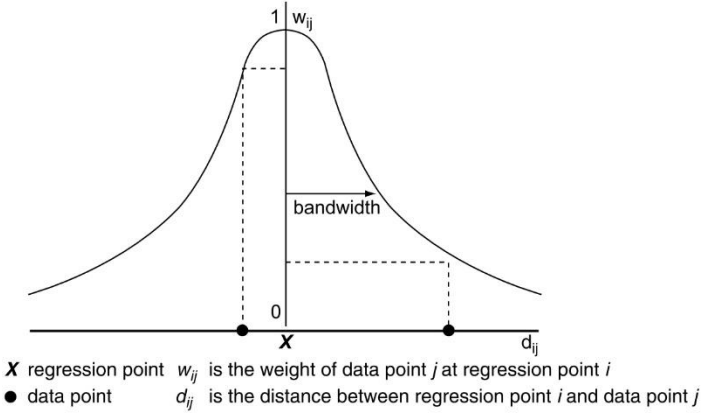
Fig. 2.7: Spatial kernel (from Fotheringham et al. 2002).

x regression point
● data point
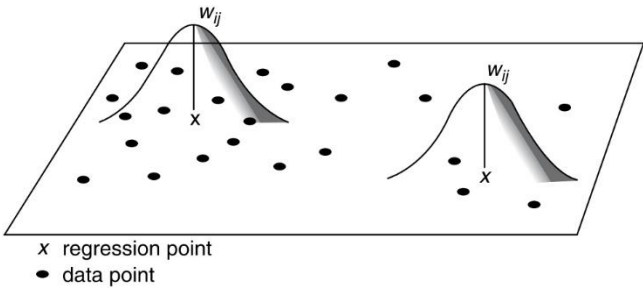
Fig. 2.8: Fixed kernel (from Fotheringham et al. 2002).
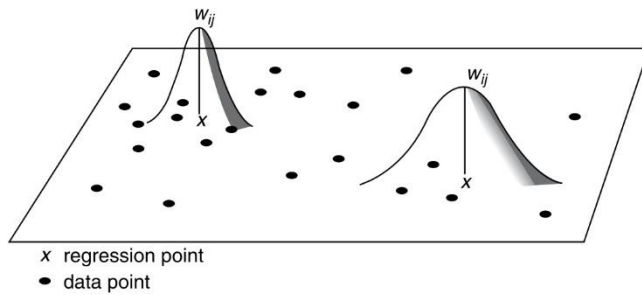
Fig. 2.9: Adaptive kernel (from Fotheringham et al. 2002).

The AICc approach was also applied for the GWR model selection. Here, an adaptive kernel was calculated again. From all possible models, for each weighting factor and the associated bandwidth the model with the lowest AICc value was chosen. With these selected GWR models, the actual geographically weighted regressions along with the adjusted coefficient of determination were calculated.

Subsequently, the GWR models were checked for validity. For the numerical investigation of the spatial variability of the predictors and thus, the justification of using a GWR, Monte Carlo randomization tests (Brunsdon et al. 1996) with 999 randomizations were performed. Once again, the intercept, which is checked by default, was considered to be irrelevant. The local correlation coefficients (r) and the local condition numbers (CN) (Brunsdon et al. 2012) were calculated for the diagnosis of potential collinearity between the predictors, again using an adaptive kernel. Local r > 0.8 as well as local CN > 30 indicate multicollinearity, whereby these thresholds are merely indicative values. As in the case of global regression, the studentized residuals of the GWR were also examined for outliers. Studentized residuals, whose value exceeded 3 or fell below -3 were mapped in red. The residuals were also checked for autocorrelation with the Moran's I statistic and the results again presented as Moran scatter plots. The residuals were finally depicted as maps using the standard deviation as scaling. Maps of the regression values with the same scaling were also generated. Since the regression values themselves have very little informative value, the t-values of the local regressions were used to derive p-values from them. Using the p-values, it was now possible to remove those regression values which were not significant. However, only those regression values are of ultimate importance that differ significantly from the others. Therefore, those regression values which do not deviate at least two standard deviations from the arithmetic mean of all regression values were deleted. The remaining regression values now depicted those points, which actually differed significantly from all the points with usual values and were again displayed in map form.

Not always all predictors are either just stationary or nonstationary. In such cases, neither a global regression nor an ordinary GWR is the adequate solution. The mixed GWR (Fotheringham et al. 2002) allows the combined use of geographically variable and stationary predictors. Therefore, the results of the Monte Carlo randomization tests were used to investigate the spatial variability of the predictors, so that the stationary predictors could be selected for the mixed GWR. As an exception, here, the level of significance was raised to $\alpha \geq 0.075$, for a too strict handling seemed to me not to be very useful in this case. Afterwards, the models from the ordinary GWR were used for the calculation of mixed GWR models, this time employing the previously determined stationary predictors. Regrettably, there is no computation of adjusted R^2, local correlation values, local condition numbers, Moran's I and p values available for the mixed GWR, so I decided to use the values of the ordinary GWR as an orientation. As in the case of the ordinary GWR, the regression values of the mixed GWR were graphically displayed in maps.

The OLS, GWR and, in case of local and global variables occurring together, also MGWR models, were compared by their respective AICc values. If the AICc of the GWR or MGWR model was at least the value of 2 below that of the OLS model, the GWR or MGWR model was considered superior. In order to reduce the number of graphics, only maps and diagrams of those models are shown that were superior to the models of the other analysis approaches on their respective analysis level.

3 Results

Table 3.1 shows the results of the model selection for the OLS regression. In four cases (a1 15, a3 14, a3 15 and g14), the regression model with the lowest AICc value was not selected for further investigation because at least one predictor was not significant or had a negative regression value. The adjusted coefficient of determination of the models ultimately chosen at the municipal level (0.25 (m1 14) - 0.37 (m2 14 and m3 14)) is lower than at the arrondissement level (0.6 (a1 15) - 0.84 (a3 14)). With 0.42 (g14), 0.32 (g15), 0.34 (f14) and 0.29 (f15), the R² values of the analyses with the hexagon grid and of Flanders are similarly low as in the analyses of the municipalities. Table 3.1 shows that the highest β-coefficients are that of the predictor "agricultural area" at the first aggregation level and in the analysis approach with the grid and that of "arable land" at the second aggregation level. The β-coefficients at the municipal level are always lower than those at the level of the arrondissements. The variance inflation factors show that multicollinearity does neither occur at the municipal nor at the arrondissement level. At both the municipal and arrondissement level, all Durbin-Watson tests negate the existence of autocorrelation. This also applies to the Moran's I statistics. According to the Breusch-Pagan tests, in the analysis of municipalities and arrondissements heteroscedasticity occurs in four cases (a2 14, a2 15, a3 14, a3 15). All Jarque-Bera and Shapiro-Wilk tests for normality of the residuals were negative. No OLS model that fulfills all conditions could be found at the analysis levels g14, g15, f14 and f15. According to the Durbin-Watson test as well as Moran's I, in all of those models autocorrelation is present. Additionally, the Breusch-Pagan test shows heteroscedasticity is a problem at analysis level f15.

Table 3.1: Results of the best OLS regression models.

Analysis level	AICc / Adjusted R² / Durbin-Watson autocorrelation / Global Moran's I	Koenker's Breusch-Pagan / Jarque-Bera / Shapiro-Wilk	Variable (VIF)	Estimate (β-coefficient)	Standard error	t-value
m1 14	16,955 / 0.34 / -0.02 / 0	0.05 / 3,845,700*** / 0.11***	int	25,830	99,800	0.26
			aar (N/A)	5.42*** (0.58)	0.33	16.69
m1 15	17,275.6 / 0.25 / -0.01 / 0	0.24 / 2,483,100*** / 0.12***	int	-24,870	135,200	-0.18
			bus (2.31)	4,491* (0.12)	2,131	2.11
			aar (2.31)	4.70*** (0.41)	0.65	7.23

Analysis level	AICc / Adjusted R² / Durbin-Watson autocorrelation / Global Moran's I	Koenker's Breusch-Pagan / Jarque-Bera / Shapiro-Wilk	Variable (VIF)	Estimate (β-coefficient)	Standard error	t-value
m2 14	16,930.9 / 0.37 / -0.02 / 0	0.55 / 4,653,000*** / 0.14***	int	-9,690.76	97,727.48	-0.10
			arl (1.07)	5.19*** (0.41)	0.45	11.60
			ppl (1.01)	25.05*** (0.23)	3.69	6.79
			pgr (1.07)	5.33*** (0.29)	0.64	8.32
m2 15	17,215.1		int	-38,880		
			arl	5.22***		
			ppl	44.88***		
			pgr	4.83***		
			pig	4.79		
	Model with 2nd lowest AICc					
	17,215.8 / 0.33 / 0 / 0.01	1.21 / 3,531,900*** / 0.16***	int	-45,760	126,100	-0.36
			arl (1.2)	5.55*** (0.36)	0.59	9.50
			ppl (1.01)	45.14*** (0.34)	4.70	9.61
			pgr (1.2)	4.93*** (0.18)	1.08	4.57
m3 14	16,931.1 / 0.37 / -0.03 / -0.01	1.27 / 4,388,600*** / 0.16***	int	-2,244.88	99,580.07	-0.02
			bus (1.6)	5,342.48*** (0.18)	1,272.13	4.20
			cer (1.29)	8.05*** (0.3)	1.04	7.75
			leg (1.39)	182.50* (0.09)	81.27	2.25
			orc (1.17)	19.51*** (0.18)	4.05	4.81
			pgr (1.59)	4.10*** (0.23)	0.78	5.25
m3 15	17,213.2 / 0.34 / -0.01 / 0	1.84 / 3,579,200*** / 0.17***	int	-8,440.76	125,779.99	-0.07
			cer (1.16)	8.94*** (0.27)	1.24	7.21
			fod (3.24)	4.90* (0.13)	2.42	2.02
			orc (1.02)	46.16*** (0.34)	4.82	9.57
			pgr (2.58)	4.43** (0.16)	1.58	2.81
			por (1.27)	12.82** (0.11)	4.58	2.81
a1 14	1,417.4 / 0.73 / -0.22 / 0.04	1.53 / 890.9*** / 0.53***	int	-774,700	1,881,000	-0.41
			aar (N/A)	5.77*** (0.86)	0.54	10.59
a1 15	1,440.9		int	-1,502,000		
			bus	5,172		
			aar	4.88***		
	Model with 2nd lowest AICc					
	1,441.8 / 0.6 / -0.18 / 0.11	1.25 / 540.2*** / 0.54***	int	521,700	2,513,000	0.21
			aar (N/A)	5.67*** (0.78)	0.72	7.88
a2 14	1,400.1 / 0.83 / -0.19 / -0.03	17.22*** / 268.3*** / 0.74***	int	-1,817,000	1,500,000	-1.21
			arl (1.02)	5.69*** (0.65)	0.56	10.09
			ppl (1.03)	32.47*** (0.4)	5.30	6.13
			pgr (1.03)	5.62*** (0.6)	0.61	9.22

Analysis level	AICc / Adjusted R² / Durbin-Watson autocorrelation / Global Moran's I	Koenker's Breusch-Pagan / Jarque-Bera / Shapiro-Wilk	Variable (VIF)	Estimate (β-coefficient)	Standard error	t-value
a2 15	1,416.1 / 0.8 / -0.13 / 0.03	17.91*** / 156.1*** / 0.75***	int	-1,311,000	1,790,000	-0.73
			arl (1.02)	5.83*** (0.61)	0.67	8.76
			ppl (1.02)	45.46*** (0.51)	6.25	7.27
			pgr (1.03)	5.03*** (0.39)	0.91	5.52
a3 14	1,398.1		int	-586,500		
			cer	10.17***		
			orc	30.58***		
			pgr	5.27***		
			cal	73.07		
			por	13.36**		
	Model with 2nd lowest AICc					
	1,399.1 / 0.84 / -0.3 / 0.04	19.16*** / 56.1*** / 0.88***	int	-692,100	1,461,000	-0.47
			cer (1.09)	10.43*** (0.67)	1.03	10.15
			orc (1.05)	30.53*** (0.36)	5.46	5.60
			pgr (1.05)	5.85*** (0.62)	0.61	9.57
			por (1.08)	18.43*** (0.28)	4.31	4.27
a3 15	1,403.3		int	-2,807,000		
			cer	11.03***		
			leg	274.4*		
			nur	68.39*		
			orc	62.25***		
			sfr	-1,657***		
			pgr	6.20***		
			por	28.58***		
	Alternative model because no model from interval M + 2 is free from concern					
	1,415.3 / 0.81 / -0.27 / 0.07	19.1*** / 35.1*** / 0.9***	int	-1,015,000	1,741,000	-0.58
			cer (1.09)	10.68*** (0.63)	1.19	8.96
			orc (1.04)	42.82*** (0.47)	6.30	6.80
			pgr (1.04)	7.02*** (0.55)	0.89	7.92
			por (1.07)	23.60*** (0.32)	5.17	4.57
g14	110,055.6		int	7,155.69		
			aar	5.63***		
			cat	-22.42*		
			pou	0.38*		
			int	6,792.41		
			aar	5.28***		
			pou	0.15		
	Model with 2nd lowest AICc					
	110,057.5 / 0.42 / 0.59*** / 0.35***	0.17 / 87,650,000*** / 0.1***	int	6,689	4,865	1.38
			aar (N/A)	5.33*** (0.65)	0.10	53.95
g15	95,654.3 / 0.32 / 0.03 / 0.36***	1.68 / 64,297,000*** / 0.09***	int	10,340	6,453	1.60
			aar (1.22)	5.18*** (0.52)	0.14	36.26

Analysis level	AICc / Adjusted R² / Durbin-Watson autocorrelation / Global Moran's I	Koenker's Breusch-Pagan / Jarque-Bera / Shapiro-Wilk	Variable (VIF)	Estimate (β-coefficient)	Standard error	t-value
			pig (1.22)	7.29*** (0.1)	1.06	6.86
f14	88,490 / 0.34 / 0.42*** / 0.17***	4.79 / 81,205,000*** / 0.11***	int	10,870	5,755	1.89
			pot (1.01)	0.04*** (0.1)	0.0064	6.98
			frn (1)	0.19*** (0.22)	0.0080	14.88
			csl (1.01)	0.05*** (0.23)	0.0033	15.55
			gra (1.01)	0.05*** (0.41)	0.0016	28.33
			vhh (1.01)	0.05*** (0.09)	0.0076	5.99
			mai (1.01)	0.05*** (0.35)	0.0020	23.83
			sug (1.01)	0.05*** (0.07)	0.01	4.58
f15	95,654.3 / 0.29 / 0.42*** / 0.15***	19.57*** / 69,826,000*** / 0.12***	int	9,536	7,098	1.34
			pot (1.01)	0.05*** (0.09)	0.0084	6.21
			frn (1)	0.20*** (0.28)	0.01	19.37
			csl (1.01)	0.06*** (0.21)	0.0039	14.27
			gra (1.01)	0.05*** (0.35)	0.0020	24.17
			vhh (1.01)	0.06*** (0.08)	0.0099	5.76
			mai (1.01)	0.05*** (0.28)	0.0027	19.46
			sug (1.01)	0.06*** (0.05)	0.02	3.70

Note: ***, significant at 0.1 % level; **, significant at 1 % level; * significant at 5 % level; N (= number of observations) = 544 at analysis levels m1 14, m1 15, m2 14, m2 15, m3 14 and m3 15; N = 42 at analysis levels a1 14, a1 15, a2 14, a2 15, a3 14 and a3 15; N = 4,001 at analysis levels g14 and g15; N = 3,166 at analysis level f14; N = 3,363 at analysis level f15; int, Intercept; for abbrevations of other variables see Table 2.2 and Table 2.3.

Table 3.2 shows the results of the GWR model selection. The calculated bandwidths do not show any recognizable pattern, the majority of them either encompasses between 40 and 60 percent or approximately 90 to almost 100 percent of all data points. According to Moran's I, apart from the analysis levels g14, g15, f14 and f15, no significant autocorrelation occurs in any GWR model. However, due to too high local correlation values or local condition numbers several models (bisquare: m1 15, m3 14, m3 15, a2 14, a2 15, a3 14, a3 15; boxcar: m3 15, a3 14, a3 15; Gaussian a3 15) had to be corrected by removing affected variables in order to prevent multicollinearity. In some cases, few local correlation values > 0.8 (bisquare m3 14) or local condition numbers > 10 (bisquare: a2 14, a3 15; boxcar a3 15) still remained even after the correction, but compared to the total number of data points their number was insignificant and the values only marginally above the threshold. In cases where the AICc of the global regression model was not significantly different from that of the GWR model, the correction of the latter was dispensed with since it only would have caused an increase in the AICc value (Gaussian: m1 14, m3 14, m3 15; bisquare: m1 15, m2 15, a1 15; boxcar: m1 14, m1 15, m3 14).

A GWR model, in which the kernel applied has a clearly better result with respect to the AICc than the other kernels exists at the analysis levels m1 15, m3 14, a2 14, a2 15, a3 14 and a3 15. In all the other cases, the AICc of the three or at least two models are too close together to be able to speak of a superior model. As in the case of the global regression models, the adjusted coefficient of determination at the municipal level is always lower than the one at the arrondissement level. At the municipal level, its value is between 0.27 and 0.47, whereas it is between 0.62 and 0.94 at the arrondissement level. As in the case of the global regression, no satisfactory model could be found for the GWR at the analysis levels g14, g15, f14 and f15. At all these four analysis levels, Moran's I reveals significant spatial autocorrelation, and what is more, multicollinearity is present in the analysis approach with the hexagon grid. Nevertheless, a correction was tried, but brought no improvement.

Table 3.2: Results of the best GWR models.

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)		
m1 14	gau	16,953.8 / 0.34 / 266 / -0.01	5.8 - 8	int (0.43)			
					bus	cat	
				aar (0.78)	0.73 - 0.8	0.72 - 0.83 (199)	
				bus (0.06)		0.7 - 0.82 (86)	
				cat (0.82)			
	bis	16,955.2 / 0.34 /536 / -0.01	N/A	int (0.71)			
				aar (0.03)			
	box	16,954 / 0.34 / 307 / -0.01	4.7 - 7.2	int (0.27)			
					bus		
				aar (0)	0.68 - 0.85 (193)		
				bus (0)			
m1 15	gau	17,199.5 / 0.38 /29 / -0.04	6 - 25.3 (470)	int (0.6)			
					pig	pou	aar
				bus (0)	0.27 - 0.88 (65)	0.21 - 0.84 (16)	0.71 - 0.97 (462)

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)		
				pig (0)	0.2 - 0.89 (64)	0.09 - 0.8	
				pou (0)		0.11 - 0.82 (6)	
				aar (0)			
		corrected model					
		17,231.9 / 0.33 / 29 / -0.03	2.9 - 5.9	int (0.33)			
					aar		
				pig (0)	0.09 - 0.8		
				aar (0)			
	bis	17,248.9 / 0.31 /266 / -0.02	6.5 - 18.5 (365)	int (0.51)			
					pig	aar	pou
				bus (0)	0.34 - 0.81 (8)	0.72 - 0.95 (392)	0.24 - 0.76
				pig (0)		0.11 - 0.73	0.2 - 0.86 (84)
				aar (0)			0.11 - 0.72
				pou (0)			
	box	17,266.8 / 0.27 /240 / -0.01	4.6 - 8.6	int (0.13)			
					aar		
				bus (0)	0.66 - 0.89 (236)		
				aar (0)			
m2 14	gau	16,931.3 / 0.37 / 542 / 0	3.1 - 3.2	int (0.68)			
					pgr	ppl	
				arl (0.03)	0.21 - 0.29	0.06 - 0.08	
				pgr (0.14)		-0.07 - -0.05	
				ppl (1)			
	bis	16,933 / 0.37 / 542 / 0	3.2 - 4.5	int (0.74)			
					pgr	ppl	
				arl (0.14)	0.1 - 0.57	0.03 - 0.12	
				pgr (0)		-0.1 - 0	
				ppl (1)			

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)						
	box	16,930.5 / 0.37 / 476 / 0	3.1 - 3.6	int (0.56)							

			Variable	pgr	ppl
			arl (0.13)	0.18 - 0.54	0.05 - 0.1
			pgr (0.02)		-0.07 - -0.01
			ppl (1)		

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable
m2 15	gau	17,215.4 / 0.33 / 539 / 0	3.5 - 3.7	int (0.52)

Variable	ppl	pgr	pig
arl (0.45)	0.04 - 0.06	0.38 - 0.43	0.33 - 0.38
ppl (1)		-0.05 - -0.03	0.04 - 0.05
pgr (0.02)			0.15 - 0.23
pig (0.92)			

	bis	17,218.4 / 0.33 / 542 / 0	7.6 - 10.9 (84)	int (0.73)

Variable	ppl	arl	pgr	pig
bus (0.45)	0.17 - 0.33	0.63 - 0.72	0.48 - 0.7	0.53 - 0.79
ppl (1)		0.01 - 0.1	-0.08 - 0.02	0.04 - 0.11
arl (0.47)			0.31 - 0.6	0.13 - 0.42
pgr (0)				0.02 - 0.44
pig (0.55)				

	box	17,213.5 / 0.34 / 480 / 0	3.3 - 3.9	int (0.51)

Variable	ppl	pgr	pig
arl (0.54)	0.04 - 0.06	0.34 - 0.56	0.14 - 0.37
ppl (1)		-0.05 - 0.01	0.03 - 0.13
pgr (0)			0.05 - 0.42
pig (0.07)			

m3 14	gau	16,922.2 / 0.39 / 275 / -0.02	13.4 - 18.9 (544)	int (0.92)

Variable	cer	leg	pgr	orc	sow	pgl	cal
bus (0.04)	0.41 - 0.46	0.15 - 0.26	0.42 - 0.64	0.13 - 0.31	0.6 - 0.76	0.59 - 0.76	0.52 - 0.59
cer (0.65)		0.14 - 0.29	0.06 - 0.4	0.08 - 0.14	0.01 - 0.16	0.01 - 0.18	0.29 - 0.27

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)				
				leg (0.03)	0.31 - 0.46	0.21 - 0.38	-0.06 - 0.03	-0.06 - 0.03	0.07 - 0.22
				pgr (0)		-0.08 - -0.01	0.02 - 0.35	0.02 - 0.35	0.37 - 0.55
				orc (0.87)			0 - 0.04	-0.01 - 0.05	-0.03 - -0.01
				sow (0)				0.98 (544)	0.34 - 0.45
				pgl (0)					0.32 - 0.43
				cal (0.76)					

	bis	16,292.1 / 0.83 / 100 / -0.14	9 - 52.1 (539)	int (0.53)						
					bus	pgl	leg	c>2	cer	orc	veg
				sow (0)	0.03 - 0.91 (64)	0.85 - 1 (544)	-0.25 - 0.69	-0.07 - 0.87 (30)	-0.15 - 0.81	- 0.13 - 0.78	-0.24 - 0.75
				bus (0.04)		0 - 0.88 (69)	-0.1 - 0.71	0.47 - 0.96 (379)	0.27 - 0.93 (72)	-0.11 - 0.83 (8)	-0.05 - 0.88 (19)
				pgl (0)			-0.26 - 0.62	-0.07 - 0.82 (9)	-0.16 - 0.76	-0.11 - 0.73	-0.25 - 0.79
				leg (0)				-0.08 - 0.66	-0.07 - 0.83 (11)	-0.11 - 0.91 (30)	-0.22 - 0.59
				c>2 (0.06)					-0.14 - 0.94 (54)	-0.16 - 0.54	-0.17 - 0.78
				cer (0.17)						-0.05 - 0.78	0.13 - 0.83 (20)
				orc (0.02)							-0.08 - 0.68
				veg (0.03)							

	corrected model	16,875.8 / 0.47 / 100 / -0.05	2.8 - 10	int (0)			
					leg	cer	veg
				pgl (0)	-0.26 - 0.62	-0.16 - 0.76	-0.25 - 0.79
				leg (0)		-0.07 - 0.83 (11)	-0.22 - 0.59
				cer (0)			0.13 - 0.83 (20)
				veg (0)			

	box	16,896.4 / 0.43 / 236 / -0.03	8.6 - 20.2 (517)	int (0.4)

| Analysis level | Kernel function | AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I | Local condition numbers (n > 10) | Variable (Monte Carlo-test for spatial variability) | Local correlation values (n > 0.8) | | | | | | | |

Correlation sub-table 1:

Variable	cer	leg	sow	por	pgl	hor	pgr	orc
bus (0.02)	0.33 - 0.68	0.09 - 0.47	0.13 - 0.78	0.42 - 0.78	0.12 - 0.78	-0.04 - 0.53	0.35 - 0.89 (228)	0.13 - 0.55
cer (0.24)		0 - 0.45	-0.12 - 0.28	-0.08 - 0.4	-0.12 - 0.3	-0.18 - 0.11	-0.1 - 0.65	0.03 - 0.28
leg (0)			-0.12 - 0.32	-0.1 - 0.25	-0.12 - 0.32	-0.12 - 0.04	0.02 - 0.56	-0.03 - 0.76
sow (0)				0.41 - 0.97 (351)	0.92 - 0.98 (544)	-0.05 - 0.66	-0.14 - 0.67	-0.01 - 0.46
por (0)					0.42 - 0.97 (362)	-0.1 - 0.62	-0.09 - 0.66	0 - 0.41
pgl (0)						-0.04 - 0.62	-0.14 - 0.63	-0.01 - 0.49
hor (0)							-0.09 - 0.43	-0.06 - 0.12
pgr (0)								-0.13 - 0.19
orc (0)								

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable
m3 15	gau	17,209.7 / 0.34 / 542 / 0	16.2 - 17.7 (544)	int (0.76)

Correlation sub-table 2:

Variable	orc	cer	c>2	sow	pgl	pgr
bus (0.28)	0.16 - 0.2	0.44 - 0.46	0.74 - 0.78	0.71 - 0.75	0.71 - 0.74	0.48 - 0.52
orc (0.99)		0.09 - 0.11	-0.06 - -0.05	0 - 0.01	0.01 - 0.02	-0.05 - -0.04
cer (0.53)			0.35 - 0.41	0.11 - 0.16	0.11 - 0.16	0.14 - 0.21
c>2 (0.66)				0.4 - 0.48	0.41 - 0.48	0.85 - 0.86 (544)
sow (0.46)					0.98 (544)	0.14 - 0.22
pgl (0.65)						0.15 - 0.23
pgr (0.36)						

	bis	17,205.8 / 0.35 / 542 / 0.01	8.4 - 12.1 (171)	int (0.7)

Correlation sub-table 3:

Variable	orc	cer	c>2	sow	por
bus (0.14)	0.12 - 0.3	0.39 - 0.51	0.69 - 0.85 (206)	0.48 - 0.77	0.53 - 0.78
orc (0.99)		0.05 - 0.15	-0.08 - -0.03	0 - 0.08	-0.01 - 0.08
cer (0.46)			0.22 - 0.52	-0.03 - 0.19	0.01 - 0.15
c>2 (0)				0.14 - 0.58	0.21 - 0.58

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo test for spatial variability)	Local correlation values (n > 0.8)				
				sow (0)	0.91 - 0.94 (544)				
				por (0)					
		corrected model 17,214.2 / 0.34 / 542 / 0	3.2 - 4.8	int (0.52)					
						cer	c>2	por	
				orc (1)		0.05 - 0.15	-0.08 - -0.03	-0.01 - 0.08	
				cer (0.23)			0.22 - 0.52	0.01 - 0.15	
				c>2 (0)				0.21 - 0.58	
				por (0.01)					
	box	17,206.3 / 0.35 / 462 / 0.01	8.4 - 10.14 (182)	int (0.7)					
					orc	cer	c>2	sow	por
				bus (0.07)	0.17 - 0.3	0.38 - 0.52	0.68 - 0.84	0.46 - 0.76	0.56 - 0.77
				orc (0.74)		0.08 - 0.12	-0.05 - -0.04	-0.01 - 0.14	-0.01 - 0.16
				cer (0.38)			0.27 - 0.47	-0.04 - 0.15	-0.03 - 0.12
				c>2 (0)				0.17 - 0.59	0.24 - 0.6
				sow (0)					0.83 - 0.95 (544)
				por (0)					
		corrected model 17210.7 / 0.34 / 462 / 0	3.1 - 4.2	int (0.63)					
					cer	c>2	por		
				orc (0.98)	0.08 - 0.12	-0.05 - -0.04	-0.01 - 0.16		
				cer (0.05)		0.27 - 0.47	-0.03 - 0.12		
				c>2 (0)			0.24 - 0.6		
				por (0)					
a1 14	gau	1,417.8 / 0.72 / 39 / 0.03	N/A	int (0.45)					
				aar (0.81)					
	bis	1,414.9 / 0.78 / 39 / 0	7.6 - 12.9 (15)	int (0.44)					

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)		
						bus	cat
				aar (0.81)		0.43 - 0.73	0.5 - 0.73
				bus (0.66)			0.32 - 0.75
				cat (0.92)			
	box	1,417.7 / 0.73 / 36 / 0.04	N/A	int (0.79)			
				aar (0.73)			
a1 15	gau	1,440.3 / 0.64 / 39 / 0.1	7.9 - 8.2	int (0.54)			
						bus	cat
				aar (0.89)		0.48 - 0.52	0.6 - 0.66
				bus (0.76)			0.61 - 0.65
				cat (0.98)			
	bis	1,438.4 / 0.68 / 39 / 0.01	7.6 - 13.7 (14)	int (0.37)			
						bus	cat
				aar (0.79)		0.4 - 0.74	0.46 - 0.74
				bus (0.62)			0.33 - 0.74
				cat (0.95)			
	box	1,441.8 / 0.62 / 38 / 0.05	5.8 - 6.3	int (0.37)			
						bus	
				aar (0.17)		0.41 - 0.52	
				bus (0.71)			
a2 14	gau	1,393.1 / 0.86 / 22 / -0.04	5.2 - 6.2	int (0.51)			
						pgr	ppl
				arl (0.79)		-0.21 - 0.35	0.05 - 0.15
				pgr (0)			-0.23 - 0.01
				ppl (0.32)			
	bis	1,382.2 / 0.94 / 22 / -0.12	8.8 - 41 (38)	int (0.53)			
					ppl	arl	pgr
				bus (0.01)	0.14 - 0.73	0.11 - 0.83 (3)	0.16 - 0.9 (11)

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo test for spatial variability)	Local correlation values (n > 0.8)		
				ppl (0.08)	-0.2 - 0.43	-0.39 - 0.52	
				arl (0.62)		-0.41 - 0.76	
				pgr (0)			
		corrected model					
		1,376.9 / 0.94 / 22 / -0.07	4.7 - 15.7 (8)	int (0.53)			
					arl	pgr	
				ppl (0.38)	-0.2 - 0.43	-0.39 - 0.52	
				arl (0.6)		-0.41 - 0.76	
				pgr (0.01)			
	box	1,400.2 / 0.83 / 40 / -0.03	5.2 - 5.6	int (0.26)			
					pgr	ppl	
				arl (0.31)	-0.14 - 0.06	0.08 - 0.1	
				pgr (0.11)		-0.18 - -0.16	
				ppl (0.58)			
a2 15	gau	1,411.5 / 0.82 / 22 / 0.03	5.3 - 5.9	int (0.37)			
					ppl	pgr	
				arl (0.74)	-0.03 - 0.11	0 - 0.37	
				ppl (0.42)		-0.2 - 0.04	
				pgr (0.01)			
	bis	1,410 / 0.89 / 24 / -0.08	8.8 - 30.8 (38)	int (0.33)			
					ppl	arl	pgr
				bus (0.01)	0.14 - 0.68	0.26 - 0.76	0.21 - 0.89 (14)
				ppl (0.21)		-0.27 - 0.37	-0.32 - 0.44
				arl (0.4)			-0.17 - 0.63
				pgr (0)			
		corrected model					
		1,405.5 / 0.89 / 24 / -0.02	4.8 - 14.6 (4)	int (0.24)			
					arl	pgr	
				ppl (0.63)	-0.27 - 0.37	-0.32 - 0.44	

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)					
				arl (0.78)	-0.17 - 0.63					
				pgr (0)						
	box	1,416.3 / 0.81 / 40 / -0.03	5.7 - 6.1	int (0.29)						
					ppl	pgr	pig			
				arl (0.37)	0.01 - 0.03	0.06 - 0.22	-0.02 - 0.04			
				ppl (0.59)		-0.15 - -0.13	0.01 - 0.05			
				pgr (0.12)			-0.09 - -0.05			
				pig (0.75)						
a3 14	gau	1,396.5 / 0.86 / 22 / 0.01	7.8 - 9.9	int (0.62)						
					cer	pgr	orc	bro		
				bus (0.29)	0.09 - 0.2	0.19 - 0.42	0.16 - 0.27	0.53 - 0.61		
				cer (0.79)		-0.2 - 0.25	0.12 - 0.26	-0.2 - -0.06		
				pgr (0.01)			-0.21 - -0.02	-0.12 - 0.18		
				orc (0.36)				-0.04 - 0.05		
				bro (0.58)						
	bis	1,395.8 / 0.93 / 29 / -0.21	13.3 - 30.4 (42)	int (0.62)						
					sfr	cer	orc	cal	hor	pot
				bus (0.1)	0.17 - 0.71	0.06 - 0.57	0.07 - 0.62	0.35 - 0.84 (4)	0.37 - 0.75	0.11 - 0.6
				sfr (0.45)		-0.02 - 0.71	0.25 - 0.8	-0.18 - 0.45	-0.35 - 0.76	-0.05 - 0.36
				cer (0.17)			-0.01 - 0.51	-0.17 - 0.44	-0.35 - 0.28	0.52 - 0.86 (15)
				orc (0.2)				-0.26 - 0.21	-0.26 - 0.54	-0.17 - 0.13
				cal (0.51)					-0.16 - 0.77	-0.4 - 0.53
				hor (0)						-0.13 - 0.29
				pot (0.03)						
	corrected model									
		1,421.1 / 0.82 / 29 / -0.17	5.4 - 13.1 (8)	int (0.8)						

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)			

					orc	cal	hor	pot
				sfr (0.59)	0.25 - 0.8	-0.18 - 0.45	-0.35 - 0.76	-0.05 - 0.36
				orc (0.41)		-0.26 - 0.21	-0.26 - 0.54	-0.17 - 0.13
				cal (0.48)			-0.16 - 0.77	-0.4 - 0.53
				hor (0)				-0.13 - 0.29
				pot (0.76)				

	box	1,394 / 0.9 / 26 / 0.01	16.1 - 45.8 (42)	int (0.28)				

					cer	hor	orc	c12	veg	cal
				bus (0.12)	-0.01 - 0.5	0.3 - 0.72	0.12 - 0.47	0.22 - 0.86 (6)	0.06 - 0.45	0.25 - 0.77
				cer (0.89)		-0.4 - 0.06	0.06 - 0.56	-0.07 - 0.25	-0.01 - 0.72	-0.15 - 0.31
				hor (0)			-0.16 - 0.15	-0.27 - 0.71	-0.18 - 0.21	-0.29 - 0.74
				orc (0.09)				-0.33 - 0.03	-0.16 - 0.16	-0.19 - 0.15
				c12 (0.71)					-0.24 - 0.27	0.66 - 0.97 (19)
				veg (0.02)						-0.26 - 0.22
				cal (0.25)						

corrected model

		1,415.5 / 0.8 / 26 / -0.2	4.9 - 8.8	int (0.59)				

					hor	orc	veg	cal
				cer (0.65)	-0.4 - 0.06	0.06 - 0.56	-0.01 - 0.72	-0.15 - 0.31
				hor (0)		-0.16 - 0.15	-0.18 - 0.21	-0.29 - 0.74
				orc (0.27)			-0.16 - 0.16	-0.19 - 0.15
				veg (0.02)				-0.26 - 0.22
				cal (0.62)				

a3 15	gau	1401.3 / 0.89 / 20 / -0.09	11.2 - 16.9 (42)	int (0.17)				

					cer	orc	c>2	sfr	leg
				bus (0.57)	0.07 - 0.2	0.15 - 0.27	0.49 - 0.76	0.2 - 0.4	-0.31 - -0.04
				cer (0.26)		0.09 - 0.26	-0.14 - 0.2	0.16 - 0.23	0.09 - 0.57
				orc (0.79)			-0.29 - -0.08	0.41 - 0.67	-0.29 - -0.02

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)						
				c>2 (0.01)	-0.17 - 0.18	0.28 - 0.43					
				sfr (0.09)		-0.17 - 0.15					
				leg (0.15)							
		corrected model									
		1,407.6 / 0.86 / 20 / 0.02	6.4 - 8.6	int (0.17)							
					orc	c>2	sfr	leg			
				cer (0.27)	0.09 - 0.26	-0.14 - 0.2	0.16 - 0.23	0.09 - 0.57			
				orc (0.88)		-0.29 - -0.08	0.41 - 0.67	-0.29 - -0.02			
				c>2 (0.03)			0.41 - 0.67	-0.29 - -0.02			
				sfr (0.07)				-0.17 - 0.18			
				leg (0.85)							
	bis	1,402 / 0.91 / 39 / -0.04	19.3 - 33.7 (42)	int (0.26)							
					cer	orc	c>2	sfr	leg	por	fal
				bus (0.46)	0.06 - 0.48	0.12 - 0.51	0.27 - 0.77	0.13 - 0.47	-0.31 - 0.04	0.56 - 0.74	-0.09 - 0.25
				cer (0.39)		0.06 - 0.31	-0.21 - 0.22	0.11 - 0.51	-0.1 - 0.64	-0.36 - 0.06	0.81 - 0.88 (42)
				orc (0.69)			-0.35 - -0.06	0.36 - 0.7	-0.4 - -0.01	-0.13 - 0.38	-0.12 - 0.08
				c>2 (0.03)				-0.23 - 0.25	0.19 - 0.56	0.15 - 0.67	0 - 0.34
				sfr (0.42)					-0.27 - 0.43	0 - 0.49	-0.02 - 0.51
				leg (0.01)						-0.26 - -0.05	0.19 - 0.9 (26)
				por (0.54)							-0.36 - 0.02
				fal (0.03)							
		corrected model									
		1,400 / 0.89 / 39 / -0.12	6 - 11.8 (2)	int (0.39)							
					orc	sfr	c>2	por			
				cer (0.69)	0.06 - 0.31	0.11 - 0.51	-0.21 - 0.22	-0.36 - 0.06			
				orc 0.9)		0.36 - 0.7	-0.35 - -0.06	-0.13 - 0.38			
				sfr (0.35)			-0.23 - 0.25	0 - 0.49			
				c>2 (0)				0.15 - 0.67			

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)					
				por (0.34)						
	box	1,398.8 / 0.91 / 28 / -0.17	15.7 - 23.6 (42)	int (0.05)						
					cer	hor	orc	sfr	leg	c12
				bus (0.19)	-0.02 - 0.51	0.27 - 0.71	0.12 - 0.45	0.14 - 0.51	-0.3 - 0.03	0.33 - 0.83
				cer (0.65)		-0.36 - -0.11	0.05 - 0.35	0.06 - 0.37	0.1 - 0.69	-0.08 - 0.29
				hor (0)			-0.13 - 0.12	-0.11 - 0.34	-0.45 - -0.13	-0.26 - 0.63
				orc (0.51)				0.33 - 0.69	-0.3 - 0.04	-0.31 - 0.01
				sfr (0.03)					-0.16 - 0.31	-0.2 - 0.28
				leg (0.3)						0.29 - 0.59
				c12 (0.57)						
	corrected model	1,412.4 / 0.86 / 28 / -0.08	6.9 - 11.6 (4)	int (0.01)						
					hor	orc	sfr	leg	c12	
				cer (0.32)	-0.36 - -0.11	0.05 - 0.35	0.06 - 0.37	0.1 - 0.69	-0.08 - 0.29	
				hor (0.18)		-0.13 - 0.12	-0.11 - 0.34	-0.45 - -0.13	-0.26 - 0.63	
				orc (0.75)			0.33 - 0.69	-0.3 - 0.04	-0.31 - 0.01	
				sfr (0.02)				-0.16 - 0.31	-0.2 - 0.28	
				leg (0.27)					0.29 - 0.59	
				c12 (0)						
g14	gau	105,710.6 / 0.82 / 17 / 0.19***	6.9 - 1,709.5 (3,982)	int (0.73)						
					bus	aar	cat	pou		
				pig (0)	0.12 - 0.99 (1,069)	0.09 - 0.98 (878)	0.03 - 0.97 (706)	-0.17 - 0.98 (412)		
				bus (0)		0.79 - 1 (3,999)	0.62 - 1 (3,772)	0.01 - 0.96 (405)		
				aar (0)			0.63 - 1 (3,847)	-0.02 - 0.97 (392)		
				cat (0)				-0.04 - 0.97 (516)		
				pou (0)						

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)						
	bis	Computation not successful									
	box	110,012.2 / 0.43 / 887 / 0.34***	6 - 17 (1,717)	int (0.01)							

	bus	cat
aar (0)	0.83 - 0.98 (4,001)	0.58 - 0.97 (3,159)
bus (0)		0.65 - 0.95 (2,457)
cat (0)		

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable
g15	gau	106,693.5 / 0.85 / 19 / 0.18***	6 - 1,782.1 (3,976)	int (0.85)

	bus	aar	pou	cat
pig (0)	0.11 - 0.98 (985)	0.12 - 0.98 (858)	-0.14 - 1 (440)	0.1 - 0.98 (625)
bus (0)		0.75 - 1 (3,977)	0.02 - 0.95 (454)	0.62 - 0.99 (3,712)
aar (0)			-0.01 - 0.96 (421)	0.63 - 1 (3,797)
pou (0)				0.01 - 0.96 (412)
cat (0)				

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable
	bis	Computation not successful		
	box	Computation not successful		
f14	gau	88,489.8 / 0.34 / 1,911 / 0.17***	1.8 - 1.9	int (0.61)

	mai	csl	frn	pot	vhh	sug	inf
gra (0.79)	-0.06	-0.05 - -0.04	-0.03 - -0.02	-0.05	-0.05 - -0.04	-0.04 - -0.03	-0.06
mai (0.56)		-0.04	-0.02	-0.05 - -0.04	-0.04	-0.04 - -0.03	-0.06 - -0.05
csl (0)			-0.02 - -0.01	-0.04 - -0.03	-0.03	-0.03 - -0.02	-0.04
frn (1)				-0.02 - -0.01	-0.02 - -0.01	-0.01	-0.02
pot (0.44)					-0.04 - -0.03	-0.03	-0.05 - -0.04
vhh (0.73)						-0.03 - -0.02	-0.04
sug (0.01)							-0.03
inf (0.7)							

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)						
	bis	88,492.9 / 0.34 / 3,165 / 0.17***	1.8 - 1.9	int (0.47)							

	mai	csl	frn	pot	vhh	sug	inf
gra (0.74)	-0.06	-0.05 - -0.04	-0.03 - -0.02	-0.05	-0.05 - -0.04	-0.04 - -0.03	-0.06
mai (0.53)		-0.05 - -0.04	-0.03 - -0.02	-0.05	-0.05 - -0.04	-0.04 - -0.03	-0.06 - -0.05
csl (0)			-0.02 - -0.01	-0.04 - -0.03	-0.04 - -0.03	-0.03 - -0.02	-0.04
frn (0.98)				-0.02 - -0.01	-0.02 - -0.01	-0.01	-0.02
pot (0.25)					-0.04 - -0.03	-0.03	-0.05 - -0.04
vhh (0.76)						-0.03 - -0.02	-0.04
sug (0.01)							-0.04 - -0.03
inf (0.56)							

	box	88,488.9 / 0.34 / 2,559 / 0.17***	1.8 - 1.9	int (0.05)

	mai	csl	frn	pot	vhh	sug	inf
gra (0.39)	-0.06	-0.05 - -0.04	-0.03 - -0.02	-0.05	-0.06 - -0.04	-0.04 - -0.03	-0.06
mai (0.41)		-0.05 - -0.04	-0.03 - -0.02	-0.05 - -0.04	-0.05 - -0.04	-0.04 - -0.03	-0.06 - -0.05
csl (0)			-0.02 - -0.01	-0.04 - -0.03	-0.04 - -0.03	-0.03 - -0.02	-0.05 - -0.04
frn (0.34)				-0.02	-0.02 - -0.01	-0.02 - -0.01	-0.02
pot (0.02)					-0.04 - -0.03	-0.03	-0.05 - -0.04
vhh (0.02)						-0.03 - -0.02	-0.05 - -0.04
sug (0)							-0.04 - -0.03
inf (0.06)							

f15	gau	95,650.7 / 0.3 / 1,795 / 0.15***	1.8	int (0.46)

	frn	mai	csl	pot	vhh	sug	inf
gra (0.75)	-0.03 - -0.02	-0.06	-0.05 - -0.04	-0.05	-0.05 - -0.04	-0.03	-0.06 - -0.05
frn (1)		-0.02	-0.02 - -0.01	-0.02 - -0.01	-0.02 - -0.01	-0.01	-0.02
mai (0.55)			-0.04	-0.05 - -0.04	-0.04	-0.03	-0.05
csl (0)				-0.04 - -0.03	-0.03	-0.03 - -0.02	-0.04
pot (0.13)					-0.03	-0.03 - -0.02	-0.04

Analysis level	Kernel function	AICc / Adjusted R² / Adaptive bandwidth / Global Moran's I	Local condition numbers (n > 10)	Variable (Monte Carlo-test for spatial variability)	Local correlation values (n > 0.8)					
					frn	mai	csl	pot	vhh	sug
				vhh (0.76)					-0.03 -0.02	-0.04
				sug (0.01)						-0.03
				inf (0.51)						
	bis	95,651.9 / 0.3 / 2,946 / 0.15***	1.7 - 1.8	int (0.34)						
					frn	mai	csl	pot	vhh	sug
				gra (0.56)	-0.03 - -0.02	-0.06	-0.05 - -0.04	-0.05	-0.05 - -0.04	-0.04 - -0.03
				frn (0.43)		-0.03 - -0.02	-0.02 - -0.01	-0.02	-0.02 - -0.01	-0.02 - -0.01
				mai (0.46)			-0.05 - -0.04	-0.05 - -0.04	-0.05 - -0.04	-0.04 - -0.03
				csl (0)				-0.04 - -0.03	-0.04 - -0.03	-0.03 - -0.02
				pot (0.07)					-0.04 - -0.03	-0.03 - -0.02
				vhh (0.17)						-0.03 - -0.02
				sug (0.01)						
	box	95,648.8 / 0.3 / 2,341 / 0.15***	1.7	int (0.03)						
					frn	mai	csl	pot	vhh	sug
				gra (0.43)	-0.03 - -0.02	-0.06 - -0.05	-0.05 - -0.04	-0.05 - -0.04	-0.05 - -0.04	-0.04 - -0.03
				frn (0.51)		-0.03 - -0.02	-0.02 - -0.01	-0.02	-0.02 - -0.01	-0.02 - -0.01
				mai (0.37)			-0.05 - -0.04	-0.05 - -0.04	-0.05 - -0.04	-0.04 - -0.03
				csl (0)				-0.04 - -0.03	-0.04 - -0.03	-0.03 - -0.02
				pot (0.07)					-0.04 - -0.03	-0.03 - -0.02
				vhh (0.02)						-0.03 - -0.02
				sug (0)						

Note: ***, significant at 0.1 % level; N (= number of observations) = 544 at analysis levels m1 14, m1 15, m2 14, m2 15, m3 14 and m3 15; N = 42 at analysis levels a1 14, a1 15, a2 14, a2 15, a3 14 and a3 15; N = 4,001 at analysis levels g14 and g15; N = 3,166 at analysis level f14; N = 3,363 at analysis level f15; gau, Gaussian; bis, bisquare; box, boxcar; int, intercept; n, number of observations beyond threshold value; for abbreviations of variables see Table 2.2 and Table 2.3.

Table 3.3 shows the results of the MGWR. At the analysis levels m1 14, m1 15, m3 14, a1 14 and a1 15 no mixed GWR was calculated at all since either all predictors were exclusively global or local respectively, or the GWR models already had an approximately equal or higher AICc than that of the global regression model. For the same reason, no MGWR model was generated with the Gaussian

kernel at analysis level m3 15, with the bisquare kernel at the analysis levels m2 15, m3 15 and a3 14 as well as with the boxcar kernel at the analysis levels a2 14, a2 15 and a3 14. At the analysis levels g14, g15, f14, and f15, an MGWR appeared to be not of any usefulness because of the spatial autocorrelation of the GWR models. In addition, the calculation of such a model simply took too much time and could not be completed. For these reasons, an MGWR was abandoned for these analysis levels.

Table 3.3: Results of the MGWR models.

Analysis level	Kernel function (Adaptive bandwidth)	AICc	Global Variables	Local Variables
m2 14	gau (542)	16,931	pgr	int
			ppl	arl
	bis (542)	16,932	arl	int
			ppl	pgr
	box (467)	16,930	arl	int
			ppl	pgr
m2 15	gau (539)	17,215	arl	int
			ppl	pgr
			pig	
	box (480)	17,213	arl	int
			ppl	pgr
				pig
m3 15	box (462)	17,211	orc	int
				cer
				c>2
				por
a2 14	gau (22)	1,401	arl	int
			ppl	pgr
	bis (22)	1,410	ppl	int
			arl	pgr
a2 15	gau (22)	1,418	arl	int
			ppl	pgr
	bis (24)	1,424	ppl	int
			arl	pgr
a3 14	gau (22)	1,402	bus	int
			cer	pgr
			orch	
			bro	
a3 15	gau (20)	1,412	cer	int
			orc	c>2
			leg	sfr
	bis (39)	1,406	cer	int
			orc	c>2
			sfr	
			por	

Analysis level	Kernel function (Adaptive bandwidth)	AICc	Global Variables	Local Variables
	box (28)	1,412	cer	int
			hor	sfr
			orc	c12
			leg	

Note: N (= number of observations) = 544 at analysis levels m2 14, m2 15 and m3 15; N = 42 at analysis levels a2 14, a2 15, a3 14 and a3 15; gau, Gaussian; bis, bisquare; box, boxcar; int, intercept; for abbrevations of variables see Table 2.2 and Table 2.3.

Table 3.4: Regression models for the interpretation of equity of the distribution of the subsidies.

Analysis level	Model type	Adjust-ed R^2	Global variable	Global Coefficient	Local variable	Local estimates				
						Minimum	1st Quantile	Median	3rd Quantile	Maximum
m1 14	OLS	0.34	int	25,830						
			aar	5.42						
m1 15	GWR (gau)	0.33			int	-1,488,000	-157,300	-37,700	100,700	545,400
					pig	-50.26	8.23	11.36	40.70	396.90
					aar	2.98	4.56	5.32	6.26	14.93
m2 14	OLS	0.37	int	-9,690.76						
			arl	5.19						
			ppl	25.05						
			pgr	5.33						
m2 15	MGWR (box)	0.34	arl	5.46	int	-95,400	-91,540	-58,170	-29,540	-3,667
			ppl	45.08	pgr	2.98	3.68	4.23	5.07	5.08
					pig	2.99	3.77	5.30	5.45	10.90
m3 14	GWR (bis)	0.47			int	-340,300	146,200	323,000	524,500	971,300
					pgl	-149.30	31.15	66.50	140.30	3,849.00
					leg	-146.50	51.54	262.80	659.50	27,270.00
					cer	-5.14	8.97	11.53	13.71	33.53
					veg	-305.80	-9.36	0.13	5.37	93.02
m3 15	MGWR (box)	0.34	orc	47.34	int	2,979	16,490	53,860	73,690	88,720
					cer	7.75	7.94	8.58	8.87	9.36
					c>2	184.60	220.50	245.00	319.90	327.60
					por	-12.67	2.39	11.13	11.80	16.78
a1 14	OLS	0.73	int	-774,700						
			aar	5.77						
a1 15	OLS	0.60	int	521,700						
			aar	5.67						
a2 14	OLS	0.83	int	-1,817,000						
			arl	5.69						
			ppl	32.47						
			pgr	5.62						
a2 15	OLS	0.80	int	-1,311,000						
			arl	5.83						
			ppl	45.46						
			pgr	5.03						
a3 14	OLS	0.84	int	-692,100						
			cer	10.43						
			orc	30.53						
			pgr	5.85						
			por	18.43						
a3 15	MGWR (gau)	0.86	cer	8.80	int	-2,141,000	-1,834,000	-1,718,000	-1,261,000	-638,200
			orc	65.33	c>2	287.50	320.10	352.00	360.10	375.70
			leg	158.61	sfr	-1,556.00	-1,448.00	-1,037.00	-3.10	620.80

Analysis level	Model type	Adjust-ed R²	Global variable	Global Coefficient	Local variable	Local estimates				
						Minimum	1st Quantile	Median	3rd Quantile	Maximum
g14	no model free from concern									
g15	no model free from concern									
f14	no model free from concern									
f15	no model free from concern									

Note: gau, Gaussian; bis, bisquare; box, boxcar; int, intercept; values of adjusted R² for MGWR models are taken from the corresponding GWR models; for abbrevations of variables see Table 2.2 and Table 2.3.

Table 3.4 shows which regression models were ultimately used to interpret the equity of the distribution of the agricultural subsidies. At seven analysis levels, including five of the six at the arrondissement level, the global model is superior or at least equal to the GWR and MGWR model. In two cases, the GWR model has been selected, at level m1 15 the one with the Gaussian kernel and at level m3 14 the one with the bisquare kernel. Both models are clearly superior to the other models, although, at level m3 14, some local correlation values between "legumes" and "cereals" or "vegetables" and "cereals" respectively are slightly above the threshold of 0.8. However, as mentioned above, this is not important because of the small number of points above the threshold value compared to the total number of data points. In three cases, the MGWR model proved to be superior, at analysis level m2 15 and m3 15 with the boxcar kernel and at analysis level a3 15 with the Gaussian kernel. For the analysis levels g14, g15, f14 and f15 no satisfactory model was found at all, neither the OLS model nor the GWR models passed the diagnostic tests.

Among the diagnostic plots of the OLS regression models selected for interpretation (Fig. 3.1) the diagrams "Residuals vs Fitted" in all cases show a more or less linear relationship. The Q-Q plots show a normal distribution of the residuals, albeit outliers cause a slight S-curvature. At the analysis levels m1 14 and m2 14 the plots "Scale-Location" indicate moderate heteroscedasticity although both studentized Breusch-Pagan tests have been negative. At the analysis level m2 14 the plot "Residuals vs Leverage" show that there is a point (Sint Truiden) that slightly distorts or could distort the result. The diagnosis plots reveal that in the case of the municipalities Sint Truiden, Leuven and often Roeselare too are almost always outliers, in the case of the arrondissements Leuven, Tongeren, Hasselt and Halle-Vilvoorde. The outliers of the OLS models are shown in Fig. 3.2 and confirm the findings of the diagnostic plots. In all cases, the municipality of Leuven and the arrondissement Leuven are among the outliers. For both OLS regression models, this also applies to the municipality of Sint Truiden at the municipal level and in two cases of the GWR models also to the arrondissement Tongeren at the arrondissement level. The Moran plots (Fig. 3.3) of the OLS regressions show that not a single model suffers from significant autocorrelation, even though the decision at the analysis level a1 15 is narrow, recognizable by the slope of the regression line. Again, the Moran plots can be used to identify the

same outliers as in the previous graphics. However, in three of the five arrondissement level plots, the arrondissement of Leuven has a high value compared to the neighboring arrondissements, while the arrondissement of Tongeren is mainly surrounded by low values, which, nevertheless, are higher than the value of Tongeren itself. The residual maps (Fig. 3.4) also confirm this picture. They very much resemble the maps of the outliers. Almost always the residuals of the municipalities of Sint Truiden, Leuven and Roeselare or the arrondissement of Leuven are much higher than those of the other municipalities or arrondissements and lower in the case of the arrondissement of Tongeren. The residual maps of the analysis levels a1 14 and a1 15 also show the high residuals of the arrondissement Hasselt.

m2 14

Residuals vs Fitted

110. Leuven

194. Roeselare

432. Sint-Truid

Normal Q-Q

110. Leuven

432. Sint-Truiden
194. Roeselare

Scale-Location

110. Leuven

432. Sint-Truid

194. Roeselare

Residuals vs Leverage

110. Leuven

432. Sint-Truid

453. Borgloon

Cook's distance

a1 14

Residuals vs Fitted

5. Leuven

32. Hasselt

4. Halle-Vilvoorde

Normal Q-Q

5. Leuven

32. Hasselt

4. Halle-Vilvoorde

Scale-Location

5. Leuven

32. Hasselt

4. Halle-Vilvoorde

Residuals vs Leverage

5. Leuven

32. Hasselt

30. Verviers

Cook's distance

44

a1 15

a2 14

Fig. 3.1: Diagnostic plots for those analysis levels where the OLS model is superior.
Note: The three outermost outliers are indicated by their names.

m1 14

m2 14

110. Leuven
432. Sint-Truiden

194. Roeselare
110. Leuven
432. Sint-Truiden

a1 14

a1 15

5. Leuven

5. Leuven

a2 14

a2 15

5. Leuven
Tongeren

5. Leuven
Tongeren
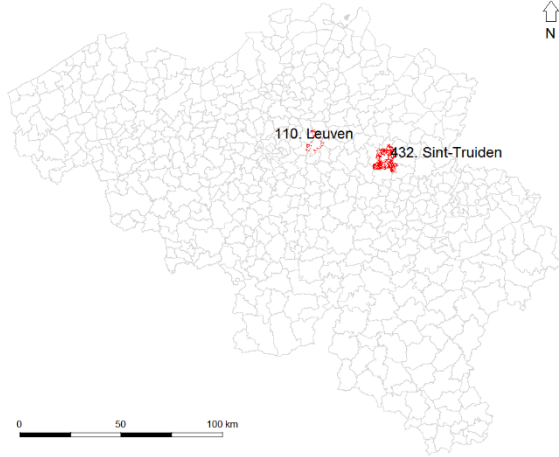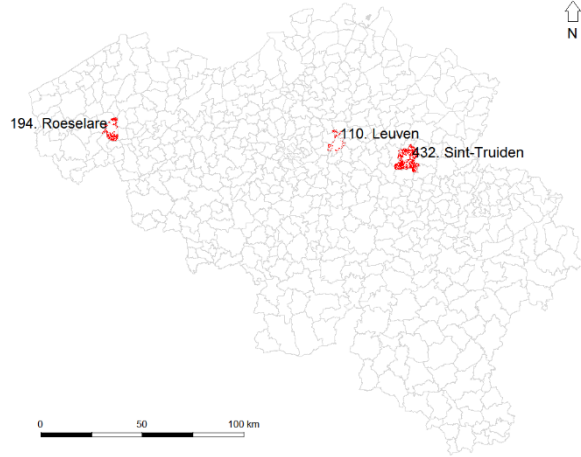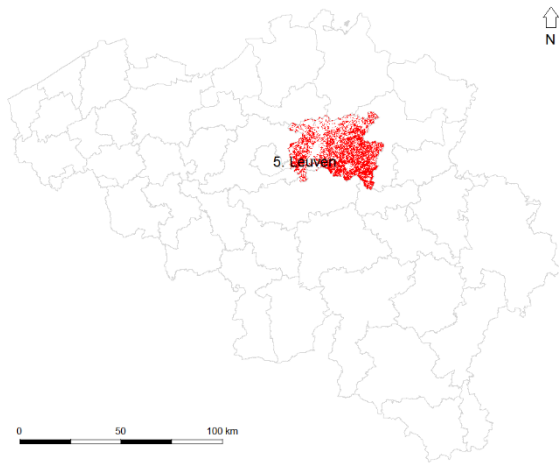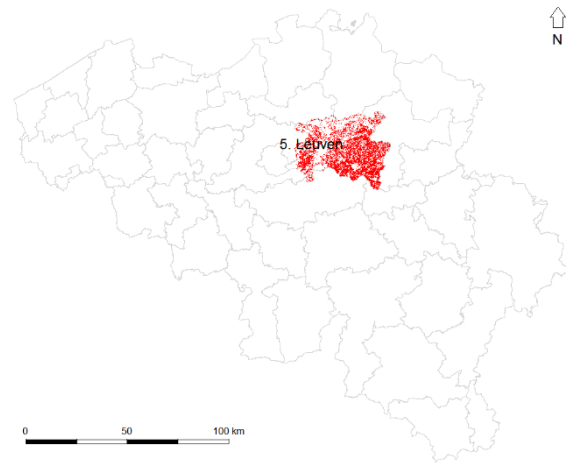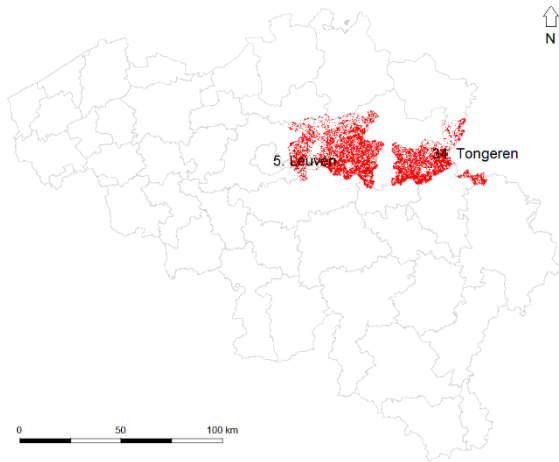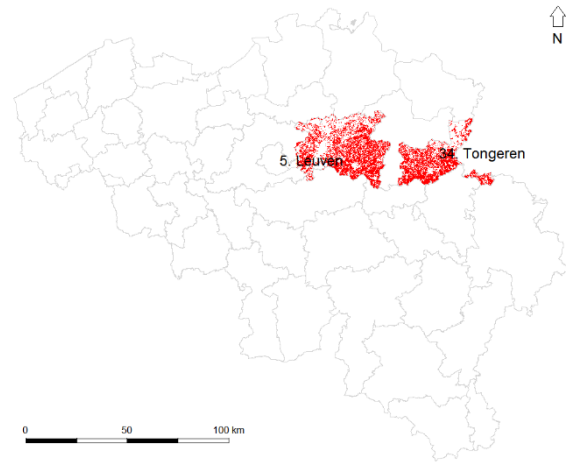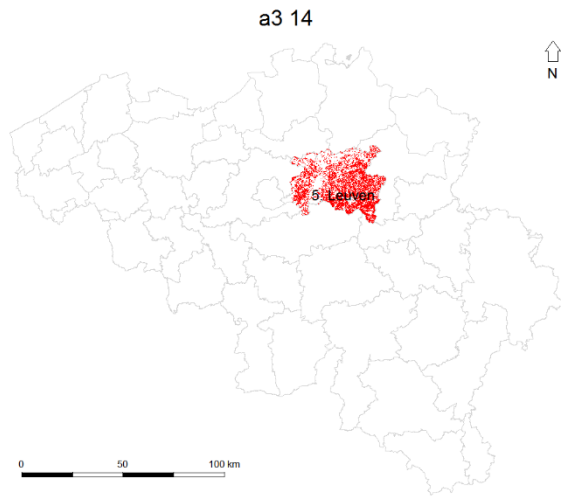
Fig. 3.2: OLS outliers for those analysis levels where the OLS model is superior.

a1 14

a1 15

a2 14

49

a2 15

Moran's I: 0.029
p-value: 0.213

spatially lagged residuals

34. Tongeren

5. Leuven

33. Maaseik

residuals

a3 14

Moran's I: 0.036
p-value: 0.149

spatially lagged residuals

2. Mechelen

34. Tongeren

5. Leuven
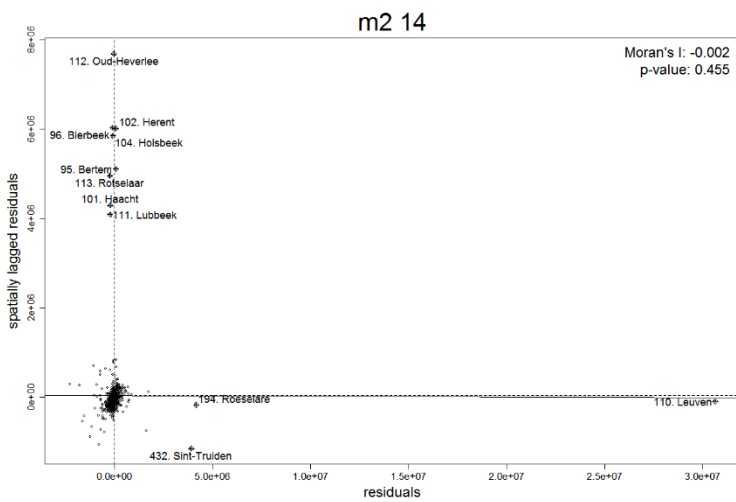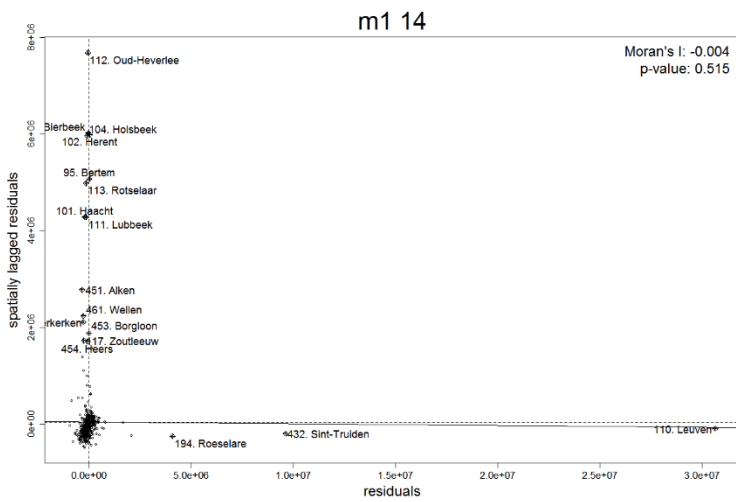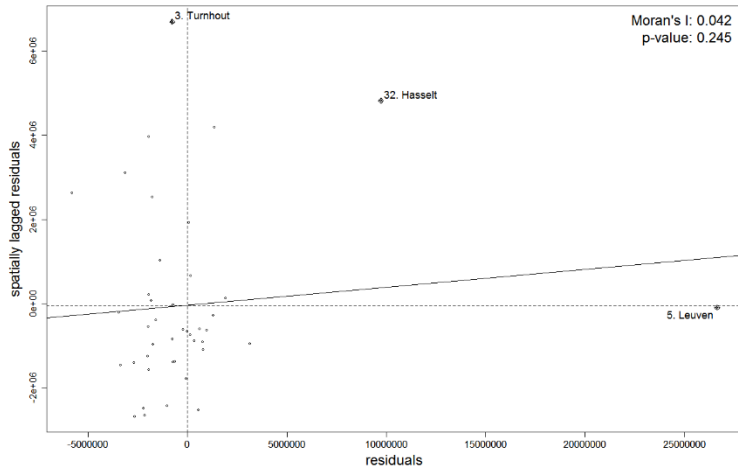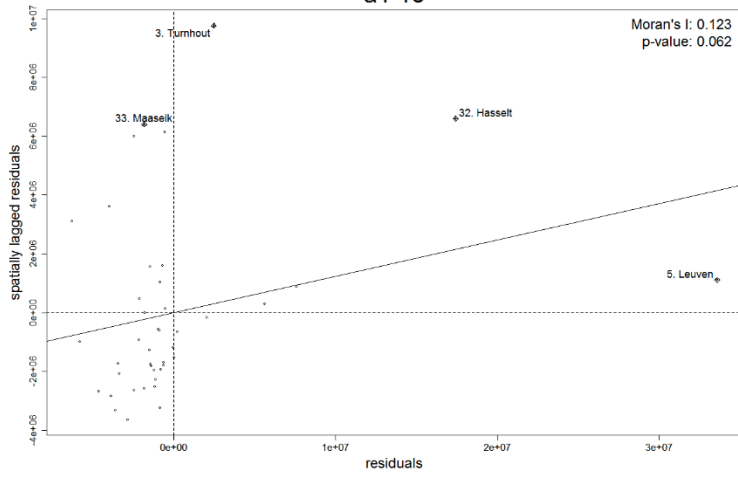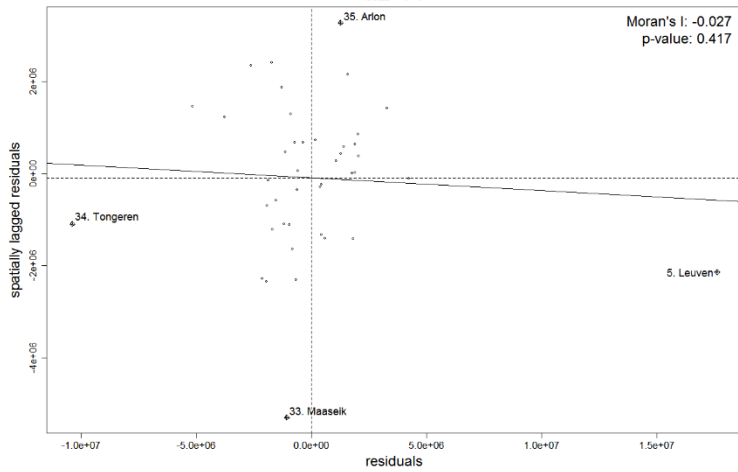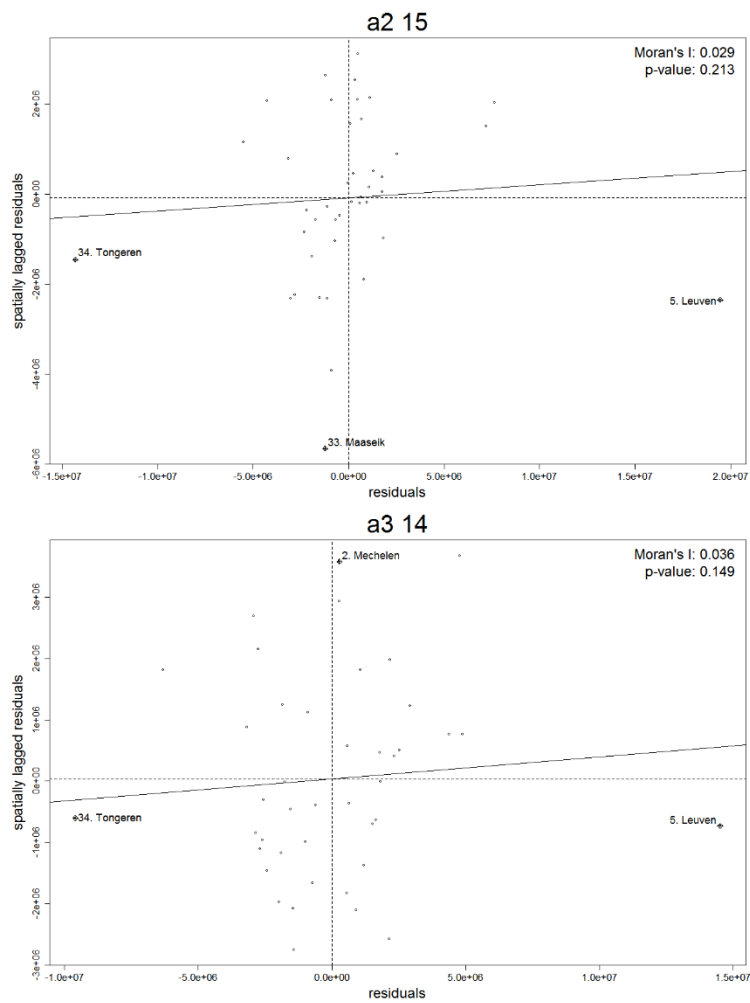
residuals

Fig.3.3: OLS Moran plots for those analysis levels where the OLS model is superior.

m1 14

N

OLS residuals
(SD from mean)
< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

m2 14

N

OLS residuals
(SD from mean)
< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km
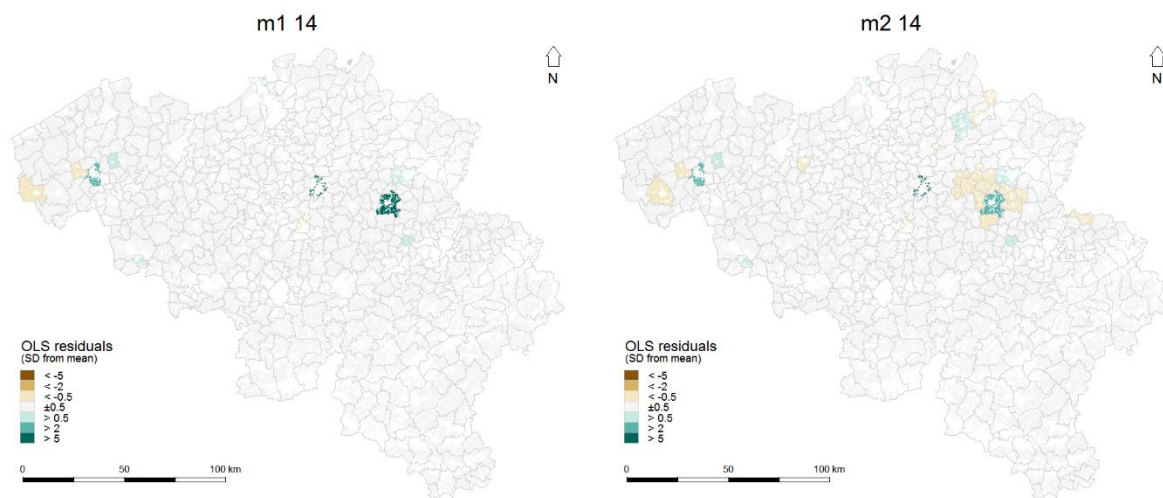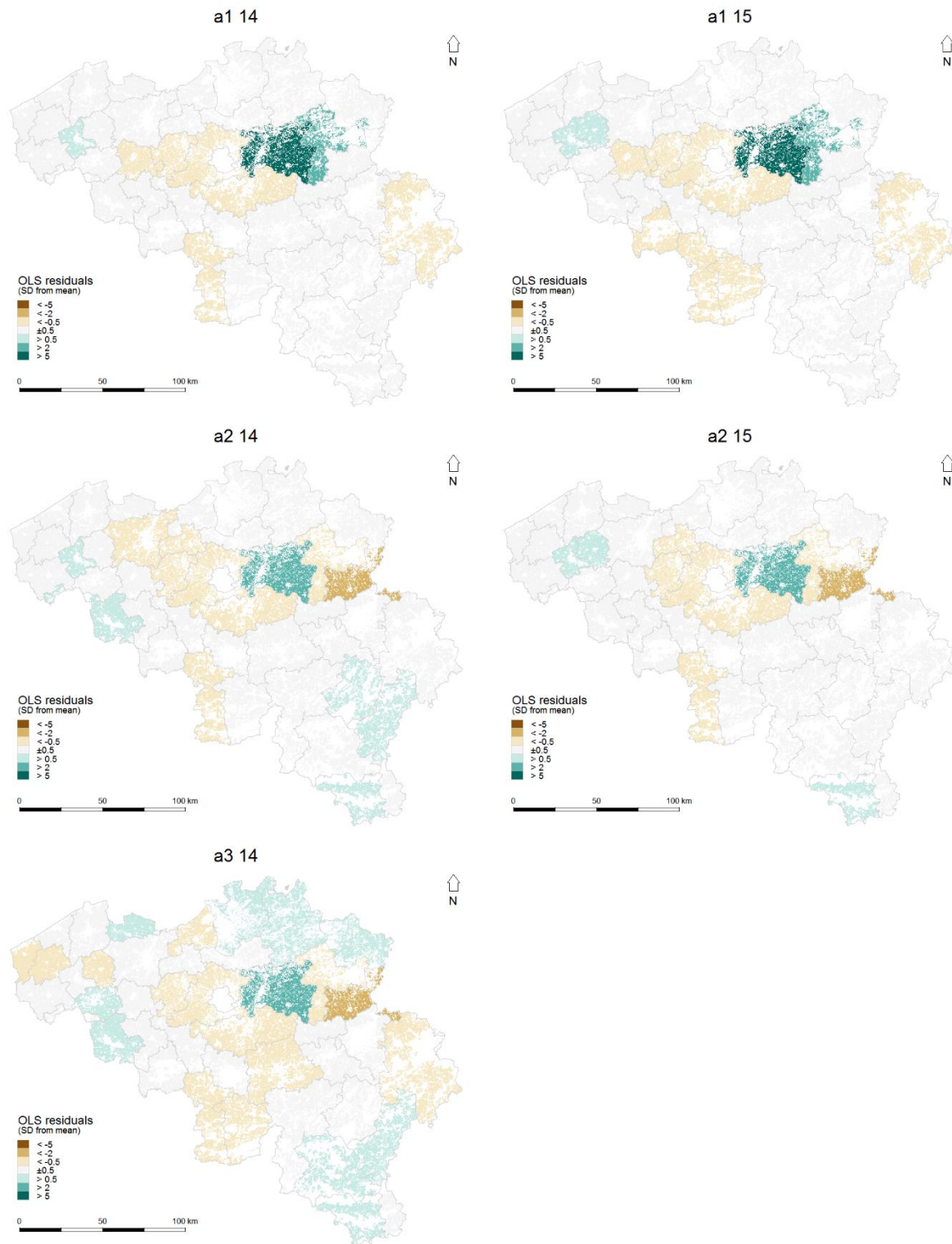
Fig. 3.4: OLS residuals for those analysis levels where the OLS model is superior.

The Moran plots of the two GWR models (Fig. 3.5) are similar to those of the global regressions and also the two respective residual maps show the same picture, which is known from the OLS regressions

already. The GWR estimates of the variables are given in Fig. 3.7, the ultimately significant regions in Fig. 3.8. At analysis level m1 15, the predictor "agricultural area" has significantly high values in the province of Limburg and the predictor "pigs" in an area from the municipality of Leuven in the north to Namur in the southeast and Nivelles in the west. At analysis level m3 14 high values exist in the predictor "piglets" in central Belgium near the capital region of Brussels. In the predictor "legumes" there is such an area north of Brussels, which extends in an elongated form to the north-east as far as Herentals. The predictor "cereals" has high values in an area along the border with Luxembourg and Germany, while the predictor "vegetables" has low values in the very same region.
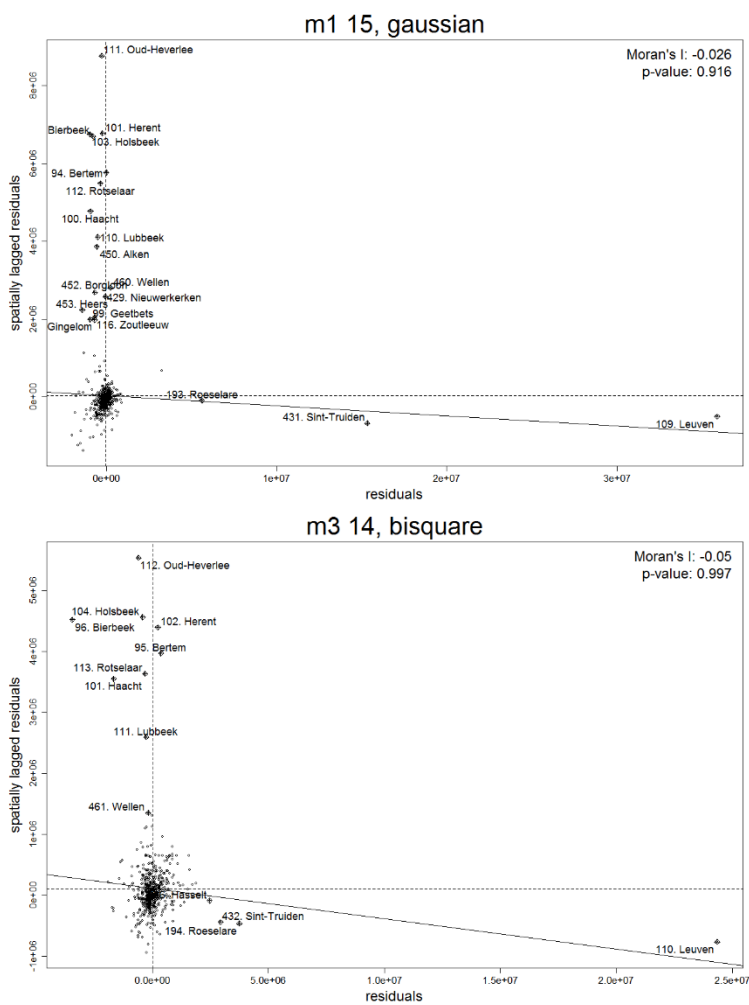
Fig. 3.5: GWR Moran plots for those analysis levels where a GWR model is superior.
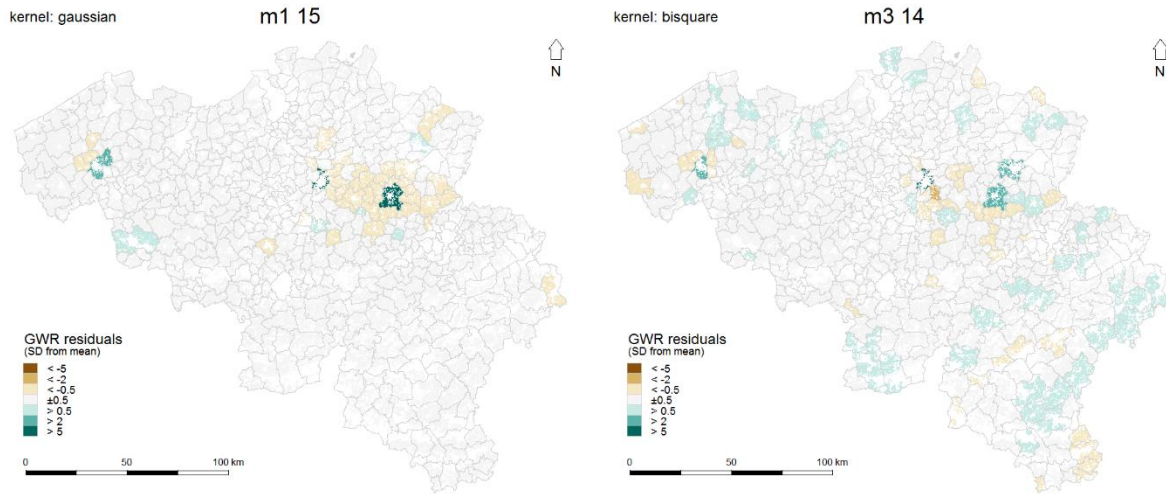
Fig. 3.6: GWR residuals for those analysis levels where a GWR model is superior.
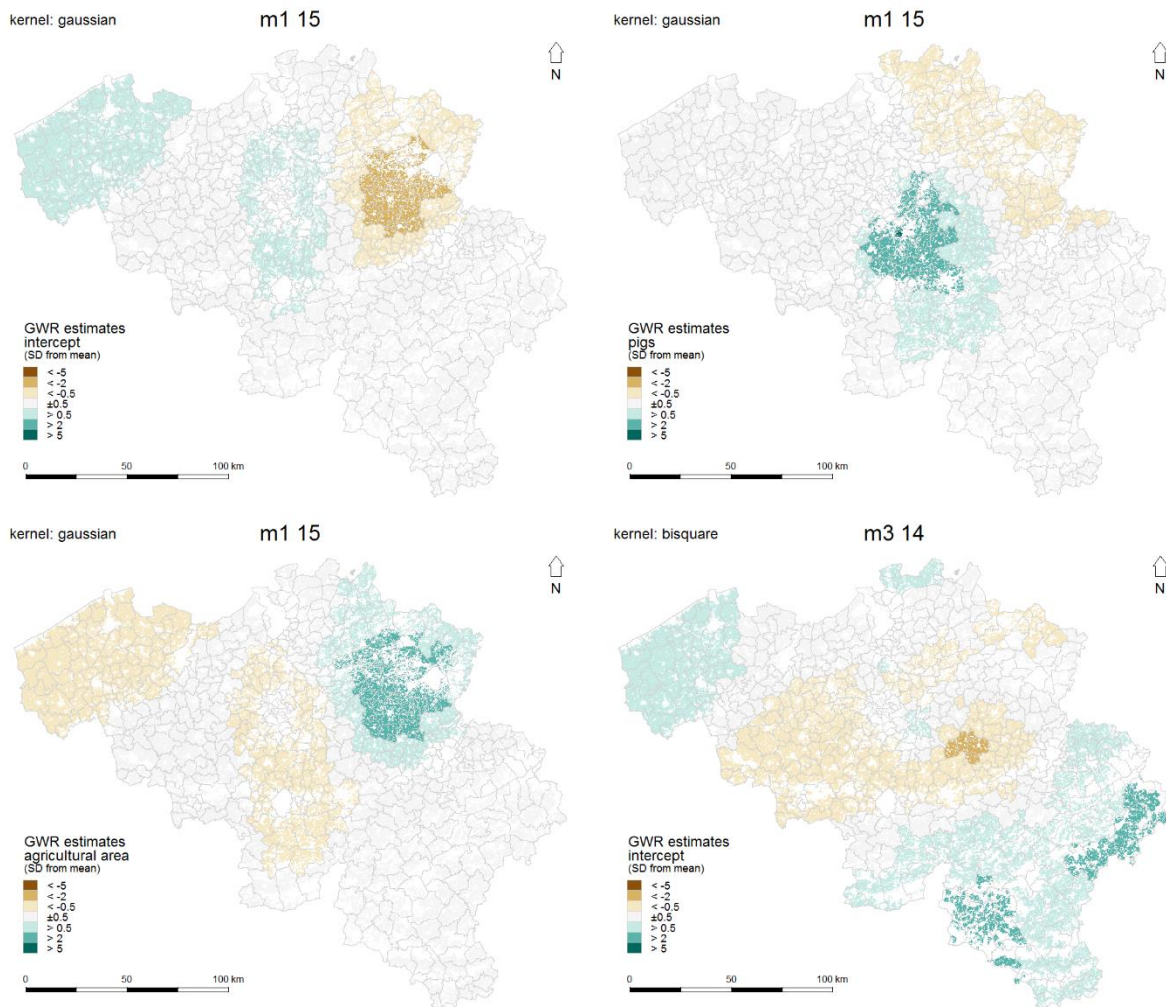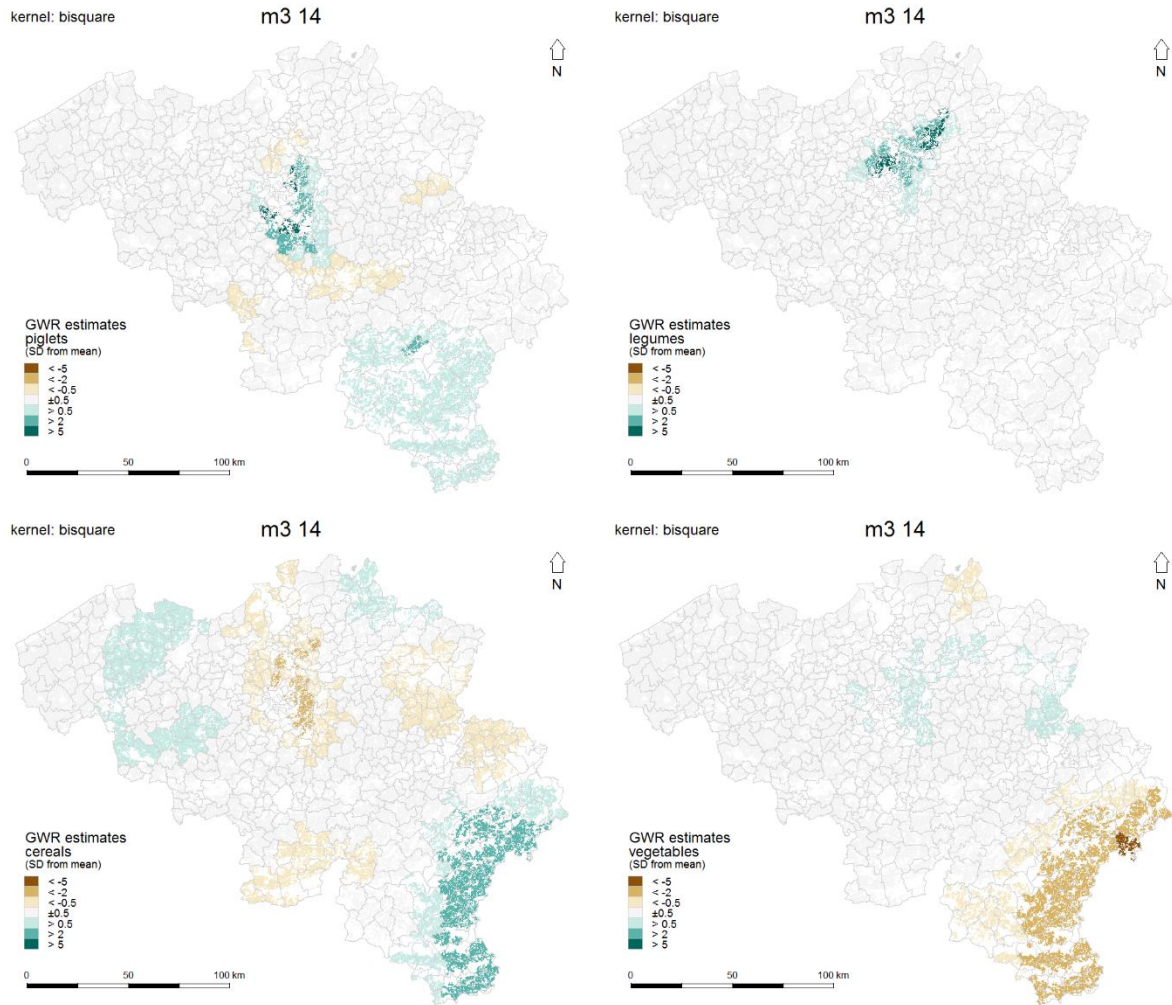
Fig. 3.7: GWR estimates for those analysis levels where a GWR model is superior.
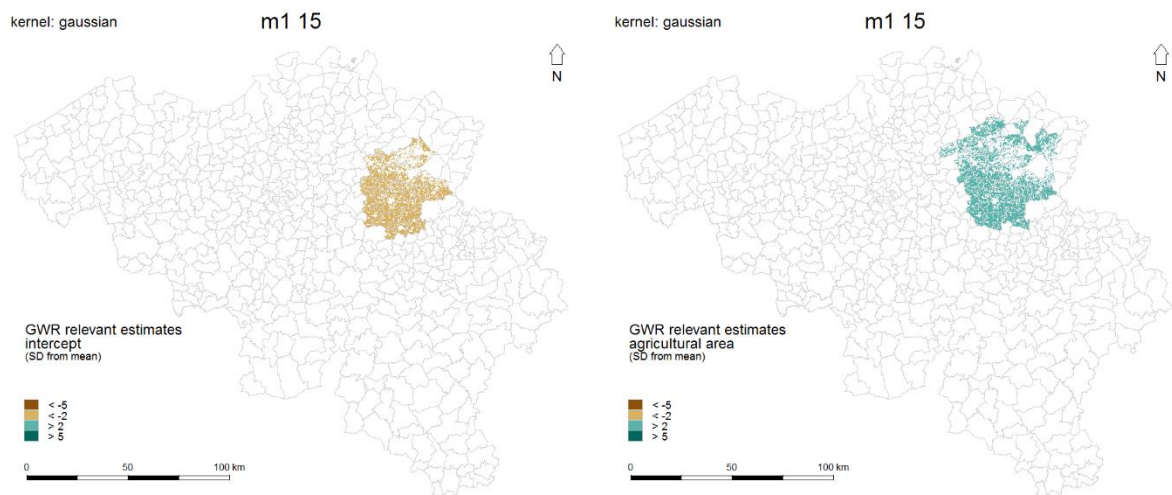
Fig. 3.8: Relevant GWR estimates for those analysis levels where a GWR model is superior.

Fig. 3.9 shows the Moran plots for those analysis levels where an MGWR model has the lowest AICc. Since the calculation of the Moran's I value for mixed GWR models is not available in any R package,

the values of the ordinary GWR have been used instead, for the sake of simplicity. With respect to the great similarity of the results between GWR and MGWR in terms of the estimates (Fig. 3.11), this step is thoroughly justifiable. The Moran plots express the picture already known at the municipal level: There is no significant autocorrelation and the municipalities Leuven, Sint Truiden and Roeselare form the outliers. The same applies to the analysis level a3 15, in which the arrondissements of Leuven, Hasselt and Roeselare can be identified as outliers. For the residual maps, the values of the GWR again were used as a substitute. Of course, all three maps confirm the findings of the Moran plots. Fig. 3.11 pictures the GWR and MGWR estimates side by side and demonstrates the almost complete consistency of the results of the local variables of the MGWR and its corresponding GWR counterparts. Most likely, the differences at analysis level a3 15 could be criticized, but they, too, are too small to justify the refusal of this approach. Therefore, the GWR estimates were also used for the representation of the relevant MGWR estimates shown in Fig. 3.12. No significant estimates exist at analysis level m2 15 for the variable "perennial grassland", at analysis level m3 15 for the variables "cereals", "cattle more than 2 years" and "pigs 20 to 50 kg and porkers" and at analysis level m3 15 for "small fruits". Significantly high values are only present at level m2 15 for the predictor "pigs" on a narrow strip reaching from the municipality of Couvin in the south to Retie in the north east, whereas significantly low values occur at level a3 15 for the predictor "cattle more than 2 years" in the arrondissement of Kortrijk.
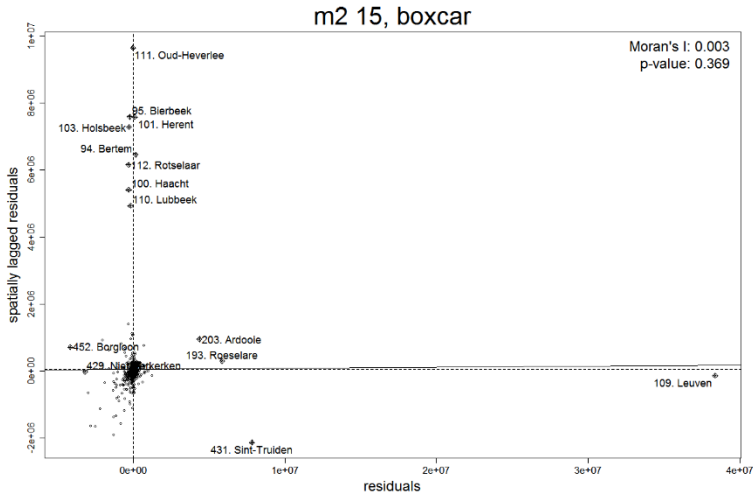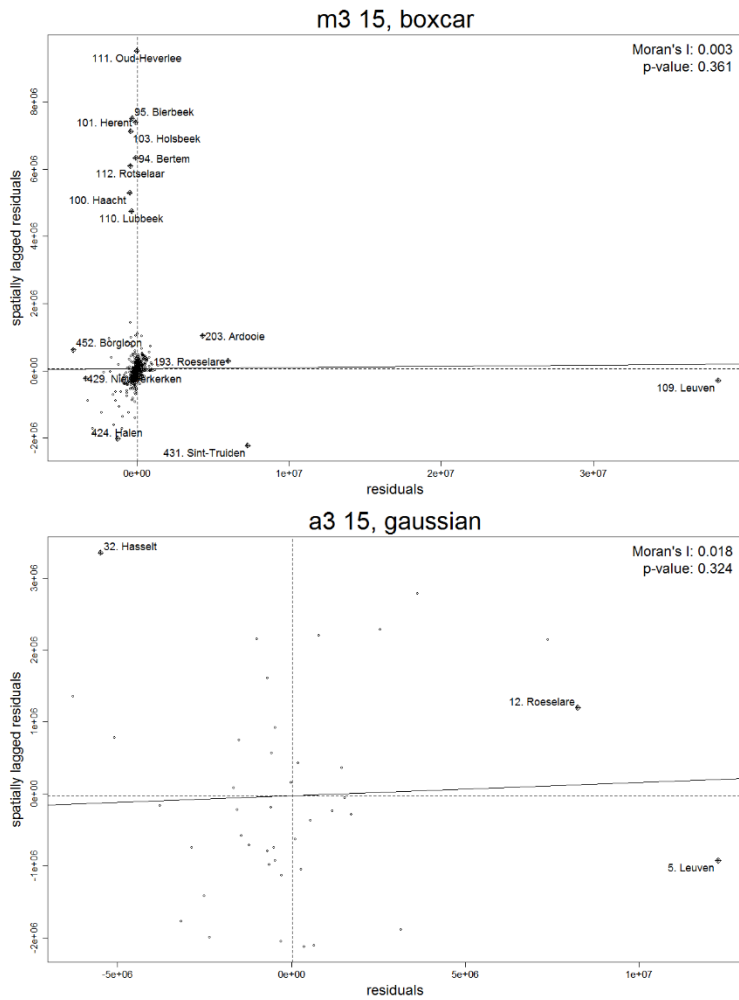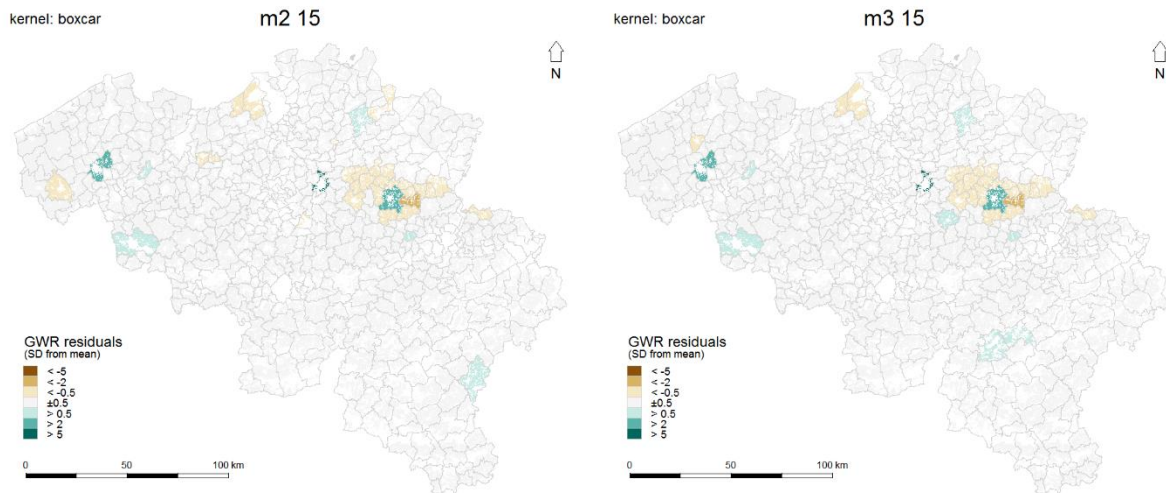
Fig. 3.9: GWR Moran plots for those analysis levels where an MGWR model is superior.

Note: Moran plots have been taken from the corresponding GWR models due to the almost identical results of GWR and MGWR.
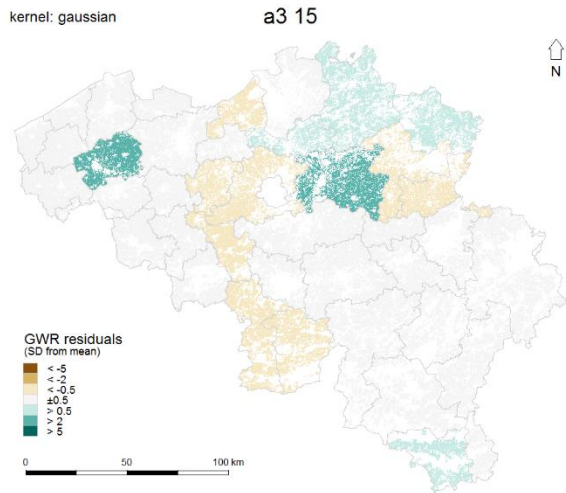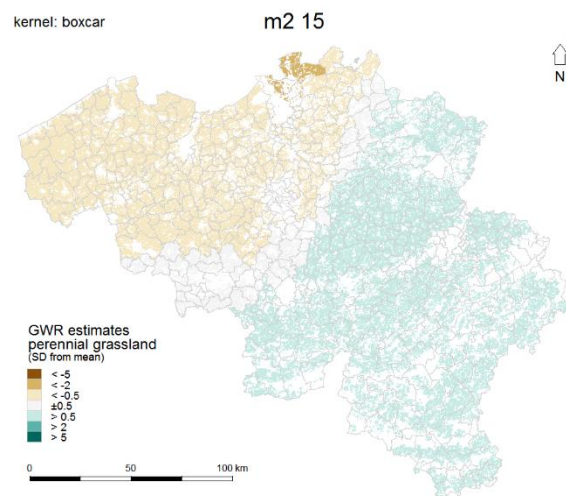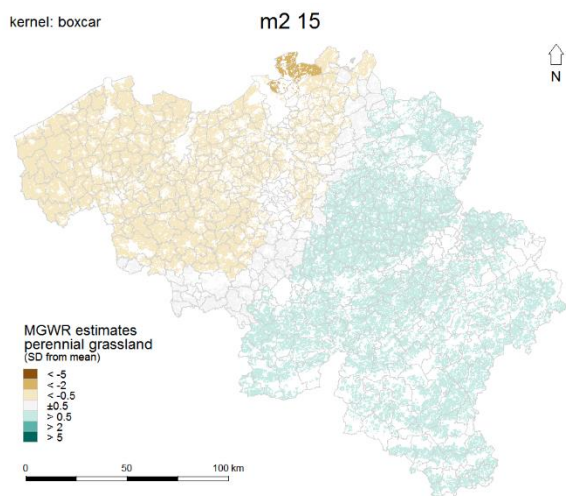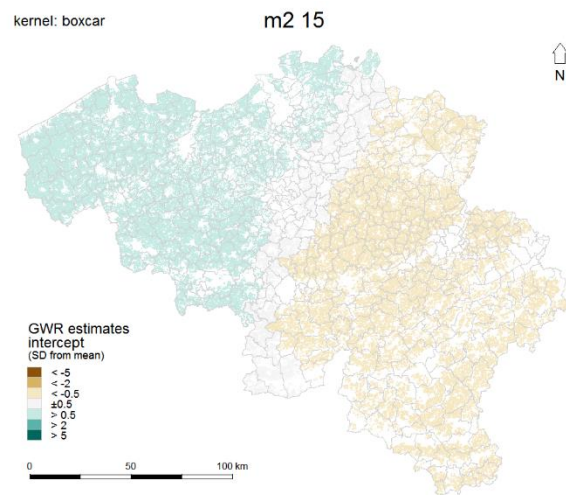
Fig. 3.10: GWR residuals for those analysis levels where a n MGWR model has the lowest AICc.
Note: Residual maps have been taken from the corresponding GWR models due to the almost identical results of GWR and MGWR.

m3 15

m3 15

MGWR estimates
cattle more than 2 yr.
(SD from mean)

< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

GWR estimates
cattle more than 2 yr.
(SD from mean)

< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

m3 15

m3 15

MGWR estimates
pigs 20 to 50 kg and porkers
(SD from mean)

< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

GWR estimates
pigs 20 to 50 kg and porkers
(SD from mean)

< -5
< -2
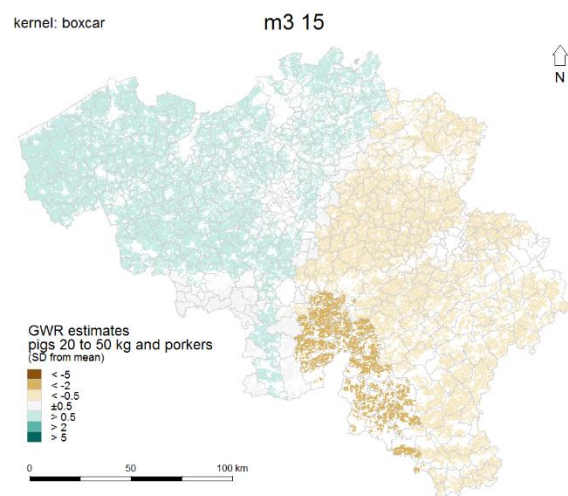< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

a3 15

a3 15

MGWR estimates
intercept
(SD from mean)

< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km

GWR estimates
intercept
(SD from mean)

< -5
< -2
< -0.5
±0.5
> 0.5
> 2
> 5

0   50   100 km
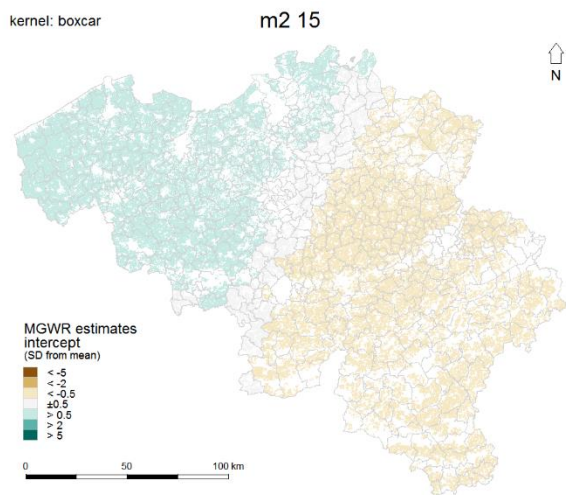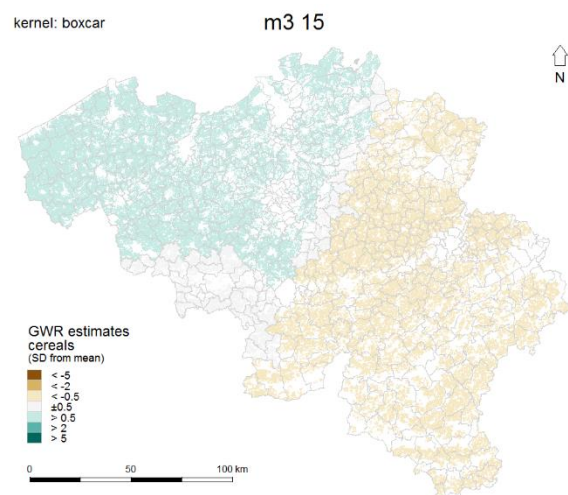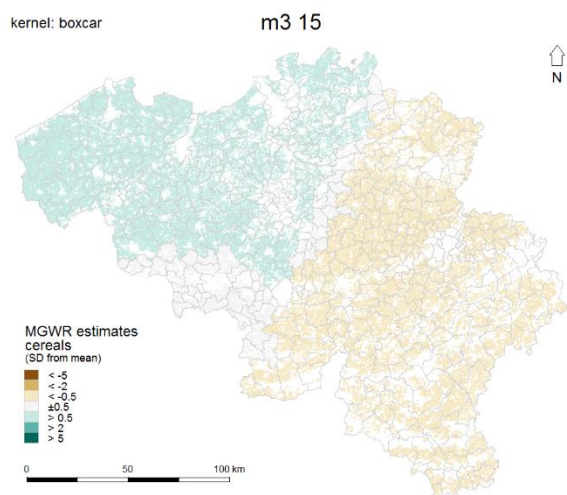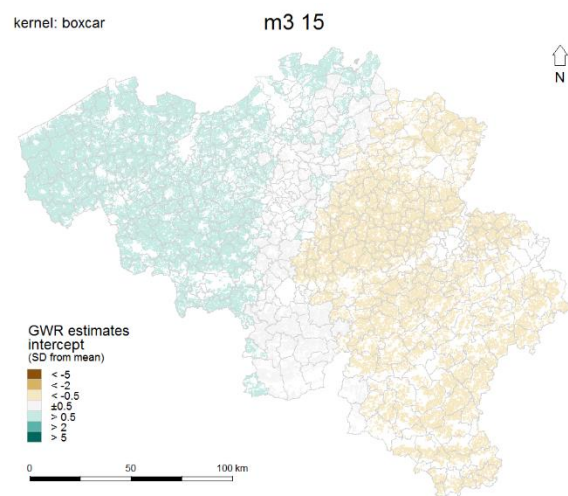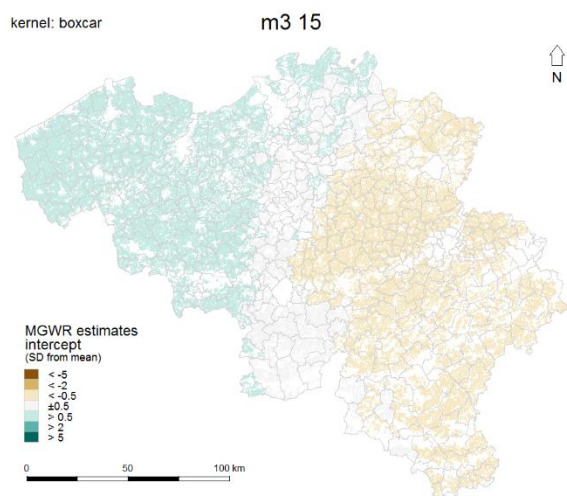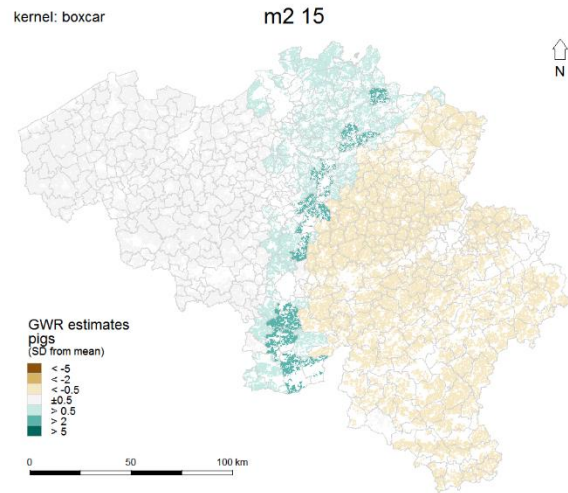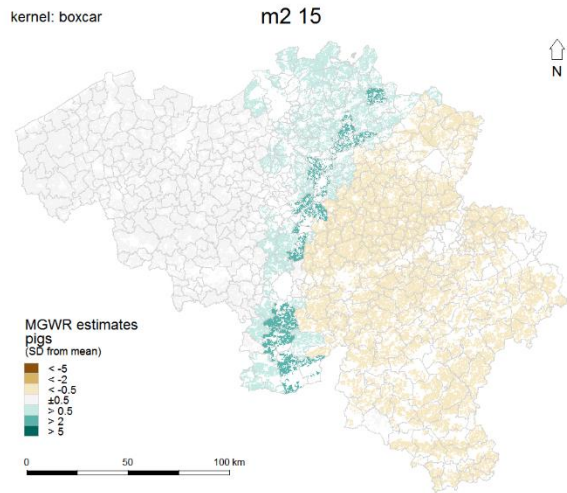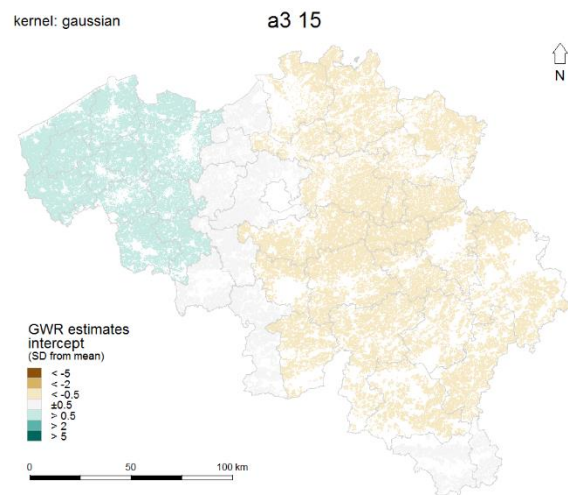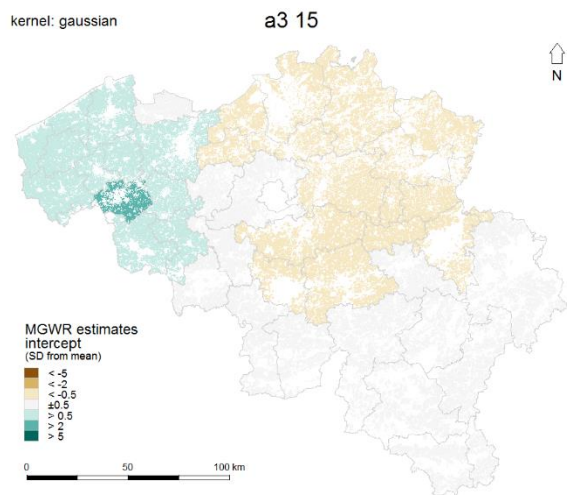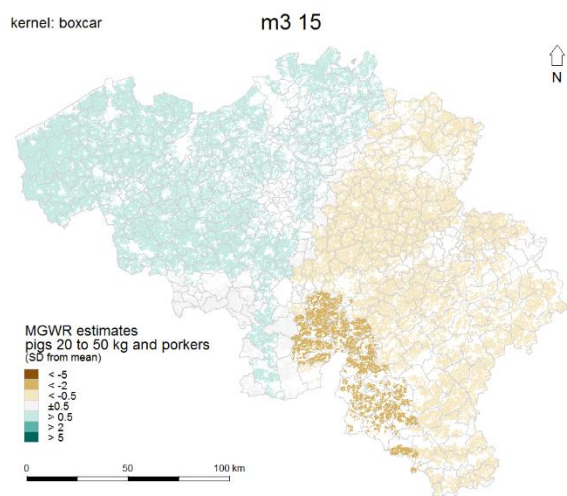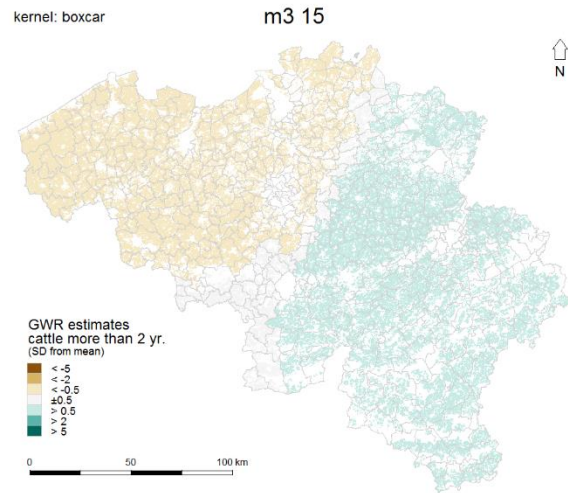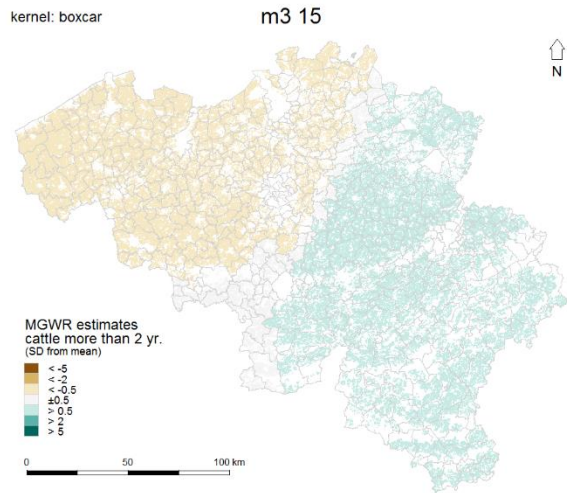
Fig. 3.11: Comparison of MGWR and GWR estimates for all those analysis levels where an MGWR model is superior.
Note: The MGWR models are pictured on the left whereas the GWR models can be found on the right side.

m2 15

N

GWR relevant estimates
pigs
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

m3 15

N

GWR relevant estimates
intercept
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

m3 15

N

GWR relevant estimates
cereals
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

m3 15

N

GWR relevant estimates
cattle more than 2 yr.
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

m3 15

N

GWR relevant estimates
pigs 20 to 50 kg and porkers
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

a3 15

N

GWR relevant estimates
intercept
(SD from mean)

< -5
< -2
> 2
> 5

0   50   100 km

Fig. 3.12: Relevant GWR estimates for those analysis levels where an MGWR model is superior.

Note: Maps of the relevant estimates have been taken from the corresponding GWR models due to the almost identical results of GWR and MGWR.
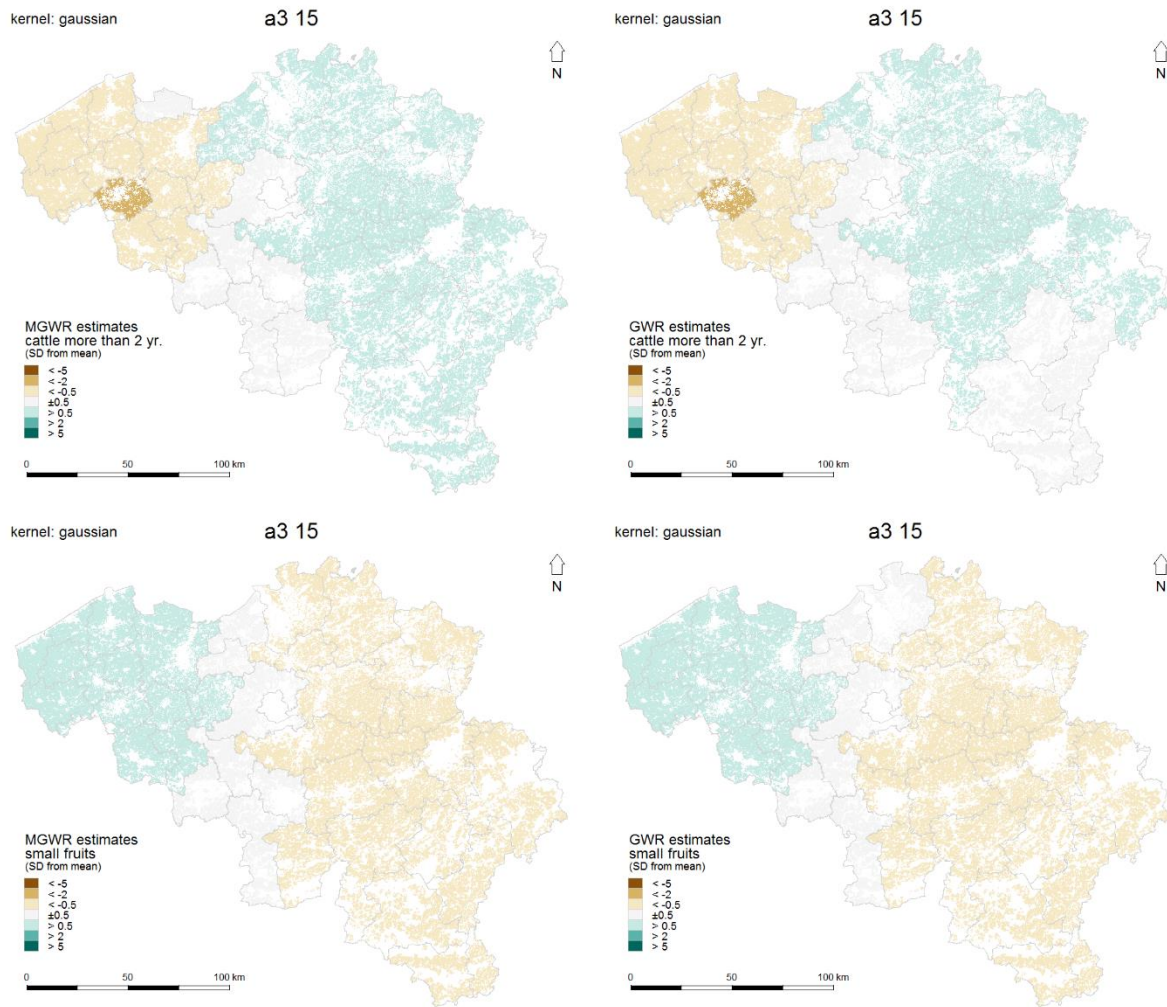
4 Discussion

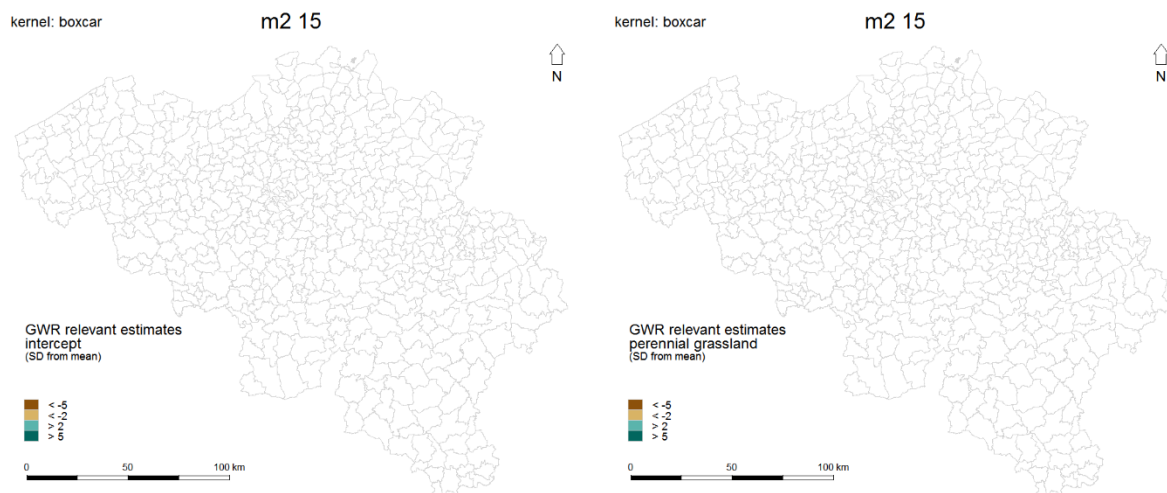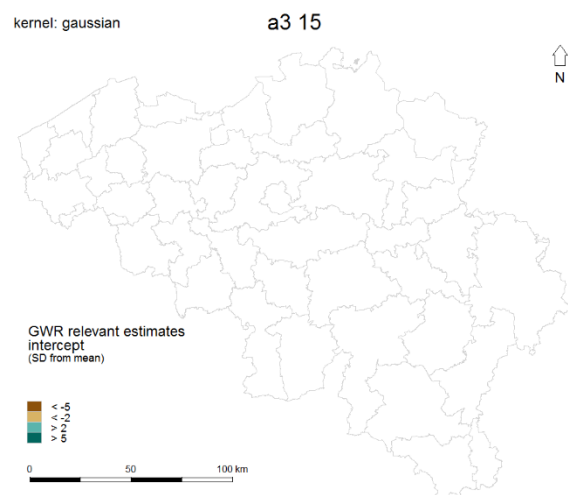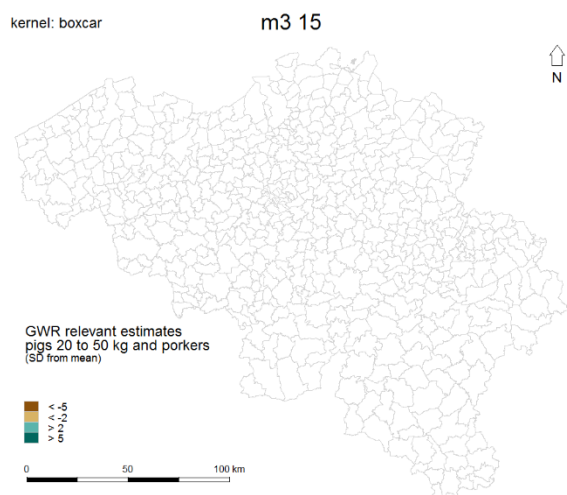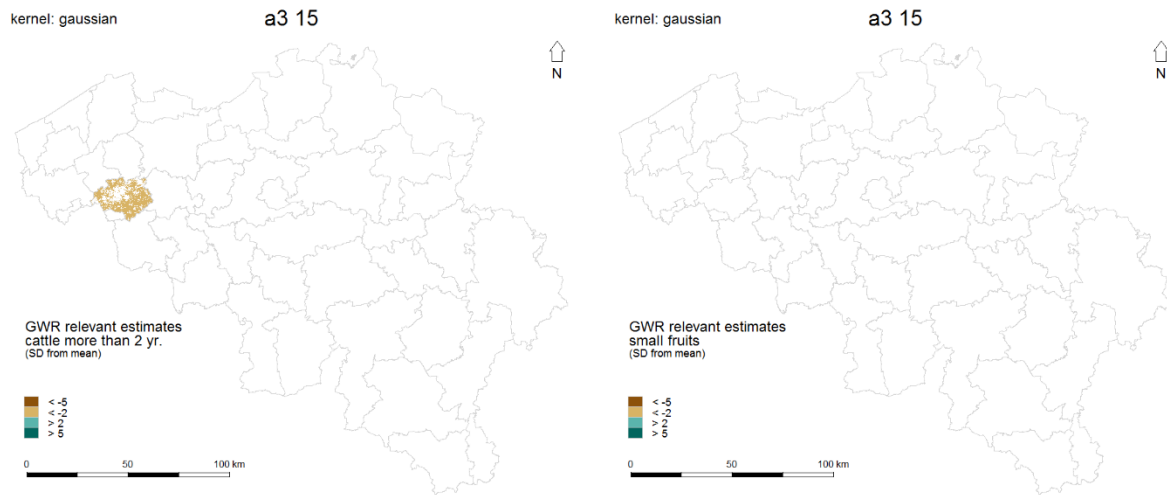## 4.1 Interpretation and Conclusions

The purpose of this work is to investigate whether the agricultural subsidies of the European Union for Belgium in 2014 and 2015 have been distributed geographically equally. However, the variety of results and maps makes an interpretation quite difficult. There is no uniform pattern for identifying certain financially disadvantaged or privileged regions. In seven cases, the global regression brought the best result which means that there are no significant differences in the distribution of the subsidies at all. Where the GWR or MGWR provides the best model, particularly high or low regression values occur only among a few predictors and their geographical distribution is not limited to specific regions, but rather a result by chance.

According to the Scale-Location plots (Fig. 3.1), moderate heteroscedasticity occurs in the two OLS models at the municipal level (m1 14 and m2 14) which is, however, less problematic since the OLS regression is relatively robust with regard to the violation of the condition of homoscedasticity. Moreover, the models only describe existing data and will not be used for predicting future events. If we look at the quantile-quantile plots (Fig. 3.1) of the OLS regression models, it is not surprising that all Jarque-Bera- and Shapiro-Wilk tests for normal distribution of the regression residuals are positive. The curvature at the ends of the line is caused by those outliers which have already been identified as such by the residual maps and the maps of the outliers of both the global and the local regression models. Both test methods react very sensitively to outliers, which is why these results are not surprising. As the graphical tests prove, there is not really any problem. In the case of outliers and / or significantly high residuals, these are above all the municipalities of Leuven, Sint Truiden and Roeselare, which are home to large agricultural cooperatives, which regularly receive very high EU agricultural subsidies. In these cases, however, it can't be assumed that the farms of these municipalities benefit excessively from the subsidies, since many agricultural businesses and farmers are organized in those cooperatives who often don't even come from one of these municipalities themselves. Of course, those municipalities have a big influence on the results at the arrondissement level: The arrondissement of Leuven, in all cases, is an upper outlier, just as is the arrondissement of

Hasselt in two cases, which the municipality of Sint Truiden belongs to. On the other hand, the significantly low residuals of the arrondissement Tongeren, which is adjacent to the arrondissement Hasselt, are likely to be a random product or statistical artefact at the analysis levels a2 14, a2 15 and a3 14. At least, no obvious reason for this can be found and the maps of the significant GWR or MGWR estimates do not indicate any abnormalities in this area.

The results of the analysis at the municipal level are to be interpreted with caution since the adjusted coefficient of determination is usually only about 0.35 which means that only about 35 percent of the variation in the data is explained by the respective model. Nonetheless, the present study does not allow the conclusion of a significantly unequal geographical distribution of the EU's agricultural subsidies among the Belgian municipalities. At the arrondissement level, the adjusted coefficient of determination is markedly higher. Here, with only one exception, the global regression models achieved the best results, indicating that there are no significant differences between the arrondissements. Therefore, the assertion can be made that there is distributive justice at the arrondissement level.

The occurrence of significant autocorrelation in the analysis approach with the hexagon grid is not surprising and shows that it is not suitable for the calculation of a GWR. The way the spatial disaggregation has been carried out automatically results in spatial autocorrelation of the residuals since many adjacent areas that are used for the calculation have been given the very same value. Autocorrelation also occurs in the analysis, which is exclusively restricted to the region of Flanders, which is why these models are not trustworthy. This is also due to the analysis design, because the respective amounts of the subsidies and predictors of each municipality were simply allocated proportionately to the agricultural areas. Thus, this approach is also not suitable.

4.2 Shortcomings

An even more interesting task surely would have been to answer the question as to whether there is a balance between small and medium-sized enterprises with regards to the EU's agricultural subsidies. However, since, except for the subsidies, there were no data freely available on the businesses themselves, only the equity of the geographical distribution could be investigated. The fact that all the cultivation areas and livestock animals of a municipality have been allocated to all the farms regardless of their actual size of course leads to statistical superposition or mutual elimination. Thus, assuming that the ratio of large to small agricultural businesses is approximately the same in each municipality, an actual existence of unequal distribution can't be detected with this analysis approach. Moreover, a

long-term investigation would be of interest. Although the parallel investigation of two budget years can catch random outliers of a year, it can't make any statement about the development within several years or decades. In addition, this work only provides information about the situation in Belgium, a country with a small agricultural sector. It does not provide any information on the transferability of the results to large countries such as Germany or the United Kingdom or to countries with a larger agricultural sector such as Portugal, Spain, France, Italy, Greece, Romania or Poland. Yet, an analysis of the whole European Union would have clearly exceeded the limits of what was possible for me in terms of time, and, what is more, I would not have had the necessary computer resources and data at my disposal. A further drawback is that only the total subsidy sum was analyzed. A much more detailed analysis would have examined the subsidies broken down by fund and purpose. These data would have been available, but again, unfortunately, no detailed data of the individual enterprises were present. Finally, it must also be mentioned that an interpretation of the results of a multiple GWR is always difficult. If it indicates differences, it is not always certain whether they are existent in reality, or whether the model is only specified insufficiently.

Bibliography

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*. In: Petrov, B. N. & Caski, F. (eds). Proceedings of the Second International Symposium on Information Theory: 267-281, Akadémiai Kiadó, Budapest.

Anselin, L. (1996). *The Moran scatterplot as an ESDA tool to assess local instability in spatial association*. In: Fischer, M., Scholten, H., & Unwin, D. (eds.). Spatial Analytical Perspectives on GIS: 111-125, Taylor and Francis, London.

Bera, A. K. & Jarque, C. M. (1980). *Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals*. Economic Letters 6: 255-259.

Breusch, T. S. & Pagan, A. R. (1979). *A Simple Test for Heteroscedasticity and Random Coefficient Variation*. Econometrica 47 (5): 1287-1294.

Brunsdon, C., Fotheringham A. S. & Charlton, M. E. (1996). *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity*. Geographical Analysis 28 (4): 281-298.

Brunsdon, C., Charlton, M. E. & Harris, P. (2012). *Living with Collinearity in Local Regression Models*. In: Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Florianópolis.

Burnham, K. P. & Anderson, D. R. (2004). *Multimodel Inference. Understanding AIC and BIC in Model Selection*. Sociological Methods & Research 33 (2): 261-304.

Calcagno, V. & de Mazancourt, C. (2010). *glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models.* Journal of Statistical Software 34 (12): 1-29.

Durbin, J. & Watson, G. S. (1950). *Testing for Serial Correlation in Least Squares Regression. I*. Biometrika 37 (3-4): 409-428.

Durbin, J. & Watson, G. S. (1951). *Testing for Serial Correlation in Least Squares Regression. II*. Biometrika 38 (1-2): 159-179.

European Parliament & Council of the European Union (2013). *Regulation (EU) No 1307/2013 establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009.* URL: `http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1494767623664&uri=CELEX:32013R1307` (retrieved May 14, 2017).

European Commission (2014). *Regulation (EU) No 639/2014 supplementing Regulation (EU) No 1307/ 2013 of the European Parliament and of the Council establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and amending Annex X to that Regulation.* URL: `http://eur-lex.europa.eu/legal-content/ EN/TXT/?qid=1494768352669&uri=CELEX:32014R0639` (retrieved May 14, 2017).

Fotheringham, A. S., Brunsdon, C. & Charlton, M. E. (2002). *Geographically Weighted Regression. The analysis of spatially varying relationships.* John Wiley & Sons, Chichester.

Fox, J. (1991). *Regression Diagnostics*. Quantitative Applications in the Social Sciences 79, Sage, Newbury Park.

Gollini, I., Lu, B., Charlton, M. E., Brunsdon, C. & Harris, P. (2015). *GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models*. Journal of Statistical Software 63 (17): 1-50.

Koenker, R. (1981). *A Note on Studentizing a Test for Heteroscedasticity*. Journal of Econometrics 17: 107-112.

Moran, P. A. P. (1950). *Notes on Continuous Stochastic Phenomena*. Biometrika 37 (1): 17-23.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. CATMOG 38. GeoBooks, Norwich.

Pardoe, I. (2006). *Applied Regression Modeling. A Business Approach*. John Wiley & Sons, Hoboken.

Pearson, K. (1895). *Note on Regression and Inheritance in the Case of Two Parents*. Proceedings of the Royal Society of London 58: 240-242.

QGIS Development Team (2016). *QGIS Geographical Information System*. Open Source Geospatial Foundation. URL: `http://www.qgis.org/` (retrieved October 13, 2016).

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. URL: `http://www.r-project.org/` (retrieved June 30, 2016).

Raftery, A. E. (1995). *Bayesian Model Selection in Social Research*. Sociological Methodology 25: 111-163.

RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Boston. URL: `http://www.rstudio.com/` (retrieved December 12, 2016).

Shapiro, S. S. & Wilk, M. B. (1965). *An analysis of variance test for normality (complete samples)*. Biometrika 52 (3-4): 591-611.

Sugiura, N. (1978). *Further Analysis of the Data by Akaike's Information Criterion and Finite Corrections*. Communications in Statistics - Theory and Methods 7 (1): 13-26.

Appendix

A R-Code

A.1 Main Script

```
##############################################################################
##################### PREPARATION OF ENVIRONMENT #####################
##############################################################################

# load packages
 library(rgdal); library(sfsmisc); library(MASS); library(spdep); library(tseries); library(GWmodel);
   library(car); library(lmtest); library(glmulti); library(QuantPsyc); library(plyr); library(rgeos);
   library(stringr); library(maptools); library(GISTools)
# source scripts
 source("R scripts/gwr.mixed (corrected).R")
 source("R scripts/plot.glmulti2.R")
 source("R scripts/moran.plot2.R")
 source("R scripts/Scalebar2.R")
 source("http://www.math.mcmaster.ca/bolker/R/misc/legendx.R")
# create lists with strings and numbers
 listLayersAgric <- c(rep(c("m14", "m15"), 3), "g14", "g15", rep(c("a14", "a15"), 3), "f14", "f15")
 listLayersAdmin <- c(rep("mun", 8), rep("arr", 6), rep("f_mun", 2))
 listVarNrs.1 <- rep(as.data.frame(c(10, 11, 42, 54, 58, 61)), 2)
 listVarNrs.2 <- rep(as.data.frame(c(10, 12, 32, 38, 39, 42, 54, 58, 61)), 2)
 listVarNrs.3 <- rep(as.data.frame(c(10, 13, 18, 23:25, 29:31, 33:34, 37:38, 40:41, 43, 46:47, 55:57,
   59:61)), 2)
 listVarNrs.4 <- rep(as.data.frame(c(10, 11, 42, 54, 58, 61) + 2), 2)
 listVarNrs.5 <- rep(as.data.frame(c(10, 11, 42, 54, 58, 61) - 6), 2)
 listVarNrs.6 <- rep(as.data.frame(c(10, 12, 32, 38, 39, 42, 54, 58, 61) - 6), 2)
 listVarNrs.7 <- rep(as.data.frame(c(10, 13, 18, 23:25, 29:31, 33:34, 37:38, 40:41, 43, 46:47, 55:57,
   59:61) - 6), 2)
 listVarNrs.8 <- rep(as.data.frame(c(3:16)), 2)
 listVarNrs <- c(listVarNrs.1, listVarNrs.2, listVarNrs.3, listVarNrs.4, listVarNrs.5, listVarNrs.6,
   listVarNrs.7, listVarNrs.8)
 listVarNames.1 <- rep(as.data.frame(c("businesses", "agricultural_area", "cattle", "pigs", "poultry",
   "subsidies")), 2)
 listVarNames.2 <- rep(as.data.frame(c("businesses", "arable_land", "perennial_plants",
   "perennial_grassland", "greenhouse_crops", "cattle", "pigs", "poultry", "subsidies")), 2)
```

```r
listVarNames.3 <- rep(as.data.frame(c("businesses", "cereals", "industrial_crops", "potatoes",
  "legumes", "fodder_plants", "vegetables", "horticultural_plants", "fallows", "plant_nurseries",
  "orchards", "small_fruits", "perennial_grassland", "greenhouse_vegetables",
  "horticultural_plants_in_greenhouses", "calves", "cattle_1_to_2_yr.", "cattle_more_than_2_yr.",
  "piglets", "breeding_sows_more_than_50_kg", "pigs_20_to_50_kg_and_porkers", "laying_hens",
  "broilers", "subsidies")), 2)
listVarNames.4 <- rep(as.data.frame(c("potatoes", "fruits_and_nuts", "cereals_seeds_and_legumes",
  "grassland", "vegetables_herbs_and_horticultural_plants", "woody_plants",
  "agricultural_infrastructure", "maize", "other_plants", "sugar_beets", "flax_and_hemp",
  "fodder_plants", "no_information", "subsidies")), 2)
listVarNames <- c(listVarNames.1, listVarNames.2, listVarNames.3, listVarNames.1, listVarNames.1,
  listVarNames.2, listVarNames.3, listVarNames.4)
listVarLabels.1 <- rep(as.data.frame(c("businesses", "agricultural area", "cattle", "pigs", "poultry",
  "subsidies", "intercept")), 2)
listVarLabels.2 <- rep(as.data.frame(c("businesses", "arable land", "perennial plants", "perennial
  grassland", "greenhouse crops", "cattle", "pigs", "poultry", "subsidies", "intercept")), 2)
listVarLabels.3 <- rep(as.data.frame(c("businesses", "cereals", "industrial crops", "potatoes",
  "legumes", "fodder plants", "vegetables", "horticultural plants", "fallows", "plant nurseries",
  "orchards", "small fruits", "perennial grassland", "greenhouse vegetables", "horticultural plants
  in greenhouses", "calves", "cattle 1 to 2 yr.", "cattle more than 2 yr.", "piglets", "breeding sows
  more than 50 kg", "pigs 20 to 50 kg and porkers", "laying hens", "broilers", "subsidies",
  "intercept")), 2)
listVarLabels.4 <- rep(as.data.frame(c("potatoes", "fruits and nuts", "cereals, seeds and legumes",
  "grassland", "vegetables, herbs and horticultural plants", "woody plants", "agricultural
  infrastructure", "maize", "other plants", "sugar beets", "flax and hemp", "fodder plants", "no
  information", "subsidies", "intercept")), 2)
listVarLabels <- c(listVarLabels.1, listVarLabels.2, listVarLabels.3, listVarLabels.1,
  listVarLabels.1, listVarLabels.2, listVarLabels.3, listVarLabels.4)
listUnit.1 <- rep(as.data.frame(c("", " (a)", "", "", "", " (€)", "")), 2)
listUnit.2 <- rep(as.data.frame(c("", " (a)", " (a)", " (a)", " (a)", "", "", "", " (€)", "")), 2)
listUnit.3 <- rep(as.data.frame(c("", " (a)", " (a)", " (a)", " (a)", " (a)", " (a)", " (a)", " (a)",
  " (a)", " (a)", " (a)", " (a)", " (a)", " (a)", "", "", "", "", "", "", "", "", " (€)", "")), 2)
listUnit.4 <- rep(as.data.frame(c(" (a)", " (a)", " (a)", " (a)", " (a)", " (a)", " (a)", " (a)", "
  (a)", " (a)", " (a)", " (a)", " (a)", " (€)", "")), 2)
listUnit <- c(listUnit.1, listUnit.2, listUnit.3, listUnit.1, listUnit.1, listUnit.2, listUnit.3,
  listUnit.4)
listTags <- c("m1 14", "m1 15", "m2 14", "m2 15", "m3 14", "m3 15", "g14", "g15", "a1 14", "a1 15",
  "a2 14", "a2 15", "a3 14", "a3 15", "f14", "f15")


############################################################################
####################### CREATION OF FUNCTIONS ######################### ##########
############################################################################

# fetch strings and numbers from data
 fetchData <- function(x) {
   assign("layerAgric", listLayersAgric[x], envir = .GlobalEnv)
   assign("layerAdmin", listLayersAdmin[x], envir = .GlobalEnv)
   assign("varNrs", listVarNrs[[x]], envir = .GlobalEnv)
```

```r
    assign("varNames", as.character(listVarNames[[x]]), envir = .GlobalEnv)
    assign("varLabels", as.character(listVarLabels[[x]]), envir = .GlobalEnv)
    assign("unit", as.character(listUnit[[x]]), envir = .GlobalEnv)
    assign("tag", listTags[x], envir = .GlobalEnv)
  }
# prompt to enter a number to choose a dataset
chooseData <- function() {
    dataSet <- readline(prompt = "enter a number from 1 - 16 to choose a dataset for the statistical
    analysis: ")
  fetchData(as.integer(dataSet))
  }
# create list with predictors that should participate in model selection
fetchPreds <- function(x) {
    if (length(x) == 0)
      assign("xr", names(preds), envir = .GlobalEnv)
    else
      assign("xr", names(subset(preds, select = -x)), envir = .GlobalEnv)
  }
# prompt to enter variable numbers to be excluded from model selection
choosePreds <- function() {
    cat("enter the number(s) (corMatNr) of the variable(s) you want to exclude from the analysis (comma-
    separated, press ENTER if you want to keep all variables): ")
    choice <- scan("", sep = ",", quiet = T, nlines = 1)
    fetchPreds(choice)
  }
# select user-defined basic regression model
fetchLmModel <- function(x) {
    assign("lmNr", x, envir = .GlobalEnv)
    assign("lmModel", lm(lmSelect@formulas[[lmNr]]), envir = .GlobalEnv)
  }
# prompt to enter a number to choose a basic regression model
chooseLmModel <- function() {
    lmChoice <- readline(prompt = "enter the number of the basic regression model you want to use for
    further analysis: ")
    fetchLmModel(as.integer(lmChoice))
  }
# draw coplot
panLm <- function(x, y, ...) {
    points(x, y, pch = 20, cex = 0.7)
    abline(lm(y ~ x))
  }
# calculate interval limits
sdInt <- function(x) {
    nr <- c(-1000, -5, -2, -0.5, 0.5, 2, 5, 1000)
    regSd <- sd(x)
    regMean <- mean(x)
    regIv <- list()
    regIv <- lapply(nr, function(x) regMean + (x * regSd))
```

```
    regIv <- unlist(regIv)
    return(regIv)
 }
# draw choropleth map from standard intervals
 sdMap <- function(data, shp, ...) {
    int <- sdInt(data)
    cat <- findInterval(data, int)
    look <- palDiv[cat]
    plot(shp, col = look, border = NA, ...)
 }
# create vector with class limits for p-values
 pVal <- c(0, 0.001, 0.01, 0.05, 0.1, 1)
# draw choropleth map from p-values
 pMap <- function(data, shp, ...) {
    cat <- findInterval(data, pVal)
    look <- palP[cat]
    plot(shp, col = look, border = NA, ...)
 }
# draw choropleth map from standard intervals for significant estimates and relative predictors
 sdMapSig <- function(cat, shp, ...) {
    look <- palDiv[cat]
    plot(shp, col = look, border = NA, ...)
 }


# check if a value is inside a range
 isBetween <- function(x, a, b) {
    x > a & x < b
 }
# calculate standardized residuals
 stand <- function(x) {
    mn <- mean(x)
    std <- sd(x)
    (x - mn) / std
 }


############################################################################
######################### LOAD AND PREPARE DATA #########################
############################################################################


# choose dataset
 chooseData()
# load shapefiles and transform them to data frames
 agric <- readOGR(dsn = "C:/master/analysis/shapes", layer = layerAgric, stringsAsFactors = F, encoding
    = "UTF-8", use_iconv = T)
 names(agric@data)[varNrs] <- varNames
 agricDf <- data.frame(agric)
 agricRows <- seq.int(nrow(agricDf))
```

```r
 admin <- readOGR(dsn = "C:/master/analysis/shapes", layer = layerAdmin, stringsAsFactors = F, encoding
    = "UTF-8", use_iconv = T)
# append sequential numbers to the spatial data frame "agric" and create a new one
 nrAgric <- 1:length(agric)
 agricNr <- spCbind(agric, nrAgric)
# create datasets with potential predictors
 vars <- agricDf[varNrs]
 preds <- subset(vars, select = -subsidies)
# calculate polygon centroids
 cents <- gCentroid(agric, byid = T)
# create topological lists of the polygons
 topoList <- poly2nb(agric)
# calculate spatial weights for the polygons
 spatWgts <- nb2listw(topoList, zero.policy = T)
# create vector with probability values
 probs <- c(0, seq(0.2, 0.5, 0.15), seq(0.6, 1, 0.1))
# create vector with names of kernel functions for the GWR
 kernels <- c("gaussian", "bisquare", "boxcar")
# create color palettes
 palDiv <- c("#8c510a", "#d8b365", "#f6e8c3", "#f5f5f5", "#c7eae5", "#5ab4ac", "#01665e")
 palDivSig <- c("#8c510a", "#d8b365", "#5ab4ac", "#01665e")
 palP <- c("#006d2c", "#2ca25f", "#66c2a4", "#b2e2e2", "#edf8fb")
# create labels for legends
 sdMapLegLab <- c("< -5", "< -2", "< -0.5", "\u00B10.5", "> 0.5", "> 2", "> 5")
 sdMapSigLegLab <- c("< -5", "< -2", "> 2", "> 5")
 pMapLegLab <- c("\u2264 0.001", "\u2264 0.01", "\u2264 0.05", "\u2264 0.1", "> 0.1")


################################################################################
##################### GLOBAL LINEAR REGRESSION #########################
################################################################################


############### LINEAR REGRESSION MODEL SELECTION #################

 attach(vars)

# correlation matrix of all potential predictors
 symnum(corMat <- round(cor(preds), 2))
 corMat[upper.tri(corMat, diag = T)] <- NA
 corMatId <- seq.int(nrow(corMat))
 (corMatNr <- cbind(corMatId, corMat))

# choose predictors to be excluded from model selection (press enter )
 choosePreds()
# calculate possible models
 lmSelect <- glmulti("subsidies", xr = xr, level = 1, crit = "aicc", plotty = F, data = vars)
 (lmTable <- weightable(lmSelect))
 img <- file.path(paste0(tag, ". OLS - 1. model selection plot.tif"))
 tiff(img, width = 1280, height = 850)
```

```
  par(mar = c(5, 5, 4, 2))
  plot.glmulti2(lmSelect, main = tag, cex.main = 3.5, font.main = 1, xaxt = "n", cex.axis = 2, cex.lab =
    2.5)
  title(ylab = "AICc", cex.lab = 2.5, line = 3)
  axis(side = 1, at = c(1, seq(5, length(lmSelect@formulas), 5)), cex.axis = 2)
  abline(v = c(1, seq(5, length(lmSelect@formulas), 5)), col = "light grey")
  dev.off()
  (lmTableBest <- lmTable[lmTable$aicc <= min(lmTable$aicc) + 2, ])
  lmBest <- lapply(lmSelect@objects, function(x) summary(x))
  (lmBest[1:nrow(lmTableBest)])
# choose model for further analysis
  chooseLmModel()
  summary(lmModel)
# calculate beta coefficients
  lm.beta(lmModel)


################# LINEAR REGRESSION DIAGNOSTICS ###################

# test predictors for multicollinearity (redundancy) (x > 10 => multicollinearity)
  vif(lmModel)
# diagnostic plots
  img <- file.path(paste0(tag, ". OLS - 2. diagnostic plots.tif"))
  tiff(img, width = 1000, height = 1000)
  par(mfrow = c(2, 2), oma = c(0, 1, 2, 1), mgp = c(2.5, 1, 0))
  if (layerAdmin == "arr") {
    plot(lmModel, cex.caption = 2, sub.caption = NA, labels.id = paste0(nrAgric, ". ", agric$ARRNAM),
    label.pos = c(1, 3, 2), cex.id = 1.6, cex.lab = 1.5, cex.axis = 1.5)
    title(tag, cex.main = 3.5, font.main = 1, outer = T, line = -1)
  } else {
    plot(lmModel, cex.caption = 2, sub.caption = NA, labels.id = paste0(nrAgric, ". ", agric$MUN),
    label.pos = c(1, 3, 2), cex.id = 1.6, cex.lab = 1.5, cex.axis = 1.5)
    title(tag, cex.main = 3.5, font.main = 1, outer = T, line = -1)
  }
  dev.off()
# check studentized residuals for outliers
  lmModelStdRes <- rstandard(lmModel)
  nrStdRes <- 1:length(lmModelStdRes)
  lmModelStdResNr <- cbind(lmModelStdRes, nrStdRes)
  lmModelStdResOutl <- subset(lmModelStdResNr, (lmModelStdResNr[, 1] > 3 | lmModelStdResNr[, 1] < -3))
  lmModelStdResOutlDf <- as.data.frame(lmModelStdResOutl)
  lmOutlierNames <- agric[[2]][lmModelStdResOutl[, 2]]
  lmOutlierNames[duplicated(lmOutlierNames)] <- ""
  lmOutlierNrsPre <- lmModelStdResOutlDf$nrStdRes
  lmOutlierNrsAux <- list()
  j <- 1
  for (i in lmOutlierNrsPre) {
    if (lmOutlierNames[j] == "")
      lmOutlierNrsAux[j] <- ""
```

74

```
    else
      lmOutlierNrsAux[j] <- lmOutlierNrsPre[j]
    j <- j + 1
  }
  lmOutlierNrs <- lmOutlierNrsAux
  lmOutlierLabs <- list()
  j <- 1
  for (i in lmOutlierNrs) {
    if (nchar(i) != 0)
      lmOutlierLabs[j] <- paste0(lmOutlierNrs[j], ". ", lmOutlierNames[j])
    else
      lmOutlierLabs[j] <- ""
    j <- j + 1
  }
  lmOutlier <- subset(agricNr, agricNr$nrAgric %in% lmModelStdResOutl[nrStdRes])
  lmOutlierCents <- gCentroid(lmOutlier, byid = T)
  img <- file.path(paste0(tag, ". OLS - 3. outliers standardized residuals.tif"))
  if (layerAdmin == "f_mun") {
    tiff(img, width = 1000, height = 550)
    par(mar = c(0, 0, 3, 0))
    plot(admin, border = "light grey")
    plot(lmOutlier, col = "red", border = NA, add = T)
    pointLabel(lmOutlierCents@coords, unlist(lmOutlierLabs), cex = 2)
    north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
    Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
    title(tag, line = 0, cex.main = 3, font.main = 1)
  } else {
    tiff(img, width = 1000, height = 850)
    par(mar = c(0, 0, 3, 0))
    plot(admin, border = "light grey")
    plot(lmOutlier, col = "red", border = NA, add = T)
    pointLabel(lmOutlierCents@coords, unlist(lmOutlierLabs), cex = 2)
    north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
    Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
    title(tag, line = 0, cex.main = 3, font.main = 1)
  }
  dev.off()


## --- test residuals for autocorrelation ---
# Durbin-Watson test
  dwt(lmModel)
# global Moran's I (H0: no autocorrelation)
  (moranLm <- lm.morantest(lmModel, spatWgts, zero.policy = T))
# Moran scatter plot
  img <- file.path(paste0(tag, ". OLS - 4. Moran plot.tif"))
  tiff(img, width = 1280, height = 850)
  par(mar = c(5, 5, 4, 2))
```

```
  moran.plot2(lmModel$residuals, spatWgts, zero.policy = T, labels = paste0(agricRows, ". ",
    agric[[2]]), main = tag, cex.main = 3.5, font.main = 1, xlab = "residuals", ylab = "spatially
    lagged residuals", cex = 0.8, cex.lab = 2, cex.axis = 1.4)
  mtext(paste0("Moran's I: ", round(moranLm$estimate[1], 3), "\np-value: ", round(moranLm$p.value[[1]],
    3)), side = 4, cex = 2, las = 1, adj = 1, padj = -6.4, line = -1.3)
  dev.off()


# Breusch-Pagan test against heteroscedasticity (H0: residuals homoskedastic)
  bptest(lmModel, studentize = T)


## --- test residuals for normality ---
# Jarque-Bera test
  jarque.bera.test(lmModel$residuals)
# Shapiro-Wilk test
  shapiro.test(lmModel$residuals)


## --- map residuals ---
  img <- file.path(paste0(tag, ". OLS - 5. residuals.tif"))
  if (layerAdmin == "f_mun") {
    tiff(img, width = 1000, height = 550)
    par(mar = c(0, 0, 3, 0))
    sdMap(lmModel$residuals, agricS)
    plot(admin, border = "light grey", add = T)
    north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
    Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
    title(tag, line = 0, cex.main = 3, font.main = 1)
    par(lheight = 0.8)
    text(15000, 252000, "OLS residuals", pos = 4, cex = 2)
    text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
    legend("topleft", inset = c(0, 0.15), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8, border =
    "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
  } else {
    tiff(img, width = 1000, height = 850)
    par(mar = c(0, 0, 3, 0))
    sdMap(lmModel$residuals, agricS)
    plot(admin, border = "light grey", add = T)
    north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
    Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
    title(tag, line = 0, cex.main = 3, font.main = 1)
    par(lheight = 0.8)
    text(25000, 97000, "OLS residuals", pos = 4, cex = 2)
    text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
    legend("bottomleft", inset = c(0.04, 0.14), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8,
    border = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
  }
  dev.off()
```

```
###########################################################################
################### GEOGRAPHICALLY WEIGHTED REGRESSION ###################
###########################################################################


##########==================== BASIC GWR ====================##########
###########################################################################


# determine optimal bandwidth for GWR
 bandw <- list()
 j <- 1
 for (i in kernels) {
   bw <- bw.gwr(lmModel, data = agric, approach = "AIC", kernel = i, adaptive = T)
   bandw[[j]] <- bw
   j <- j + 1
 }
 bandw


## --- GWR model selection ---
 gwrList <- list()
 j <- 1
 for (i in bandw) {
   gwrList[[j]] <- model.selection.gwr(DeVar = "subsidies", InDeVars = names(preds), data = agric, bw =
   i, kernel = kernels[j], adaptive = T, approach = "AIC")
   j <- j + 1
 }
# choose rows with lowest AICc
 gwrListNr <- list()
 j <- 1
 for (i in gwrList) {
   gwrListNr[[j]] <- which(i[[2]][, 3] == min(i[[2]][, 3]), arr.ind = T)
   j <- j + 1
 }
# extract models with lowest AICc
 gwrListBest <- list()
 j <- 1
 for (i in gwrListNr) {
   k <- as.integer(i[[1]][1])
   gwrListBest[[j]] <- gwrList[[j]][[1]][[k]][[1]]
   j <- j + 1
 }


## --- perform basic GWR ---
 gwrModel <- list()
 j <- 1
 for (i in bandw) {
   gwrModel[[j]] <- gwr.basic(gwrListBest[[j]], data = agric, bw = i, adaptive = T, kernel =
   kernels[[j]])
   print(gwrModel[[j]])
```

77

```
    j <- j + 1
 }


####################### BASIC GWR DIAGNOSTICS #######################


# test spatial variability of predictors (H0: no variability)
 gwrMCNames <- lapply(gwrModel, function(x) names(x$lm$model[-1]))
 gwrMCVars <- lapply(gwrMCNames, function(x) append(x, "(Intercept)", 0))
 gwrMC <- list(); gwrMCList <- list()
 set.seed(25042011)
 j <- 1
 for (i in bandw) {
   gwrMC[[j]] <- montecarlo.gwr(gwrListBest[[j]], data = agric, nsims = 999, bw = i, kernel =
   kernels[j], adaptive = T)
   gwrMCList[[j]]  <- cbind(gwrMC[[j]], gwrMCVars[[j]])
   j <- j + 1
 }


## --- local collinearity diagnostics ---
 gwrColDiag <- list(); gwrLocCor <- list(); gwrVif <- list(); gwrLocCon <- list()
 j <- 1
 for (i in bandw) {
   tryCatch( {
     gwrColDiag[[j]] <- gwr.collin.diagno(gwrListBest[[j]], data = agric, bw = i, adaptive = T, kernel
   = kernels[j])
     gwrLocCor[[j]] <- round(gwrColDiag[[j]]$corr.mat, 2)
     gwrVif[[j]] <- round(gwrColDiag[[j]]$VIF, 1)
     gwrLocCon[[j]] <- round(gwrColDiag[[j]]$local_CN, 1)
   }
   , error = function(e) cat("ERROR: ", conditionMessage(e), "\n"))
   j <- j + 1
 }
 j <- 1
 for (i in gwrColDiag) {
   tryCatch( {
     if (layerAdmin == "arr") {
       gwrLocCorCol <- gsub("Corr_", "", names(i$SDF)[startsWith(names(i$SDF), "Corr_")])
       colnames(gwrLocCor[[j]]) <- gwrLocCorCol
       rownames(gwrLocCor[[j]]) <- paste0(agric$ARRNIS, " ", agric$ARRNAM)
       gwrVifCol <- gsub("_VIF", "", names(i$SDF)[endsWith(names(i$SDF), "_VIF")])
       colnames(gwrVif[[j]]) <- gwrVifCol
       rownames(gwrVif[[j]]) <- paste0(agric$ARRNIS, " ", agric$ARRNAM)
       gwrLocCon[[j]] <- cbind(paste0(agric$ARRNIS, " ", agric$ARRNAM), gwrLocCon[[j]])
       write.table(gwrLocCor[[j]], paste0(tag, " GWR local correlation values ", kernels[[j]], ".csv"),
   sep = ";")
       write.table(gwrVif[[j]], paste0(tag, " GWR local VIF ", kernels[[j]], ".csv"), sep = ";")
       write.table(gwrLocCon[[j]], paste0(tag, " GWR local condition numbers ", kernels[[j]], ".csv"),
   sep = ";")
```

78

```
      }
      else {
        gwrLocCorCol <- gsub("Corr_", "", names(i$SDF)[startsWith(names(i$SDF), "Corr_")])
        colnames(gwrLocCor[[j]]) <- gwrLocCorCol
        rownames(gwrLocCor[[j]]) <- paste0(agric$NIS, " ", agric$MUN)
        gwrVifCol <- gsub("_VIF", "", names(i$SDF)[endsWith(names(i$SDF), "_VIF")])
        colnames(gwrVif[[j]]) <- gwrVifCol
        rownames(gwrVif[[j]]) <- paste0(agric$NIS, " ", agric$MUN)
        gwrLocCon[[j]] <- cbind(paste0(agric$NIS, " ", agric$MUN), gwrLocCon[[j]])
        write.table(gwrLocCor[[j]], paste0(tag, " GWR local correlation values ", kernels[[j]], ".csv"),
    sep = ";")
        write.table(gwrVif[[j]], paste0(tag, " GWR local VIF ", kernels[[j]], ".csv"), sep = ";")
        write.table(gwrLocCon[[j]], paste0(tag, " GWR local condition numbers ", kernels[[j]], ".csv"),
    sep = ";")
      }
    }
    , error = function(e) cat("ERROR:", conditionMessage(e), "\n"))
    j <- j + 1
  }

# check studentized residuals for outliers
 gwrOutl <- list()
 j <- 1
 for (i in gwrModel) {
   gwrOutl[[j]] <- count(i$SDF$Stud_residual > 3 | i$SDF$Stud_residual < -3)
   j <- j + 1
 }
 gwrOutl

## --- test residuals for autocorrelation ---
# Moran's I (H0: no autocorrelation)
 (moranBasGwr <- lapply(gwrModel, function(x) moran.test(x$SDF$residual, spatWgts, zero.policy = T)))
# Moran scatter plot
 j <- 1
 morplotlist <- list()
 for (i in gwrModel) {
   img <- file.path(paste0(tag, ". GWR - 1. Moran plot, ", kernels[j], ".tif"))
   tiff(img, width = 1280, height = 850)
   par(mar = c(5, 5, 4, 2))
   moran.plot2(i$SDF$residual, spatWgts, zero.policy = T, labels = paste0(agricRows, ". ", agric[[2]]),
   main = paste0(tag, ", ", kernels[j]), cex.main = 3.5, font.main = 1, xlab = "residuals", ylab =
   "spatially lagged residuals", cex = 0.8, cex.lab = 2, cex.axis = 1.4)
   mtext(paste0("Moran's I: ", round(moranBasGwr[[j]]$estimate[1], 3), "\np-value: ",
   round(moranBasGwr[[j]]$p.value[[1]], 3)), side = 4, cex = 2, las = 1, adj = 1, padj = -6.4, line =
   -1.3)
   dev.off()
   j <- j + 1
 }
```

```
## --- map residuals ---
 j <- 1
 for (i in gwrModel) {
   img <- file.path(paste0(tag, ". GWR - 2. residuals, ", kernels[j], ".tif"))
   if (layerAdmin == "f_mun") {
     tiff(img, width = 1000, height = 550)
     par(mar = c(0, 0, 3, 0))
     sdMap(i$SDF$residual, agricS)
     plot(admin, border = "light grey", add = T)
     north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
     Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
     title(tag, line = 0, cex.main = 3, font.main = 1)
     mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
     par(lheight = 0.8)
     text(15000, 252000, "GWR residuals", pos = 4, cex = 2)
     text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
     legend("topleft", inset = c(0, 0.15), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8, border
   = "light g  rey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
   } else {
     tiff(img, width = 1000, height = 850)
     par(mar = c(0, 0, 3, 0))
     sdMap(i$SDF$residual, agricS)
     plot(admin, border = "light grey", add = T)
     north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
     Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
     title(tag, line = 0, cex.main = 3, font.main = 1)
     mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
     par(lheight = 0.8)
     text(25000, 97000, "GWR residuals", pos = 4, cex = 2)
     text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
     legend("bottomleft", inset = c(0.04, 0.14), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8,
   border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
   }
   dev.off()
   j <- j + 1
 }


########################## BASIC GWR RESULTS ##########################

 gwrVars <- list()
 gwrT <- list()
 gwrKernel <- list()
 j <- 1
 for (i in gwrModel) {
   gwrVars[[j]] <- names(i$lm$model[2:length(i$lm$model)])
   gwrVars[[j]] <- append(gwrVars[[j]], "intercept", after = 0)
   gwrT[[j]] <- paste0(gwrVars[[j]], "_TV")
```

```
    gwrKernel[[j]] <- rep(kernels[j], length(gwrVars[[j]]))
    j <- j + 1
 }
# show ranges of estimates
 gwrEst <- list()
 j <- 1
 for (i in gwrModel) {
   gwrEst[[j]] <- cbind(i$SDF@data[names(i$SDF@data) %in% gwrVars[[j]] | names(i$SDF@data) %in%
   "Intercept"])
   j <- j + 1
 }
 gwrEstRngAux <- list()
 gwrEstRng <- list()
 j <- 1
 for (i in gwrEst) {
   l <- 1
   for (k in i) {
     gwrEstRngAux[[l]] <- range(k)
     gwrEstRngAux[[l]] <- append(gwrEstRngAux[[l]], c(gwrVars[[j]][l], gwrKernel[[j]][l]))
     l <- l + 1
   }
   gwrEstRng[[j]] <- gwrEstRngAux
   gwrEstRngAux <- list()
   j <- j + 1
 }
 gwrEstRng
# show ranges of t-values
 gwrT <- list()
 j <- 1
 for (i in gwrModel) {
   gwrT[[j]] <- cbind(i$SDF@data[names(i$SDF@data) %in% paste0(gwrVars[[j]], "_TV") | names(i$SDF@data)
   %in% paste0("Intercept", "_TV")])
   j <- j + 1
 }
 gwrTRngAux <- list()
 gwrTRng <- list()
 j <- 1
 for (i in gwrT) {
   l <- 1
   for (k in i) {
     gwrTRngAux[[l]] <- range(k)
     gwrTRngAux[[l]] <- append(gwrTRngAux[[l]], c(gwrVars[[j]][l], gwrKernel[[j]][l]))
     l <- l + 1
   }
   gwrTRng[[j]] <- gwrTRngAux
   gwrTRngAux <- list()
   j <- j + 1
 }
```

```
 gwrTRng
# create labels
 varGwrLabAux <- list()
 varGwrLab <- list()
 j <- 1
 for (i in gwrEst) {
   k <- 1
   for (l in (names(i))) {
     pattern <- l
     pattReplace <- agrep(pattern, varLabels, max.distance = 0.2, ignore.case = T)
     for (m in varLabels[pattReplace]) {
       if (isBetween((nchar(m) / nchar(pattern)), 0.9, 1.1)) {
         varGwrLabAux[[k]] <- m
         k <- k + 1
       }
     }
   }
   varGwrLab[[j]] <- varGwrLabAux
   varGwrLabAux <- list()
   j <- j + 1
 }
# map estimates
 j <- 1
 for (i in gwrEst) {
   l <- 1
   for (k in i) {
     img <- file.path(paste0(tag, ". GWR - 3. estimates, ", kernels[j], ", ", varGwrLab[[j]][l],
   ".tif"))
     if (layerAdmin == "f_mun") {
       tiff(img, width = 1000, height = 550)
       par(mar = c(0, 0, 3, 0))
       sdMap(k, agricS)
       plot(admin, border = "light grey", add = T)
       north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
       Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
       title(tag, line = 0, cex.main = 3, font.main = 1)
       mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
       par(lheight = 0.8)
       text(15000, 254000, paste0("GWR estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
       text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
       legend("topleft", inset = c(0, 0.15), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8,
     border = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
     } else {
       tiff(img, width = 1000, height = 850)
       par(mar = c(0, 0, 3, 0))
       sdMap(k, agricS)
       plot(admin, border = "light grey", add = T)
       north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
```

82

```
        Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
        title(tag, line = 0, cex.main = 3, font.main = 1)
        mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
        par(lheight = 0.8)
        text(25000, 100000, paste0("GWR estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
        text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
        legend("bottomleft", inset = c(0.04, 0.14), fill = palDiv, legend = sdMapLegLab, y.intersp =
    0.8, border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
      }
      dev.off()
      l <- l + 1
    }
    j <- j + 1
  }
# calculate p-values from t-values
 gwrPPre <- list()
 gwrPAux <- list()
 gwrP <- list()
 j <- 1
 for (i in gwrT) {
    m <- 1
    for (k in i) {
      l <- 1
      for (t in k) {
        gwrPPre[[l]] <- 2 * pt(abs(t), unlist(bandw[j]), lower = F)
        l <- l + 1
      }
      gwrPAux[[m]] <- unlist(gwrPPre)
       gwrPPre <- list()
      m <- m + 1
    }
    gwrP[[j]] <- gwrPAux
    gwrPAux <- list()
    j <- j + 1
  }
# map p-values
 j <- 1
 for (i in gwrP) {
    l <- 1
    for (k in i) {
      img <- file.path(paste0(tag, ". GWR - 4. p-values, ", kernels[j], ", ", varGwrLab[[j]][l],
    ".tif"))
      if (layerAdmin == "f_mun") {
        tiff(img, width = 1000, height = 550)
        par(mar = c(0, 0, 3, 0))
        pMap(k, agricS)
        plot(admin, border = "light grey", add = T)
        north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
```

```
      Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
      par(lheight = 0.8)
      text(15000, 249000, paste0("GWR p-values\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
      legend("topleft", inset = c(0, 0.15), fill = palP, legend = pMapLegLab, y.intersp = 0.8, border
  = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    } else {
      tiff(img, width = 1000, height = 850)
      par(mar = c(0, 0, 3, 0))
      pMap(k, agricS)
      plot(admin, border = "light grey", add = T)
      north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
      Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
      par(lheight = 0.8)
      text(25000, 84000, paste0("GWR p-values\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
      legend("bottomleft", inset = c(0.04, 0.14), fill = palP, legend = pMapLegLab, y.intersp = 0.8,
  border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    }
    dev.off()
    l <- l + 1
  }
  j <- j + 1
}
# calculate interval limits
gwrEstInt <- list()
gwrEstIntAux <- list()
k <- 1
for (i in gwrEst) {
  l <- 1
  for (j in i) {
    gwrEstIntAux[[l]] <- sdInt(j)
    l <- l + 1
  }
  gwrEstInt[[k]] <- gwrEstIntAux
  gwrEstIntAux <- list()
  k <- k + 1
}
# assign estimates to intervals
gwrCat <- list()
gwrCatPre <- list()
gwrCatAux <- list()
l <- 1
for (i in gwrEst) {
  m <- 1
  for (j in i) {
```

```
      n <- 1
      for (k in j) {
        gwrCatPre[[n]] <- findInterval(k, gwrEstInt[[l]][[m]])
        n <- n + 1
      }
      gwrCatAux[[m]] <- unlist(gwrCatPre)
      gwrCatPre <- list()
      m <- m + 1
    }
    gwrCat[[l]] <- gwrCatAux
    gwrCatAux <- list()
    l <- l + 1
 }
# set non-significant estimates NA
 gwrCatSig <- gwrCat
 gwrEstSig <- gwrEst
 l <- 1
 for (i in gwrP) {
   m <- 1
   for (j in i) {
     n <- 1
     for (k in j) {
       if (k > 0.05) {
         gwrCatSig[[l]][[m]][n] <- NA
         gwrEstSig[[l]][[m]][n] <- NA
       }
       n <- n + 1
     }
     m <- m + 1
   }
   l <- l + 1
 }
# map significant estimates
 j <- 1
 for (i in gwrCatSig) {
   l <- 1
   for (k in i) {
     img <- file.path(paste0(tag, ". GWR - 5. significant estimates, ", kernels[j], ", ",
   varGwrLab[[j]][l], ".tif"))
     if (layerAdmin == "f_mun") {
       tiff(img, width = 1000, height = 550)
       par(mar = c(0, 0, 3, 0))
       sdMapSig(k, agricS)
       plot(admin, border = "light grey", add = T)
       north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
       Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
       title(tag, line = 0, cex.main = 3, font.main = 1)
       mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
```

```r
        par(lheight = 0.8)
        text(15000, 254000, paste0("GWR significant estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
        text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
        legend("topleft", inset = c(0, 0.15), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8,
    border = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
      } else {
        tiff(img, width = 1000, height = 850)
        par(mar = c(0, 0, 3, 0))
        sdMapSig(k, agricS)
        plot(admin, border = "light grey", add = T)
        north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
        Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
        title(tag, line = 0, cex.main = 3, font.main = 1)
        mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
        par(lheight = 0.8)
        text(25000, 100000, paste0("GWR significant estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
        text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
        legend("bottomleft", inset = c(0.04, 0.14), fill = palDiv, legend = sdMapLegLab, y.intersp =
    0.8, border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
      }
      dev.off()
      l <- l + 1
    }
    j <- j + 1
  }
# filter out estimates with standard deviation < +/-2
  gwrCatSigFil <- gwrCatSig
  l <- 1
  for (i in gwrCatSigFil) {
    m <- 1
    for (j in i) {
      n <- 1
      for (k in j) {
        if (isBetween(k, 2, 6) || is.na(k) == T) {
          gwrCatSigFil[[l]][[m]][n] <- NA
        }
        n <- n + 1
      }
      m <- m + 1
    }
    l <- l + 1
  }
# map filtered significant estimates
  j <- 1
  for (i in gwrCatSigFil) {
    l <- 1
    for (k in i) {
```

```
    img <- file.path(paste0(tag, ". GWR - 6. filtered significant estimates, ", kernels[j], ", ",
varGwrLab[[j]][l], ".tif"))
    if (layerAdmin == "f_mun") {
      tiff(img, width = 1000, height = 550)
      par(mar = c(0, 0, 3, 0))
      sdMapSig(k, agricS)
      plot(admin, border = "light grey", add = T)
      north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
      Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
      par(lheight = 0.8)
      text(15000, 254000, paste0("GWR relevant estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
      text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
      legend("topleft", inset = c(0, 0.15), fill = palDivSig, legend = sdMapSigLegLab, y.intersp =
0.8, border = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    } else {
      tiff(img, width = 1000, height = 850)
      par(mar = c(0, 0, 3, 0))
      sdMapSig(k, agricS)
      plot(admin, border = "light grey", add = T)
      north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
      Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
      par(lheight = 0.8)
      text(25000, 100000, paste0("GWR relevant estimates\n", varGwrLab[[j]][l]), pos = 4, cex = 2)
      text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
      legend("bottomleft", inset = c(0.04, 0.14), fill = palDivSig, legend = sdMapSigLegLab, y.intersp
= 0.8, border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    }
    dev.off()
    l <- l + 1
  }
  j <- j + 1
}


##########==================== MIXED GWR ====================##########
####################################################################

# determine global variables
(varsFixed <- lapply(gwrMCList, function(x) x[-1, 2][x[-1, 1] >= 0.075]))

## --- perform mixed GWR ---
mgwrModel <- list()
j <- 1
for (i in bandw) {
  tryCatch( {
```

```
      mgwrModel[[j]] <- gwr.mixed(gwrListBest[[j]], data = agric, bw = i, adaptive = T, kernel =
   kernels[[j]], fixed.vars = varsFixed[[j]])
      print(mgwrModel[[j]])
   }
   , error = function(e) cat("ERROR: ", conditionMessage(e), "\n"))
   j <- j + 1
}


######################### MIXED GWR RESULTS #########################

# show ranges of estimates
 mgwrEst <- list()
 j <- 1
 for (i in mgwrModel) {
   mgwrEst[[j]] <- cbind(i$SDF@data[!names(i$SDF@data) %in% paste0(varsFixed[[j]], "_F")])
   j <- j + 1
 }
 mgwrEstRngAux <- list()
 mgwrEstRng <- list()
 j <- 1
 for (i in mgwrEst) {
   l <- 1
   for (k in i) {
     mgwrEstRngAux[[l]] <- range(k)
     mgwrEstRngAux[[l]] <- append(mgwrEstRngAux[[l]], c(gwrVars[[j]][l], gwrKernel[[j]][l]))
     l <- l + 1
   }
   mgwrEstRng[[j]] <- mgwrEstRngAux
   mgwrEstRngAux <- list()
   j <- j + 1
 }
 mgwrEstRng
# create labels
 mgwrEstMod <- list()
 j <- 1
 for (i in mgwrEst) {
   mgwrEstMod[[j]] <- substr(names(i), 1, nchar(names(i)) - 2)
   j <- j + 1
 }
 varMgwrLabAux <- list()
 varMgwrLab <- list()
 j <- 1
 for (i in mgwrEstMod) {
   k <- 1
   for (l in i) {
     pattern <- l
     pattReplace <- agrep(pattern, varLabels, max.distance = 0.2, ignore.case = T)
     for (m in varLabels[pattReplace]) {
```

```
      if (isBetween((nchar(m) / nchar(pattern)), 0.9, 1.1)) {
        varMgwrLabAux[[k]] <- m
        k <- k + 1
      }
    }
  }
  varMgwrLab[[j]] <- varMgwrLabAux
  varMgwrLabAux <- list()
  j <- j + 1
}
# map estimates
j <- 1
for (i in mgwrEst) {
  l <- 1
  for (k in i) {
    img <- file.path(paste0(tag, ". MGWR - estimates, ", kernels[j], ", ", varMgwrLab[[j]][l],
  ".tif"))
    if (layerAdmin == "f_mun") {
      tiff(img, width = 1000, height = 550)
      par(mar = c(0, 0, 3, 0))
      sdMap(k, agricS)
      plot(admin, border = "light grey", add = T)
      north.arrow(250000, 250000, 2500, lab = "N", cex.lab = 2)
      Scalebar2(18000, 140000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 17000, cex = 2)
      par(lheight = 0.8)
      text(15000, 254000, paste0("MGWR estimates\n", varMgwrLab[[j]][l]), pos = 4, cex = 2)
      text(15000, 246000, "(SD from mean)", pos = 4, cex = 1.5)
      legend("topleft", inset = c(0, 0.15), fill = palDiv, legend = sdMapLegLab, y.intersp = 0.8,
    border = "light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    } else {
      tiff(img, width = 1000, height = 850)
      par(mar = c(0, 0, 3, 0))
      sdMap(k, agricS)
      plot(admin, border = "light grey", add = T)
      north.arrow(290000, 235000, 2500, lab = "N", cex.lab = 2)
      Scalebar2(28000, 30000, distance = 100000, scale = 0.001, t.cex = 1.3)
      title(tag, line = 0, cex.main = 3, font.main = 1)
      mtext(paste0("kernel: ", kernels[j]), side = 3, line = 0, adj = 0, at = 20000, cex = 2)
      par(lheight = 0.8)
      text(25000, 100000, paste0("MGWR estimates\n", varMgwrLab[[j]][l]), pos = 4, cex = 2)
      text(25000, 91000, "(SD from mean)", pos = 4, cex = 1.5)
      legend("bottomleft", inset = c(0.04, 0.14), fill = palDiv, legend = sdMapLegLab, y.intersp =
    0.8, border = "  light grey", bty = "n", cex = 1.5, box.cex = c(1.5, 0.8))
    }
    dev.off()
    l <- l + 1
```

```
    }
    j <- j + 1
 }
```

A.2 Mixed GWR

```
gwr.mixed <- function(formula, data, regression.points, fixed.vars, intercept.fixed = F, bw, diagnostic
    = T, kernel = "bisquare", adaptive = F, p = 2, theta = 0, longlat = F, dMat) {
    timings <- list()
    timings[["start"]] <- Sys.time()
    this.call <- match.call()
    p4s <- as.character(NA)
    if (missing(regression.points)) {
      rp.given <- F
      regression.points <- data
      rp.locat <- coordinates(data)
      hatmatrix <- T
    }
    else {
      rp.given <- T
      hatmatrix <- F
      if (is(regression.points, "Spatial")) {
        rp.locat <- coordinates(regression.points)
      }
      else if (is.numeric(regression.points) && dim(regression.points)[2] == 2)
        rp.locat <- regression.points
      else {
        warning("Output locations are not packed in a Spatial object, and it has to be a two-column
    numeric vector")
        rp.locat <- dp.locat
      }
    }
    if (is(data, "Spatial")) {
      p4s <- proj4string(data)
      dp.locat <- coordinates(data)
      data <- as(data, "data.frame")
    }
    else {
      stop("Given regression data must be Spatial*DataFrame")
    }
    dp.n <- nrow(dp.locat)
    rp.n <- nrow(rp.locat)
    if (missing(dMat)) {
      DM.given <- F
      DM1.given <- F
      if (dp.n + rp.n <= 10000) {
        dMat <- gw.dist(dp.locat = dp.locat, rp.locat = rp.locat, p = p, theta = theta, longlat =
    longlat)
```

```
        DM.given <- T
    }
}
else {
    DM.given <- T
    DM1.given <- T
    dim.dMat <- dim(dMat)
    if (dim.dMat[1] != dp.n || dim.dMat[2] != rp.n)
        stop("Dimensions of dMat are not correct")
}
mf <- match.call(expand.dots = F)
    m <- match(c("formula", "data"), names(mf), 0L)
    mf <- mf[c(1L, m)]
    mf$drop.unused.levels <- T
    mf[[1L]] <- as.name("model.frame")
    mf <- eval(mf, parent.frame())
    mt <- attr(mf, "terms")
    y <- model.extract(mf, "response")
    x <- model.matrix(mt, mf)
    idx1 <- match("(Intercept)", colnames(x))
    if (!is.na(idx1))
        colnames(x)[idx1] <- "Intercept"
if (missing(fixed.vars)) {
    warning("No independent variables in the formula is specified as fixed terms!")
    if (!intercept.fixed)
        stop("Please use basic GWR function to calibrate this model")
}
else {
    if (intercept.fixed)
        fixed.vars <- c("Intercept", fixed.vars)
}
idx.fixed <- match(fixed.vars, colnames(x))
x1 <- x[, -idx.fixed]
x2 <- x[, idx.fixed]
if (!is.null(x1)) x1 <- as.matrix(x1, nrow = dp.n)
if (!is.null(x2)) x2 <- as.matrix(x2, nrow = dp.n)
colnames(x1) <- colnames(x)[-idx.fixed]
colnames(x2) <- colnames(x)[idx.fixed]
y <- as.matrix(y, nrow = dp.n)
model <- gwr.mixed.2(x1, x2, y, dp.locat, out.loc = rp.locat, adaptive = adaptive, bw = bw, kernel =
kernel, p = p, theta = theta, longlat = longlat, dMat)
res <- list()
 res$local <- model$local
 res$global <- apply(model$global, 2, mean, na.rm = T)
 mgwr.df <- data.frame(model$local, model$global)
 colnames(mgwr.df) <- c(paste(colnames(x1), "L", sep = "_"), paste(colnames(x2), "F", sep = "_"))
 rownames(rp.locat) <- rownames(mgwr.df)
griddedObj <- F
```

```r
    if (is(regression.points, "Spatial")) {
      if (is(regression.points, "SpatialPolygonsDataFrame")) {
        polygons <- polygons(regression.points)
        SDF <- SpatialPolygonsDataFrame(Sr = polygons, data = mgwr.df, match.ID = F)
      }
      else {
        griddedObj <- gridded(regression.points)
        SDF <- SpatialPointsDataFrame(coords = rp.locat, data = mgwr.df, proj4string = CRS(p4s),
  match.ID = F)
        gridded(SDF) <- griddedObj
      }
    }
    else
      SDF <- SpatialPointsDataFrame(coords = rp.locat, data = mgwr.df, proj4string = CRS(p4s),
  match.ID = F)

  res$SDF <- SDF
  if (diagnostic) {
    gwr.fitted <- function(x, b) apply(x * b, 1, sum)
    edf <- gwr.mixed.trace(x1, x2, y, dp.locat, adaptive = adaptive, bw = bw, kernel = kernel, p = p,
theta = theta, longlat = longlat,dMat = dMat)
    model2 <- gwr.mixed.2(x1, x2, y, dp.locat, adaptive = adaptive, bw = bw, kernel = kernel, p = p,
theta = theta, longlat = longlat, dMat = dMat)
    rss <- sum((y - gwr.fitted(model2$global, x2) - gwr.fitted(model2$local, x1))^2)
    n1 <- length(y)
    sigma.aic <- rss / n1
    aic <- log(sigma.aic * 2 * pi) + 1 + 2 * (edf + 1) / (n1 - edf - 2)
    aic <- n1 * aic
    res$aic <- aic
    res$df.used <- edf
    res$rss <- rss
  }
  GW.arguments <- list(formula = formula, rp.given = rp.given, hatmatrix = hatmatrix, bw = bw, kernel
  = kernel, adaptive = adaptive, p = p, theta = theta, longlat = longlat, DM.given = DM1.given,
  diagnostic = diagnostic)
  res$GW.arguments <- GW.arguments
  res$this.call <- this.call
  timings[["stop"]] <- Sys.time()
  res$timings <- timings
  class(res) <- "mgwr"
  res
}
```

A.3 Moran Scatter Plot

```r
moran.plot2 <- function(x, listw, zero.policy = NULL, spChk = NULL, labels = NULL, xlab = NULL, ylab =
  NULL, quiet = NULL, ...) {
  if (!inherits(listw, "listw"))
```

```
      stop(paste(deparse(substitute(listw)), "is not a listw object"))
if (is.null(quiet))
    quiet <- !get("verbose", envir = spdep:::.spdepOptions)
stopifnot(is.vector(x))
stopifnot(is.logical(quiet))
if (is.null(zero.policy))
    zero.policy <- get("zeroPolicy", envir = spdep:::.spdepOptions)
stopifnot(is.logical(zero.policy))
xname <- deparse(substitute(x))
if (!is.numeric(x))
    stop(paste(xname, "is not a numeric vector"))
if (any(is.na(x)))
    stop("NA in X")
n <- length(listw$neighbours)
if (n != length(x))
    stop("objects of different length")
if (is.null(spChk))
    spChk <- get.spChkOption()
if (spChk && !chkIDs(x, listw))
    stop("Check of data and weights ID integrity failed")
labs <- T
if (is.logical(labels) && !labels)
    labs <- F
if (is.null(labels) || length(labels) != n)
    labels <- as.character(attr(listw, "region.id"))
wx <- lag.listw(listw, x, zero.policy = zero.policy)
if (is.null(xlab))
    xlab <- xname
if (is.null(ylab))
    ylab <- paste("spatially lagged", xname)
plot(x, wx, xlab = xlab, ylab = ylab, ...)
if (zero.policy) {
  n0 <- wx == 0
  if (any(n0))
    symbols(x[n0], wx[n0], inches = F, circles = rep(diff(range(x)) / 50, length(which(n0))), bg =
"grey", add = T)
}
xwx.lm <- lm(wx ~ x)
abline(xwx.lm)
abline(h = mean(wx), lty = 2)
abline(v = mean(x), lty = 2)
infl.xwx <- influence.measures(xwx.lm)
is.inf <- which(apply(infl.xwx$is.inf, 1, any))
points(x[is.inf], wx[is.inf], pch = 9, cex = 1.2)
if (labs)
    pointLabel(x[is.inf], wx[is.inf], labels = labels[is.inf], cex = 1.5)
rownames(infl.xwx$infmat) <- labels
if (!quiet)
```

```
      summary(infl.xwx)
    invisible(infl.xwx)
}
```

A.4 Model Selection Plot

```
plot.glmulti2 <- function(x, type = "p", ...) {
  if (type == "p") {
    plot(x@crits, xlab = "model no.", ylab = "", pch = 19, cex = 2, ...)
    abline(h = x@crits[1] + 2, col = "red")
  }  else if (type == "w") {
    ww <- exp(-(x@crits - x@crits[1]) / 2)
    ww <- ww / sum(ww)
    plot(ww, xlab = "model no.", ylab = paste("evidence weight (", x@params$crit, ")"), pch = 19, ...)
    cucu <- function(i) sum(ww[1:i])
    wwc <- lapply(1:length(ww), cucu)
    abline(v = min(which(wwc >= 0.95)), col = "red")
  }
  else if (type == "r") {
    if (length(x@objects)) {
      windows(21, 7)
      par(mfrow = c(2, min(length(x@crits), 5)))
      for (k in 1:min(length(x@crits), 5)) plot(x@objects[[k]], which = c(1), main =
  deparse(x@formulas[[k]]), ...)
      for (k in 1:min(length(x@crits), 5)) plot(x@objects[[k]], which = c(2), ...)
    }
    else warning("unavailable: use includeobjects = T when calling glmulti.")
  }
  else if (type == "s") {
    ww <- exp(-(x@crits - x@crits[1]) / 2)
    ww <- ww / sum(ww)
    clartou = function(x) {
      pieces <- sort(strsplit(x, ":")[[1]])
      if (length(pieces) > 1)
        paste(pieces[1], ":", pieces[2], sep = "")
      else x
    }
    tet <- lapply(x@formulas, function(x) sapply(attr(delete.response(terms(x)), "term.labels"),
  clartou))
    allt <- unique(unlist(tet))
    imp <- sapply(allt, function(x) sum(ww[sapply(tet, function(t) x %in% t)]))
    par(oma = c(0, 3, 0, 0))
    barplot(sort(imp), xlab = "", xlim = c(0, 1), ylab = "", horiz = T, las = 2, names.arg =
  allt[order(imp)], main = "model-averaged importance of terms", ...)
    abline(v = 0.8, col = "red")
  }
  else warning("plot: invalid type argument for plotting glmulti objects.")
}
```

A.5 Scale Bar

```
Scalebar2 <- function(x, y, distance, unit = "km", scale = 1, t.cex = 0.8)
 {
   xvals = distance * c(0, 0.25, 0.5, 0.75, 1) + x
   yvals = c(0, distance / c(50, 40, 15)) + y
   cols <- c("black", "white", "black", "white")
   for (i in 1:4) rect(xvals[i], yvals[1], xvals[i + 1], yvals[2], col = cols[i])
   for (i in 1:5) segments(xvals[i], yvals[2], xvals[i], yvals[3])
   labels <- c((xvals[c(1, 3)] - xvals[1]) * scale, paste((xvals[5] - xvals[1]) * scale, unit))
   text(xvals[c(1, 3, 5)], yvals[4], labels = labels, adj = 0.5, cex = t.cex)
 }
```