FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Change in Net Primary Production and land cover due to the eastward enlargement of the European Union

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Business Informatics

eingereicht von

Dipl.-Ing. Peter Aufner
Matrikelnummer 0825073

an der Fakultät für Informatik
der Technischen Universität Wien
Betreuung: Ass.Prof. Dipl.-Math. Dr. Timo Trimborn

Wien, 1. Jänner 2001

Peter Aufner	Timo Trimborn

Change in Net Primary Production and land cover due to the eastward enlargement of the European Union

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Dipl.-Ing. Peter Aufner
Registration Number 0825073

to the Faculty of Informatics

at the TU Wien

Advisor: Ass.Prof. Dipl.-Math. Dr. Timo Trimborn

Vienna, 1st January, 2001

Peter Aufner | Timo Trimborn

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Peter Aufner
Attemsgasse 5/2/101, 1220 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 1. Jänner 2001

Peter Aufner

Danksagung

Als Erstes möchte ich mich bei meinem Betreuer, Dr. Timo Trimborn, für seine hervorragende Unterstützung während der Forschung für und dem Schreiben an meiner Diplomarbeit bedanken.

Ich möchte mich ebenso bei Dr. Hubert Hasenauer und Dr. Mathias Neumann vom Institut für Waldbau der BOKU Wien bedanken, da sie mir die wichtigsten Konzepte, die als Basis für diese Arbeit dienen, vorgestellt haben und mich mit wertvollem Feedback während der Erstellung der Arbeit versorgt haben.

Mein Dank geht auch an meine Eltern, die mich finanziell, moralisch und auf jede andere Art, die sie bislang konnten, in meinem Leben unterstützt haben.

Abschließend möchte ich auch meiner Partnerin Patricia meine Dankbarkeit für ihre liebevolle Unterstützung während meines Masterstudiums und meiner Arbeit an dieser Diplomarbeit aussprechen.

Acknowledgements

First, I would like to thank my advisor, Dr. Timo Trimborn, for his expectational support throughout the research for and writing of this theses.

I would like to thank Dr. Hubert Hasenauer and Dr. Mathias Neumann from the department of silviculture of the University of Earth Sciences (BOKU) Vienna as well for introducing me to the important concepts, that served as a base for this work and for providing me with valuable feedback throughout the writing of this work.

My gratitude also goes to my parents, who supported me financially, morally and any other way they could throughout my life thus far.

Finally, I want to express my gratitude to my spouse Patricia for her loving support during my master studies and my work on this thesis.

Kurzfassung

Diese Arbeit befasst sich mit der Auswirkung der EU Osterweiterung auf die Natur. Genauer gesagt, werden die folgenden drei Fragen beantwortet:

1. Besteht eine direkte Auswirkung des EU Beitritts auf die Werte der Netto Primär Produktion?
2. Welche Änderungen an der Landüberdeckung führen zu den beobachteten Veränderungen der Netto Primär Produktion?
3. Kann ein exogener Einfluss des BIP auf die Umwelt Kuznets Kurve für Europa beobachtet werden?

Um diese Fragen zu beantworten wurden Satellitendaten der Jahre 2000 bis 2012 genutzt. Die Daten zur Landüberdeckung, bereitgestellt von dem MODIS Satelliten, wurden als Maß für die Auswirkung vor und nach der Erweiterung genutzt. Die Daten zur Netto Primär Produktion (NPP), bereitgestellt von der Universität für Bodenkultur (BOKU) Wien, wurden als Stellvertreter für den Gesamtzustand der Natur genutzt. NPP misst wieviel Kohlenstoff in Pflanzen über eine gegebene Zeit gespeichert wird und wird als angemessener Indikator für den Gesundheitszustand der Natur betrachtet.

Um die Entwicklung über die Zeit zu messen wurden zwei Dimensionen genutzt. Zum einen, eine Difference-in-Differences Regression, die die direkte Auswirkung der EU Osterweiterung auf die Waldfläche, Landwirtschaft, Verstädterung und NPP misst. Zum anderen, wurde die Umwelt Kuznets Kurve für die Regionen abgeleitet um die Entwicklung im Verhältnis zueinander darzustellen. Für diese Berechnungen wurden die Grenzregionen zwischen den alten EU Ländern (Deutschland, Österreich und Italien) und den neuen EU Ländern (Polen, Tschechische Republik, Ungarn, Slowenien und Slowakei) in benachbarte NUTS 3 Regionen unterteilt. Dann wurde Paare von NUTS 3 Regionen gebildet um die Entwicklung in den Unterschieden westlich und östlich der Grenze zu beobachten.

So wurde gezeigt, dass die EU Osterweiterung den Unterschied im NPP zwischen den alten EU Ländern und den neuen EU Ländern verringerte. Es gab einen Anstieg im Unterschied der Bewaldung zwischen den alten EU Ländern und den neuen Eu Ländern. Es konnte keine signifikante Auswirkung auf die landwirtschaftlich genutzte Fläche oder

die Urbanisierung beobachtet werden. Trotz des höheren Unterschied in der Bewaldung zwischen alten und neuen EU Ländern, ist die Bewaldung in der Europäischen Union insgesamt gestiegen. So könnte dennoch eine positive Auswirkung auf die NPP erklärt werden.

Bezüglich der Umwelt Kuznets Kurve: Sowohl die alten EU Länder als auch die neuen EU Länder sind über den tiefsten Punkt hinweg. Das bedeutet, dass der Zustand der Natur in beiden Teilen der EU besser wird. Klar ersichtlich ist die Natur in den alten EU Ländern in einem besseren Zustand wegen der höheren wirtschaftlichen Entwicklung und stärkeren Umweltregulierungen. Jedoch konnten die neuen EU Ländern im 13 jährigen Beobachtungszeitraum von 2000 bis 2012 aufholen.

Abstract

This work examines the impact of the eastward enlargement of the European Union on nature. More precisely, the following three questions will be answered:

1. Is there a direct impact of joining the European Union on Net Primary Production values?
2. What kind of changes in land cover causes the observed changes in Net Primary Production values?
3. Can the exogenous impact of GDP on Net Primary Production be observed in the environmental Kuznets curve for Europe?

To answer these questions satellite data from the years 2000 to 2012 were used. Land cover data generated by the MODIS satellite was taken to measure the impact on the changes before and after the enlargement. Net Primary Production (NPP) data derived by the University of Earth Sciences (BOKU) Vienna was taken as a proxy for the overall well being of nature. NPP measures how much carbon is stored in plants in a given time and is regarded as a well suited overall proxy for environmental well being.

To measure the development over time two dimensions were used. First, a difference-in-differences regression aimed to measure the direct impact of the EU eastward enlargement on the forest cover, agriculture, urbanization and NPP. Second, the Environmental Kuznets Curve was derived for the regions to visualize how they developed in comparison to each other. For these calculations the border regions between the old EU countries (Germany, Austria and Italy) and the new EU countries (Poland, Czech Republic, Hungary, Slovenia and Slovakia) were split into the neighbouring NUTS 3 regions. Then pairs of NUTS 3 regions were built to observe the development in the differences west and east from the border.

It was shown, that the EU eastward enlargement decreased the difference in NPP between the old EU countries and the new EU countries. There was an increase in the difference in forest cover between the old EU countries and the new EU countries. No significant impact on area used for agriculture or urban area was observed. Despite the increase in the difference of forest cover between the old and new EU countries, there is still more

forest cover overall in the European Union, thus it may still explain a positive impact on the NPP.

Concerning the Environmental Kuznets Curve: Both the old EU countries and the new EU countries are beyond the lowest point, meaning that the state of the environment is getting better in both parts of the EU. Clearly, the old EU countries are in a better environmental shape already due to higher economic well being and stronger environmental regulations. However in the course of the observed 13 years from 2000 to 2012, the new EU countries did catch up.

Contents

CHAPTER 1

Introduction

Fast economic growth has been postulated to lead to a time of environmental degradation for a certain period of time, before per capita income has grown enough for people to start becoming conscious about destruction of the environment. A prominent example is the deforestation in the middle ages in Europe, which caused shortage in wood resources as well as natural catastrophes such as landslides and avalanches. The destruction of tropical forests worldwide represents a more contemporary example. The connection of economic development and environmental degradation is described, for example, with the environmental Kuznets curve. This may occur both due to intensification of industry, but also due to land-use changes, like converting forests into agricultural areas or reduction of productive land area for industry or urban development.

Net Primary Production (NPP) is the amount of carbon uptake by plants per unit area per unit of time. It is a good indicator for the biomass production by vegetation or the vitality of plants. A low NPP value indicates low vegetation cover in a given region and/or low productivity, which is for instance the case in cities or on barren ground. Decreasing NPP could result in reduced air quality or shortage of resources that could be gained from plants. Increased exploitation could even result in a change in land cover, for instance from forests to shrub dominated systems. Fast economic growth, i.e. through changes in political factors as favored by the eastward enlargement of the EU, could promote such processes.

The aim of this work is to analyze the effect of political changes on NPP. NPP information is derived from MODIS satellite data for the years from 2000 to 2012, enriched with climate information from E-OBS and WorldClim. Changes in NPP will be examined primarily in boarder regions of countries that joined the European Union during the eastward enlargement (i.e. Poland, Czech Republic, Hungary, Romania) as compared to countries that have already been part of the European Union before such as Austria, Germany or Italy. This should give insights whether a trend as predicted by the Kuznets curve is actually visible over the given time regarding various economic indicators for the regions such as Gross domestic product (GDP) from EUROSTAT data or country statistics.

To sum it up, this work answers the following three questions:

1. Is there a direct impact of joining the European Union on NPP values?
2. What kind of changes in land cover causes the potential changes in NPP values?
3. Can the exogenous impact of GDP on NPP be observed in the environmental Kuznets curve for Europe?

In order to answer these questions: First the state of the art concerning all relevant topics this work is related to is briefly introduced in chapter 2. Then an analysis and explanation of the sources used is given in chapter 3. This is followed by an explanation of the methodological approach in chapter 4 and a detailed analysis of the results in chapter 5. Finally, the conclusion summarizes the results in chapter 6.

CHAPTER 2

State of the Art

This work in its very nature touches on several important aspects of current scientific research. The state of the art in eco-system service research, Environmental Kuznets Curve, the effects of the eastward enlargement of the European Union, land cover changes, Net Primary Production and remote sensing will be examined in this chapter.

Since many western countries have reached a certain level of financial wealth, research has started focusing on other important factors for a population. Knoke et al. in their work [26] utilize portfolio theory to show, that environmental diversity is more important than only maximizing measurable economic benefit. Bateman et al. analysed in their work [4] the impact of taking account all eco-system services on suggested land-use for the United Kingdom. They revealed, that when taking all monetary values of nature into account it may be beneficial to choose a nature-at-work approach that emphasizes bio-diversity instead of strict cultivation. These monetary values include not only earnings from agriculture or lumber but also money people are willing to pay to visit a well-maintained natural park. This scenario would cause an enormous reduction in carbon emissions while improving overall well-being of the population. Further evidence is provided in [5] that the necessity for change is rising due to CO2 emission rates rising while CO2 sinks are decreasing in efficiency. This shows that they way the environment is treated in an economically growing country becomes more and more important the wealthier the population becomes.

The connection between economic wealth and environmental condition has been explored before. One popular concept is the Environmental Kuznets Curve. It describes a U-shaped relationship between economic wealth and the state of the environment. In the end, it describes, that economic growth is indefinitely possible without damaging the environment. This situation is assumed since, once income has reached a certain level, people will automatically start to keep the environment in better shape. Stern et al. in their work [43] critically examine the concept of the Environmental Kuznets Curve. They compared several previous studies that focused on various aspects of connection between the environment, approximated by measures like SO2 emisisons, and economic wealth, estimated either by GDP or per-capita income. All of them show that there is indeed a U-shaped connection between the two dimensions. However, the Environmental Kuznets Curve presumes that there is only a one-way impact of the economy on the environment and no feedback, it is therefore suggested that it can only be used as a descriptive instrument. Panayotou has taken another look at the Environmental Kuznets Curve in the work [39] and suggested that by including more parameters in the base regression, it is possible to turn it from a solely descriptive tool into a policy making tool. He derived a term that includes the GDP growth rate as well as population density and industry share. The result is, that policies can be included into the equation and show that there is an impact on the curve, thus making it useful to devise governmental policies for environmental protection under economic consideration. This concept will also be followed in this work by including the EU eastward enlargement into the regression as a factor.

The eastward enlargement of the European Union was an important event discussed by science. Jehlicka et al. raised the question, whether the eastward enlargement would lead to the end of a progressive environmental policy in the European Union in their work [24]. They saw, that the mere chance of joining the European Union at some point lead to an improvement in the way the environment was treated in the candidate countries. However, it was feared, that once they joined the European Union, they would be free to stick mainly to their own agenda again and thus relax their focus on environmental policy again, which would obviously harm the environment. Schimmelfennig et al. in their work [42] concluded, that the eastward enlargement indeed had a positive impact on the environment in the newly joined countries. It is debated though whether the European Unions positive reaches beyond its borders as discussed for example in [25]

Without a doubt, changes in the economy of a country lead to changes in the country's land cover. DeFries et al. examined carbon emissions in the tropical regions in their work [11]. They showed that both increase in urbanization as well as more intense agricultural use of land cause an increase in carbon emissions and thus suggest to stronger protect (tropical) forests to avoid environmental degradation. Droughts over the last years have done their own share to altering both land cover and the environment further. Allen et al. in their work [2] observed that climate change is already causing an increase in forest fires which in turn release carbon into the atmosphere further encouraging climate change. Another study by Lehsten et al. [29] suggests that there will be a reforestation in Europe that will most likely influence the climate to become better, yet lead to challenges in finding habitats. They suggest that despite reforestation being a welcome change as forests are considered an important carbon sink [38], it still needs to be managed properly to be good for the environment as a whole. Neumann et al. observed that due to the increasing variability in climate, a higher tree mortality can already be observed in Europe, making environmental management not only a policy problem but also one of fighting against the changes humanity has already caused [33].

Net Primary Production is a measure for the overall production of plants in a given region. Melillo et al. did a global survey in [31]. They estimated that about half of the global net primary production occurs in the tropic regions. It also strongly depends on the interactions in the dynamics of carbon, nitrogen and water how well plants are able to incorporate CO2 into their stems. Waring et al. have explored the impact of geographic location on the actual NPP values [47]. Despite the need for specific correction values, this measure still allows for easy estimation of environmental health. This is also shown in [48] where a global trend in the environment getting worse is shown by NPP values. This work will strongly use NPP as a proxy for environmental health, the most important challenge here, was how to obtain the data, which was done via remote sensing technologies.

Remote sensing technologies, in other words using satellite imagery, is popular to gain an overview of what is happening in large areas of land. The satellite data provides the base of further research here. It is provided at a given spatial resolution, i.e. km^2, and a temporal resolution, that might be daily, weekly, etc.. In the case of generating NPP values, the satellites provide leaf area indices that must then be inserted into the proper formula to derive the actual NPP values [22]. There is application in agriculture where the rising availability of imagery may help in managing soil, crop and even pests [32]. Remote sensing data can also be used to observe deforestation and reforestation around the world over several years to reveal global trends as done in [8].

CHAPTER 3

Sources

In this chapter, the sources of data used in this work are described. First the important algorithms, including the derivation of the Net Primary Production, to understand the background of this work, will be discussed. Next the map data obtained from MODIS satellite imagery will be discussed and finally the statistical data obtained from EuroStat will be explained. The quality of all these data sources is important to draw proper conclusions: For example, the availability of NPP data in raster form allowed to properly estimate the the averages in all given political regions, removing the necessity to harmonize various data inputs in order to allow for proper correlation of the inputs.

3.1 Net Primary Production

Net Primary Production (NPP) is the difference between Gross Primary Production (GPP) and plant autotrophic respiration. In otherwords, it is the net carbon fixed by vegetation through photosynthesis. NPP may also be interpreted as a measure of goods provided by the ecosystem. This makes it on the one hand a measure for the stress of plants (including trees) in a certain environment but also a key indicator of environmental changes[34][31]. In short, NPP tells how much carbon is stored in plants per square kilometre per a given time, in this case a year. To understand the data, that will be introduced later on, a little better, a brief introduction to how the NPP is derived, will be given. To to so, first, the Gross Primary Production(GPP) must be calculated:

$$GPP = LUE_{max} * f_{Tmin} * f_{vpd} * 0.45 * SW_{rad} * FPAR \tag{3.1}$$

"LUE_{max} is the maximum light use efficiency, which gets adjusted by f_{Tmin} and f_{vpd} to address water stress due to low temperature (f_{Tmin}) and vapor pressure deficit (f_{vpd}). SWrad is short wave solar radiation load, of which 45% is photosynthetically active. FPAR is the fraction of absorbed photosynthetic active radiation."[34]

The NPP is defined as follows:

$$NPP = GPP - R_M - R_G \tag{3.2}$$

This is the GPP minus the energy the plant needs to grow and stay alive. By performing just a few steps, it is possible to derive NPP values from satellite imagery. The MOD17 algorithm is capable of deriving the NPP values from images generated by the MODIS satellite, as discussed in [41]. Using these images it is possible to first derive the land cover type and by further taking into account the amount of sun exposure at a specific point on earth, the final production of plants is calculated.

For the scope of this work, it is not necessary to understand all these formulae in full detail. What is important is to remember the following:

1. NPP depends on how much light is available. (There is less light available, the closer one gets to the poles)
2. Temperature and availability of water are important to determine the NPP
3. Plants do respire just like humans do, but the main part of what they exhale is oxygen that we can breath

This explains why in the northern region of Europe, the NPP will in general be lower, whereas in moderate regions it is most likely to be relatively high.

3.2 Map data

In this section, the sources for the analysis in this paper will be enumerated. Specifically, which resolution for the map material was used and where it can be obtained.

3.2.1 GeoTiff Format

GeoTiff is a simple file format to store raster data. It is based on the tiff file format augmented with georeference information necessary to directly work with the information in Geo Information Systems(GIS). [17] There is no compression applied to the captured information. One pixel in the image corresponds to one unit in whatever resolution the image was taken in. For example, at a 1 kilometer resolution, 1 pixel equals 1 square kilometer.

3.2.2 Land Cover

The data for the land cover in all relevant regions was downloaded from the ftp server of the Global Land Cover Facility.[18] It was created by the MODIS satellite and is also known by the product code, MCD12Q1. To cover the whole area of Europe, it was necessary to download several tiles and combine them afterwards. The tiles are available in the geotiff format and are split by the UTM zones. [46] To stitch together the area necessary for this work, the latitude codes S through W and the longitude codes 29 to 40 had to be downloaded. There are in total 16 types of land cover identified in the images with two additional values for unclassified points.

The classification is too fine for what is interesting for this kind of work, therefore the relevant categories were grouped together as shown in table 3.1 together with their original values in the images provided by the satellite. All land cover types not mentioned in this table are grouped together in the 'other' category and not relevant for this analysis.

Value	Label	Group
1	Evergreen Needleleaf forest	
2	Evergreen Broadleaf forest	
3	Deciduous Needleleaf forest	Forest
4	Deciduous Broadleaf forest	
5	Mixed forest	
12	Croplands	Agriculture
14	Cropland/Natural vegetation mosaic	
13	Urban and built-up	Urban

Table 3.1: Groups of land cover types used in this work

3.2.3 Net Primary Production

The map data containing information about the average NPP values for the years 2000 to 2012 was provided by the University for Earth Sciences (BOKU) in Vienna and are publicly available from the university's ftp server[23]. The files are available in geotiff format and contain data at a 1 square km resolution for all of Europe. It is important to note, that the creators of the map chose to provide the data in $10gC/m^2/year$. This choice was made so the data could be expressed in integers instead of floating point numbers and thus safe space in the creation of the geotiff files. The numbers were used exactly as is in this work, thus all occurrences of NPP refer to the value being measured in $10gC/m^2/year$.

3.2.4 Area polygons

To split the geotiffs into sections for further analysis, shapefiles, provided by the European Union, were used[36]. The shapefile contains polygons for all NUTS regions at all three levels. The file was split into single polygons for each NUTS 3 region for easier processing.

3.3 Statistical Data

In this section the source for the statistical data used is explained as well as the concept of the NUTS.

3.3.1 NUTS

NUTS stands for Classification of Territorial Units for Statistics coming from French: Nomenclature des unitès territoriales statistiques. The standard was developed by the European Union to simplify the work of the structural funds. There are three NUTS levels for each country, that is part of the European Union. Each of them aims to correspond to a given governmental section of the country. For example, in Austria, NUTS 1 would be groups of states, NUTS 2 are the states, i.e. Vienna, and NUTS 3 are the districts. For this work, the 2013 standard is used, as it fits best with the data, that was available to gain the necessary economic values.[35]

3.3.2 Gross Domestic Product

The Gross Domestic Product GDP is defined as $GDP = PrivateConsumption + PrivateInvestment + GovernmentPurchases + Exports - Imports$. It is therefore a good proxy to estimate the economic well being in a certain region. It does so more than just examining the per capita income, since this measure does not show what actually happens with the money in the given region. [20]

Data for the Gross Domestic Product(GDP) of the different countries at NUTS 3 level, was obtained from EuroStat.[9] The table used is called nama_10r_3gdp, which is part of the National accounts dataset ESA 2010. Here, the per habitat GDP was obtained for the years from 2000 to 2012 at the NUTS 3 level. The choice for per habitat GDP was made, because it gives a better representation of how much value is created in an area in relation to how many people live there instead of just absolute numbers. EuroStat provides a convenient JSON Web Download Service at [14]. It allows to specify the table one wants to obtain data from and offers guidance, via a web interface, to generate the proper query to download the data programmatically.

CHAPTER 4

Methodology

In this chapter, the methodological choices made for the paper are explained in detail. First, the choice of NUTS regions along the borders is explained, then how the data was acquired from EuroStat and as products from the MODIS satellite. There will also be a discussion of choices made in the implementation of software to work with the data starting with acquisition and storage followed by further aggregation. This leads to an analysis of the gathered data of all types including graphical representations as well as the most significant numbers for each kind. Finally, the statistical evaluation is performed by creating a fixed effects model interpreting the gathered data as a panel to do both the differences in differences estimation and examine whether the Environmental Kuznets Curve can be observed.

4.1 Choice of NUTS regions

Since we want to get a thorough understanding of the impact of the EU eastward enlargement, the choice was made, to focus on the border regions of the old and new EU countries. The EU is providing a standardized framework with the NUTS regions to gain comparable regions, as explained before. The NUTS 3 regions, that were examined in this work, are shown in figure 4.1. As can be seen in the figure, all border regions between the old EU countries: Germany, Austria and Italy and the new EU countries: Czech Republic, Slovakia, Hungary and Slovenia (north to south) were selected. Maps with translations of NUTS region codes to the countries own names for each region can be obtained from Eurostat [37]. A list of NUTS 3 regions with their respective areas is provided in the appendix 6.

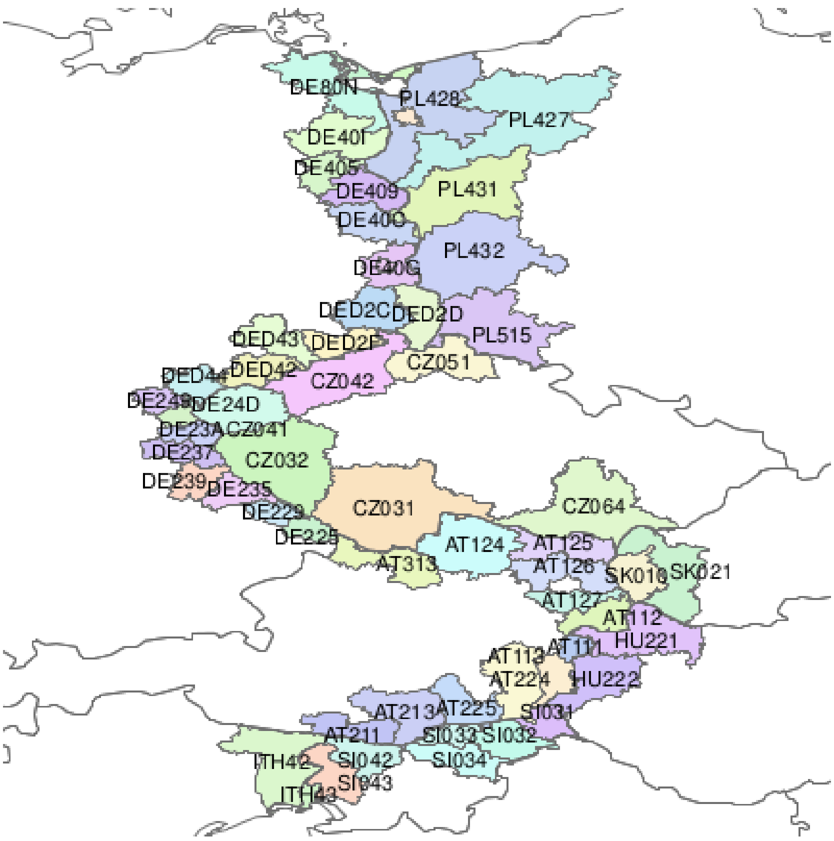

Figure 4.1: Map of the chosen NUTS regions

4.2 Preparation of Data

In this section, the process of acquiring the data and preparing it for analysis will be explained in greater detail. The main focus is on why the implementation choices were made. For full sources, please have a look at the appendix 6

4.2.1 Acquisition

Acquiring the data was, for the most part, easy. As explained in chapter 3 all information was easily obtainable programmatically either via ftp, or http download from the given sources. The only challenging part was, that the land cover information was only available in small tiles, that had to be stitched together before they could be used. In order to do so, the GDAL library, available at [16] proved to be very useful. While not part of the main program, there is an additional python script, that is part of the library, called "gdal_merge.py". By gathering a list of the tiles in plain text files, one file per line, it was easy to stitch the tiles together with the call as shown in algorithm 4.1.

Algorithm 4.1: Call to gdal_merge

python gdal_merge.py -o <location of merged.tif> -q -v –optfile <list of tiles.txt>

Table	Contents
countries	Name of country and whether they are new to the EU or an old country
gdp	Per habitat GDP per NUTS 3 region and year
gdp_normalized	Content of the table gdp normalized to the min-/max- range of each NUTS 3 region
landcover	Land cover classification for each pixel in the geotiff per year and NUTS 3 region
landcover_type	Mapping of land cover number to phonetic name
npp	Net Primary Production for each pixel in the geotiff per year and NUTS 3 region
npp_normalized	Content of the table npp normalized to the min-/max- range of each NUTS 3 region
nuts	List of all relevant NUTS 3 regions
nuts_part	Mapping of which NUTS 3 region belongs to which country
touches	List of all relevant neighbours for each NUTS 3 region

Table 4.1: List of tables in the sqlite database

4.2.2 Choice of Database

For this project, a simple sqlite database was chosen. At first, it seemed interesting to put all the data into a full blown Geo-Database like PostGIS but it proved to be difficult to work with and also far too complex for the type of queries that are actually needed as part of this work. The final layout of the database is discussed in table 4.1.

In hindsight, it would have been considered a better choice to at least use a regular DBMS like Postgres, mainly for performance reasons. Yet, the decision was made to stick with sqlite, since the file could be copied from one computer to another and had no setup time other than to create the tables once.

4.2.3 Splitting data into NUTS regions

Now, that all data is acquired and a database is ready, it remains to store the data in the database. Obtaining the GDP data from EuroStat via HTTP requests was very easy and straight forward. It took a little more work to prepare the map data for insertion to the database. At this point, the informations is just available as large images for the whole of Europe for each year. It would be rather cumbersome to check for each pixel, whether it is within relevant coordinates and then assign it to the proper NUTS 3 region. Therefore, the maps were split into small chunks, according to NUTS 3 regions and inserted into the database individually.

The fact, that the maps for NPP and land cover were provided as gridded data is also a huge advantage. It allows to split them in any way necessary for a given application. In this case based on NUTS 3 regions.

As already explained in section 3.2.4, a file with polygons for all NUTS regions in Europe was obtained. To extract the polygons from the single file, the Python pyshp library, available at [40], proved to be very helpful. Care had to be taken, if a polygon was constructed of multiple parts. More details can be found in the appendix 6.

To slice the geotiff files into NUTS regions, the GDAL library was used again, as shown in algorithm 4.2. The arguments "crop_to_cutline" and "dstnodata" are both very important, since the first reduces the resulting tif to a rectangle, that encompasses the resulting polygon as tightly as possible, thus saving space. The latter specifies which value should be used for the blank space resulting around the polygon. The value would normally be zero, making it indistinguishable from water for the land cover maps, or an urban region, since they usually have an NPP of 0, for NPP maps.

Algorithm 4.2: Slicing a geotiff down to one NUTS region

```
gdalwarp <Path to source.tif> <output.tif> -cutline <Path to shapefile.shp>
  -crop_to_cutline -dstnodata "65535"
```

4.2.4 Import to the Database

With all the files ready, the task of importing them into the database is straight forward. For working with geotiff files in Python, the GDAL library binding is available at [15]. Utilizing the geo information provided by each geotiff, one can iterate over all pixels and import their value one-by-one into the database.

From that point on, all sorts of analysis on the acquired data can be executed as well as to generate csv files to transform them in any way necessary.

4.3 Analysis of the Data

In this section the data, that has been gathered thus far, will be analysed. This will be done in various levels of detail starting at a very coarse view of the old countries versus the newly joined ones and then on a per country basis. The dimensions of the data used in this section are described in table 4.2.

Value	Dimension
NPP	average 10*gC/m^2/year
GDP	EUR/Habitat
Forest	Pixels (approx 1km^2)
Agriculture	Pixels (approx 1km^2)
Urbanization	Pixels (approx 1km^2)
Road_Density	km/1000km^2
Area	km^2

Table 4.2: Dimensions of data

4.3.1 Land Cover

Following are the number of pixels per group and year identified as a certain type of land cover for the old EU countries, shown in table 4.3, and the new EU countries, shown in table 4.4. These two tables summarize all relevant NUTS 3 regions respectively. When looking at the data, it is important to remember, that the numbers are based on satellite imagery, so a little deviation from one year to the other may be explained by atmospheric influences, that could lead to a different classification of land cover type by the algorithm.

	2000	2001	2002	2003	2004	2005	2006
Forest	211440	211440	213688	205049	202449	190547	205118
Agriculture	255444	255444	257263	265992	265856	275913	264578
Urban	17949	17949	17949	17956	17961	17963	17955
Other	15971	15971	11904	11807	14538	16381	13153
	2007	2008	2009	2010	2011	2012	Average
Forest	210565	226741	232352	225264	228452	223153	214327,538461538
Agriculture	259993	244626	239246	248042	244621	250482	255961,538461538
Urban	17938	17953	17940	17958	17963	17963	17953,6153846154
Other	12308	11484	11266	9540	9768	9206	12561,3076923077

Table 4.3: Land Cover for old EU countries by year in Pixels (approx 1 km^2)

	2000	2001	2002	2003	2004	2005	2006
Forest	587242	587242	592174	571720	567426	544347	573850
Agriculture	851828	851828	860135	882488	881558	901216	877054
Urban	55671	55671	55666	55699	55683	55715	55646
Other	35674	35674	22440	20508	25748	29137	23865
	2007	2008	2009	2010	2011	2012	Average
Forest	591924	622668	635984	621001	634629	626652	596681,461538461
Agriculture	861329	833337	819695	839358	824024	832312	855089,384615385
Urban	55582	55675	55585	55711	55714	55708	55671,2307692308
Other	21580	18735	19151	14345	16048	15743	22972,9230769231

Table 4.4: Land Cover for new EU countries by year in Pixels (approx 1 km^2)

There are additional tables with percentage distributions of land cover types to make trends in land cover a little more easily visible. These are provided for the old countries in table 4.5 and the new countries in table 4.6 separately. Also, a graphical representation is shown in figure 4.2 for agriculture and in figure 4.3 for forest cover.

Let us examine the numbers in greater detail:

- The border region consist mainly of forest and agricultural land in both the old and new EU countries.

	2000	2001	2002	2003	2004	2005	2006
Forest	42,22 %	42,22 %	42,67 %	40,94 %	40,42 %	38,05 %	40,96 %
Agriculture	51,01 %	51,01 %	51,37 %	53,11 %	53,09 %	55,09 %	52,83 %
Urban	3,58 %	3,58 %	3,58 %	3,59 %	3,59 %	3,59 %	3,59 %
Other	3,19 %	3,19 %	2,38 %	2,36 %	2,90 %	3,27 %	2,63 %
	2007	2008	2009	2010	2011	2012	
Forest	42,05 %	45,28 %	46,40 %	44,98 %	45,62 %	44,56 %	
Agriculture	51,92 %	48,85 %	47,77 %	49,53 %	48,85 %	50,02 %	
Urban	3,58 %	3,58 %	3,58 %	3,59 %	3,59 %	3,59 %	
Other	2,46 %	2,29 %	2,25 %	1,90 %	1,95 %	1,84 %	

Table 4.5: Land Cover in percent for old EU countries by year

	2000	2001	2002	2003	2004	2005	2006
Forest	38,37 %	38,37 %	38,69 %	37,36 %	37,08 %	35,57 %	37,50 %
Agriculture	55,66 %	55,66 %	56,20 %	57,66 %	57,60 %	58,89 %	57,31 %
Urban	3,64 %	3,64 %	3,64 %	3,64 %	3,64 %	3,64 %	3,64 %
Other	2,33 %	2,33 %	1,47 %	1,34 %	1,68 %	1,90 %	1,56 %
	2007	2008	2009	2010	2011	2012	
Forest	38,68 %	40,69 %	41,56 %	40,58 %	41,47 %	40,95 %	
Agriculture	56,28 %	54,45 %	53,56 %	54,85 %	53,84 %	54,38 %	
Urban	3,63 %	3,64 %	3,63 %	3,64 %	3,64 %	3,64 %	
Other	1,41 %	1,22 %	1,25 %	0,94 %	1,05 %	1,03 %	

Table 4.6: Land Cover in percent for new EU countries by year

- The degree of urbanization remains constant before and after the entry into the EU.
- There is a slight decrease in agricultural use of the land in the old EU countries.
- Agricultural use increases for a few years following the entry of the EU in the new countries, but decreases afterwards.
- The forest cover dips in both the old and new EU countries after entry but a trend towards more forest cover remains. This dip is however not related to the entry in the EU but hurricane "Gudrun", which caused damages to forests all over Europe in the size of the whole yearly timber harvest of Sweden. [12]

To further make sure, that changes in single countries, are not just evened out by contrary changes in other countries, table 4.7 lists the median values for individual countries before 2005 and starting with 2005. This overview supports the afore mentioned trends. In all countries the forest cover has grown. Agriculturally used land on the other hand has declined in general. There was very little change in the degree of urbanization.

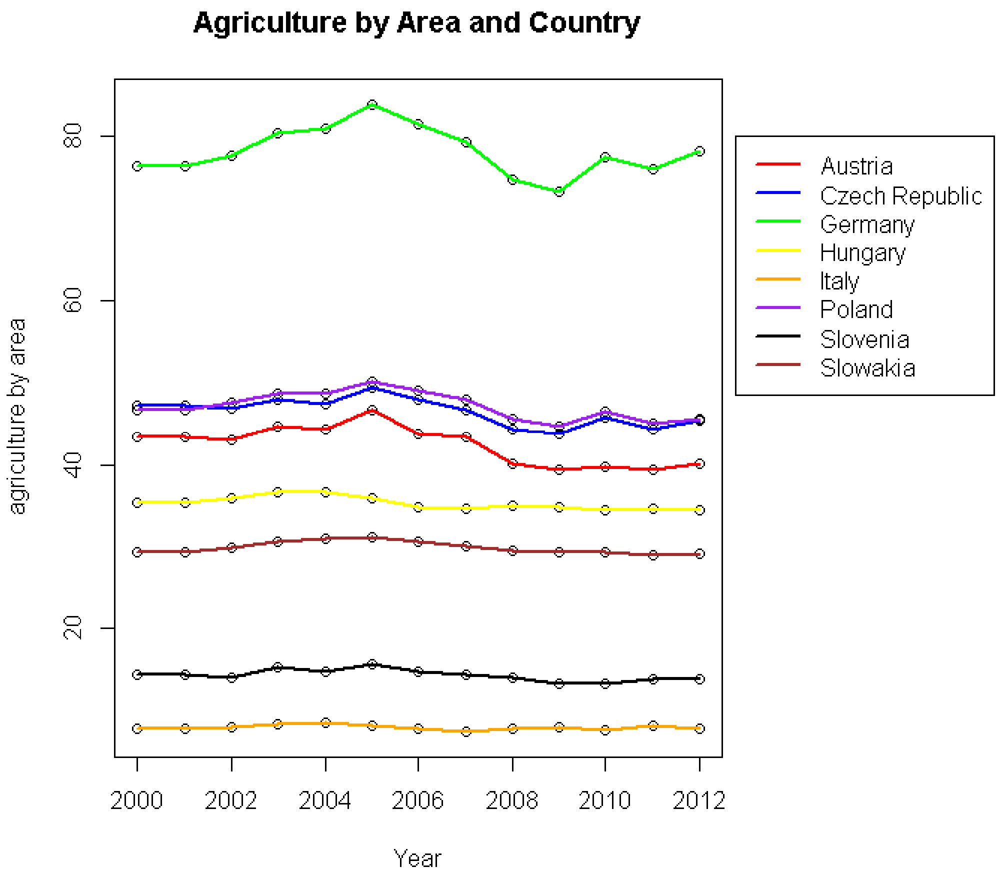

Figure 4.2: Land used for agriculture by country and year as sum of percent pixels identified as agriculture

4.3.2 NPP

Table 4.8 contains a list of the average NPP values per area for the relevant NUTS regions, grouped by country. It was chosen to put the NPP in perspective to the overall area of each country used, to achieve better comparability. To aid understanding the data, figure 4.4 contains a visual representation. It shows, that there is a relatively stable trend for most of the years of interest. There is a distinct drop in NPP values for the year 2003. This was caused by an enormous heatwave that posed significant stress on plants.[13] [6] At this point the importance of forests could also be observed as they mitigate overall heat better than grasslands as described in [44]. Other than that a slight decrease in NPP may be observed starting with the year 2008, which is mostly caused by the general trend towards a warmer climate over the last centuries.[19]

It might be debatable whether it is appropriate to use an average NPP by area, since NPP

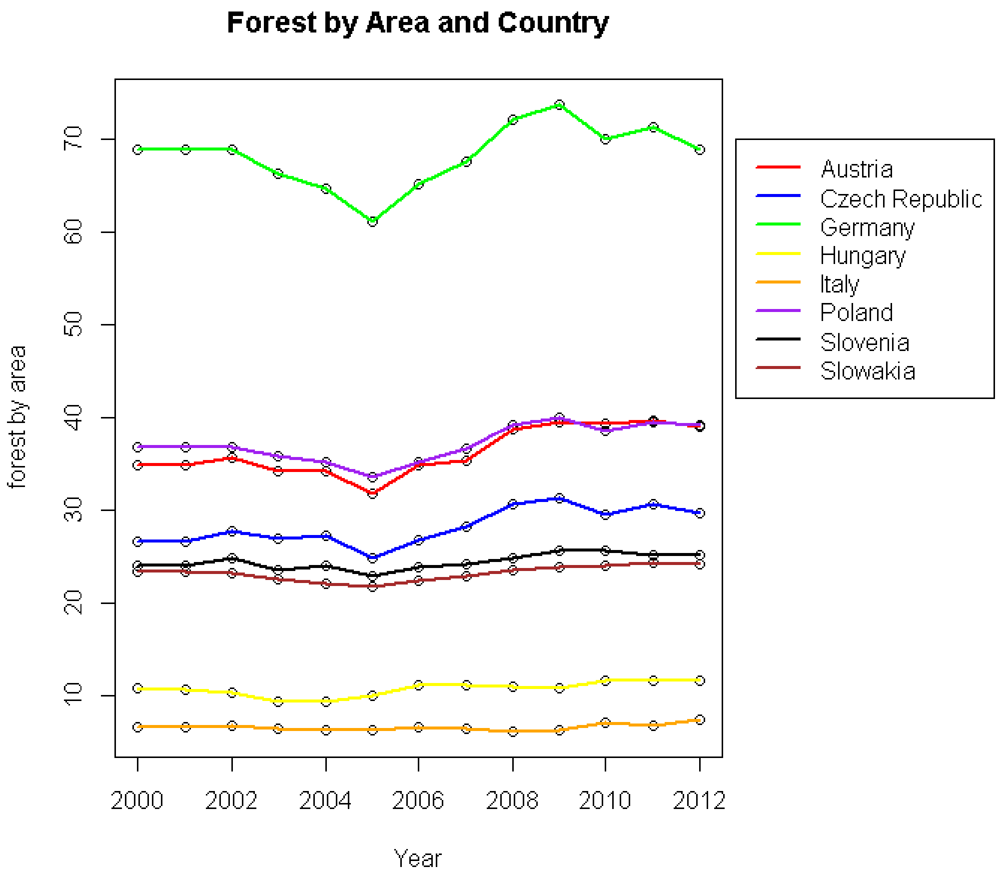

Figure 4.3: Land covered by forest by country and year as sum of percent pixels identified as forest

is already a measure by area. However, it seemed like a sensible choice in this context to put the various sizes of NUTS 3 regions into perspective, evening out differences in very large regions compared to smaller ones.

4.3.3 GDP

Table 4.9 contains a list of the average GDP for the relevant NUTS regions grouped by country. To aid understanding the data, figure 4.5 contains a visual representation. It can clearly be seen, that the average GDP of the old EU countries: Germany, Austria and Italy, is above those of the countries, that joined the EU in 2004. While there is an upward trend for all countries, the trend is stronger manifested in the old EU countries, whereas it stagnates around 2008 for the countries, that newly joined the EU.

	Forest		Agriculture	
Country	before 2005	starting 2005	before 2005	starting 2005
Germany	105296,5	108610,5	142658	140196
Austria	86944,5	98376,5	102987	93277,5
Italy	15892,5	16079,5	15982	15894
Poland	196584,5	207891,5	247077,5	237693,5
Czech Republic	146574	160438	262830,5	252868
Slowakia	150922	156692	178238	173220
Hungary	41223	46606,5	155337	149573
Slovenia	45057	46831,5	27377	26390,5
	Urbanization		Other	
Country	before 2005	starting 2005	before 2005	starting 2005
Germany	11179	11175,5	6760	5517,5
Austria	4993	5001,5	3529	2999
Italy	1779,5	1779	2969	2729,5
Poland	10402	10399	8585,5	6185,5
Czech Republic	18393,5	18397	6403	4082,5
Slowakia	11829	11829,5	4878,5	3724,5
Hungary	13874	13883,5	1840	1900,5
Slovenia	1181	1181	2387	2114

Table 4.7: Median values before 2005 and starting with 2005 for all countries as pixels identified as given type. 1 pixel is approximately $1km^2$.

4.4 Introduction to statistical methods used in this work

In this section, a brief overview of the important techniques: differences in differences estimation as well as fixed effects models, that will be used to analyse the data, will be given.

4.4.1 Difference in Differences

Difference-in-Differences (DiD) is a technique used in econometric research that aims to use observational study data as a set up for an experiment. This is done by dividing the data, that was recorded into occurrences of a treatment group and a control group. It is assumed, that the average outcomes of the two groups, would have followed a parallel path, had the treatment not been applied. At least one observation before the treatment and one observation after the treatment are necessary. In the case of this work, the treatment is the joining of the Eastern European countries to the European Union. After the introduction of two dummy variables, one denoting the treatment status: either applied or not, and one representing the group of which the observation is, it is possible to perform an Ordinary Least Squares regression on the data, to determine whether a treatment effect exists.[1]

Country	2000	2001	2002	2003	2004	2005	2006
Germany	0,16	0,15	0,16	0,14	0,16	0,15	0,15
Austria	0,21	0,21	0,21	0,17	0,20	0,20	0,19
Italy	0,97	0,95	0,99	0,81	0,95	0,94	0,88
Poland	0,09	0,09	0,09	0,08	0,09	0,09	0,08
Czech Republic	0,09	0,09	0,09	0,08	0,09	0,09	0,09
Slowakia	0,46	0,11	0,11	0,10	0,12	0,11	0,10
Hungary	0,17	0,19	0,18	0,15	0,19	0,18	0,18
Slovenia	0,13	0,57	0,58	0,45	0,56	0,53	0,52
Country	2007	2008	2009	2010	2011	2012	
Germany	0,16	0,16	0,16	0,15	0,16	0,15	
Austria	0,20	0,21	0,21	0,20	0,20	0,19	
Italy	0,99	0,93	0,94	0,95	0,94	0,91	
Poland	0,09	0,09	0,09	0,09	0,09	0,09	
Czech Republic	0,09	0,09	0,09	0,09	0,09	0,09	
Slowakia	0,10	0,11	0,11	0,11	0,11	0,10	
Hungary	0,17	0,19	0,18	0,18	0,17	0,15	
Slovenia	0,56	0,55	0,57	0,55	0,55	0,52	

Table 4.8: Average percent NPP per area and year

Table 4.9: Average GDP per year

Country	2000	2001	2002	2003	2004	2005	2006
Germany	16919,05	17409,52	17904,76	17847,62	18247,62	18671,43	19571,43
Austria	18850,00	19275,00	19733,33	20266,67	21341,67	21775,00	22716,67
Italy	22733,33	23933,33	24633,33	25033,33	25800,00	26966,67	27966,67
Poland	3950,00	4483,33	4350,00	3950,00	4225,00	5033,33	5541,67
Czech Republic	5618,18	6300,00	7245,45	7290,91	7909,09	8963,64	10054,55
Slowakia	4225,00	4562,50	5025,00	5762,50	6650,00	7500,00	8712,50
Hungary	4328,57	5000,00	6014,29	6657,14	7342,86	7885,71	7785,71
Slovenia	9433,33	9983,33	10700,00	11016,67	11550,00	12116,67	12866,67
Country	2007	2008	2009	2010	2011	2012	
Germany	20661,90	21680,95	21442,86	22685,71	23728,57	24447,62	
Austria	24258,33	24966,67	24408,33	25216,67	26508,33	27133,33	
Italy	28966,67	28800,00	27533,33	28666,67	29133,33	28200,00	
Poland	6325,00	7408,33	6291,67	7116,67	7433,33	7583,33	
Czech Republic	10990,91	12590,91	11690,91	12290,91	12936,36	12672,73	
Slowakia	10762,50	12537,50	12125,00	12750,00	13525,00	13837,50	
Hungary	8600,00	8914,29	7514,29	8171,43	8414,29	8314,29	
Slovenia	14283,33	15433,33	14316,67	14383,33	14766,67	14433,33	

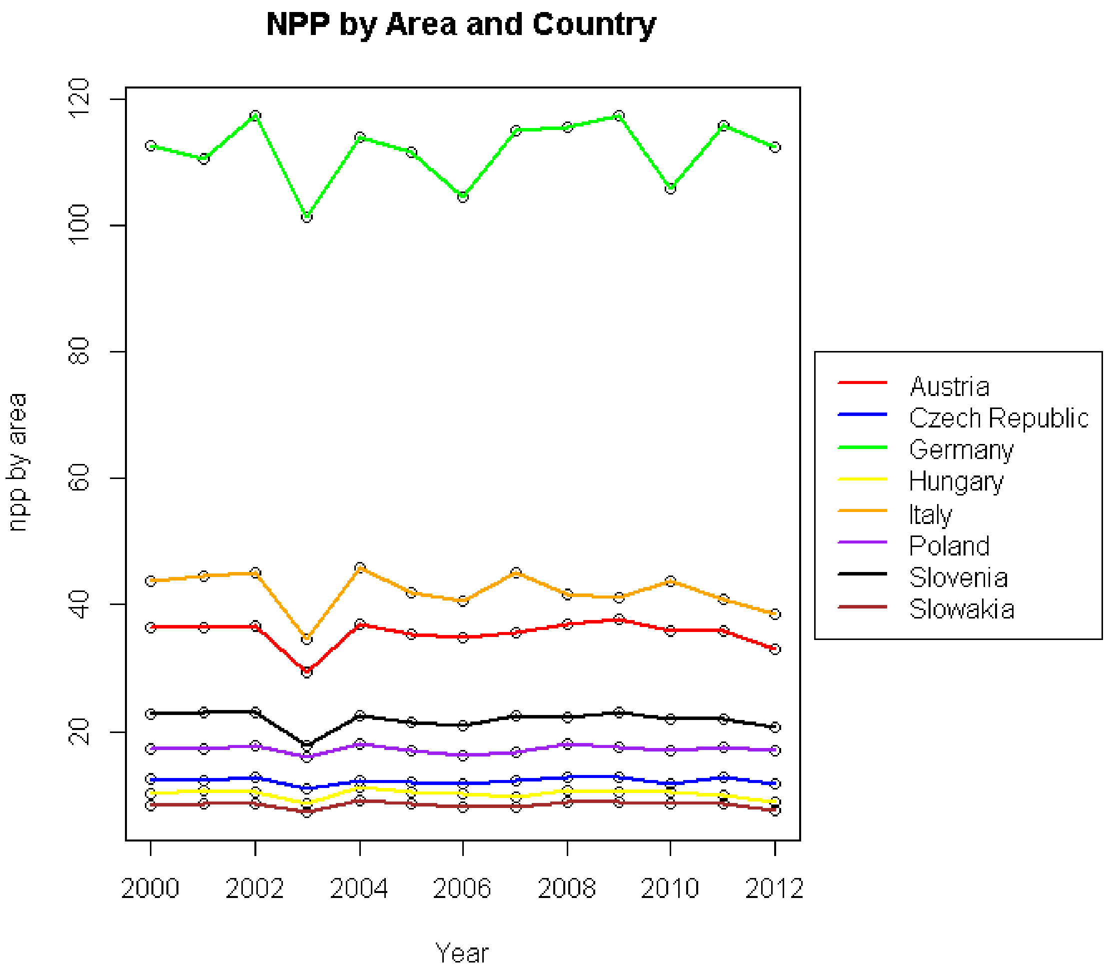

Figure 4.4: Line chart of NPP development over time

4.4.2 Fixed Effects

Fixed effects models help controlling unobserved heterogeneity. This is achieved by introducing an additional variable into the regression equation, that absorbs the individual effect. [3] The fixed effects models for this work were calculated using the R library plm [7]. This library allows to handle panel data natively in R without manually introducing dummy variables into the model.

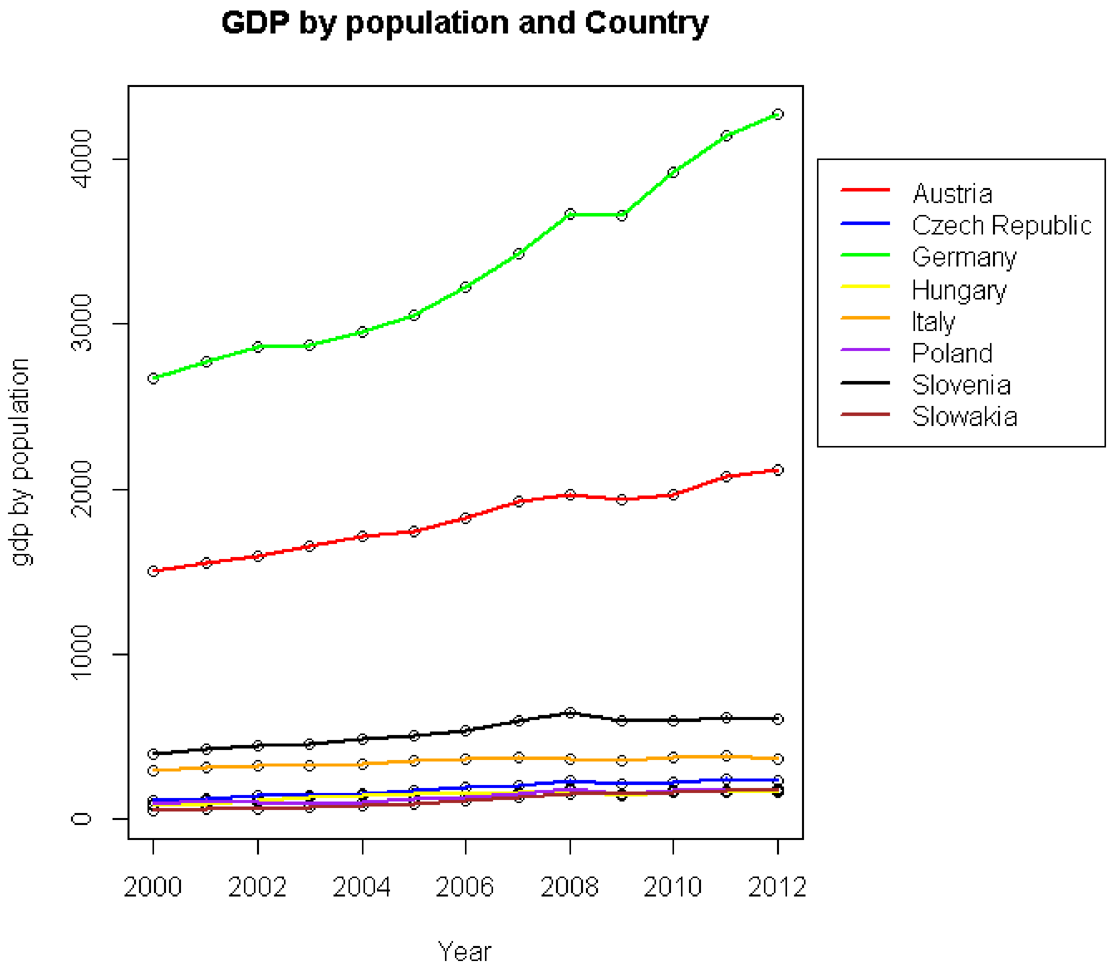

Figure 4.5: Line chart of GDP development over time

CHAPTER 5

Results

In this chapter the results of this work will be discussed. First, the results concerning the Environmental Kuznets Curve will be explained followed by the regression analysis that examines the impact of joining the European Union along with the proof of relevance of the Environmental Kuznets curve. Furthermore, the results will be quantified and translated to illustrate the environmental implications.

5.1 Environmental Kuznets Curve

The environmental Kuznets curve aims to show a dependency of the state of the environment based upon the economy of a country using either per capita income or GDP, based on a reduced-form regression. The curve is derived from panel data using a fixed-effects model. The base equation contains a quadratic function of the income or GDP. Additional variables may be introduced for stronger explanatory power. The Environmental Kuznets Curve is based on many empirically derived assumptions, such as, that when the economy develops, it usually follows a path from simple agriculture to industrialization and further to strong environmental regulation. It also illustrates, that the environment is in good health while the GDP is low in a given region. Then there is a time where GDP and income rise while the environment suffers and finally, when the GDP is high enough, the population becomes more environmentally focused again, leading to a regain in the health of the environment. [43] [10]

In this section, the aim is to show the position of the old EU countries and the new EU countries on the environmental Kuznets curve over time. Since there are various definitions of what the environmental Kuznets curve should exactly look like, it will be used as a comparative tool. Usually it is assumed that there is a U-shaped correlation between state of the environment and an economic dimension.

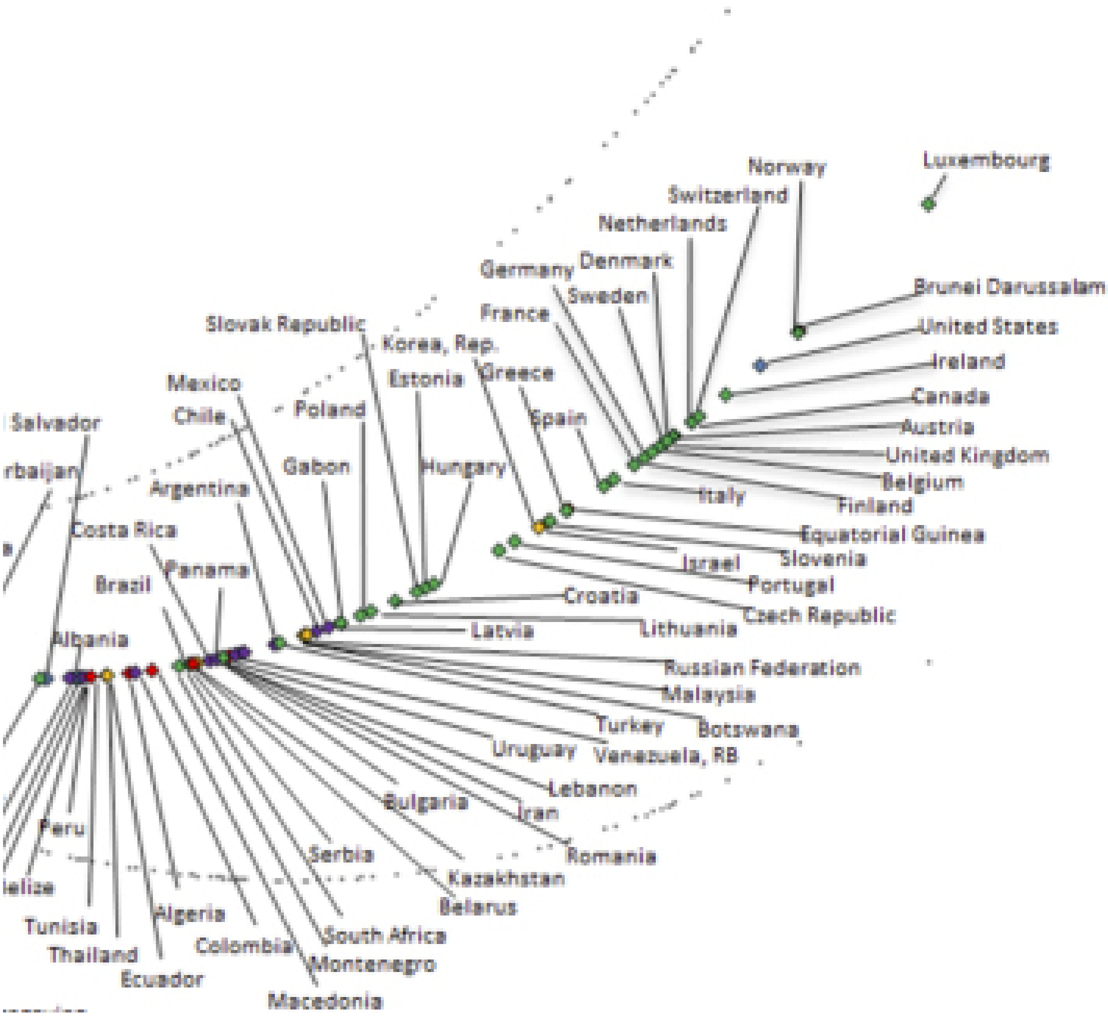

Figure 5.1: Excerpt from the environmental Kuznets Curve for deforestation from [8] It shows the cross border deforestation index on the y-axis from countries around the world compared to the per-capita income on the x-axis. All European countries are past the lowest point of the curve. It also shows similar results to those of this work.

It is safe to assume at least two things: First, that both old EU countries and new EU countries must be beyond the downward slope of the U-curve, since industrialization has already taken place in both parts of Europe. Very poor countries like Guinea can be found in the downward slope. Furthermore, it is also safe to assume, that the old EU countries are farther up the right leg of the U. This is mainly due to their higher development and also due to the fact, that the European Union enforces relatively strict environmental standards. Results of this study support the assumption, since the Environmental Kuznets Curve for deforestation can also be observed as shown in figure 5.1. All European countries are clearly beyond the lowest point of the curve and the old EU countries are farther up than the new ones as of 2005.

To build the Environmental Kuznets Curve, the NPP was taken as a proxy for the state of the environment. Since, as already discussed, this number provides an overall statement of the environmental health in a region, it can be assumed to be a good measure. Instead of the commonly used income per capita, the GDP was used to illustrate the state of the economy in the regions. This also provides a clear picture of the overall economy in a given region, since it describes how much value is actually created instead of just the income of people living there.

There are three figures for the years 2000, 2007 and 2012. The year 2000 shown in figure 5.2 marks the first year of observation and clearly shows the countries lie on a curve as expected. The three old EU countries are close together in one group, while the countries yet to join the EU are behind, both in the state of environment and in GDP. In the year 2007 shown in figure 5.3, three years after joining the EU, it is apparent that the freshly joined countries are starting to catch up with the old EU countries, that remained relatively stable at their position. Taking a look at the year 2012 in figure 5.4, the final year of the observation period, gives a more skewed picture. This may be due to different levels of adherence to EU regulations in the new EU countries but also due to the general degradation of NPP values as discussed above.

The observations also agree with the Kuznets curve drawn in [8]. In this paper the deforestation of many countries around the world was used as a proxy for the state of the environment. The curve, they have drawn, shows the state of the year 2005, which closely resembles the state on this curve in 2007.

Another observation in the three figures is, that Italy is always far ahead of the other countries. While also [8] has shown Italy above countries like Germany and Austria, the difference here may be explained by the circumstance, that only a relatively small region in the north of Italy is used as a data source as shown in section 4.1.

5.2 Statistical evaluation

In order to answer the questions posed at the beginning of this work, a regression analysis was performed.

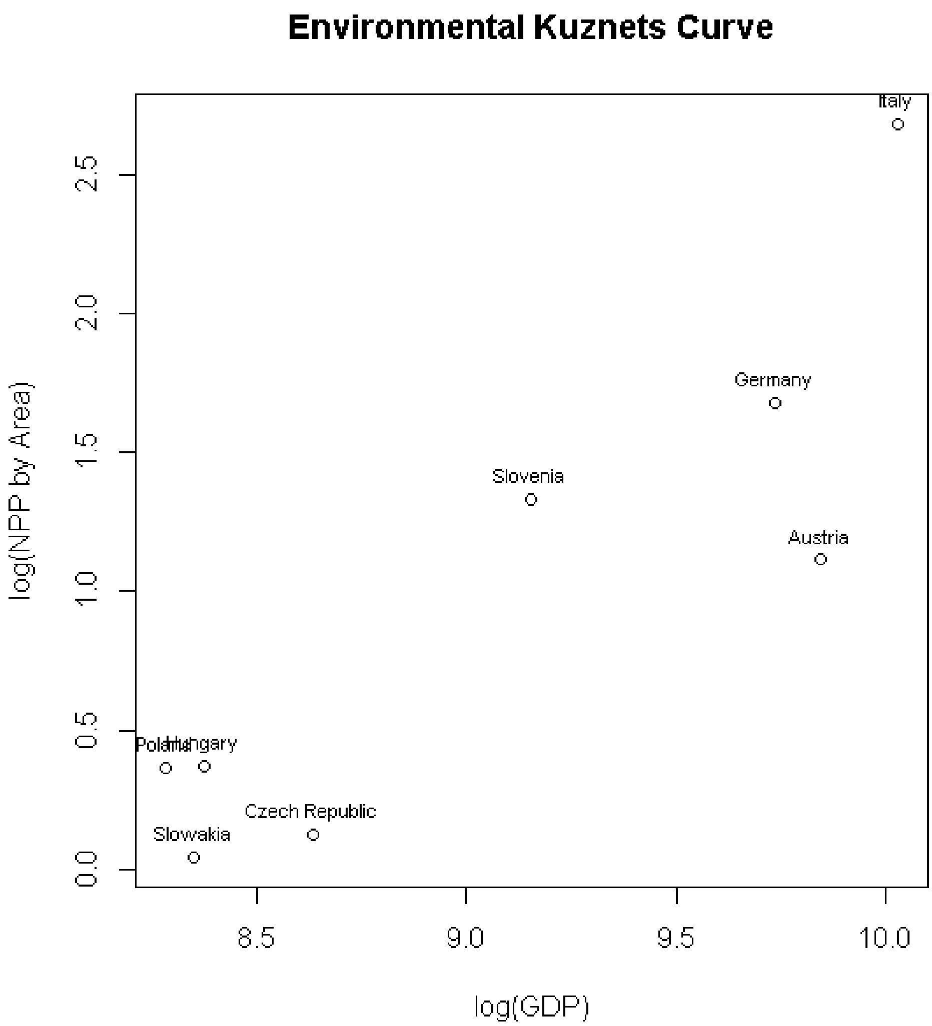

Figure 5.2: Relationship between NPP and GDP in the year 2000

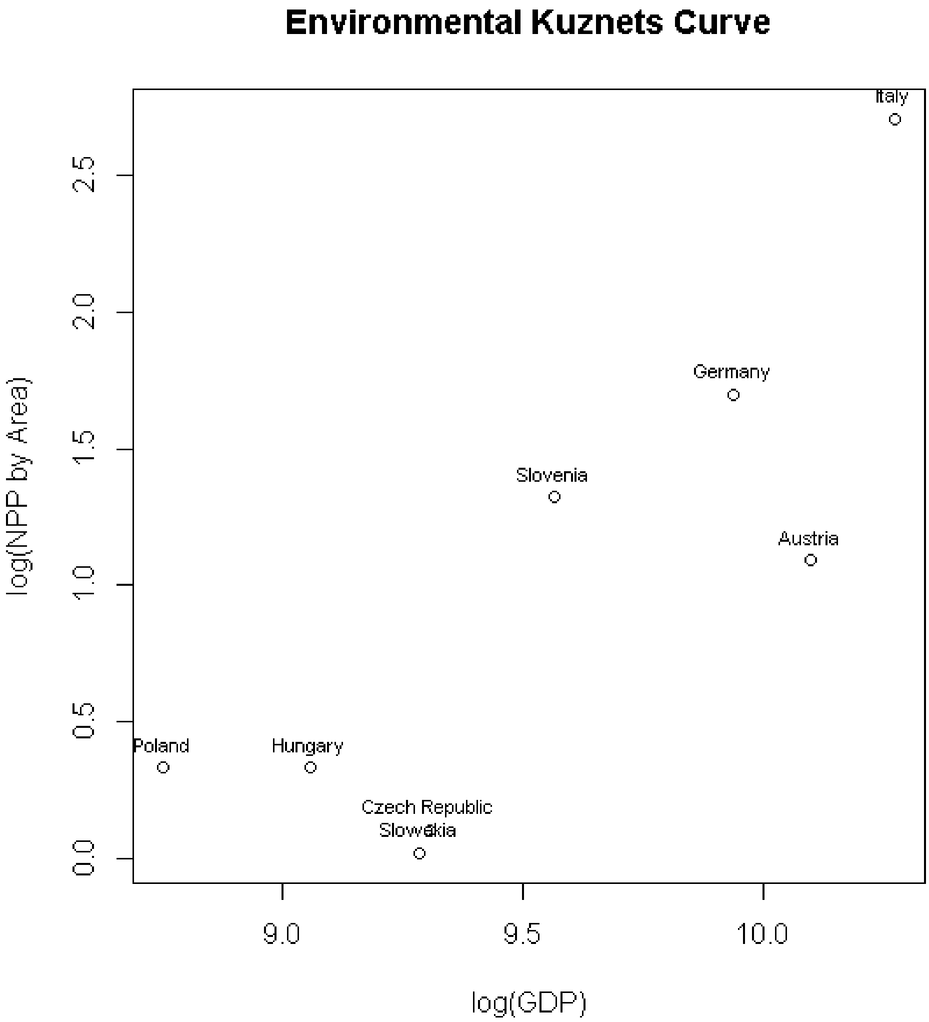

Figure 5.3: Relationship between NPP and GDP in the year 2007

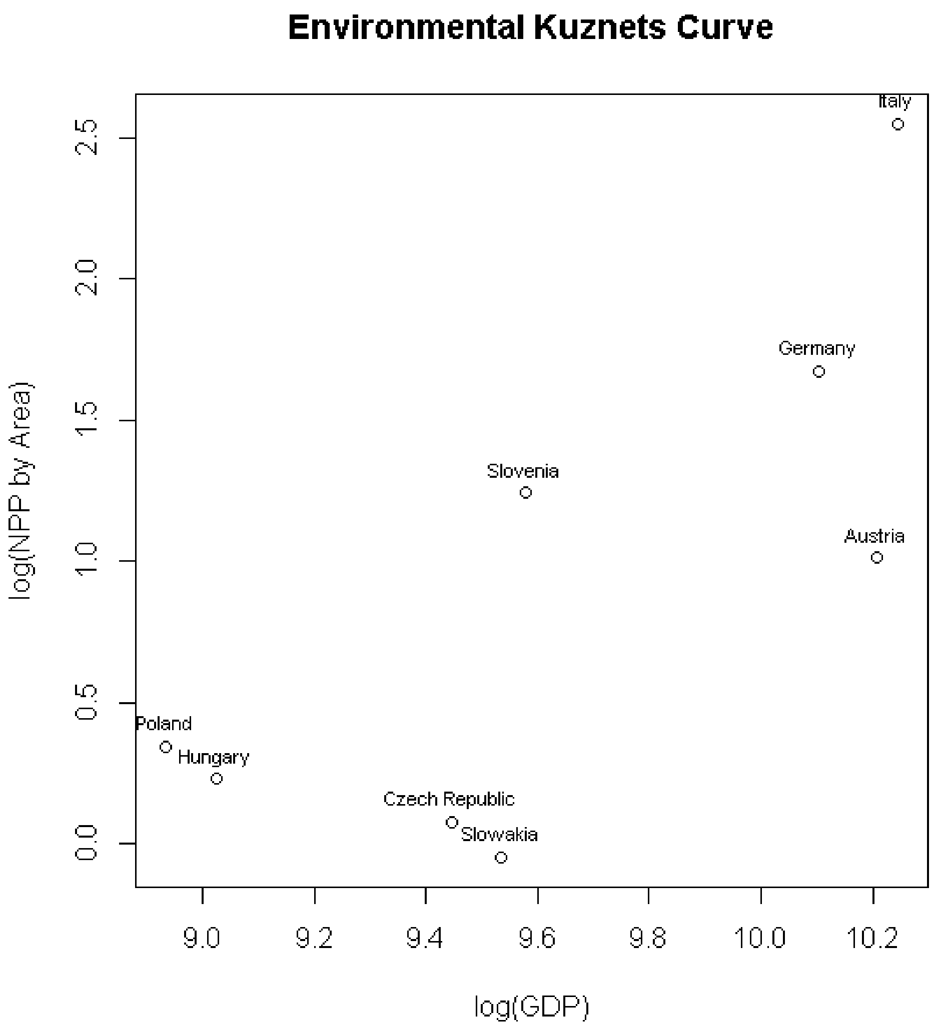

Figure 5.4: Relationship between NPP and GDP in the year 2012

5.2.1 Preparation of the data

The data gathered for this work was used to form a panel of combinations of NUTS 3 regions for each year. To this end, all NUTS 3 regions from the old EU countries were combined with each NUTS 3 region of a new EU country they touch, this again was done for all the years from 2000 to 2012. The result was a table with 676 entries. Table 5.1 summarizes the data in the way it will be used in the regression model later on.

Variable	Observations	Mean	Std.Dev.	Minimum	Maximum
log NPP Dif.	676	-0,7636279	0,8856657	-2,516079	1,245759
log Forest Dif.	676	0,02288287	0,5703672	-1,375922	1,764275
GDP (NUTS 3 region pair differences)	676	-0,7440546	0,3202281	-1,53393	0,5368011
GDP^2 (NUTS 3 region pair differences)	676	-14,06142	5,869858	-28,39974	10,86549
Population (NUTS 3 region pair differences)	676	0,8389097	0,8533507	-1,508735	2,493875

Table 5.1: Descriptive statistics for variables used in regression models

5.2.2 Derivation of the formula

The model was derived based on the work presented in [8]. Yet, for this work a simpler base formula was adopted and also the income per capita was substituted for the GDP in a given region. While the basic idea of the Environmental Kuznets Curve is to show the dependence of the environment based on the income per capita, the GDP seems to be a good proxy. Both values aim to show the wealth in a region, yet, the income per capita could be easily skewed by few people earning large amounts of money, while the broad mass might be poor. The GDP on the other hand reflects the true value created in the region, diminishing income differences in the equation. Furthermore, since the NUTS 3 regions used in this work are not equal in size, the explained variables: NPP, forest cover and agriculture were related to the area of the NUTS 3 region.

While a normal differences in differences regression would be set up to explain a variable per region under treatment, it was necessary to form pairs of regions for this model. This was necessary due to the endogenous nature of the GDP per region. Since this work aims to explain values of nature, like the NPP, that would link back to the GDP, it is only possible to do this comparative analysis.

$$\begin{aligned} log(NPP\ Comparison\ factor) &= \alpha * Enlargement + \beta * (log(GDP_2) - log(GDP_1)) \\ &+ \gamma * (log(GDP_2)^2 - log(GDP_1)^2) + \delta * X_{nuts(Combination,Year)} \end{aligned} \tag{5.1}$$

The final equation is presented in equation 5.1. The NPP Comparison Factor is defined as $NPP_{2byarea}/NPP_{1byarea}$ where $NPP_{xbyarea}$ translates to $NPP_x/Area_x$. This new value should only be regarded as an NPP based coefficient instead of NPP. It was used in the course of this work, to correct potential errors introduced by comparing the average NPP values of differently sized regions. The index 2 is only used for NUTS 3 regions of the new EU countries, while the index 1 is used only for NUTS 3 regions of the old EU countries. The enlargement coefficient is 0 before 2004 and 1 from 2004 on, as it is the dummy variable for the Differences in Differences estimate as described in section 4.4.1. The coefficient $X_{nuts(Combination,Year)}$ is introduced since the model contains fixed effects that aim to catch the regional differences of the many combinations of NUTS 3 regions used for the model. The concept of fixed effects models was described in section 4.4.2.

The same regression including the proper comparison factor has been used to attempt explaining the difference between land used for agriculture as $log(Agriculture_{2byarea}/Agriculture_{1byarea})$ and forest cover as $log(Forest_{2byarea}/Forest_{1byarea})$.

5.2.3 Results and Evaluation

In this section, the results of the regression analysis are discussed. Resulting from the linear model based on panel data, it can be clearly seen, that both the entry into the European Union as well as the GDP play a role in environmental health. It can be observed, that the entry into the EU has a positive impact on the difference in NPP between old and new EU countries, meaning it diminishes the gap slightly, whereas the U-shaped influence of GDP can also be observed. Table 5.2 summarizes the results in the column base. The standard error is shown below each coefficient as well as the t-value that shows the significance of each term. A lower t-value means that a variable is more significant.

Several other coefficients have been introduced, one at a time, into the base model, to check the robustness. The column marked as "1" of table 5.2 shows, that the results are robust to the inclusion of an interaction term between GDP and Population. The columns marked as 2 and 3 show robustness in interaction terms between the eastward Enlargement and the GDP and GDP-squared respectively. 4 suggests that the population in a given area may also be a strong indicator for the environmental health, while 5 reveals, that road density is insignificant in border regions.

The results of all models show, that there is a U-shaped relationship between the state of nature represented by the NPP and economic wealth represented by the GDP.

5 can be interpreted as a proxy for the importance of agriculture, as done by [8]. Since this coefficient is insignificant in the model, it can be assumed that agriculture does not play a huge role in the interaction between economic well being and the state of the environment.

The same analysis as for NPP has also been made for forest cover. The results are shown in table 5.3. The base model is the same as for NPP, yet it shows slightly different results: The entry in the EU did increase the difference in forest cover, suggesting deforestation in the new EU countries, while the coefficients for GDP again suggest the U-shaped relation between forest cover and GDP.

The fixed effects for both models are listed in the appendix 6.

	base	1	2	3	4	5
Difference in Difference	-0,0078951	-0,008052	-0,0096289	-0,0081839	-0,0075661	-0,008551
std.err	0,0029072	0,002912	0,0067125	0,0039531	0,0029091	0,0031575
Pr > |t|	0,006797	0,005861	0,15194	0,03877	0,009522	0,006974
GDP	0,2022396	0,200674	0,1933608	0,19767	0,1084563	0,1249574
std.err	0,0711173	0,0711412	0,0776196	0,082619	0,089787	0,0871205
Pr > |t|	0,004605	0,004944	0,01299	0,01703	0,227535	0,152044
GDP^2	-0,0106188	-0,0100186	-0,0100411	-0,010356	-0,0056335	-0,0063281
std.err	0,0037975	0,0038494	0,0043018	0,0045031	0,0047864	0,0046323
Pr > |t|	0,00533	0,009472	0,01991	0,0218	0,23965	0,172458
GDP_Population		-0,0081094				
std.err		0,0084917				
Pr > |t|		0,339963				
DID_GDP			-0,0022329			
std.err			0,0077907			
Pr > |t|			0,7745			
DID_GDP^2				-9,3443E-13		
std.err				8,5716E-12		
Pr > |t|				0,91323		
Population					0,0559718	
std.err					0,0327957	
Pr > |t|					0,088382	
Road Density						0,0011637
std.err						0,0056901
Pr > |t|						0,838033
No. Observations	676	676	676	676	676	676
R-squared	0,016394	0,017838	0,016524	0,016412	0,020993	0,014162

Table 5.2: Estimation results for Difference in NPP

No significant relation between agricultural land cover or urbanization and the entry into the European Union and the GDP could be found.

5.3 Interpretation

As already discussed in previous work, the forest is an important carbon sink. [38] Therefore this section will focus on the implied changes in forest cover before and after the EU eastward enlargement and how they translate to Diesel in terms of stored energy. To keep this discussion simple, the average forest cover of all regions used in this work was used to express the final numbers.

First, we need to understand how NPP translates to biomass. According to [28] $500gC * m^{-2} * year^{-1}$ translate to $1000gBiomass * m^{-2} * year^{-1}$ meaning that the carbon content of said biomass is at 50% [45]. The biomass so far entails: wood, canopy and fine roots. [30] has examined that 39% of the biomass becomes wood. Now we know how to translate NPP into wood per area and time.

Based on [27] the energy density of 1kg oven-dry wood is 18,5MJ and 48MJ of the equal

	base	1	3	4	5	6
Difference in Difference	0,0201831	0,0212317	-0,0289695	0,015146	0,0191036	0,0186242
std.err	0,0065863	0,0065597	0,0150497	0,0089508	0,0065737	0,0071522
Pr > |t|	0,002276	0,0012738	0,0546968	0,09112	0,0037905	0,009459
GDP	0,4779663	0,4884349	0,226257	0,399	0,7856728	0,5629777
std.err	0,1611174	0,1602549	0,1740264	0,18707	0,2028889	0,1973372
Pr > |t|	0,003127	0,0024028	0,1940396	0,03333	0,0001192	0,004493
GDP^2	-0,0265241	-0,0305376	-0,0101451	-0,021978	-0,0428811	-0,0320102
std.err	0,0086033	0,0086713	0,0096449	0,010196	0,0108156	0,0104925
Pr > |t|	0,00214	0,0004603	0,2932709	0,03151	8,204E-05	0,002391
GDP_Population		0,0542271				
std.err		0,191287				
Pr > |t|		0,004734				
DID_GDP			-0,0633028			
std.err			0,0174671			
Pr > |t|			0,0003137			
DID_GDP^2				-1,6134E-11		
std.err				1,9408E-11		
Pr > |t|				0,40613		
Population					-0,1836457	
std.err					0,0741074	
Pr > |t|					0,0134734	
Road Density						0,039075
std.err						0,0128887
Pr > |t|						0,002544
No. Observations	676	676	676	676	676	676
R-squared	0,075995	0,087818	0,095163	0,077023	0,085057	0,11417

Table 5.3: Estimation results for Difference in forest cover

amount diesel [21]. In other words, 1kg wood is equal to 0,39kg Diesel or, to translate it into a temporal and areal dimension: $390gWood * m^{-2} * year^{-1} = 0,15kgDiesel * m^{-2} * year^{-1}$.

With the basics explained, the calculations are as follows: The whole area of NUTS 3 regions from old and new EU countries examined in this work together is $281376,4km^2$. The average forest cover before the eastward enlargement was 40,11% and after the enlargement it was 41,24%.

By using only the coefficient for the EU eastward enlargement with the inputs from the model, the table 5.4 was derived.

	Difference	Biomass (g/m^2/year)	Wood (g/m^2/year)	Wood (g/year)	Wood (t/year)	Diesel (t/year)
Before	1	2	0,78	88021774247,4367	88021,7742474367	34328,4919565003
After	0,992136	1,984272	0,77386608	89803590349,9483	89803,5903499483	35023,4002364799
Difference	-0,0078951	-0,0157902	-0,006158178	-714628162,038145	-714,628162038145	-278,704983194877

Table 5.4: Difference in Differences of NPP caused by the EU eastward enlarement

The table 5.4 shows how the numbers were derived. The column Wood (g/year) already takes the average forest cover before and after the EU eastward enlargement into account. The final two columns compare the difference in wood between the old EU countries and the new EU countries before and after the enlargement. The final row expresses the differences in differences and shows that while not too huge, there is a significant amount of wood that could be burnt instead of relying on fossil fuels like Diesel, if the overall amount of wood should be kept constant.

	Old Countries before Enlargement	Old Countries after Enlargement	New Countries before Enlargement	New Countries after Enlargement
Forest Cover	42,01 %	43,14 %	38,20 %	39,34 %
Difference	1,13 %	1,13 %	1,14 %	1,14 %

Table 5.5: Average Percent of Forest Cover before and after the EU eastward enlargement

The decrease of difference in NPP before and after the EU eastward enlargement cannot be clearly attributed to an increase in NPP in the new EU countries, or a decrease in the old EU countries. However, when comparing the average forest cover before and after the EU enlargement as done in table 5.5, it can be observed, that the forest cover increased slightly more in the new EU countries than in the old countries. Therefore it may be assumed, that the decrease in difference of NPP can be attributed to more forest growth in the new EU countries.

The model for forest cover may seem contradictory to the above theory at first, as it suggests that the EU eastward enlargement caused an increase in difference of forest cover between old EU countries and new EU countries. However, the mere area covered by forests is not the only indicator for the quality of the forests. If there is a smaller area covered with a denser forest, it may still have a higher NPP than a larger area that is sparsely populated with trees.

CHAPTER 6

Conclusion

This work has shown the use of Net Primary Production (NPP) as an environmental indicator in an economic context. It was also possible to show the environmental impact of political and economic changes over a 12 year period of time. This period of time had the important event of the first EU eastward enlargement. Thus the countries Germany, Austria and Italy were used to represent the old EU, while Poland, the Czech Republic, Hungary, Slovenia and Slovakia were the countries that newly joined. The available data was interpreted as a panel to be used for the regression analysis.

The NPP data was generated from satellite imagery alone, thus making it a viable option to investigate changes on the large scale of up to whole continents. For instance, change in NPP observed in many parts of the world could actually relate to changes in the economic situation of the local communities. These economic changes may overlay climatic effects due to global warming. Since the data is readily available in the GeoTIFF format, it is also feasible to import the data into databases and run the necessary analysis on them. The same also holds true for land cover classification which is actually part of the input to generate NPP data.

It was possible to observe a change in NPP after the EU eastward enlargement. Indeed, the difference in the border regions between the old EU countries and the new EU countries became smaller. This may be attributed to changes in regulative law in the new EU countries. It was also observed, that there was an increase in forest area in the new EU countries after joining the European Union. Due to the important role of the forests as carbon sinks, much of the change may be attributed to the increase in forest cover and possibly also to an increase in density of forests.

The EU eastward enlargement explained an increase in the difference of forest cover of the border regions. However, as the analysis of the data showed, in absolute terms, the forest cover increased in all observed countries. There was a higher increase in the new EU countries after the EU eastward enlargement. No significant impact on the land used for urbanization or agriculture could be found. However, this only means that there was no change in the area used by either. This does not mean, that in the meantime cities could not have become denser or agricultural land use more intense.

Concerning all countries, that were observed in this work, it is visible, that they are on the upward leg of the Environmental Kuznets Curve. Since, with increasing economic development, the health of the natural systems increase as well. For this work the NPP was used to illustrate the state of the environment and the GDP was taken to represent economic development. Over the 12 years of observation, it became clear, that the new EU countries are behind the old EU countries both in GDP and NPP values. This would be expected since their economic development is behind that of the old EU countries. Looking at later years, it is also visible, that the new EU countries were starting to catch up. This did however not happen equally for all the observed countries. Thereby this work has shown, that there is an impact of GDP on NPP.

The results of this work encourage further economic analysis in a similar manner to examine the impact of other economic events on the environment.

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] Alberto Abadie. „Semiparametric difference-in-differences estimators“. In: *The Review of Economic Studies* 72.1 (2005), pp. 1–19.

[2] Craig D Allen et al. „A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests“. In: *Forest Ecology and Management* 259 (2010), pp. 660–684. DOI: 10.1016/j.foreco.2009.09.001. URL: http://ac.els-cdn.com/S037811270900615X/1-s2.0-S037811270900615X-main.pdf?%7B%5C_%7Dtid=38c9879a-3d42-11e7-9e76-00000aacb362%7B%5C&%7Dacdnat=1495274209%7B%5C_%7D.

[3] Paul D Allison. *Fixed effects regression models.* Vol. 160. SAGE publications, 2009.

[4] Ian J. Bateman et al. „Bringing Ecosystem Services into Economic Decision-Making: Land Use in the United Kingdom“. In: *Science* 341.6141 (2013). URL: http://science.sciencemag.org/content/341/6141/45.full.

[5] Josep G Canadell et al. „Contributions to accelerating atmospheric CO 2 growth from economic activity, carbon intensity, and efficiency of natural sinks“. In: (). URL: http://www.pnas.org/content/104/47/18866.full.pdf.

[6] Ph Ciais et al. „Europe-wide reduction in primary productivity caused by the heat and drought in 2003“. In: (). DOI: 10.1038/nature03972. URL: https://www.nature.com/nature/journal/v437/n7058/pdf/nature03972.pdf.

[7] *CRAN - Package plm.* URL: https://cran.r-project.org/web/packages/plm/index.html.

[8] Jesús Crespo Cuaresma et al. „Economic Development and Forest Cover: Evidence from Satellite Data“. In: *Scientific Reports* 7.December 2016 (2017), p. 40678. ISSN: 2045-2322. DOI: 10.1038/srep40678. URL: http://www.nature.com/articles/srep40678.

[9] *Database - Eurostat.* URL: http://ec.europa.eu/eurostat/data/database.

[10] Stern David. „environmental Kuznets curve“. In: *Encyclopedia of energy* 2 (2004), pp. 1–3.

[11] Ruth S Defries et al. „Deforestation driven by urban population growth and agricultural trade in the twenty-first century“. In: (2010). DOI: 10.1038/NGEO756. URL: https://www.nature.com/ngeo/journal/v3/n3/pdf/ngeo756.pdf.

[12] *EUR-Lex - 52010DC0066 - EN - EUR-Lex*. URL: http://eur-lex.europa.eu/legal-content/de/ALL/?uri=CELEX%3A52010DC0066.

[13] *European Heatwave 2003*. 2011. URL: https://www.britannica.com/event/European-heat-wave-of-2003 (visited on 05/02/2017).

[14] *EuroStat - Json and Unicode Web Services*. 2017. URL: http://ec.europa.eu/eurostat/web/json-and-unicode-web-services/about-this-service (visited on 04/23/2017).

[15] *GDAL 2.1.3 : Python Package Index*. 2017. URL: https://pypi.python.org/pypi/GDAL/ (visited on 04/23/2017).

[16] *GDAL: GDAL - Geospatial Data Abstraction Library*. 2017. URL: http://www.gdal.org/ (visited on 04/23/2017).

[17] *GeoTIFF*. 2017. URL: https://trac.osgeo.org/geotiff/ (visited on 04/22/2017).

[18] *GLCF: MODIS Land Cover*. URL: http://glcf.umd.edu/data/lc/.

[19] *Global and European temperature Assessment*. 2017. URL: https://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature-3/assessment (visited on 05/02/2017).

[20] *gross domestic product | Definition and Formula | Britannica.com*. URL: https://www.britannica.com/topic/gross-domestic-product.

[21] Ali Hasan. „Analyzing Performance and Emissions of Diesel and Hydrogen Fuels in a Combustion Nozzle -Using Computer Numerical Techniques“. In: *International Journal of Engineering Innovation & Research* 3.6 (), pp. 2277–5668. URL: http://www.ijeir.org/administrator/components/com%7B%5C_%7Djresearch/files/publications/IJEIR%7B%5C_%7D1219%7B%5C_%7DFinal.pdf.

[22] Hubert Hasenauer et al. „Forest Ecology and Management Reconciling satellite with ground data to estimate forest productivity at national scales“. In: *Forest Ecology and Management* 276 (2012), pp. 196–208. ISSN: 0378-1127. DOI: 10.1016/j.foreco.2012.03.022. URL: http://dx.doi.org/10.1016/j.foreco.2012.03.022.

[23] *Index von ftp://palantir.boku.ac.at/Public/MODIS_EURO/*. URL: ftp://palantir.boku.ac.at/Public/MODIS%5C_EURO/.

[24] Petr Jehlička and Andrew Tickle. „Environmental Implications of Eastern Enlargement: The End of Progressive EU Environmental Policy?“ In: (). DOI: 10.1080/0964401041000168514 6. URL: http://www.tandfonline.com/doi/pdf/10.1080/09644010410001685146.

[25] Christoph Knill and Jale Tosun. „Chair of Comparative Public Policy and Administration Hierarchy, Networks, or Markets: How Does the EU Shape Environmental Policy Adoptions Within and Beyond Its Borders? Hierarchy, Networks, or Markets: How Does the EU Shape Environmental Policy Adoption". In: (2009). URL: http://kops.ub.uni-konstanz.de/volltexte/2009/7972/.

[26] Thomas Knoke et al. „Optimizing agricultural land-use portfolios with scarce data-A non-stochastic model". In: *Ecological Economics* 120 (2015), pp. 250–259. ISSN: 09218009. DOI: 10.1016/j.ecolecon.2015.10.021. URL: http://dx.doi.org/10.1016/j.ecolecon.2015.10.021.

[27] Nike. Krajnc. *Wood fuels handbook.* Food and agricultural organization of the United Nations, 2015. ISBN: 9789251087282. URL: http://www.fao.org/documents/card/en/c/42655f7f-3bdb-4695-a7d8-39763219693e/.

[28] S H Lamlom and R A Savidge. „A reassessment of carbon content in wood : variation within and between 41 North American species". In: *Biomass and Bioenergy* 25 (2003), pp. 381–388. DOI: 10.1016/S0961-9534(03)00033-3. URL: https://www.researchgate.net/profile/Rodney%7B%5C_%7DSavidge/publication/242094365%7B%5C_%7DA%7B%5C_%7Dreassessment%7B%5C_%7Dof%7B%5C_%7Dcarbon%7B%5C_%7Dcontent%7B%5C_%7Din%7B%5C_%7Dwood%7B%5C_%7Dvariation%7B%5C_%7Dwithin%7B%5C_%7Dand%7B%5C_%7Dbetween%7B%5C_%7D41%7B%5C_%7DNorth%7B%5C_%7DAmerican%7B%5C_%7Dspecies/links/5851e85b08ae0c0f3220000f/A-reassessment-of-carbon-content-in-wood-variation-within-and-between-41-North-American-species.pdf.

[29] Veiko Lehsten et al. „Disentangling the effects of land-use change, climate and CO2 on projected future European habitat types". In: *Global Ecology and Biogeography* (2015), n/a–n/a. ISSN: 1466-8238. DOI: 10.1111/geb.12291. URL: http://dx.doi.org/10.1111/geb.12291%7B%5C%%7D5Cnhttp://onlinelibrary.wiley.com/store/10.1111/geb.12291/asset/geb12291.pdf?v=1%7B%5C&%7Dt=i5q1sa6a%7B%5C&%7Ds=b6d9ac3a.

[30] Yadvinder Malhi, Christopher Doughty, and David Galbraith. „The allocation of ecosystem net primary productivity in tropical forests". In: (). DOI: 10.1098/rstb.2011.0062. URL: http://rstb.royalsocietypublishing.org/content/royptb/366/1582/3225.full.pdf.

[31] Jerry M. Melillo et al. „Global climate change and terrestrial net primary production". In: *Nature* 363.6426 (May 1993), pp. 234–240. ISSN: 0028-0836. DOI: 10.1038/363234a0. URL: http://www.nature.com/doifinder/10.1038/363234a0%20http://www.as.wvu.edu/biology/bio463/Melillo%20et%20al%201993TEM%20NPP%20Estimations.pdf.

[32] David J Mulla. „Special Issue: Sensing in Agriculture Review Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps“. In: *Biosystems Engineering* 114 (2013), pp. 358–371. DOI: 10.1016/j.biosystemseng.2012.08.009. URL: http://ac.els-cdn.com/S1537511012001419/1-s2.0-S1537511012001419-main.pdf?%7B%5C_%7Dtid=d1e31c72-3d54-11e7-996b-00000aab0f26%7B%5C&%7Dacdnat=1495282196%7B%5C_%7D.

[33] Mathias Neumann et al. „Climate variability drives recent tree mortality in Europe“. In: *Global Change Biology* April (2017), pp. 1–10. ISSN: 13541013. DOI: 10.1111/gcb.13724. URL: http://doi.wiley.com/10.1111/gcb.13724.

[34] Mathias Neumann et al. „Creating a regional MODIS satellite-driven net primary production dataset for European forests“. In: *Remote Sensing* 8.7 (2016), p. 554.

[35] *NUTS*. URL: http://ec.europa.eu/eurostat/web/nuts/overview.

[36] *NUTS - Eurostat - shapefiles*. URL: http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts.

[37] *NUTS - Maps as PDF*. 2017. URL: http://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf- (visited on 04/23/2017).

[38] Yude Pan et al. „A Large and Persistent Carbon Sink in the World's Forests“. In: *Science* 333.6045 (2011). URL: http://science.sciencemag.org/content/333/6045/988.

[39] Theodore Panayotou. „Demystifying the Environmental Kuznets Curve: Turning a Black Box into a Policy Tool“. In: (). URL: http://www.cid.harvard.edu/archive/esd/pdfs/iep/643.pdf.

[40] *pyshp 1.2.10 : Python Package Index*. 2017. URL: https://pypi.python.org/pypi/pyshp/ (visited on 04/23/2017).

[41] Steven W Running et al. „A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production“. In: (). URL: http://www.bioone.org/doi/pdf/10.1641/0006-3568%7B%5C%%7D282004%7B%5C%%7D29054%7B%5C%%7D5B0547%7B%5C%%7D3AACSMOG%7B%5C%%7D5D2.0.CO%7B%5C%%7D3B2.

[42] Frank Schimmelfennig and Ulrich Sedelmeier. „Governance by conditionality: EU rule transfer to the candidate countries of Central and Eastern Europe“. In: (). DOI: 10.1080/1350176042000248089. URL: https://www.researchgate.net/profile/Ulrich%7B%5C_%7DSedelmeier/publication/40865491%7B%5C_%7D%7B%5C%%7D27Governance%7B%5C_%7Dby%7B%5C_%7DConditionality%7B%5C_%7DEU%7B%5C_%7DRule%7B%5C_%7DTransfer%7B%5C_%7Dto%7B%5C_%7Dthe%7B%5C_%7DCandidate%7B%5C_%7DCountries%7B%5C_%7Dof%7B%5C_%7DCentral%7B%5C_%7Dand%7B%5C_%7DEastern%7B%5C_%7DEurope%7B%5C%%7D27/links/0deec5183dc0d35952000000/Governance-by-Conditionality-EU-Rule-Tr.

[43] David I. Stern, Michael S. Common, and Edward B. Barbier. „Economic growth and environmental degradation: The environmental Kuznets curve and sustainable development“. In: *World Development* 24.7 (1996), pp. 1151–1160. ISSN: 0305750X. DOI: 10.1016/0305-750X(96)00032-0. URL: http://ac.els-cdn.com/0305750X96000320/1-s2.0-0305750X96000320-main.pdf?%7B%5C_%7Dtid=d120a60c-3d3f-11e7-9e0c-00000aacb362%7B%5C&%7Dacdnat=1495273176%7B%5C_%7D.

[44] Adriaan J Teuling et al. „Contrasting response of European forest and grassland energy exchange to heatwaves“. In: *Nature Geoscience* 3 (2010). DOI: 10.1038/NGEO950. URL: https://www.nature.com/ngeo/journal/v3/n10/pdf/ngeo950.pdf.

[45] Sean C. Thomas and Adam R. Martin. „Carbon Content of Tree Tissues: A Synthesis“. In: *Forests* 3.4 (June 2012), pp. 332–352. ISSN: 1999-4907. DOI: 10.3390/f3020332. URL: http://www.mdpi.com/1999-4907/3/2/332/.

[46] *UTM Rows and Zones – tmackinnon.com.* URL: https://tmackinnon.com/utm-rows-and-zones-map.php.

[47] R. H. Waring, J. J. Landsberg, and M. Williams. „Net primary production of forests: a constant fraction of gross primary production?“ In: *Tree Physiology* 18.2 (Feb. 1998), pp. 129–134. ISSN: 0829-318X. DOI: 10.1093/treephys/18.2.129. URL: https://academic.oup.com/treephys/article-lookup/doi/10.1093/treephys/18.2.129.

[48] Maosheng Zhao and Steven W. Running. „Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009“. In: *Science* 329.5994 (2010). URL: http://science.sciencemag.org/content/329/5994/940.

Fixed Effects

This chapter lists all the fixed effects in the models.

Fixed effects for the Net Primary Production

Old NUTS3 Region	New NUTS3 Region	Fixed Effect
DE80N	PL428	-1.23329516475825
DE40I	PL428	-0.751720417111654
DE405	PL428	-0.131820008635206
DE409	PL427	0.924264524466017
DE409	PL428	-0.760258960182249
DE409	PL431	-0.847779160102951
DE40C	PL431	-0.113179943033839
DE40C	PL432	-0.686074031652148
DE40G	PL432	0.902909453513329
DED2D	CZ051	1.30451978527871
DED2D	PL432	0.283116039733725
DED2D	PL515	0.479549512373903
DED2C	CZ042	0.173728013021209
DED2F	CZ042	-0.0408826869595153
DED43	CZ042	0.127612882907045
DED42	CZ041	-0.296870013622059
DED42	CZ042	-0.0804977179376495
DED44	CZ041	-1.12138103502525
AT313	CZ031	-0.482106284080332
AT124	CZ031	-0.785881025826721
AT124	CZ064	0.201669782834111
AT125	CZ064	-0.0526840365806983
AT125	SK010	-0.376926392042235
AT125	SK021	-0.620884101191001
AT126	SK010	-0.238600136688428
AT127	SK010	-0.362807701652365

AT112	HU221	-0.277737560664166
AT112	SK010	-0.311442270512794
AT111	HU221	-0.106676244187719
AT111	HU222	0.266602621678532
AT113	HU222	-0.0708468213193999
AT113	SI031	-0.411687018951486
AT224	SI031	0.54812541531299
AT224	SI032	0.326182806802103
AT225	SI032	0.785439206379672
AT225	SI033	-0.0242296336323366
AT213	SI033	-0.0828584277762787
AT213	SI034	-0.0639256801346
AT213	SI042	0.221233900204657
AT211	SI042	0.411113974898121
ITH42	SI042	-0.39080392380986
ITH42	SI043	0.175050847738061
ITH43	SI043	0.411195636581147
DE249	CZ041	0.0293908304747685
DE24D	CZ041	-0.269465336312636
DE23A	CZ032	-0.0229780350769465
DE23A	CZ041	-0.575072787591392
DE237	CZ032	0.956889915404336
DE239	CZ032	0.118996919389098
DE235	CZ032	0.480393686049453
DE229	CZ032	0.430631678379829
DE225	CZ031	1.59476655965771

Table 1: List of all fixed effects for the model of NPP

Fixed effects forforest cover

Old NUTS3 Region	New NUTS3 Region	Fixed Effect
DE80N	PL428	-1.23329516475825
DE40I	PL428	-0.751720417111654
DE405	PL428	-0.131820008635206
DE409	PL427	0.924264524466017
DE409	PL428	-0.760258960182249
DE409	PL431	-0.847779160102951
DE40C	PL431	-0.113179943033839
DE40C	PL432	-0.686074031652148
DE40G	PL432	0.902909453513329

DED2D	CZ051	1.30451978527871
DED2D	PL432	0.283116039733725
DED2D	PL515	0.479549512373903
DED2C	CZ042	0.173728013021209
DED2F	CZ042	-0.0408826869595153
DED43	CZ042	0.127612882907045
DED42	CZ041	-0.296870013622059
DED42	CZ042	-0.0804977179376495
DED44	CZ041	-1.12138103502525
AT313	CZ031	-0.482106284080332
AT124	CZ031	-0.785881025826721
AT124	CZ064	0.201669782834111
AT125	CZ064	-0.0526840365806983
AT125	SK010	-0.376926392042235
AT125	SK021	-0.620884101191001
AT126	SK010	-0.238600136688428
AT127	SK010	-0.362807701652365
AT112	HU221	-0.277737560664166
AT112	SK010	-0.311442270512794
AT111	HU221	-0.106676244187719
AT111	HU222	0.266602621678532
AT113	HU222	-0.0708468213193999
AT113	SI031	-0.411687018951486
AT224	SI031	0.54812541531299
AT224	SI032	0.326182806802103
AT225	SI032	0.785439206379672
AT225	SI033	-0.0242296336323366
AT213	SI033	-0.0828584277762787
AT213	SI034	-0.0639256801346
AT213	SI042	0.221233900204657
AT211	SI042	0.411113974898121
ITH42	SI042	-0.39080392380986
ITH42	SI043	0.175050847738061
ITH43	SI043	0.411195636581147
DE249	CZ041	0.0293908304747685
DE24D	CZ041	-0.269465336312636
DE23A	CZ032	-0.0229780350769465
DE23A	CZ041	-0.575072787591392
DE237	CZ032	0.956889915404336
DE239	CZ032	0.118996919389098
DE235	CZ032	0.480393686049453

DE229	CZ032	0.430631678379829
DE225	CZ031	1.59476655965771

Table 2: List of all fixed effects for the model of forest cover

NUTS 3 Areas

NUTS 3 Region	Area (km^2)	Country
DE80N	3930.0	Germany
DE40I	3058.1	Germany
DE405	1494.4	Germany
DE409	2127.7	Germany
DE403	147.6	Germany
DE40C	2242.4	Germany
DE40G	1661.7	Germany
DED2D	2106.0	Germany
DED2C	2391.0	Germany
DED2F	1654.0	Germany
DED43	2114.0	Germany
DED42	1828.0	Germany
DED44	1412.0	Germany
AT313	2659.5	Austria
AT124	4614.1	Austria
AT125	2411.9	Austria
AT126	2721.6	Austria
AT127	1474.4	Austria
AT112	1792.6	Austria
AT111	701.5	Austria
AT113	1471.8	Austria
AT224	3352.5	Austria
AT225	2222.9	Austria
AT213	3374.0	Austria
AT211	2029.3	Austria
ITH42	4904.0	Italy
ITH43	466.0	Italy
ITH44	212.0	Italy
PL427	10339.0	Poland
PL428	7888.0	Poland
PL431	6107.0	Poland

PL432	7877.0	Poland
PL515	5571.0	Poland
PL517	4179.0	Poland
PL523	4092.0	Poland
PL227	1354.0	Poland
PL225	2352.0	Poland
PL219	2632.0	Poland
PL218	3525.0	Poland
PL323	5538.0	Poland
CZ041	3314.4	Czech Republic
CZ042	5335.1	Czech Republic
CZ051	3163.1	Czech Republic
CZ052	4758.1	Czech Republic
CZ053	4518.5	Czech Republic
CZ071	5139.3	Czech Republic
CZ080	5554.2	Czech Republic
CZ072	3964.1	Czech Republic
CZ064	7065.7	Czech Republic
CZ031	10055.6	Czech Republic
CZ032	7560.8	Czech Republic
SK010	2053.0	Slowakia
SK021	4148.0	Slowakia
SK022	4502.0	Slowakia
SK023	6343.0	Slowakia
SK031	6788.0	Slowakia
SK032	9455.0	Slowakia
SK041	8993.0	Slowakia
SK042	6753.0	Slowakia
HU221	4062.0	Hungary
HU222	3336.0	Hungary
HU223	3784.0	Hungary
HU212	2251.0	Hungary
HU102	6393.0	Hungary
HU313	2544.0	Hungary
HU311	7247.0	Hungary
SI031	1337.2	Slovenia
SI032	2169.7	Slovenia
SI033	1040.8	Slovenia
SI034	2301.0	Slovenia
SI042	2136.5	Slovenia
SI043	2325.1	Slovenia

DE249	892.6	Germany
DE24D	606.4	Germany
DE23A	1085.1	Germany
DE237	1429.9	Germany
DE239	1472.9	Germany
DE235	1510.0	Germany
DE229	975.1	Germany
DE225	984.2	Germany

Table 3: Areas of all NUTS 3 regions used in this work

Sources

This chapter contains all the sources and related files to fetch the necessary data, prepare it, fill the database and finally generate output files for analysis in R.

Database

This section lists the necessary inputs to prepare the database.

Listing 1: SQLite database schema

```
BEGIN TRANSACTION;
CREATE TABLE `touches` (
        `nuts1` TEXT,
        `nuts2` TEXT,
        PRIMARY KEY(`nuts1`,`nuts2`),
        FOREIGN KEY(`nuts1`) REFERENCES nuts(name),
        FOREIGN KEY(`nuts2`) REFERENCES nuts(name)
);
CREATE TABLE `road_dens` ( `nuts` TEXT, `year` INTEGER, `value`
    REAL, PRIMARY KEY(`nuts`,`year`), FOREIGN KEY(`nuts`)
   REFERENCES nuts(name) );
CREATE TABLE "population" ( `nuts` TEXT, `year` INTEGER, `value
   ` REAL, PRIMARY KEY(`nuts`,`year`), FOREIGN KEY(`nuts`)
   REFERENCES nuts(name) );
CREATE TABLE `nuts_part` (
        `nuts` TEXT,
        `country`         TEXT,
        PRIMARY KEY(`nuts`,`country`),
        FOREIGN KEY(`nuts`) REFERENCES nuts(name),
        FOREIGN KEY(`country`) REFERENCES countries(name)
);
CREATE TABLE `nuts` (
        `name` TEXT,
        PRIMARY KEY(`name`)
);
```

```
CREATE TABLE `npp_normalized` ( `x` INTEGER, `y` INTEGER, `
    value` REAL, `nuts` TEXT, `year` INTEGER, PRIMARY KEY(`x`,`y
    `,`year`), FOREIGN KEY(`nuts`) REFERENCES nuts(name) );
CREATE TABLE `npp` (
        `x`      INTEGER,
        `y`      INTEGER,
        `value`  INTEGER,
        `nuts`   TEXT,
        `year`   INTEGER,
        PRIMARY KEY(`x`,`y`,`year`),
        FOREIGN KEY(`nuts`) REFERENCES nuts(name)
);
CREATE TABLE `landcover_type` (
        `type`   INTEGER,
        `name`   TEXT,
        PRIMARY KEY(`type`)
);
CREATE TABLE `landcover` (
        `x`      INTEGER,
        `y`      INTEGER,
        `type`   INTEGER,
        `nuts`   TEXT,
        `year`   INTEGER,
        PRIMARY KEY(`x`,`y`,`year`),
        FOREIGN KEY(`type`) REFERENCES landcover_type(type),
        FOREIGN KEY(`nuts`) REFERENCES nuts(name)
);
CREATE TABLE `gdp_normalized` ( `nuts` TEXT, `year` INTEGER, `
    value` REAL, PRIMARY KEY(`nuts`,`year`), FOREIGN KEY(`nuts`)
     REFERENCES nuts(name) );
CREATE TABLE `gdp` (
        `nuts`   TEXT,
        `year`   INTEGER,
        `value`  REAL,
        PRIMARY KEY(`nuts`,`year`),
        FOREIGN KEY(`nuts`) REFERENCES nuts(name)
);
CREATE TABLE `employment` ( `nuts` TEXT, `year` INTEGER, `value
    ` REAL, PRIMARY KEY(`nuts`,`year`), FOREIGN KEY(`nuts`)
    REFERENCES nuts(name) );
CREATE TABLE "countries" (
        `name`   TEXT,
        `joined`         INTEGER,
```

```
        PRIMARY KEY(`name`)
);
CREATE TABLE `area` ( `nuts` TEXT, `year` INTEGER, `value` REAL
    , PRIMARY KEY(`nuts`,`year`), FOREIGN KEY(`nuts`) REFERENCES
    nuts(name) );
COMMIT;
```

Listing 2: List of countries and their joining date if they are new EU countries

```
Germany 0
Austria 0
Italy   0
Poland  2004
Czech Republic  2004
Slowakia        2004
Hungary 2004
Slovenia        2004
```

Listing 3: List of landcover classifications

```
0       Water
1       Evergreen Needleleaf forest
2       Evergreen Broadleaf forest
3       Deciduous Needleleaf forest
4       Deciduous Broadleaf forest
5       Mixed forest
6       Closed shrublands
7       Open shrublands
8       Woody savannas
9       Savannas
10      Grasslands
11      Permanent wetlands
12      Croplands
13      Urban and built-up
14      Cropland/Natural vegetation mosaic
15      Snow and ice
16      Barren or sparsely vegetated
254     Unclassified
255     Fill Value
```

Listing 4: List of all NUTS3 regions and the country they are part of

```
DE80N   Germany
DE40I   Germany
DE405   Germany
DE409   Germany
```

DE403	Germany
DE40C	Germany
DE40G	Germany
DED2D	Germany
DED2C	Germany
DED2F	Germany
DED43	Germany
DED42	Germany
DED44	Germany
DE249	Germany
DE24D	Germany
DE23A	Germany
DE237	Germany
DE239	Germany
DE235	Germany
DE229	Germany
DE225	Germany
AT313	Austria
AT124	Austria
AT125	Austria
AT126	Austria
AT127	Austria
AT112	Austria
AT111	Austria
AT113	Austria
AT224	Austria
AT225	Austria
AT213	Austria
AT211	Austria
ITH42	Italy
ITH43	Italy
ITH44	Italy
PL427	Poland
PL428	Poland
PL431	Poland
PL432	Poland
PL515	Poland
PL517	Poland
PL523	Poland
PL227	Poland
PL225	Poland
PL219	Poland
PL218	Poland

```
PL323    Poland
CZ041    Czech Republic
CZ042    Czech Republic
CZ051    Czech Republic
CZ052    Czech Republic
CZ053    Czech Republic
CZ071    Czech Republic
CZ080    Czech Republic
CZ072    Czech Republic
CZ064    Czech Republic
CZ031    Czech Republic
CZ032    Czech Republic
SK010    Slowakia
SK021    Slowakia
SK022    Slowakia
SK023    Slowakia
SK031    Slowakia
SK032    Slowakia
SK041    Slowakia
SK042    Slowakia
HU221    Hungary
HU222    Hungary
HU223    Hungary
HU212    Hungary
HU102    Hungary
HU313    Hungary
HU311    Hungary
SI031    Slovenia
SI032    Slovenia
SI033    Slovenia
SI034    Slovenia
SI042    Slovenia
SI043    Slovenia
```

Listing 5: List of which NUTS3 regions touch each other

```
DE40I:  DE80N
DE405:  DE40I
DE409:  DE405
DE40C:  DE409
DE40G:  DE40C
DED2D:  DE40G
DED2C:  DED2D, DE40G
DED2F:  DED2C
```

DED43: DED2F
DED42: DED43
DED44: DED42
SI042: SI043
SI034: SI042
SI033: SI034
SI032: SI033, SI034
SI031: SI032
HU222: HU223
HU221: HU222
HU212: HU221
HU102: HU212
HU313: HU102
HU311: HU313
SK021: SK010
SK022: SK021
SK023: SK022, SK021
SK031: SK022
SK032: SK031, SK023
SK041: SK031, SK032
SK042: SK041, SK032
CZ042: CZ041
CZ051: CZ042
CZ052: CZ051
CZ053: CZ052
CZ071: CZ053
CZ080: CZ071
CZ072: CZ080, CZ071
CZ064: CZ071, CZ053
CZ031: CZ064
CZ032: CZ031, CZ041
PL427: PL428
PL431: PL427
PL432: PL431
PL515: PL432
PL517: PL515
PL523: PL517
PL227: PL523
PL225: PL227
PL219: PL225
PL218: PL219
PL523: PL218
ITH43: ITH42

ITH44: ITH43
AT124: AT313
AT125: AT124
AT126: AT125, AT124
AT127: AT126
AT112: AT127
AT111: AT112
AT113: AT111
AT224: AT113
AT225: AT224
AT213: AT225
AT211: AT213
SI031: HU222, HU223, AT224, AT113
SI042: ITH42, AT211, AT213
SI043: ITH42, ITH43
SI034: AT213
SI033: AT213, AT225
SI032: AT225, AT224
DED44: CZ041
DED42: CZ041, CZ042
DED43: CZ042
DED2F: CZ042
DED2C: CZ042
DED2D: CZ051
DED2D: PL515, PL432
DE40G: PL432
DE40C: PL432, PL431
DE409: PL431, PL427, PL428
DE405: PL428
DE40I: PL428
DE80N: PL428
HU221: SK010, SK021, AT112, AT111
HU212: SK023
HU102: SK023
HU313: SK032
HU311: SK032, SK042
HU222: AT111, AT113
SK021: CZ064, AT125
SK022: CZ064, CZ072
SK031: CZ080
SK010: AT125, AT126, AT127, AT112
CZ051: PL515
CZ052: PL515, PL517

```
CZ053: PL517
CZ071: PL517, PL523
CZ080: PL523, PL227, PL225
CZ031: AT313, AT124
CZ064: AT124, AT125
DED44: DE249
DE249: DE24D, CZ041
DE24D: DE23A, CZ041
DE23A: DE237, CZ041, CZ032
DE237: DE239, CZ032
DE239: DE235, CZ032
DE235: DE229, CZ032
DE229: DE225, CZ032
DE225: CZ031
```

Listing 6: Import countries into the database

```
import sqlite3 as lite
import os, sys, os.path

db = "F:\Diplomarbeit\Mapdata\data.db"
countries = "F:\Diplomarbeit\Mapdata\countries.txt"

con = lite.connect(db)

with con:
    cur = con.cursor()
    with open(countries, 'r') as lc:
        for line in lc:
            input = line.split("\t")
            print input
            cur.execute("INSERT INTO countries VALUES('" +
               input[0] + "'," + input[1].replace('\r\n', '') +
                ")")
```

Listing 7: Insert landcover types into the database

```
import sqlite3 as lite
import os, sys, os.path

db = "F:\Diplomarbeit\Mapdata\data.db"
landcoverclasses = "F:\Diplomarbeit\Mapdata\
   landcover_classification.txt"

con = lite.connect(db)
```

```
with con:
    cur = con.cursor()
    with open(landcoverclasses, 'r') as lc:
        for line in lc:
            input = line.split('\t')
            cur.execute("INSERT INTO landcover_type VALUES('" +
                input[0] + "','" + input[1] + "')")
```

Listing 8: Insert NUTS 3 regions into the database

```
import sqlite3 as lite

db = "F:\Diplomarbeit\Mapdata\data.db"
nuts = "../whitelist.txt"

whitelist = []

with open(nuts, 'r') as nu:
    for line in nu:
        whitelist.append(line.replace('\n',''))

con = lite.connect(db)

with con:
    for nuts in whitelist:
        cur = con.cursor()
        cur.execute("INSERT INTO nuts VALUES('"+ nuts +"')")
```

Listing 9: Insert which NUTS 3 regions are part of which country into the database

```
import sqlite3 as lite
import os, sys, os.path

db = "F:\Diplomarbeit\Mapdata\data.db"
parts = "F:\Diplomarbeit\Mapdata\\relevant_nuts.txt"

con = lite.connect(db)

with con:
    cur = con.cursor()
    with open(parts, 'r') as lc:
        for line in lc:
            input = line.split('\t')
            print input
```

```
            cur.execute("INSERT␣INTO␣nuts_part␣VALUES('" +
                input[0] + "','" + input[1].replace('\r\n', '').
                replace('\n', '') + "')")
```

Listing 10: Insert which NUTS 3 regions touch each other into the database

```
import sqlite3 as lite
import os, sys, os.path

db = "F:\Diplomarbeit\Mapdata\data.db"
parts = "F:\Diplomarbeit\Mapdata\\touchpoints.txt"
con = lite.connect(db)

with con:
    cur = con.cursor()
    with open(parts, 'r') as lc:
        for line in lc:
            input = line.split(':')
            master = input[0]
            others = []
            o = input[1].split(',␣')
            for i in o:
                others.append(i.replace('\r', '').replace('\n',
                    ''))
            for o in others:
                oi = o.replace('␣', '').replace('\n', '').
                    replace('\r', '')
                cur.execute("INSERT␣INTO␣touches␣VALUES('" +
                    master + "','" + oi + "')")
                cur.execute("INSERT␣INTO␣touches␣VALUES('" + oi
                    + "','" + master + "')")
```

Fetching and importing data

This section focuses on the sources to obtain the data.

Map Data

Following are all the necessary files to fetch and prepare the map data.

Listing 11: Download the land cover data for the appropriate regions

```
from ftplib import FTP
import os, sys, os.path
```

```
source = 'glcf/Global_LNDCVR/UMD_TILES/Version_5.1/'
destination = 'tiles'
os.chdir(destination)

ftp = FTP('ftp.glcf.umd.edu')
ftp.login()

ftp.cwd(source)
years = ftp.nlst()

letters = ['S', 'T', 'U', 'V', 'W']
numbers = [29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40]

for year in years:
    if not os.path.exists(year.split('.')[0]):
        os.makedirs(year.split('.')[0])
    os.chdir(year.split('.')[0])
    ftp.cwd(year)
    tiles = ftp.nlst()
    for tile in tiles:
        identifier = tile.split('.')[2]
        if 'jpg' not in identifier:
            lettercode = identifier[0:2]
            numberone = int(identifier[2:4])
            numbertwo = int(identifier[4:6])
            if (lettercode[0] in letters or lettercode[1] in
               letters) and (numberone in numbers and numbertwo
                in numbers):
                print year + ": " + lettercode + str(numberone)
                    + str(numbertwo)
                ftp.cwd(tile)
                filename = tile + ".tif.gz"
                file = open(filename, 'wb')
                ftp.retrbinary('RETR ' + filename, file.write)
                file.close()
                ftp.cwd('..')
    os.chdir('..')
    ftp.cwd('..')

ftp.quit()
```

Listing 12: Programatically extract the downloaded gzip files

```
import gzip
```

```
import os, sys, os.path

destination = 'tiles'
os.chdir(destination)
extractdir = 'extracted'

for year in os.listdir('.'):
    os.chdir(year)
    if not os.path.exists(extractdir):
        os.makedirs(extractdir)
    for tile in os.listdir('.'):
        if '.gz' in tile:
            print tile
            inF = gzip.open(tile, 'rb')
            outF = open(extractdir + '\\' + tile[:-3], 'wb')
            outF.write(inF.read())
            inF.close()
            outF.close()
    os.chdir('..')
```

Listing 13: Merge the tiles into one large map

```
import os, sys, os.path
from subprocess import call

destination = 'tiles'
os.chdir(destination)
gdal_call = "C:\\Program Files (x86)\\GDAL\\gdal_merge.py"
merged_dir = "D:\\PythonProjects\\MasterWork\\tiles\\merged\\"

extractdir = 'extracted'
for year in os.listdir('.'):
    os.chdir(os.path.join(year, extractdir))
    file = open("list.txt", "w")
    for f in os.listdir('.'):
        if '.tif' in f and 'MCD12Q1' in f:
            file.write(f + "\n")
    file.close()
    call(['python', gdal_call, '-o', merged_dir + "LC_" + year.
        split('.')[0] + "_merged.tif", '-q', '-v', '--optfile',
        'list.txt'])
    os.chdir('../..')
```

Listing 14: Generate shapes to clip maps to single NUTS 3 regions

```
import shapefile

whitelist = []
with open('../whitelist.txt', 'r') as bl:
    for line in bl:
        whitelist.append(line.replace('\n', ''))

sf = shapefile.Reader("../NUTS_RG_01M_2013.shp")

shapes = sf.shapes()
records = sf.records()
fields = sf.fields

shplen = len(list(sf.iterShapes()))

myf = []

for i in range(1, len(fields)):
    myf.append(fields[i])

def wrapper2(func, args): # without star
    func(*args)

for i in range(0, shplen):
    if len(records[i][0]) >= 5 and records[i][0] in whitelist:
        print "THIS IS: " + records[i][0]
        partcount = len(shapes[i].parts)

        if partcount is 1:
            w = shapefile.Writer(shapeType=shapes[i].shapeType)
            w.autoBalance = 1
            pointlist = []
            for p in shapes[i].points:
                pointlist.append([p[0], p[1]])
            w.poly(parts=[pointlist])
            for f in myf:
                wrapper2(w.field, f)
            wrapper2(w.record, records[i])
            w.save('shapefiles/' + records[i][0] + ".shp")
        else:
            for k in range(0, partcount):
                w = shapefile.Writer(shapeType=shapes[i].
                    shapeType)
```

```
            w.autoBalance = 1
            pointlist = []
            curpoints = []
            start = shapes[i].parts[k]
            end = 0
            if k < partcount - 1:
                end = shapes[i].parts[k+1]
            else:
                end = len(shapes[i].points)

            for po in range(start, end):
                curpoints.append(shapes[i].points[po])

            for p in curpoints:
                pointlist.append([p[0], p[1]])
            w.poly(parts=[pointlist])
            for f in myf:
                wrapper2(w.field, f)
            wrapper2(w.record, records[i])
            w.save('shapefiles/' + records[i][0] + "_" +
               str(k) + ".shp")
```

Listing 15: Clip land cover maps to NUTS 3 regions

```
import subprocess
import os

basedir = "shapefiles"
indir = "F:\\Diplomarbeit\\Mapdata\\LC_merged\\"
outdirbase = "LC_NUTS"

for shape in os.listdir(basedir):
    if shape[-3:] in 'shp':
        shapename = shape[:-4]
        if not os.path.exists(os.path.join(outdirbase,shapename
           )):
            os.makedirs(os.path.join(outdirbase,shapename))
        outdir = os.path.join(outdirbase,shapename)
        for infile in os.listdir(indir):
            #gdalwarp -cutline shapefiles\AT111.shp -
               crop_to_cutline EU_NPP_mean_2000_2012.tif test.
               tif
            subprocess.call(['gdalwarp', os.path.join(indir,
               infile) , os.path.join(outdir, infile.split('
```

```
                merged')[0] + shape[:-4] + '.tif'), '-cutline',
                os.path.join(basedir, shape), '-crop_to_cutline'
                , '-dstnodata', '"65535"'])
```

Listing 16: Clip NPP maps to NUTS 3 regions

```
import subprocess
import os

basedir = "shapefiles"
indir = "F:\\Diplomarbeit\\Mapdata\\NPP\\"
outdirbase = "NPP_NUTS"

for shape in os.listdir(basedir):
    if shape[-3:] in 'shp':
        shapename = shape[:-4]
        if not os.path.exists(os.path.join(outdirbase,shapename
            )):
            os.makedirs(os.path.join(outdirbase,shapename))
        outdir = os.path.join(outdirbase,shapename)
        for infile in os.listdir(indir):
            if '.tif' in infile:
                #gdalwarp -cutline shapefiles\AT111.shp -
                    crop_to_cutline EU_NPP_mean_2000_2012.tif
                    test.tif
                subprocess.call(['gdalwarp', os.path.join(indir
                    ,infile) , os.path.join(outdir, infile.split
                    ('.')[0] + shape[:-4] + '.tif'), '-cutline',
                     os.path.join(basedir, shape), '-
                    crop_to_cutline', '-dstnodata', '"65535"'])
```

Listing 17: Import land cover data to the database

```
import sqlite3 as lite
import os, sys, os.path
from osgeo import gdal
import ogr

db = "F:\Diplomarbeit\Mapdata\data.db"
nutsdir = "../map_preparation/LC_nuts"

con = lite.connect(db)

with con:
    cur = con.cursor()
```

```
    for nuts in os.listdir(nutsdir):
        curnuts = os.path.join(nutsdir,nuts)
        nutsname = nuts
        if "_" in nutsname:
            nutsname = nutsname[0:nutsname.find('_')]
        for year in os.listdir(curnuts):
            print year
            yearstring = year.split('_')[1]
            src_ds = gdal.Open(os.path.join(curnuts,year))
            srcband = src_ds.GetRasterBand(1)
            arr = srcband.ReadAsArray()
            gt = src_ds.GetGeoTransform()

            yoffset = 0
            for y in arr:
                xoffset = 0
                cury = gt[3] + yoffset * gt[5]
                yoffset += 1
                for x in y:
                    curx = gt[0] + xoffset * gt[1]
                    if x < 254:
                        cur.execute("INSERT␣INTO␣landcover␣
                           VALUES('" + str(curx) + "','" + str(
                           cury) + "','" + str(x) + "','" +
                           nutsname + "','" + yearstring +"')")
                    xoffset += 1
```

Listing 18: Import NPP data to the database

```
import sqlite3 as lite
import os, sys, os.path
from osgeo import gdal
import ogr

db = "F:\Diplomarbeit\Mapdata\data.db"
nutsdir = "../map_preparation/NPP_nuts"

con = lite.connect(db)

with con:
    cur = con.cursor()
    for nuts in os.listdir(nutsdir):
        print nuts
        curnuts = os.path.join(nutsdir, nuts)
```

```
        nutsname = nuts
        if "_" in nutsname:
            nutsname = nutsname[0:nutsname.find('_')]
        for year in os.listdir(curnuts):
            if 'mean' not in year:
                print year
                yearstring = year.split('_')[2][:4]
                print yearstring
                src_ds = gdal.Open(os.path.join(curnuts, year))
                srcband = src_ds.GetRasterBand(1)
                arr = srcband.ReadAsArray()
                gt = src_ds.GetGeoTransform()

                yoffset = 0
                for y in arr:
                    xoffset = 0
                    cury = gt[3] + yoffset * gt[5]
                    yoffset += 1
                    for x in y:
                        curx = gt[0] + xoffset * gt[1]
                        if x < 65535:
                            cur.execute("INSERT INTO npp VALUES
                                ('" + str(curx) + "','" + str(
                                cury) + "','" + str(x) + "','" +
                                 nutsname + "','" + yearstring +
                                "')")
                        xoffset += 1
```

Statistical Data

These are the files to obtain statistical data from Eurostat.

Listing 19: NUTS 3 region area

```
#millions of euros
import json, urllib
import sqlite3 as lite

baseurl = "http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/
   json/en/"
query = "demo_r_d3area?precision=2&unit=KM2&time=2000&time
    =2001&time=2002&time=2003&time=2004&time=2005&time=2006&time
    =2007&time=2008&time=2009&time=2010&time=2011&time=2012&time
    =2013&time=2014&time=2015&landuse=TOTAL&geo="
```

```python
db = "D:\\backup\Diplomarbeit\Mapdata\data.db"

con = lite.connect(db)

with con:
    cur = con.cursor()
    cur.execute("SELECT name FROM nuts;")
    nutsnames = []
    for result in cur:
        nutsnames.append(result[0])
    area = {}
    for n in nutsnames:
        area[n] = {}

    for nuts in nutsnames:
        url = baseurl + query + nuts
        print nuts
        response = urllib.urlopen(url)
        data = json.loads(response.read())
        for d in range(0, 16):
            year = str(2000+d)
            value = -1
            if 'value' in data and str(d) in data['value']:
                value = str(data['value'][str(d)])
            area[nuts][year] = value

    for nuts, years in area.iteritems():
        prevval = -1
        for key, value in years.iteritems():
            if value >= 0:
                prevval = value
                break
        for key, value in years.iteritems():
            if value == -1:
                area[nuts][key] = prevval

    for nuts in nutsnames:
        years = area[nuts]
        for year, value in years.iteritems():
            if int(year) <= 2012:
                cur.execute("INSERT INTO area VALUES('" + nuts
                   + "'," + str(year) + "," + str(value) + ");"
                   )
```

```
        else:
            print("mah")
```

Listing 20: Employment rates

```
#thousands of employees
import json, urllib
import sqlite3 as lite

baseurl = "http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/
    json/en/"
query = "nama_10r_3empers?wstatus=EMP&precision=2&unit=THS&time
    =2000&time=2001&time=2002&time=2003&time=2004&time=2005&time
    =2006&time=2007&time=2008&time=2009&time=2010&time=2011&time
    =2012&geo="

db = "F:\Diplomarbeit\Mapdata\data.db"

con = lite.connect(db)

with con:
    cur = con.cursor()
    cur.execute("SELECT␣name␣FROM␣nuts;")
    nutsnames = []
    for result in cur:
        nutsnames.append(result[0])

    for nuts in nutsnames:
        url = baseurl + query + nuts
        print nuts
        response = urllib.urlopen(url)
        data = json.loads(response.read())
        for d in range(0, 13):
            year = str(2000+d)
            if str(d) in data['value']:
                value = str(data['value'][str(d)])
                cur.execute("INSERT␣INTO␣employment␣VALUES('" +
                    nuts + "'," + year + "," + value + ");")
            else:
                cur.execute("INSERT␣INTO␣employment␣VALUES('" +
                    nuts + "'," + year + "," + str(-1) + ");")
```

Listing 21: GDP per capita

```
#millions of euros
```

```
import json, urllib
import sqlite3 as lite

baseurl = "http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/
    json/en/"
query = "nama_10r_3gdp?unit=EUR_HAB&filterNonGeo=1&time=2000&
    time=2001&time=2002&time=2003&time=2004&time=2005&time=2006&
    time=2007&time=2008&time=2009&time=2010&time=2011&time=2012&
    precision=2&geo="

db = "F:\Diplomarbeit\Mapdata\data.db"

con = lite.connect(db)

with con:
    cur = con.cursor()
    cur.execute("SELECT name FROM nuts;")
    nutsnames = []
    for result in cur:
        nutsnames.append(result[0])

    for nuts in nutsnames:
        url = baseurl + query + nuts
        print nuts
        response = urllib.urlopen(url)
        data = json.loads(response.read())
        for d in range(0, 13):
            year = str(2000+d)
            value = str(data['value'][str(d)])
            cur.execute("INSERT INTO gdp VALUES('" + nuts + "',
                " + year + "," + value + ");")
```

Listing 22: Population density

```
#millions of euros
import json, urllib
import sqlite3 as lite

baseurl = "http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/
    json/en/"
query = "nama_10r_3popgdp?unit=THS&time=2000&time=2001&time
    =2002&time=2003&time=2004&time=2005&time=2006&time=2007&time
    =2008&time=2009&time=2010&time=2011&time=2012&precision=2&
    geo="
```

```
db = "D:\\backup\Diplomarbeit\Mapdata\data.db"

con = lite.connect(db)

with con:
    cur = con.cursor()
    cur.execute("SELECT name FROM nuts;")
    nutsnames = []
    for result in cur:
        nutsnames.append(result[0])

    for nuts in nutsnames:
        url = baseurl + query + nuts
        print nuts
        print url
        response = urllib.urlopen(url)
        data = json.loads(response.read())
        for d in range(0, 13):
            year = str(2000+d)
            value = str(data['value'][str(d)])
            cur.execute("INSERT INTO population VALUES('" +
               nuts + "'," + year + "," + value + ");")
```

Listing 23: Road density

```
#tran_r_net
#km/thousand sq km
import json, urllib
import sqlite3 as lite

baseurl = "http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/
   json/en/"
query = "tran_r_net?precision=2&tra_infr=MWAY&unit=KM_TKM2&time
   =2000&time=2001&time=2002&time=2003&time=2004&time=2005&time
   =2006&time=2007&time=2008&time=2009&time=2010&time=2011&time
   =2012&geo="


db = "D:\\backup\Diplomarbeit\Mapdata\data.db"

con = lite.connect(db)

with con:
```

```
cur = con.cursor()
cur.execute("SELECT␣name␣FROM␣nuts;")
nutsnames = []
nutslongnames = []
for result in cur:
    nutsnames.append(result[0][:4])
    nutslongnames.append(result[0])
nuts2names = list(set(nutsnames))
road_dens = {}
for n in nuts2names:
    road_dens[n] = {}

for nuts in nuts2names:
    url = baseurl + query + nuts
    print nuts
    response = urllib.urlopen(url)
    data = json.loads(response.read())
    for d in range(0, 13):
        year = str(2000+d)
        value = -1
        if 'value' in data and str(d) in data['value']:
            value = str(data['value'][str(d)])
        road_dens[nuts][year] = value
for nuts, years in road_dens.iteritems():
    prevval = -1
    for key, value in years.iteritems():
        if value >= 0:
            prevval = value
            break
    for key, value in years.iteritems():
        if value == -1:
            road_dens[nuts][key] = prevval

for nuts in nutslongnames:
    ntwo = nuts[:4]
    years = road_dens[ntwo]
    for year, value in years.iteritems():
        print nuts + "␣" + year + "␣" + value
        cur.execute("INSERT␣INTO␣road_dens␣VALUES('" + nuts
            + "'," + year + "," + value + ");")
```

Data export

Following is the code to prepare a csv file for use with R:

Listing 24: Csv export

```
import sqlite3 as lite
import unicodecsv

db = "D:\\backup\Diplomarbeit\Mapdata\data.db"
con = lite.connect(db)
outputprefix = "D:/backup/Diplomarbeit/regressionfiles/"

with con:
    cur_list = con.cursor()
    cur_joined = con.cursor()
    cur_lc_first = con.cursor()

    cur_list.execute("SELECT␣nuts1,␣nuts2,␣nuts_part_1.country␣
       nuts1_country,␣nuts_part_2.country␣nuts2_country,␣
       countries_2.joined␣from␣touches␣join␣nuts_part␣
       nuts_part_1␣on␣nuts_part_1.nuts␣=␣touches.nuts1␣join␣
       nuts_part␣nuts_part_2␣on␣nuts_part_2.nuts␣=␣touches.
       nuts2␣join␣countries␣countries_1␣on␣countries_1.name␣=␣
       nuts_part_1.country␣join␣countries␣countries_2␣on␣
       countries_2.name␣=␣nuts_part_2.country␣WHERE␣
       nuts1_country␣<>␣nuts2_country␣and␣countries_1.joined␣<␣
       2004")

    with open(outputprefix + '
       combination_no_and_normalize_pop_road_area.csv', 'w') as
        csvfile:
        fieldnames = ['year', 'country1', 'nuts1', 'npp1', '
           npp1_norm', 'urbanization1', 'agriculture1', '
           forest1', 'other1', 'urbanization1_norm', '
           agriculture1_norm', 'forest1_norm', 'other1_norm', '
           gdp1', 'gdp1_norm', 'population1', 'road_density1',
           'area1',
                     'country2', 'nuts2', 'npp2', 'npp2_norm',
                        'urbanization2', 'agriculture2', '
                       forest2', 'other2', '
                       urbanization2_norm', '
                       agriculture2_norm', 'forest2_norm', '
                       other2_norm', 'gdp2', 'gdp2_norm', '
                       population2', 'road_density2', 'area2'
```

```
                  ]
writer = unicodecsv.DictWriter(csvfile, fieldnames=
   fieldnames, dialect=unicodecsv.excel_tab,
                                lineterminator='\n')
writer.writeheader()
for line in cur_list:
    nuts1 = line[0]
    nuts2 = line[1]

    print "Working␣on:␣" + nuts1 + "␣and␣" + nuts2

    country1 = line[2]
    country2 = line[3]

    cur_npp_1 = con.cursor()
    cur_gdp_1 = con.cursor()
    cur_pop_1 = con.cursor()
    cur_road_1 = con.cursor()
    cur_area_1 = con.cursor()

    cur_npp_2 = con.cursor()
    cur_gdp_2 = con.cursor()
    cur_pop_2 = con.cursor()
    cur_road_2 = con.cursor()
    cur_area_2 = con.cursor()

    cur_npp_1.execute("SELECT␣avg(value)␣FROM␣npp␣WHERE
       ␣nuts='" + nuts1 + "'␣GROUP␣BY␣year␣ORDER␣BY␣
       year␣ASC;")
    cur_gdp_1.execute("SELECT␣year,␣value␣FROM␣gdp␣
       WHERE␣nuts='" + nuts1 + "'␣ORDER␣BY␣year␣ASC;")
    cur_pop_1.execute("SELECT␣year,␣value␣FROM␣
       population␣WHERE␣nuts='" + nuts1 + "'␣ORDER␣BY␣
       year␣ASC;")
    cur_road_1.execute("SELECT␣year,␣value␣FROM␣
       road_dens␣WHERE␣nuts='" + nuts1 + "'␣ORDER␣BY␣
       year␣ASC;")
    cur_area_1.execute("SELECT␣year,␣value␣FROM␣area␣
       WHERE␣nuts='" + nuts1 + "'␣ORDER␣BY␣year␣ASC;")

    cur_npp_2.execute("SELECT␣avg(value)␣FROM␣npp␣WHERE
       ␣nuts='" + nuts2 + "'␣GROUP␣BY␣year␣ORDER␣BY␣
       year␣ASC;")
```

```
cur_gdp_2.execute("SELECT year, value FROM gdp 
    WHERE nuts='" + nuts2 + "' ORDER BY year ASC;")
cur_pop_2.execute("SELECT year, value FROM 
    population WHERE nuts='" + nuts2 + "' ORDER BY 
    year ASC;")
cur_road_2.execute("SELECT year, value FROM 
    road_dens WHERE nuts='" + nuts2 + "' ORDER BY 
    year ASC;")
cur_area_2.execute("SELECT year, value FROM area 
    WHERE nuts='" + nuts2 + "' ORDER BY year ASC;")


cur_npp_norm_1 = con.cursor()
cur_gdp_norm_1 = con.cursor()

cur_npp_norm_2 = con.cursor()
cur_gdp_norm_2 = con.cursor()

cur_npp_norm_1.execute("SELECT avg(value) FROM 
    npp_normalized WHERE nuts='" + nuts1 + "' GROUP 
    BY year ORDER BY year ASC;")
cur_gdp_norm_1.execute("SELECT year, value FROM 
    gdp_normalized WHERE nuts='" + nuts1 + "' ORDER 
    BY year ASC;")

cur_npp_norm_2.execute("SELECT avg(value) FROM 
    npp_normalized WHERE nuts='" + nuts2 + "' GROUP 
    BY year ORDER BY year ASC;")
cur_gdp_norm_2.execute("SELECT year, value FROM 
    gdp_normalized WHERE nuts='" + nuts2 + "' ORDER 
    BY year ASC;")



forest_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
     0.0, 0.0, 0.0, 0.0, 0.0]
agriculture_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
     0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
urban_1 = [0.0]
other_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0]
```

```
forest_norm_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
agriculture_norm_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
urban_norm_1 = [0.0]
other_norm_1 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
   0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
for y in range(2001, 2013):
    totalcount = float(cur_lc_first.execute(
        "SELECT count(type) from landcover where 
           nuts='" + nuts1 + "' and year=" + str(
            y) + " and type >= 1 and type <= 16").
               next()[0])

    typecounts = cur_lc_first.execute(
        "SELECT count(type), type from landcover 
           where nuts='" + nuts1 + "' and year=" +
           str(
            y) + " GROUP BY type")
    for count in typecounts:
        if count[1] >= 1 and count[1] <= 5:
            forest_1[y - 2000] += float(count[0])
            forest_norm_1[y - 2000] += float(count
               [0]) / totalcount
        if count[1] == 12 or count[1] == 14:
            agriculture_1[y - 2000] += float(count
               [0])
            agriculture_norm_1[y - 2000] += float(
               count[0]) / totalcount
        if count[1] == 13:
            urban_1.append(float(count[0]))
            urban_norm_1.append(float(count[0]) /
               totalcount)
        if count[1] >= 6 and count[1] <= 16 and (
           count[1] != 12 or count[1] != 13 or
           count[1] != 14):
            other_1[y - 2000] += float(count[0])
            other_norm_1[y - 2000] += float(count
               [0]) / totalcount
    forest_1[0] = forest_1[1]
    agriculture_1[0] = agriculture_1[1]
    urban_1[0] = urban_1[1]
    other_1[0] = other_1[1]
```

```
    forest_norm_1[0] = forest_norm_1[1]
    agriculture_norm_1[0] = agriculture_norm_1[1]
    urban_norm_1[0] = urban_norm_1[1]
    other_norm_1[0] = other_norm_1[1]

forest_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0]
agriculture_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
urban_2 = [0.0]
other_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
   0.0, 0.0, 0.0, 0.0, 0.0]

forest_norm_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
agriculture_norm_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
    0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
urban_norm_2 = [0.0]
other_norm_2 = [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
   0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
for y in range(2001, 2013):
    totalcount = float(cur_lc_first.execute(
        "SELECT count(type) from landcover where
           nuts='" + nuts2 + "' and year=" + str(
            y) + " and type >= 1 and type <= 16").
               next()[0])

    typecounts = cur_lc_first.execute(
        "SELECT count(type), type from landcover
           where nuts='" + nuts2 + "' and year=" +
           str(
            y) + " GROUP BY type")
    for count in typecounts:
        if count[1] >= 1 and count[1] <= 5:
            forest_2[y - 2000] += float(count[0])
            forest_norm_2[y - 2000] += float(count
               [0]) / totalcount
        if count[1] == 12 or count[1] == 14:
            agriculture_2[y - 2000] += float(count
               [0])
            agriculture_norm_2[y - 2000] += float(
               count[0]) / totalcount
```

```
            if count[1] == 13:
                urban_2.append(float(count[0]))
                urban_norm_2.append(float(count[0]) /
                    totalcount)
            if count[1] >= 6 and count[1] <= 16 and (
                count[1] != 12 or count[1] != 13 or
                count[1] != 14):
                other_2[y - 2000] += float(count[0])
                other_norm_2[y - 2000] += float(count
                    [0]) / totalcount
        forest_2[0] = forest_2[1]
        agriculture_2[0] = agriculture_2[1]
        urban_2[0] = urban_2[1]
        other_2[0] = other_2[1]

        forest_norm_2[0] = forest_norm_2[1]
        agriculture_norm_2[0] = agriculture_norm_2[1]
        urban_norm_2[0] = urban_norm_2[1]
        other_norm_2[0] = other_norm_2[1]

    for line in range(2000, 2013):
        npp_1 = cur_npp_1.next()[0]
        gdp_1 = cur_gdp_1.next()[1]
        pop_1 = cur_pop_1.next()[1]
        road_1 = cur_road_1.next()[1]
        area_1 = cur_area_1.next()[1]
        npp_norm_1 = cur_npp_norm_1.next()[0]
        gdp_norm_1 = cur_gdp_norm_1.next()[1]
        npp_2 = cur_npp_2.next()[0]
        gdp_2 = cur_gdp_2.next()[1]
        pop_2 = cur_pop_2.next()[1]
        road_2 = cur_road_2.next()[1]
        area_2 = cur_area_2.next()[1]
        npp_norm_2 = cur_npp_norm_2.next()[0]
        gdp_norm_2 = cur_gdp_norm_2.next()[1]

        row = {'year': line, 'country1': country1, '
            nuts1': nuts1, 'npp1': npp_1, 'npp1_norm':
            npp_norm_1, 'urbanization1': urban_1[line
            -2000], 'agriculture1': agriculture_1[line
            -2000], 'forest1': forest_1[line-2000], '
            other1': other_1[line-2000], '
            urbanization1_norm': urban_norm_1[line
```

```
            -2000], 'agriculture1_norm':
            agriculture_norm_1[line -2000], 'forest1_norm
            ': forest_norm_1[line -2000], 'other1_norm':
            other_norm_1[line -2000], 'gdp1': gdp_1, '
            gdp1_norm': gdp_norm_1, 'population1': pop_1
            , 'road_density1': road_1, 'area1': area_1,
            'country2': country2, 'nuts2': nuts2, 'npp2'
            : npp_2, 'npp2_norm': npp_norm_2, '
            urbanization2': urban_2[line -2000], '
            agriculture2': agriculture_2[line -2000], '
            forest2': forest_2[line -2000], 'other2':
            other_2[line -2000], 'urbanization2_norm':
            urban_norm_2[line -2000], 'agriculture2_norm'
            : agriculture_norm_2[line -2000], '
            forest2_norm': forest_norm_2[line -2000], '
            other2_norm': other_norm_2[line -2000], 'gdp2
            ': gdp_2, 'gdp2_norm': gdp_norm_2, '
            population2': pop_2, 'road_density2': road_2
            , 'area2': area_2}
        writer.writerow(row)
print "Finished working on: " + nuts1 + " and " +
    nuts2
```