Deliverable 2.2

What do night satellite images and small-scale grid data tell us about functional changes in the rural-urban environment and the economy?

Case studies Frankfurt-Rhein/Main and Ljubljana Urban Region

Rolf Bergs

PRAC, Bad Soden

Moneim Issa

PRAC, Bad Soden

PRAC - Bergs & Issa Partnership Co. , Im Hopfengarten 19b, D-65812 Bad Soden, Germany

October 2018

ROBUST receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727988.

Table of Contents

Table of figures

List of tables

List of abbreviations

CORINE	Coordination of Information on the Environment
DMSP-OLS	Defence Meteorological Satellite Program – Operational Line Scanner
DN	Digital number
EEA	European Environmental Agency
GDP	Gross Domestic Product
GIS	Geographical Information System
GRUMP	Global Rural Urban Mapping
LED	Light emitting diode
LUR	Ljubljana Urban Region
MAUP	Modifiable Areal Unit Problem
NaN	Not a Number
NOAA	National Oceanic and Atmospheric Administration
NUTS	Nomenclature des unités territoriales statistiques
OECD	Organization of Economic Co-operation and Development
OLS	Ordinary least squares
QGIS	Quantum GIS
RVFRM	Regionalverband Frankfurt/Rhein-Main (Regional municipal association Frankfurt/Rhein-Main)
TRL	Technology readiness level
VIIRS	Visible Infrared Imaging Radiometer Suite

Summary

The study at hand addresses the use of nocturnal satellite imagery in spatial analysis of rural-urban relations. It shows how a novel approach can be used to complement the prevalent traditional survey methods below the data resolution level of official regional statistics. The actual intention is to examine a method to analyse the spatial evolution of rural–peri-urban–urban settings over time and to open opportunities to base local policy and planning decisions on a more precise empirical foundation. A better informational foundation of rural-urban planning and policies essentially stipulates a more precise empirical insight, both related to spatial (rural-urban) development processes as well as to a more precise functional differentiation of space. A sufficiently high spatial resolution of data is an important precondition for that. More importantly, we explore the information content of recent issues of nocturnal satellite imagery (VIIRS) in terms of spatial segmentation (urban, non-urban), the association with socio-economic variables at small spatial scale and the change of spatial dependence (relational space). This approach of analysis essentially combines spatial heterogeneity and spatial dependence as effects of different orders. The overarching aim is to contribute with a novel database and different empirical tools to broaden spatial information for decision making in policy and planning at small spatial scale among municipalities. With other words, the accuracy of spatial information is expected to be substantially enhanced, paving ways for better rural-urban planning coordination and synergies.

The core empirical study covers the area of the Regionalverband Frankfurt/Rhein-Main (RVFRM) as an established European growth pole. A second empirical study looks at a different spatial setting, namely the Ljubljana Urban Region, as a dynamically growing economy with important national relevance for a smaller EU member state.

Keywords: Satellite imagery; spatial heterogeneity; spatial dependence; natural (functional) cities; rural-urban synergies; Zipf's law

JEL-Classification: O 13; O 18; R 12

1 Introduction

Territorial space and spatial change are determined by natural and social factors. Natural factors comprise natural resources, climate, topographical relief, accessibility, or natural disasters like floods or earthquakes. Social factors comprise push and pull factors that imply migration and the evolution of markets for goods, services and factors, hence places where people exchange physically or virtually. Space at any scale is thus heterogeneous and functionally individual; synergic spatial functions can complement each other at the level of a region and foster sustainable growth. The functional difference between urban and rural space and the continuum in between is an important example for observation of space and its change. Traditionally, observation can be done by using socio-economic and environmental data available and to explore certain patterns quantitatively, or it can be done by action research (e.g. a qualitative monitoring by dialogue with people). While available official data are usually provided at a rather high aggregation level (NUTS 2, NUTS 3) and thus do not allow a sufficiently differentiated view and focus, qualitative research at grassroots level is mostly burdensome and always dependent of people's perceptions that do not necessarily reflect the truth. Hence, both threads of empirical research are affected by important constraints. In this paper we explore a more novel source of information to shed light on rural and urban space and the related spatial change. Satellite imagery, used in this paper, is essentially a database not elaborated and fed by people but rather automatically generated by sensor technology. The images are digital and thus represent a transformable numerical matrix database. Further to that, the entire space of habitable climate zones is recorded, containing important information of space and spatial change with a high resolution. Of course, this should not suggest an interpretation that sensor generated data are systematically faultless (see further below).

Related to methodologies and analytical instruments, in Work Package 2 of the ROBUST project (Deliverable D 2.2) one task is to pilot the use of a new approach in spatial analysis in two or three case study areas. The actual intention is to examine a method to analyse the spatial evolution of rural–peri-urban–urban settings over time and to open opportunities to base policy and planning decisions on a more precise empirical foundation.1 It is envisaged to test the usefulness of such novel data in detecting rural-urban heterogeneity, respective developments and spatial dependencies. A more precise empirical foundation is essentially related to a sufficiently high spatial resolution of data. Below NUTS 2 level, the availability of socio-economic and environmental data has been limited, but even at NUTS 3 level (e.g. district) the true variation of variables might be largely concealed by mean values in the official statistics. This might hamper prudent planning and policies at the level of neighborhoods that are basic elements of any area subject to regional planning or local policies (demographic change, communication and transport, open space etc.). The use of nocturnal satellite imagery in viewing the distribution of functional space in terms of light emission and its spatial distribution is definitively a novel and yet still largely under-researched field of regional science. In this study we do not only regard light emission as a small-scale grid variable, it is moreover aimed to merge this novel dataset with other novel grid data, such as CORINE land use data or research databases such as the IÖR-Monitor. In the second case study dealing with socio-economic change at small spatial scale, the VIIRS data are also merged with the grid data from Microm®.

[1] "Continuing studies of the dynamics of urbanization through time using nighttime lights enables the comprehension of adverse effects of urbanization related to loss of vegetation, increase in temperatures, air and water pollution, loss of species' habitats, growth and spread of urban slums, poverty, and unemployment. Nighttime light images have also informed studies of urban dynamics ". (cf. Ghosh et al. 2013)

The intention of this study is a pure experimental one. It is aimed to explore the potential information content of nocturnal satellite imagery for the analysis of rural-urban heterogeneity and dependence rather than to establish empirical facts that may guide future policy and planning decisions in the two case study regions covered.

It has been shown that nocturnal satellite images contain important ecological and socio-economic information (see the literature review further below). Recent research shows that the distribution and value of pixels in nocturnal satellite imagery (DN or radiance[2]), such as DMSP-OLS since 1992 and notably the monthly composites from the Visible Infrared Imaging Radiometer Suite (VIIRS[3]) have been closely associated with socio-economic and environmental variables at global scales (Proville et al. 2016). Thus, social, economic and environmental patterns of space appear well represented by light emission and suggest those images to be a potentially meaningful database for the purpose of ROBUST. The important advantage of such digital images has been the property of a numerically transformable matrix.

In this study, we explore the information content of recent issues of nocturnal satellite imagery (VIIRS) in terms of spatial segmentation (urban, non-urban), the association with socio-economic variables at small spatial scale and the change of spatial dependence (relational space). This approach of analysis essentially combines spatial heterogeneity and spatial dependence as effects of different orders[4]. Heterogeneity and dependence can be conceived as stand-alone categories to describe and analyse space. They are both results of processes; while heterogeneity is the result of a deterministic process and cause of a further continuing process, spatial dependence is rather a stochastic process in which spatial variables in one territorial area are influenced from other contiguous or distant areas. The overarching aim is to contribute with a tool to broaden spatial information for decision making in policy and planning at small spatial scale among municipalities. With other words, the resolution of spatial information is expected to be substantially enhanced, paving ways for better rural-urban planning coordination and synergies. **This is the wider theoretical purpose, why to explore the information content of night satellite images.**

Improved spatial information is an important value added for prudent policy and planning decisions; it might notably help to relieve conflict of interest over planning and policy objectives. Especially along the urban fringe conflicts of interest may materialize more pronounced than elsewhere:

"The rural-urban fringe is the countryside surrounding the edge of a city's built-up area. It is a transition zone between the urban environment of the city and areas that are predominantly rural in character. In between these two extremes there is a gradual change in the pattern of land use, with urban settlements and

---

2 DN means the digital number. In night satellite images it is a value (integer) assigned to a pixel. It depends on the quantization of the sensor. The lower the quantization, the lower is the possible differentiation. The simple DMSP stable lights have a 6-bit quantization, thus showing pixels in the range between zero and 63, while 14-bit images are in a range between zero and 16,284. Radiance is defined as the electrical power (Watt) per steradian per area unit ($nWcm^{-2}\ sr^{-1}$). In the International System of Units (SI) the steradian is the unit of the solid angle. Radiance is derived by the raw DN.

3 The Visible Infrared Imaging Radiometer Suite (VIIRS) is a sensor designed and manufactured by the Raytheon Company (Waltham, Mass.). Since 2011 it operates on the Suomi National Polar-orbiting Partnership (Suomi NPP) weather satellite and has replaced and improved upon older sensing technology. VIIRS provides unique data for the monitoring of global weather patterns and other predictive information critical to industries as diverse as agriculture and transportation, insurance and energy (cf.: https://www.raytheon.com/capabilities/products/viirs/ .

4 Jiang (2015) argues that "…Spatial heterogeneity is a kind of hidden order, which appears disordered on the surface, but possesses a deep order beneath. This kind of hidden order can be characterized by a power law or a heavy-tailed distribution in general. …". So if just viewing the visible heterogeneity at the surface, then the dependence between spatial items is a higher order effect, but if looking at underlying regular patterns of heterogeneity, spatial dependence constitutes a lower (first) order property.

functions giving way to agricultural activities and open space. It is best thought of as a 'continuum', in which the influence of the city declines as distance from it increases. There is a corresponding fall in population density away from the city, while the number and range of services available are also reduced. If the city is growing, the rural-urban fringe will change over time and land use patterns within it will reflect the greater influence of the city." (Morrish 2015)

Those urban fringe zones have important ecological functions as they often represent green belts of a city. Green belts have a well-defined function, namely to avoid urban sprawl, to prevent neighbouring towns from merging, to provide for recreation, to maintain fresh air passage and to assist in urban regeneration. However, there are common misconceptions about green belts, a major source of conflict over land use. Those areas are never immune to soil sealing and development. Revisions of land use plans and expropriation procedures are often sources of conflict (Morrish 2015). Another example of conflict to imagine could be a booming city with a high in-migration pressure, tempted to develop residual open space at the outskirts and hence to destroy natural, recreational or agricultural space within the city. Farmers and dwellers of that city area might have no important voice in planning and policy decisions of the city council. Only in neighbouring rural (peri-urban) villages those dwellers and farmers might probably find partners with congruent interests. Such kind of conflict is eventually the result of historically defined boundaries.

In a comprehensive research paper Caffyn and Dahlström (2005) sketched out the rural-urban interdependencies, but they did not just look at the normative part of cooperation, they also highlight the potential conflicts among stakeholders with different interests, either from the rural or the urban viewpoint. They also make clear, that there are such fuzzy spatial boundaries that often make it difficult for rural or urban stakeholders to have clear-cut positions to be negotiated in policy and planning cooperation. Fuzzy spatial boundaries are nothing than the outcome of historically defined administrative boundaries that do not anymore coincide with the functional extent of urban and non-urban space.

Repp et al. 2012 report that 'knowledge gaps' can be found especially with respect to particularly complex rural-urban interrelations. By means of a matrix illustration, 'governance gaps' are revealed, especially with regard to energy and material flows, action spaces as well as knowledge flows and innovation processes in urban-rural contexts. A fully homogenous rural or urban area is quite rare. Since the major parameter of inter-municipal policy coordination is the fixed administrative boundary, frictions are a normal consequence. More insight into the true differentiation of space could contribute to relieving such conflicts and to make policy and planning coordination more effective.

Apart from a more precise differentiation of space independent of administrative boundaries, we have to acknowledge that the evolution of neighborhoods within a larger region has never been autonomous but always been influenced by effects, declining in intensity from contiguous to distant space. Hence, if a rural village with solely agricultural production suffers from air pollution and a high prevalence of respiratory illness it can be hardly explained by predictors from within the municipality but rather by influences perturbing from contiguous or even distant municipalities. This spatial dependence is very much existent at neighborhood level. It can be negative or positive (e.g. a neighborhood close to a wealthy neighborhood is more likely to be wealthy too - rather than poor). Since light emission is closely associated with economic activity, energy consumption and carbon emissions it is possible to (i) shed light on small scale spatial dependencies and (ii) to explore the association between light emission and socio-economic or environmental variables at different spatial types of size and land use.

In the context of the ROBUST deliverable 1.1, Michael Woods and Jesse Haley (2017) have already sketched out some features of this novel thread of geo-spatial analysis including its possible limits and shortcomings.

They notably refer to the Global Rural Urban Mapping (GRUMP) project and some studies using the regular DMSP-OLS[5] images with 6 bit quantization (a variation limited to the range between 0 to < 2^6). In fact, a pixel variation in the range between DN 0 and 63 is insufficient for a precise segmentation of urban and rural space, even though some critical statements cited, such as Pritchard (2017), who deplores the implication of a moral geography when distinguishing between white (positive) and dark (negative) areas is also hardly convincing.

Limitations of spatial analysis based on nocturnal satellite imagery are much more related to the specific quality and information content of the images. Especially in highly urbanized countries, a right-censored database, such as the simple DMSP-OLS composites since 1992, is incapable to capture the true variance of luminosity in urbanized countries and to precisely distinguish between natural urban and non-urban space. In many cases, even rural areas are masked by top-coded pixels (e.g. in urbanized countries such as the UK, Belgium or the Netherlands)[6]. To overcome this issue, the National Oceanic and Atmospheric Administration (NOAA) has also produced a series of so-called radiance-calibrated images, but only for irregular periods between 1996 and 2011. The radiance-calibrated composites are based on a limited set of observations where the gain of the detector was set much lower than its typical operational setting; this was possible for some years only. The combination of those sparsely acquirable data at low gain settings with the operational data retrieved at high gain settings made it at least possible to produce a set of global nighttime lights products with no sensor saturation. However, the pixel values of those radiance-calibrated images are neither digital numbers (DN) nor radiance ($nWcm^{-2}\ sr^{-1}$); they are deemed unitless because of the lacking on-flight calibration. The images thus show the true variation of luminosity but not the absolute radiance[7]. The information content of those images for the analysis of urban extent is substantially stronger than that of the regular DMSP-OLS stable lights composites (cf. Bergs 2018).

Since 2012, the 14-bit onboard-calibrated VIIRS composites (Visible Infrared Imaging Radiometer Suite) have been available (cf. footnote above); these images will be used for the study at hand. VIIRS composites are much more precise and actual; the resolution is considerably higher and error stemming from blooming effects is already removed to a large extent in the raw images. The extremely high variation of detected radiance (between 0 and up to five-digit values) enables us to much better differentiate between natural urban extent outside the administrative city boundaries and possible rural space inside a city. A differentiated (dynamic) analysis of those distributional patterns might offer an important information base for policy and planning decisions. The only disadvantage of using the VIIRS database is the fact that a long-run analysis of urban and rural evolution is not yet possible. A direct comparison between VIIRS and the older radiance-calibrated images is hampered by the different technologies and calibration approaches; it is only possible to roughly estimate the change rate of light emission over the timespan between 1996 and 2017.

---

5 Satellite images provided by the Defense Meteorological Satellite Program (DMSP) based on the Operational Linescan System (OLS).

6 There have been three usable datasets from NASA: Stable lights ("Average Visible, Stable Lights, & Cloud Free Coverages"), unfiltered lights and radiance calibrated lights. Most papers work with the stable lights as they show the luminosity of cities and towns and are available for a period of 22 years, between 1992 and 2013. The unfiltered lights are available for the same period and can potentially give more values at low luminosity, which is helpful for studies in areas with low levels of development, but it is often hard to disentangle it from natural illumination of the soil. Radiance calibrated lights give more variation at high luminosity levels, as they are not top coded. Radiance calibrated lights are available for a number or years, namely 1996, 1999, 2000, 2002, 2004, 2005 and 2010. (cf. Jamila Nigmatulina: Nighttime Lights - How are they useful?
https://urbanisation.econ.ox.ac.uk/blog/nighttime-lights-how-are-they-useful-dzhamilya-nigmatulina)

7 https://ngdc.noaa.gov/eog/dmsp/radcal_readme.txt

The following overview shows the different products of night satellite imagery provided by the NOAA.

Table 1: Nocturnal satellite imagery provided by the National Oceanic and Atmospheric Administration

Product	Availability	Unit	Range	Product differentiation
DMSP-OLS (annual composites)	1992-2012	DN	6-bit	(i) Cloud-free coverage, (ii) Stable lights, (iii)Average visible lights and (iv)Average lights multiplied by percent frequency of light detection
DMSP-OLS (radiance-calibrated composites)	1996, 1999, 2000, 2002, 2004, 2005, 2010/11	Unit-free	infinite	none
VIIRS	Monthly since 2012	Radiance ($nWcm^{-2}$ sr^{-1})	14-bit	none

Source of information: NOAA

By using the new VIIRS composites there are three major analytical components we address to explore urban and rural space and their inter-dependencies, briefly sketched out in this introduction (for more technical details see chapter 4 further below). The first analytic component describes the exploration of spatial heterogeneity with a view to detect the natural (functional) extent of urban and rural space and to compare it with historically/ administratively delineated space (municipal boundaries). The purpose of that is to bypass the „modifiable areal unit problem", such that natural urban space evolves secularly while administrative boundaries may either remain fixed or change any time by territorial reforms with far reaching implications on regional statistics and administrative representation of urban and rural areas (see below).

It is to be noted that the simple differentiation between urban and non-urban should not at all suggest a strictly binary definition of space, such that there were exclusively rural and urban areas. In fact, the simple differentiation of urban and non-urban has a technical reason and should help to determine the upper functional segment, namely the urban space. Anything not belonging to urban space is defined within the continuum of peri-urban, rural and untenantable space. The problem in understanding this continuum in quantitative terms is simply the continuum itself. The scale of it is continuous and there is essentially no discrete differentiation, hence any differentiation induced by geo-statistical analysis would be artificial and arbitrary.

The second analytic component comprises the exploration of spatial dependence with a view to observe relational space at small scale in terms of light emission over time. Due to the strong association between radiance and several socio-economic and environmental variables it is expected to obtain insight from different viewpoints how urban and non-urban (notably rural) units interact socio-economically and environmentally.

The third analytic component is related to the empirical association between radiance and socio-economic and environmental variables at small spatial scale (grid data). The purpose of this more experimental analysis is to shed light on strength of correlation for different spatial categories (i.e. urban, rural, mixed, peri-urban) and to recognize possible systematic differences. But apart from the association with socio-economic and environmental variables, light emission is an important environmental variable itself. Light pollution is directly associated with public health and with its negative impact on flora and fauna. There is e.g. a direct relationship between light pollution and insects' demise that might merit more prominent notice. **In the end, this is the practical purpose, why to explore the information content of night satellite images.**

The core empirical study will address the area of the Regionalverband Frankfurt/Rhein-Main (RVFRM) as an established European growth pole. A second empirical study will look at a different spatial setting. We selected the Ljubljana Urban Region as a dynamically growing economy with important national relevance for a smaller EU member state.

The study is structured along the following sections:

In chapter 2 the recent and seminal literature of the specific economic, geographic, political science and sociological realms of remote sensing is reviewed. Chapter 3 addresses theoretical underpinnings and expected contribution to new insight, while chapter 4 deals with data and methodology. The two case studies are dealt with in chapters 5 and 6. A discussion and conclusion is subject to chapter 7.

2 Literature on night satellite imagery in regional science

So far, numerous international papers have been published dealing with the association between light emission and various socio-economic and environmental variables such as GDP, pollution, population density, infant mortality, poverty etc. at global and national scales (Henderson et al. 2011; Chen and Nordhaus 2011, Ghosh et al. 2010; Mellander et al. 2013; Sutton 1998, Amaral et al. 2006, Bagan et al. 2015, Chen 2015, Proville et al. 2017). Another thread of analysis is Pinkovskiy (2013) who explores discontinuities of the political economy along national borders, thus looking at spatial change in borderland regions, which are sub-national and international at the same time. More importantly, border discontinuities are supposed to shed light on the impact of the political economy on economic activity because these determinants are directly influenced by government activity and might become visible in areas with a politically determined discontinuity (the border). A well-known example is the harsh discontinuity of lights along the intra-Korean border, well recognizable even by visual inspection alone. Night satellite imagery has been also used from other viewpoints of political science. One interesting aspect has been the analysis of democracy in developing countries and its relationship with the provision of electricity that can be demonstrated by the distribution of light emission (cf. Min et al. 2013).

Apart from such socio-economic studies there are various publications based on night satellite images, dealing with classification of space. These are mostly from geo-spatial analysis and remote sensing. This thread of research is relevant when empirically dealing with administratively defined urban, peri-urban and rural space, such as the case study regions of the ROBUST project. In fact, results of analyses dealing with official spatial datasets are always affected by historically grown, but artificially determined territorial boundaries. Research within the ROBUST project is affected by that issue as well because rural-urban interaction is predominantly an administrative interaction via planning and local policies. Administratively

delineated boundaries are thus a central parameter in that kind of administrative interactions. However, do boundaries really coincide with the true rural, peri-urban and urban extent? Isn't there a mismatch between arbitrarily defined boundaries and the real physiognomy of space? Boundaries defined are (i) invisible, (ii) non-natural and (iii) can be moved at any time by political decision. More importantly, any such decision does not have an influence on the individual but it may have a major one on the aggregate, so that regional statistics are always at the risk to be distorted by modified areal units. The issue addressed here is called the "Modifiable Areal Unit Problem" (MAUP); its academic debate dates back to the 1930s (Biehl and Gehlke 1934) and its political debate is closely linked to the Functional Area approach of the OECD (OECD 2015). Likewise, the definition of what is rural, urban or peri-urban is arbitrary and also closely related to the MAUP. Any administrative definitions (e.g. Eurostat) or academic classifications (cf. Woods and Haley 2017, pp. 10 ff.) of such spatial categories, and there are many of them, are derived from different concepts and point of views, ranging from population density, share of the agricultural sector, to land use. What is urban in one concept may be categorized as rural in another one. Further to that, the continuum between urban and rural is not constant. Over time it moves continuously, while boundaries are moved discretionary, if at all.

To avoid such distortions, Jiang and Jia (2011) have first adopted the concept of "natural cities" for the United States. Those cities are not defined by administrative boundaries, but rather by the autonomous evolution of settlement and economic activity. Later, Jiang and Miao (2014) also defined the term in the context of social media users' locations that allow insights similar to those of night lights, also with respect to Zipf's law (Wu 2015).[8]

Meanwhile there is abundant research using nocturnal satellite imagery as a database. This also comprises the analysis of spatial change, or more specifically the urban and rural evolution from the viewpoint of socio-economic and environmental development.

Proville et al. (2017) address the relationship between light emission and socio-economic variables from a global viewpoint. It is a longitudinal analysis supporting several former studies exploring the usefulness of such images for the monitoring of regional economic and environmental development. The authors, however, found significant temporal and spatial effects implying inter-regional variation of correlation. Heterogeneity across regions is likely explained different forms of governance and global economic cycles. More specific former studies have been Henderson et al. (2011), Chen and Nordhaus (2011), Mellander et al. (2013) Nordhaus and Chen (2015) looking at the association of light emission with socio-economic variables, such as GDP, GDP change, infant mortality and poverty.

Cauwels et al. (2014) have used the DMSP-OLS composites to monitor the global economic regime shift. They find the planetary center of light emission moving eastwards at an annual pace of around 60 kilometers. An important statistical measure is the dynamic of a spatial light Gini coefficient. The study shows the geographical shift of economic activity from the Western economic core to Asia.

Miller et al. (2012) outline the major advantage of the new VIIRS composites, specifically for regional environmental analysis in terms of climate assessment, weather and hazards monitoring and the observation of interactions between the lower and the upper atmosphere because the sensor detects starlight and airglow that could not be differentiated from artificial light before. More generally, Kyba et al. (2017) address

8 In a wider sense, the term had been introduced long before. From an architectural perspective, Alexander (1965) discussed in a renowned essay how "natural cities" had evolved in contrast to planned artificial cities. Care is definitively needed to use the term in an interdisciplinary dialogue as it had been also used in philosophical debates, such as the issue of reintegrating the natural with the urban (e.g. Stefanovic and Scharper 2012).

some specific limitations of interpreting the information content of the VIIRS composites, that should be mentioned here. Even though the VIIRS images contain substantially more information due to the higher resolution, the detection of blooming and starlight and the true variance of radiance, a specific external factor might reduce the usability at least to a minor extent. This has to do with the virtually simultaneous introduction of LED light at a broader scale (also since 2012). Since LED light is characterized by a different color spectrum compared to traditional light sources, there is some risk of error, when comparing the evolution of light emission over time.

The meta-study by Li and Zhou (2017) addresses urban mapping by using the DMSP-OLS composites. The authors confirm the great potential of those images for mapping urban and rural space at different scales and during a long-run observation period. They also systematically point out the limitations of the 6-bit images by resolution, blooming, different sensors and other light disturbances. Roychowdhury et al. (2011) contributed with a case study by comparing and integrating daylight (Landsat) and DMSP-OLS images to detect the extent of urban space. By applying supervised classification for the urban area of Hyderabad (India) with both data sources they show that nocturnal satellite imagery is not less precise to delineate urban and sub-urban space. The improved precision of the newly introduced VIIRS composites suggests those data sources to be of superior relevance as compared to the Landsat images.

A rather recent approach is the exploration of the relationship between the distribution and intensity of night lights and the rank size distribution of cities, empirically confirmed by a quasi-natural law (Zipf's law). While Jiang et al. (2015) have confirmed Zipf's law at a global, continental and national scale by using the simple 6-bit images, Bergs (2018) and Wu et al. (2018) used DMSP radiance calibrated images (or VIIRS composites respectively) to delineate urban space. Both studies address spatial segmentation by using Zipf's law. While Wu et al. (2018) use a direct Zipf's law based method to bootstrap the optimal threshold for China, Bergs (2018), addressing the Netherlands, estimates the threshold between urban and non-urban by traditional k-means clustering and subsequently testing the resulting segmentation by Zipf's law. The latter approach is also subject to further examination for the study at hand.

Apart from using nocturnal satellite imagery for the delineation of natural cities by Zipf's law, some authors have also used location-based social media to delineate natural cities. However, it seems that night light images are a superior database for such purposes. More importantly, behaviour and regional specificities have an influence on the use of social media. As Wu (2015, p. 3) finds: "Compared to the city numbers extracted from nightlight imagery, the study found out the reason why Zipf's law does not hold for location-based social media data, i.e. due to bias of customer behaviour and regional limitations. The bias mainly resulted in the emergence of natural cities in certain regions … thus the emergence of natural cities cannot to be exhibited objectively." For the study at hand, we therefore did not use location-based social media data as another automatically generated database. But our focus is not at all restricted on night satellite images. In fact we addressed the statistical association of those digital images with other small-scale grid data and maps, such as CORINE and the IÖR-Monitor (after geo-referenced alignment with GIS).

3 Theory and expected contribution to new insight

So far, little scientific effort can be recognized in finding a more objective approach to recognize natural urban and rural space. For an analysis of the rural and the urban, understood as interlinked spatio-social

constructions and historically pre-defined by fixed boundaries, it is difficult to address the true functional differentiation of space and particularly its evolution over time. In fact, the historically defined rural and urban is simply discretionary and hardly representing the true functional differentiation. Further to that, since space evolves dynamically, any fixed delineation of rural-urban boundaries is inconsistent with the reality and consequently represents an obstacle to empirical and comparative research. In regional science and urban economics it has been deplored that the detection of true functional or natural boundaries between the urban and the rural is hardly possible with the data available. Only the more recent availability of micro-data and small-scale grid data has opened doors to a more realistic analysis of rural and urban change. In this study we seize on one particular thread of recent approaches to detect natural cities and non-urban space. Global nocturnal satellite imagery from the National Oceanic and Atmospheric Administration (NOAA) appears worth to explore for urban and rural studies. It is the simple association that light emission from cities is much stronger than from non-urban or untenantable space. Cities on the night satellite images can be easily detected when comparing the image with a map.

Intuitively, nocturnal luminosity must be closely associated with anthropogenic activity, like the level of production, service provision and distribution by transport infrastructure (i.e. economic activity), energy consumption and the related level of greenhouse emissions. A direct environmental relationship is assumed between light emission and light pollution, the latter being associated with hive death or disappearance of other insect species. Proville et al. (2016), as mentioned earlier, reviewed relationships by correlation analysis between night luminosity and several other variables as displayed in the following table:

Table 2: Global correlation with the socio-economy and environmental variables

	log (lights)
log (lights)	1.00
log (electricity consumption)	0.93***
log (GDP)	0.91***
log (population)	0.72***
log (CO2)	0.93***
poverty	-0.42***

Source: Proville et al. 2016, p.4

All Pearson-r coefficients are highly significant at the $p<0.01$ level (***). As expected, electricity consumption and CO_2 emissions are most strongly correlated with light emission. But the correlation with GDP is only slightly lower (0.91). The coefficient for population would be higher if not population but population density were estimated. The negative relationship with poverty might appear a bit weaker than expected, especially with respect to the high positive correlation with GDP. But it has to be borne in mind that poverty is also widespread in urban regions, notably in largely urbanized developing countries, so that this might superimpose the relationship. Those close relationships between light emission and socio-economic and environmental variables make ubiquitously available nocturnal satellite imagery an appealing holistic proxy for regions with a lack of official data across a broad range of different topics. Especially in developing countries, the use of night light imagery to estimate the social and economic development level has been widely used (e.g. by the World Bank). It is to be borne in mind that the strong relationship between all those environmental and socio-economic variables is only established at global level. At national and particularly sub-national level it might be lower, however still varying just depending on the region or the type of land use.

For the specific purpose of the study at hand, the relationship with population density and economic activity suggests nocturnal luminosity to be a particularly meaningful proxy to describe “natural” active population. The images clearly reflect the association between luminosity and employment at workplace, i.e. variation of population density during work time. The functional division of urban and rural space essentially has an economic foundation because of the uneven spatial distribution of production factors and accessibility (thus the non-homogeneity of space) and the resulting incentive for people to move (cf. Starrett 1978, p. 36; Brakman et al. 2009, pp. 51 ff.). The functional difference between urban and rural space (production, energy consumption and infrastructure, settlement, open space) is thus well reflected in the distribution of light (light clusters side-by-side with dark clusters of pixels). But so far, as mentioned above, there is little knowledge about the relevance of light emission as a proxy for sustainable economic activity or environmental pollution at smaller spatial levels. This qualifies nocturnal satellite imagery as a relevant experimental database for further socio-economic analysis of rural-urban evolution and change.

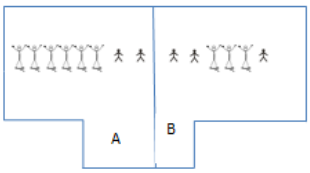
Our expectation is that if we find systematic patterns of relationship between light emission and certain socio-economic and environmental variables for different spatial categories at small-scale spatial level, luminosity could be not only useful for the segmentation of natural space but also useful to fill possible gaps of local data availability.

3.1 Some theoretical categories addressed by the study at hand

Two theoretical categories addressed by this study deserve a closer consideration. One (MAUP) is a well-known problem of spatial statistics largely affecting spatial classification, the other (Zipf’s law) is a rare example of social physics to be more comprehensively discussed further below, because for this study it is used as a tool.

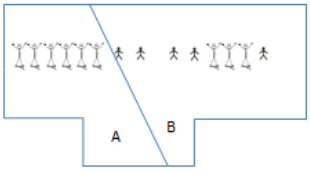
MAUP

Original administrative boundary



A: 25 percent unemployment B: 50 percent unemployment

Changed administrative boundary



A: zero percent unemployment B: 62.5 percent unemployment

The modifiable areal unit problem (MAUP) is a source of statistical bias that can considerably impact the results of statistical estimation of spatial relationships or influences including the test statistics (simplified description in the box). The MAUP affects estimates when the individual level of analysis is aggregated into higher level entities. This can affect any socio-economic variable of a region or municipality. The aggregate statistics (totals, means, spread, rates, proportions, densities) are influenced by both the shape and scale of the aggregation unit. For instance, individual census data may be aggregated into municipalities, postcode areas or any other arbitrary administrative or natural spatial partition; further to that spatial boundaries can be changed over time without changing anything at the level of the individual. Thus the results of data aggregation always depend on the choice of administrative or statistical boundary and which areal unit to use in their analysis. (cf. Wikipedia)

Exactly this problem may also affect spatial analysis of rural and urban interaction within the ROBUST project. The major parameter in inter-municipal cooperation is the administratively defined territorial boundary between municipal entities. If research in this project is to be understood as practice-oriented (in terms of policy relevance) it might be expedient to tackle the a.m. MAUP problem directly and to recognize

cities and rural areas more independently from discretionary administrative delineations and territorial reforms. The earlier mentioned approach of "natural cities" might help to detect the true extent of cities and non-urban areas. Night satellite imagery is a promising tool for that purpose (Jiang et al. 2015, Bergs 2018). To avoid statistical distortions implied by such discretionary definitions of administrative boundaries, Jiang et al. (2015) have adopted the concept of "natural cities". Those are not defined by administrative boundaries, but rather by the autonomous evolution of settlement and economic activity. Thus, the functional differences between urban and non-urban space is made visible.

The analytical approach of natural cities is closely related to spatial heterogeneity, as mentioned earlier. The simple understanding of spatial heterogeneity describes the variation of relationships over space, hence the uneven distribution of spatial variables. Topography, economic assets, urban and rural areas, population densities, converging and diverging regions or maritime versus continental climate are all examples of spatial heterogeneity. But this is not the whole story of spatial heterogeneity: The wider concept of it comprises a more holistic insight into less visible spatial structures of amazing regularity, based on fractal geometry, like the self-similar forms of coast lines, topographic reliefs, river networks or the peculiar rank-size distribution of cities. The concept of "natural cities", their detection by nighttime imagery and the subsequent estimation of their characteristic distributions are closely related to this interpretation of spatial heterogeneity.[9]

Zipf's law represents such a fractal dimension of space and is thus a candidate for testing the statistical segmentation of natural urban and non-urban space. The underlying explanation for Zipf's law as a test tool is its ubiquitously uniform property of the rank-size distribution of cities worldwide at continental, national and subnational scale. The phenomenon has been so far subject to many theoretical and empirical studies trying to explain underpinnings of it from different viewpoints (economic, statistical, even physical). The still not yet fully understood property of Zipf's law as a quasi-natural law in social evolution of space moved Paul Krugman to attribute it as something "spooky" (Krugman 1996).

In using Zipf's law for testing the derivation of segmented natural cities it is needed to briefly introduce the important underpinnings of this law to understand its relevance for the analysis of rural-urban interaction (cf. Bergs 2018).

3.2 Underpinnings of Zipf's law

To understand Zipf's law as a test of spatial segmentation, its theoretical and empirical underpinnings are first to be discussed (cf. Bergs 2018).

Zipf's law is defined as

$$S = CR^{-\alpha}$$

or in its logarithmic transformation:

[9] While spatial heterogeneity usually represents the first-order moments of space, spatial dependence is conceived as its second-order property. However, there are arguments that the relationship is contrary, namely that spatial heterogeneity is of higher order. Jiang (2015, p. 6) argues that "…Spatial heterogeneity is a kind of hidden order, which appears disordered on the surface, but possesses a deep order beneath. This kind of hidden order can be characterized by a power law or a heavy-tailed distribution in general. …". So if just viewing the visible heterogeneity at the surface only part of its complex underpinnings is regarded.

$$log(S) = log(C) - \alpha \cdot log(R),$$

where S means city size, R is rank of a city and C is a constant. It is expected that $\alpha \approx 1$. The graph of the logarithmic form is then a straight line with a gradient of -1. If α<1 a more even distribution of urban settlements than predicted by Zipf's law is found; the distribution is thus compressed. If α>1 the larger cities are larger than predicted and/or the smaller are smaller than predicted; the distribution is then stretched. It is also usual to work with the reversed form, i.e. the Pareto distribution. Then α behaves contrary to the above described pattern.

$$R = BS^{-\alpha}$$

Here rank is regressed on size. B is a constant. The probability $\Pr(X \geq x) = BS^{-\alpha}$ is then equivalent to the number of cities X larger than city x (cumulative distribution function). The probability density function (PDF) of city sizes *(X=x)* is derived by differentiation of the CDF: $\Pr(X = x) = \alpha BS^{-\alpha-1}$. In case of a perfect rank-size distribution in accordance with Zipf, the exponent is in each case the same (both -1) just as the constants are identical (B=C), while deviations from the expected exponent tend into the opposite direction, just depending on the distribution form selected.

The derivation of Zipf's law for cities and the respective preconditions for it to hold are well explained by Gabaix (1999b): To imagine it, it is useful to take normalized sizes of cities. The normalized size of city population S^i is the population of city I divided by the entire urban population, hence: $\sum_{i=1}^{N} S_t^i = 1$. The change of size of a city i is $S_{t+1}^i = \gamma_{t+1}^i \cdot S_t^i$, where the "shock" coefficient γ_{t+1}^i is assumed i.i.d. In case of γ_{t+1}^i>1, the city is growing, otherwise constant or shrinking, but the average normalized growth for all cities must be 0, i.e. $\gamma - 1$, where $\gamma = 1$ (the sum of size changes multiplied by their probabilities for any i at any t).

A process based on a zero normalized growth rate of cities must result in a steady state of an urban settlement distribution, and Zipf's law represents such a steady state. "If cities grow randomly, with the same expected growth rate and the same standard deviation, the limit distribution will converge to Zipf's law..." (Gabaix 1999b, pp. 743 f.). Size-independent city growth with the same variance (Gibrat's law) as a determinant of power law behavior (Zipf's law) implies scale invariance. The distribution has thus a fractal dimension property.

There are important economic interpretations of Zipf's law. Brakman et al. (1999) and Duranton (2007) interpret the Pareto exponent in terms of increasing returns and agglomeration economies and address it from different viewpoints. Reggiani and Nijkamp (2015) explore the secular evolution Zipf's law under consideration of the related socio-economic connectivity in urban networks. Krugman (1997, pp. 42 ff.) describes Zipf's law as a rare example of "social physics". Fujita et al (1999, p. 224) ponder on the fact that Gibrat's law also requires variance of city growth independent of city size. But with neither positive nor negative industrial spillovers the variance of the growth rate should still decline implied by diversification. Instead they suggest a theory of random connections rather than random city growth and also argue on a theoretical thread of physics. These connections may correspondingly evolve as spatial industry linkages along distribution channels. Explained by percolation theory, the behavior of connected clusters in random connections (random graphs) generates power law distributions such as traffic capacities on roads (in analogy to the distribution of river sizes measured by the volume of flow). Such percolation patterns may then also lead to a corresponding rank-size distribution of cities. In addition to that, there are further various empirical economic studies such as Rosen and Resnick (1980), Soo (2005) and Suedekum and Giesen (2011) to mention only a few.

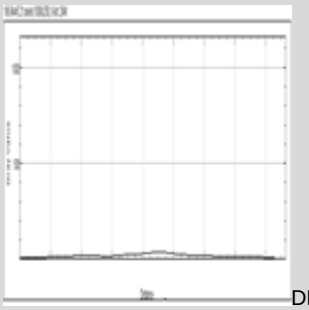
The expected insight of that experimental study is to integrate spatial heterogeneity and spatial dependence by using a quasi-holistic variable, to detect true rural and urban extent and to monitor the dependencies between those categories over time. It is a further research effort based on Bergs (2018) with a practical purpose for local policy and planning decisions.

This experimental use is in correspondence with Technology Readiness Level (TRL) 3 – experimental proof of concept. Furthermore, we will examine how commonly used data sets can be merged with more novel data collection methods, such as satellite night images. For instance, an analysis of changes over time in the different radial sections around a city will shed light on integration and disintegration processes; further merging of those data with other socio-economic or environmental grid data can largely improve spatial information. Nocturnal satellite images can provide additional insight to share and discuss with stakeholders.

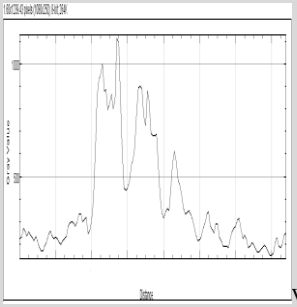
The study at hand aims to further engross and (reflect on) night satellite imagery as a novel data source for spatial economic and social change of urban and rural regions spelled out by Woods and Haley (2017, pp. 19-20.).

4 Data and Methodology

The VIIRS night satellite images of the case study regions are the basic databases of this study. They are cropped from the respective composites by GIS (based on shapefiles defining the geographic boundaries). They are precisely geo-referenced and defined in the regularly applied Lambert-Azimuthal projection being in accordance with the EU-INSPIRE directive. Any mapping or statistical analysis that is possible with GIS could be run directly (e.g. spatial dependence by Moran's I). For more detailed image analysis and for segmentation of space the software ImageJ was applied. Originally, ImageJ (a powerful imaging software), has been mainly applied for medical and biological purposes; ImageJ is not automatically geo-sensitive, hence for precise measuring of distances geographic coordinates and scales have to be defined manually.



DMSP-OLS
Frankfurt: DN (2009)



VIIRS
Frankfurt: Radiance (2012)

For this study, the core database consist of the first (2012) and last (2017) available VIIRS composites. So far, there is only one annual composite (2015). Therefore, it was necessary to select monthly composites. For spatial analysis, we follow Kyba et al. (2017, p. 6) who suggest to select the respective October composites for Europe as those are particularly suitable for comparisons of nighttime lights data for several reasons (such as little straylight reflection stemming from snow at high altitudes and cloud coverage). The 14-bit VIIRS images are straylight adjusted; the nominal spatial resolution is 750 meters, thus better than the former DMSP-OLS composites. A major problem stems from the limitations implied the short longitudinal analysis 2012-2017. It is thus not possible to detect genuinely secular spatial change in terms of spatial heterogeneity and dependence. Sure, nocturnal satellite images are available since 1992 (DMSP-OLS), but neither the simple annual images with 6-bit quantization nor the few adjusted "radiance-calibrated" annual composites are comparable (see overview of the image products presented earlier).

The 6-bit images showing digital numbers (integers) for the pixels in the range between 0 and 63 are hardly useful for urban analyses because of its rigid right-censored spectrum that does not cover the full variance of urban space. The exemplary comparison of plot profiles, showing means of vertical (column) pixel vectors in the east-west direction (left), reveals the major difference in variance between DMSP-OLS and VIIRS pixel values (DN or Radiance) for Frankfurt (range 0 to 1,100). The few further DMSP radiance-calibrated composites between 1996 and 2011 show the true variance of light emission but neither the absolute value of DN nor radiance ($nWcm^{-2}\ sr^{-1}$). Due to missing on-flight calibration the pixel values of the radiance-calibrated images are unit-free. It is thus not possible to directly compare the three images types, however it is possible to roughly estimate change rates between 1996 and 2017.

But also this requires a proper prior inter-calibration of images, because of the use of different sensors and the improved scanning technology over time.

Statistical segmentation of radiance-calibrated DMSP-OLS images is also not comparable with segmentation of VIIRS images because of a different resolution and depth of focus. The major reason for that is the fact that in contrast to VIIRS the radiance-calibrated images are not corrected for blooming effects. These effects originate in light reflection at sites without own light source. Cities detected on DMSP-OLS images usually appear larger than they are in reality. Such blooming effects can be statistically removed from the radiance calibrated DMSP-OLS images, based on e.g. a global correction factor (Small et al. 201). This can be done to test blooming-adjusted rank-size distribution of segmented space, but even though blooming can be statistically removed (reducing the pixel size by the correction factor) it cannot be precisely removed from the segmented patches as such. Thus, with radiance-calibrated images it is possible to estimate the correct city size but not its precise shape. Similarly, analyses of spatial autocorrelation would be highly sensitive to blooming effects in case of the radiance-calibrated images. Hence, for the major geostatistical analyses we are forced to use the VIIRS composites and have to confine on a rather short time horizon (2012-2017).

The technique used for segmentation of the satellite images in the study at hand is k-means clustering. It is a rather simple approach, largely used in remote sensing. There are several alternative techniques, one is Isodata leading to similarly reliable results but with considerably more computing effort (Pandya et al. 2013, Yang et al. 2017). A third interesting and rather simple segmentation tool is the Head-Tail index, actually not a clustering approach but rather an approach of differentiation of shapes along the power distribution, such as the pixel values (Jiang et al. 2015). It works well for images with a small range of pixel values (e.g. 6-bit images). For VIIRS composites the resulting number of observations is too low for sound estimates. Supervised classification is a different tool using the maximum likelihood method. It is based on proper knowledge of the region to define training areas for different land use types from those the differentiation of the entire space is estimated (e.g. Roychowdhury et al. 2011). The methodological challenge of this approach is to unambiguously define training areas as perfectly rural, peri-urban or rural. Therefore, the approach we pursue is that of a pure statistical segmentation that is tested by a globally validated empirical relationship.

A proper segmentation of urban and non-urban space thus allows a comparison of administrative and natural space over time; since the VIIRS images are available on a monthly base it is even potentially possible to have a permanent monitoring of natural space. Nocturnal satellite imagery is not only useful for a segmentation of urban and non-urban space, it is also useful to monitor spatial dependence over time, i.e. how neighborhoods are influencing each other, both in terms of municipalities and grids as well as concentric circles around cities.

Finally the last essential purpose of using nocturnal satellite imagery is the close association of light emission with various socio-economic or environmental variables. This association is only significantly established at

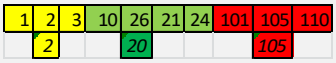
global scales (Proville 2016). As far as environmental and socio-economic variables are available at small scale grid level it could be an exciting task to estimate correlation between light emission and such other variables at different types of space at micro-scales and to identify systematic patterns.

In the following section the methodological steps are further explained.

4.1 Methodological procedure I: K-means Segmentation and Zipf's law

The study comprises case analyses of the Regionalverband Frankfurt/Rhein-Main and the Ljubljana Urban Region. Here we derive the regional classification of space from the national (German and Slovene) VIIRS extractions.

Cluster analysis



Cluster analysis is based on two inherent kinds of variance of a given distribution. The aim is to statistically recognize different groups of data with a respectively low variance and to keep the respective variance of the resulting group means as high as possible. K-means clustering is a specific procedure in which the number K of clusters is defined prior to the estimation. An optimum clustering is never guaranteed by prior fixing K. Therefore tests like the Calinski-Harabasz-Pseudo-F test help to estimate the ad-equacy. Increase of K or removal of outliers can contribute to a better cluster model shown by a higher Pseudo-F test statistic.

Segmentation of national space will be done by cluster analysis (k-means) to be subsequently tested (see box). While the former DMSP-OLS composites were largely distorted by blooming effects with a major influence on the size of segmented natural urban space (former satellite sensors could not differentiate sufficiently), the new straylight adjusted VIIRS images are much more precise. But this precision alone is not yet a sufficient precondition for a precise detection of urban and non-urban space. A remaining problem is the outlier sensitivity of cluster-based segmentation analysis. But this problem can be solved statistically.

To avoid a statistical segmentation affected by highly skewed distributions of light emission (many more dark pixels than light ones) and badly derived threshold values, the night light databases need to be carefully explored, because any kind of non-supervised segmentation is highly sensitive to outliers. Such outliers, that are often affected by non-stable light (e.g. illuminated building sites), need to be appropriately removed. This can be done by applying an adjusted box-plot approach for skewed distributions. In this study we apply the statistical correction tool proposed by Vandervieren and Huber (Vandervieren and Huber 2004). Both, k-means clustering as well as outlier removal was executed by ImageJ and Stata®. The Stata commands are **kmeans**, **medcouple**, **iqr** and **cluster stop**.

Since we work with unsupervised classification techniques the segmentation of urban space is a result that should still not be uncritically taken as the true extent of urban and non-urban space. It should be appropriately tested. To confirm the adequacy of spatial segmentation, we test whether the well-known Zipf's law on the rank-size distribution of cities can help to confirm a statistically determined "natural" segmentation of urban and non-urban space at a national level by using an empirically determined global shape parameter (Bergs 2018). The idea behind that approach is the ubiquitous empirical evidence of a uniform rank-size distribution of cities (at national, continental and global level) in accordance to a power law behavior with the exponent of minus one. With other words: If, in accordance with Zipf's law, the exponent of rank-size distribution of the upper tail of the distribution of cities in a country is close to -1, as empirically confirmed for the administrative cities in the majority of countries, it can be assumed that cities are properly separated from non-urban space. Empirically, Zipf's law holds universally, but mostly for cities above a certain population size (e.g. Gabaix 1999a and 1999b). It is to be expected that natural cities may

better represent urban space because there is no interference by discretionary territorial reforms. Natural cities thus may thus cope with "more carefully defined" metropolitan areas (Fujita et al 1999, p. 217).

So far, Zipf exponents have been mostly estimated on administrative population data. Only a few studies looked at a more natural size and appearance of cities, such as satellite images, social media density or spatial grid data. Budde and Neumann (2016) analyzed Zipf's law for population density based on small-scale grid data in Germany. They show that population density rather than official size of city population determines Zipf's law. In a master thesis, Wu (2015) analysed Zipf's law on natural cities determined by location-based social media data. Few studies also tested Zipf's law on nocturnal satellite images. Jiang et al. (2015) tested the global validity of Zipf's law with the uncalibrated 6-bit "stable lights" for 1992, 2001 and 2010 applying maximum likelihood estimation. In another master thesis under supervision of Bin Jiang, Liu (2014) explored US natural cities extracted from the 6-bit images between 1992 and 2010 to test whether Zipf's law holds. In a very recent paper, Egger et al. (2017) looked at the natural city growth in China, also based on the 6-bit composites since 1992. Even though the paper does not include a reference to the previous work of Jiang and his working group, the definition of natural cities is more or less identical. The authors do not specifically address Zipf's law, but just the comparison of size and rank between administrative cities and the clustered natural cities in China. The widely 6-bit stable light images do not allow proper segmentation of urban areas, because even in rural areas, true luminosity is often masked by top-coded pixels (DN>63). Bergs (2018) therefore used two of the few available radiance-calibrated composites (1996 and 2011) and tested segmentation of natural urban and non-urban space in a highly urbanized country10. Research in this direction is also at its very beginning. With this much more practice-oriented case study exercise it is aimed to contribute to the debate from the viewpoint of planning and local policies.

An important issue of estimating Zipf coefficients has been the choice of the estimator. The usefulness of the traditional simple OLS estimate regressing log of size on log of rank (see earlier) is constrained by a major bias implied by the standard errors in case of small samples. Therefore, many researchers use maximum likelihood estimations (Hill estimator). But this alternative approach is not really necessary because a simple change of the traditional OLS model (i.e. reverting the relation in the regression equation and subtracting ½ from the respective vector of ranks) leads to a very dependable unbiased solution, widely used for such studies (Gabaix and Ibragimov 2011). In fact, this simple approach is not only deemed superior to usual OLS but even to the widely used Hill estimator (maximum likelihood), because of the lower bias in the standard error in the Pareto estimate when samples are small, as for small countries (cf. Brakman et al. 2009, p. 302 f.). Therefore, in order to obtain the respective Pareto coefficients (reversed Zipf coefficient) and their significance levels the Gabaix-Ibragimov estimator is applied in this study. This is especially important for the case study on Slovenia.

4.2 Methodological procedure II: Testing and simulating spatial dependence

Tobler's law is a major constant in geography. It says that spatial evolution always depends on other space, but always more on contiguous or closer one than on distant space. Increasing or decreasing spatial dependence may explain changing spatial functionality of different types of areas over time. Especially the dynamic of natural space can be observed by that. The influence of neighbor space on the development of a regarded region is important to consider when it comes to policy, cooperation and planning. Light emission

[10] In this study the Netherlands are addressed: while Zipf's law does not hold for Dutch administrative cities between 1996 and 2011, the rank-size distribution of natural cities derived from nocturnal satellite imagery well confirms Zipf's law.

is also a meaningful variable to shed light on environmental and socio-economic dependencies among neighboring or more distant spatial units. In this study we focus on different selections of space (typically urban, typically rural, radial segments around a city), analyzing spatial dependence by estimating Moran's I coefficients with QGIS (SAGA command Global Moran's I). Since images can be also mathematically manipulated it is also possible to simulate development scenarios and to estimate changing spatial autocorrelation. Eventually this insight can be of major importance for land use planning and local policies.

4.3 Methodological procedure III: Merging VIIRS with other grid variables (IÖR-Monitor and CORINE)

How do socio-economic and environmental grid variables correlate with light emission? So far, research has focused on larger scale spatial units (countries or global perspectives). For developing countries the World Bank has used DMSP-OLS composites as a database to estimate socio-economic aggregates when official data are missing. For most industrialized countries (OECD countries) the official data infrastructure is mostly standardized and of high and reliable quality (e.g. Eurostat regional statistics). The problem, however, is the low maximum resolution of data (NUTS 2 or NUTS 3, in some cases at city level like Urban Audit). There is no full coverage of official micro-spatial or grid data at smallest scale. Meanwhile there is access to either restricted or open grid data that could be used for Germany. In estimating the association between light emission and such other variables at local levels we hope to find spatial levels with systematically lower or higher association between variables. For regions without access to micro scale grid data, the VIIRS could partly serve as a proxy database. The socio-economic and environmental databases used in this study are IÖR-Monitor and CORINE for land use, nature and environment. Correlation will be certainly weaker at smaller spatial levels and much more varying, but it is perhaps possible to detect systematic differences depending on land use types of areas. This may open a perspective to determine pixel values (radiance) as a proxy for certain environmental and socio-economic variables in regions where small scale grid data are missing. This is also tested in this study by merging IÖR-Monitor grid data with VIIRS images at a resolution of one square kilometer.

5 Case study “Regionalverband Frankfurt/Rhein-Main”

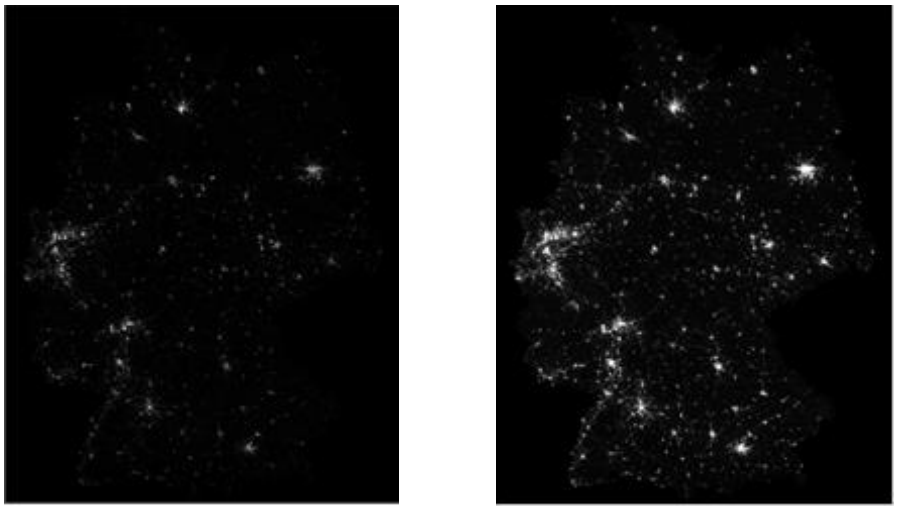
The following chapter deals with the case study on the Regionalverband Frankfurt/Rhein-Main (RVFRM). The first part addresses the statistical segmentation of urban space within the regarded area based on the national German VIIRS image extracted from the global composites (October 2012 and October 2017) by GIS and the EEA shapefile for Germany. The second part deals with an analysis of urban and non-urban change by considering spatial dependence and the positive and negative variation of light emission at the level of one square kilometer. We also include a simulation forecast to estimate the effect of a planned new suburb on the growth of the natural city of Frankfurt. In the last part we explore the relationship between radiance and other small-scale environmental grid data (IÖR-Monitor, CORINE).

5.1 Empirical analysis: National Segmentation of space - natural cities (Germany)

The RVFRM region is part of Germany; thus a classification of urban and non-urban space in this area should be first determined from the national distribution of pixel values and then to be focused on the case study region. This two-step approach is reasonable because with the procedure chosen in this study the entire segmentation of urban space is to be tested by Zipf’s law at a national level.

The two images (raw radiance data) for Germany are the following:

Figure 1: Germany 2012 and 2017: radiance x 100



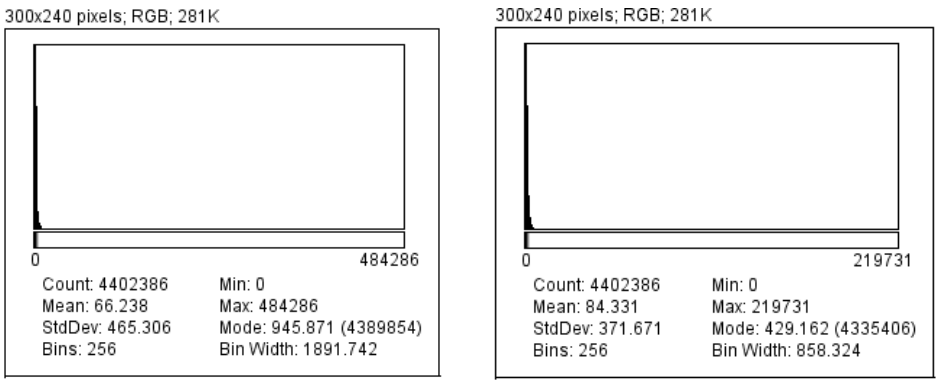
2012 *2017*

Source: NOAA

To capture the precise distribution within German borders, the zero-pixel space outside is set NaN. In these images, radiance is multiplied by the factor 100, simply to better recognize the cities by visual inspection. But visual inspection alone just allows a differentiation between somehow white, grey and black. The information by visual inspection is thus very little. Basic moments of the distribution of pixels at national

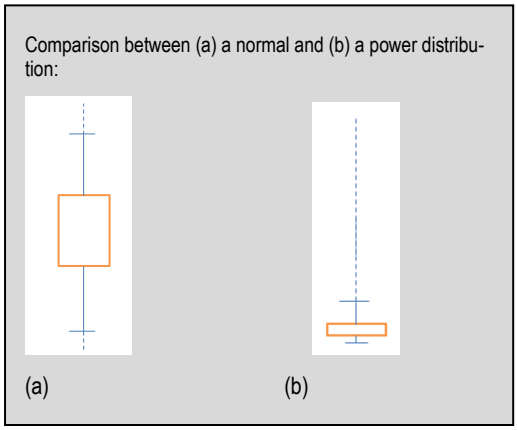
level (raw pixel data) are substantially more informative. They are illustrated by Fig. 2 and reveal the true information content of those images. The histograms and the minimum and maximum pixel values show the extremely skewed distribution of the 4.4 million pixels in both images.

Figure 2: Germany: Histogram and descriptive statistics 2012 and 2017 (radiance x 100)



2012 *2017*

Hence, there is no simple grading in just "white – grey –black" but rather a grading in the range between zero and 484,286 (in 2012) and 219,731 (in 2017) categorized into 2^{14} levels (14 bit)[11].



There are many dark pixels and very few bright ones, thus representing a power distribution (see box) as shown by the histograms. Very interesting is also the mean pixel value that has increased from 66 (2012) to 84 despite the smaller range in 2017. At the upper tail of the distribution of pixel values there are major outliers, most of them implied by the characteristic power distribution but also partly induced by scanning errors. All these outliers have an important influence on the subsequent spatial segmentation of any kind (k-means clustering, Isodata clustering, maximum entropy etc.) and need to be removed adequately to find robust thresholds for spatial segmentation.

Therefore, before classifying the national German space for both years with cluster analysis (k-means) it is important to trim the highly skewed pixel database and remove outliers adequately. Since the distribution is not normal, a simple *Tukey* box-plot analysis for normal distributions is not useful; the inherent property of a power distribution would get lost. Instead a special approach for trimming skewed power distributions is needed to estimate the right cut-off point for upper tail outliers. Such an approach needs to consider the fact that outliers are characteristic rather than exceptional in power distributions.

Vandervieren and Huber (2004) have offered a method to solve that problem. In a simulation study, they estimated shares of outlier removal for Pareto and other skewed distributions. In their estimations on Pareto

[11] This variation can also be shown by surface plots (see below). Here the raw radiance data are sufficient to visualize the variation.

distributions with location and shape parameters (3,1) the share of outliers slumps from more than eight percent (in a Tukey box-plot) to only around 0.5 to 1 percent in the adjusted version; for Pareto (1,3) distributions from more than 12 percent (Tukey box-plot) to some 2.1 to 2.5 percent (in both cases declining with increasing n). These are remarkable discrepancies. The removal of outliers is thus less than one percent, while for normal distributions it is often more than ten percent. The approach for such a specific trimming of a power distribution is proposed by Vandervieren and Huber as follows:

$$W_u = Q3 + 1.5e^{b \cdot MC}\, IQR,$$

where W_u means the upper limit of the whisker, *Q3* the third quartile, *MC* the medcouple[12] and *IQR* the inter-quartile range. The exponent *b* (for the upper tail) was estimated at ≈4.0, based on 10,000 observations generated in the simulation study (Vandervieren and Huber 2004, p. 1937; cf. also Bergs 2018).

The estimates for trimming the databases 2012 and 2017 are shown in Table 3:

Table 3: Germany: Cut-off estimates of pixel values (radiance x 100)

Year	Upper whisker cut-off point [(nWcm^{-2} sr^{-1}) x 100]	Percentage of removed pixels
2012	5,682	0.50
2017	5,169	0.51

Estimation in accordance to the approach of Vandervieren-Huber 2004: Adjusted Box-plot for Skewed Distributions (in: Antoch J (2004) Compstat 2004, Springer Heidelberg)

The cut-off estimates clearly reveal the highly skewed original distribution also shown in the histograms above. By removing only 0.5 percent of pixels of the whole distribution in the upper tail, the largest pixel value (radiance x 100) slumps from a six-digit number to just 5,169 in 2017. This has an enormous impact on the clustering and the cluster parameters and reduces the overly leverage of extreme outliers. Based on the trimmed range of pixel values the segmentation thresholds are eventually shown in the following Table:

Table 4: Germany: Segmentation thresholds and cluster centroids 2012 and 2017

Year	Threshold (radiance)	Cluster centroids
2012	735.32	45.82 | 1,574.61
2017	810.82	58.57 | 1,703.63

Automatic thresholding by ImageJ based on the outlier-adjusted images

[12] The medcouple is a robust measure of distribution skewness of a distribution function *F*:

$MC(F) = \underset{x_1 < m_F < x_2}{med} \cdot h(x_1, x_2)$, where m_F is the median of F and the kernel function h is $h(x_i, x_j) = \frac{(x_j - m_F) - (m_F - x_i)}{x_j - x_i}$

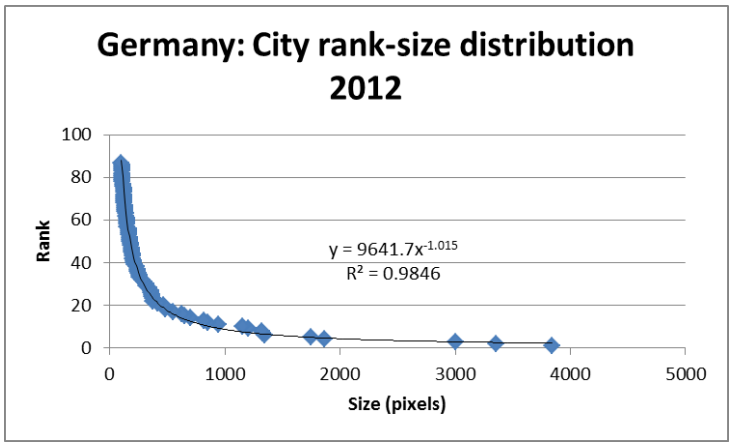
Finally, the following segmentation of national space for 2012 and 2017 is found by the specification defined:

Figure 3: Natural cities in Germany 2012 and 2017 (segmentation results)

2012 *2017*

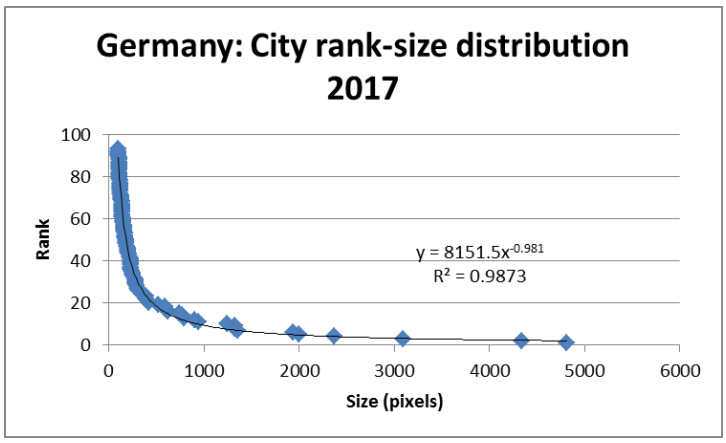
The two images showing Germany in 2012 and 2017 after spatial segmentation clearly reveal the first-order spatial heterogeneity represented by cities and urban space as distributed on a map of Germany. By visual inspection no systematic order of the patches can be recognized, the distribution of urban space – whether in Germany or any other country – appears disordered and chaotic. This impression changes when including a higher-order understanding of spatial heterogeneity (cf. Jiang et al. 2015), as revealed by the rank-size distribution of those patches (natural cities). This rank-size distribution also follows a power law behavior (many small towns versus few big ones). In ordering the segmented larger patches (>100 pixels) along their size and estimating the resulting power distribution by a simple nonlinear regression (Pareto) we find estimates for α (2012 and 2017):

Figure 4: Germany: Testing the segmentation with Zipf's law 2012 (simple power estimate for patches >100 px)



Source: NOAA

Figure 5: Germany: Testing the segmentation with Zipf's law 2017 (simple power estimate for patches >100 px)



Source: NOAA

For Germany the exponents of both annual distributions (-0.981 and -1.015) are remarkably close to -1 and are thus confirmed by Zipf's law, so that (i) there is a well-defined hidden order beneath the disordered patches and (ii) the segmentation of natural urban and non-urban space of Germany can be sufficiently justified.[13]

The rank-size distribution of the larger patches, well in accordance to Zipf's law, confirms a proper segmentation of the whole space.

Due to the high number of observations in the case of bigger countries (such as Germany), there is no risk of a major bias due to underestimated standard errors. A simple Pareto regression is usually sufficient. For smaller countries with a smaller number of observations, such as Slovenia (see below), either a maximum

[13] The segmentation is specified to capture all contiguous pixel patches; thus holes within patches segmented are included. This helps to recognize areas at the natural edge between rural and urban.

likelihood estimation (*Hill* estimator) or the widely used *Gabaix-Ibragimov* estimator should be applied. This estimate is found when transforming size and rank-1/2 into their natural logarithms.

We nevertheless also ran two *Gabaix-Ibragimov* procedures, but as expected, results for Germany still support Zipf's law rather well, both for patches larger than 100 pixels as well as for those larger than 50 pixels, testing how robust the estimated rank-size distribution is when adding observations downwards the upper tail.

Table 5: Gabaix-Ibragimov estimates 2012 and 2017 for Germany

	2012	**2017**
>100 px	y = -1.0746x + 9.4729 (0.0208) (0.1176) R^2 = 0.9692 Obs.: 87	y = -1.0376x + 9.2884 (0.0172) (0.0969) R^2 = 0.9755 Obs: 94
>50 px	y = -1.086x + 9.5391 (0.0090) (0.0442) R^2 = 0.9865 Obs: 201	y = -1.0622x + 9.4349 (0.0080) (0.0392) R^2 = 0.9882 Obs.: 211

Source of data: NOAA; y means log of rank-0.5, x means log of size; Standard errors in parentheses; For all estimates: p<0.001

This highly significant result of all four estimates ($\alpha \approx 1$) clearly supports a proper detection of German natural cities in accordance with Zipf's law and can be thus directly related to the area of the case study by focusing the respective segmentation at local level. The segmentation of the area of the Regionalverband Frankfurt/Rhein-Main is further explored in the next sub-sections.

5.2 The study area: Region Frankfurt

The study area consists of 75 municipalities belonging to seven rural districts or independent urban municipalities (Frankfurt and Offenbach). The urban center of this region is Frankfurt am Main with about 740,000 inhabitants. The areal surface of the Regionalverband Frankfurt/Rhein-Main is 2,458.45 square kilometers (within the edge coordinates the area is exactly 4,000 square kilometers). By 2015, 2.3 million people have lived in the area. The population density is about 945 inhabitants per square kilometer. Gross value added generated in the region of the RVFRM was 118 billion Euros by 2013 (Further details can be found in the data section of: www.region-frankfurt.de).

Figure 6: The area of the Regionalverband Frankfurt/Rhein-Main

For a more detailed spatial analysis the RVFRM area was cropped from the national image of segmented space. The following extraction of space is in accordance with the edge coordinates of the RVFRM area.[14]

Table 6: RV-Frankfurt-Rhein-Main: Edge coordinates (Lambert-Azimuthal EPSG: 3035)

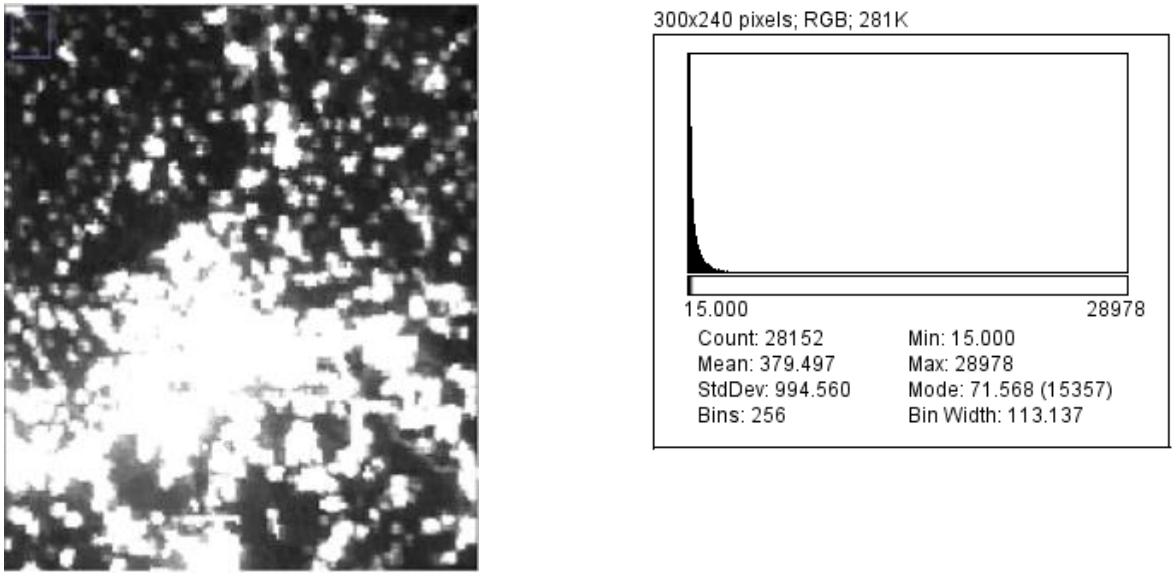
Municipality	Projection-specific coordinates
West: Ginsheim-Gustavsburg	4198662
East: Langenselbold	4255469
South: Groß-Gerau	2974253
North: Münzenberg	3042612

Source of data: RVFRM

From the digital images it is possible to estimate basic moments, such as mean radiance, the spread the distribution and a visualization of the light emission by illustrating it with surface plots for both years regarded. Compared to the national basic statistical moments we see that the range of radiance is substantially smaller across the area of the RVFRM: 15 to 28,978 versus 0 to 484,286 (2012) and 36 to 27,392 versus 0 to 429,162 (2017).

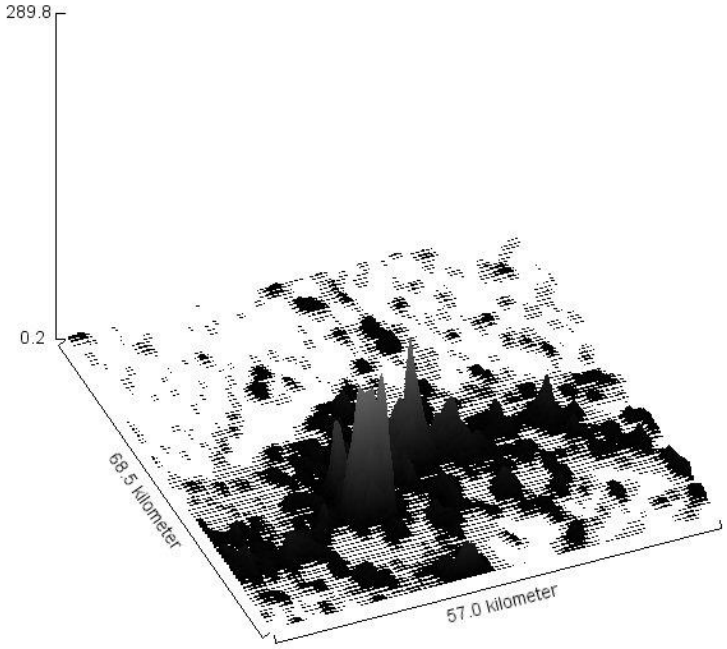
[14] Since the area of the RVFRM comes close to a rectangular shape the use of simple edge coordinates is possible.

Figure 7: Raw pixels (radiance x 100) of the RVFRM area and histogram 2012 with basic descriptive statistics



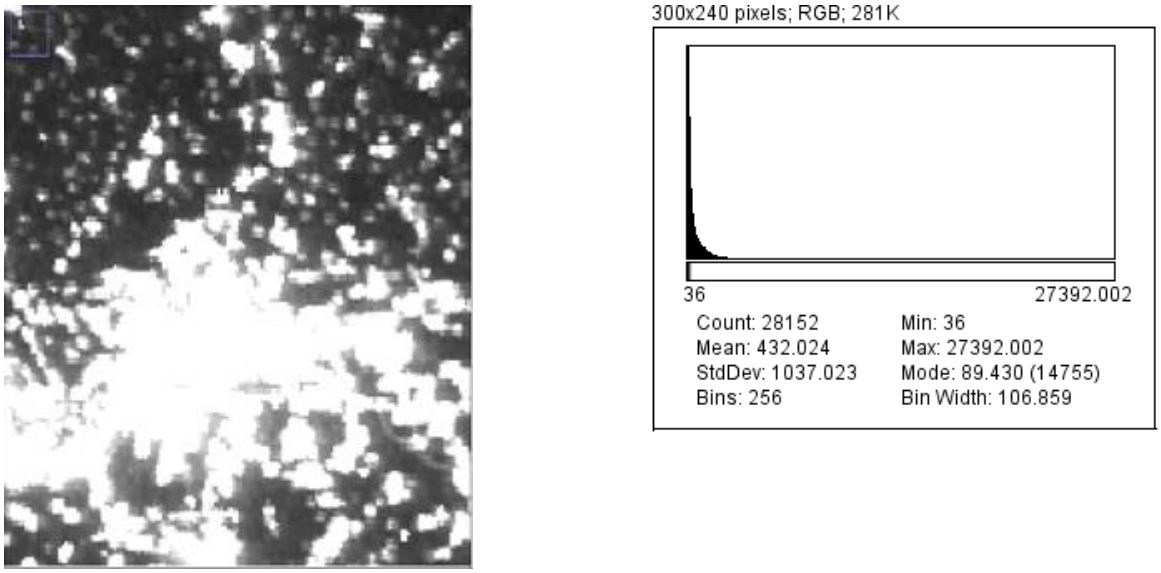
Source of data: NOAA

Figure 8: RVFRM: Surface plot Radiance ($nWcm^{-2}\ sr^{-1}$) in 2012



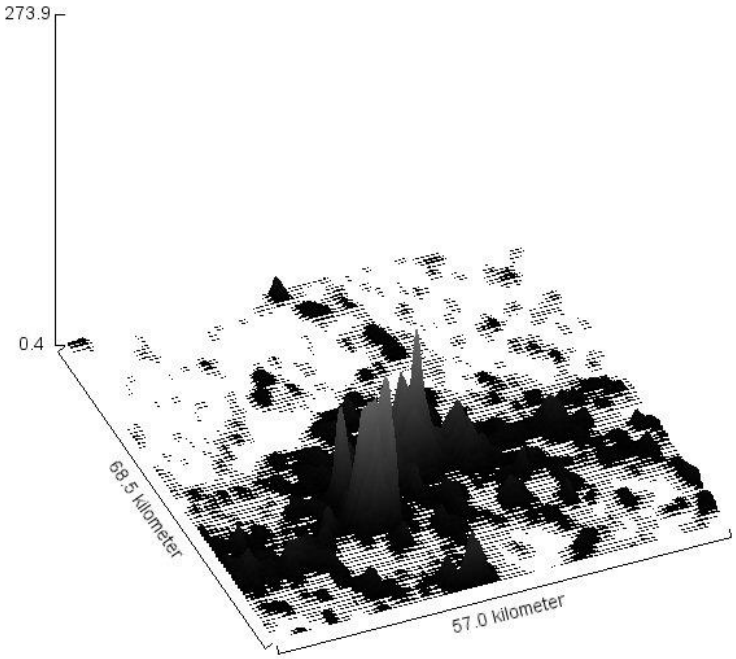
Source of data: NOAA

Figure 9: Raw pixels (radiance x 100) of the RVFRM area and histogram 2017 with descriptive statistics



Source of data: NOAA

Figure 10: RVFRM: Surface plot Radiance ($nWcm^{-2}\ sr^{-1}$) in 2017



Source of data: NOAA

The increase of light emission between 2012 and 2017 is around 14 percent (increase from mean radiance 379.497 to 432.024). Even though radiance is not directly comparable to the unit-free measure of the former radiance-calibrated DMSP-OLS images, it is at least possible to set the last available radiance-calibrated

image (2011) equal to the first VIIRS image (2012) and to roughly estimate the change rates since 1996[15] (until 2017):

Table 7: Estimated change of light emission since 1996

Year / Change	Values (RVFRM)	Values (national)
1996 (rad.cal.)*	44.10	9.66
2011 (rad.cal)*	53.74	11.77
% change	21.86	21.84
2012 (radiance)	379.50	66.24
2017 (radiance)	432.02	84.33
% change	13.98	27.31
Total change %	35.84	49.15

*Source of data: RVFRM *The radiance-calibrated images are also inter-calibrated to remove error implied by changing sensor technologies over time (Hsu 2015)*

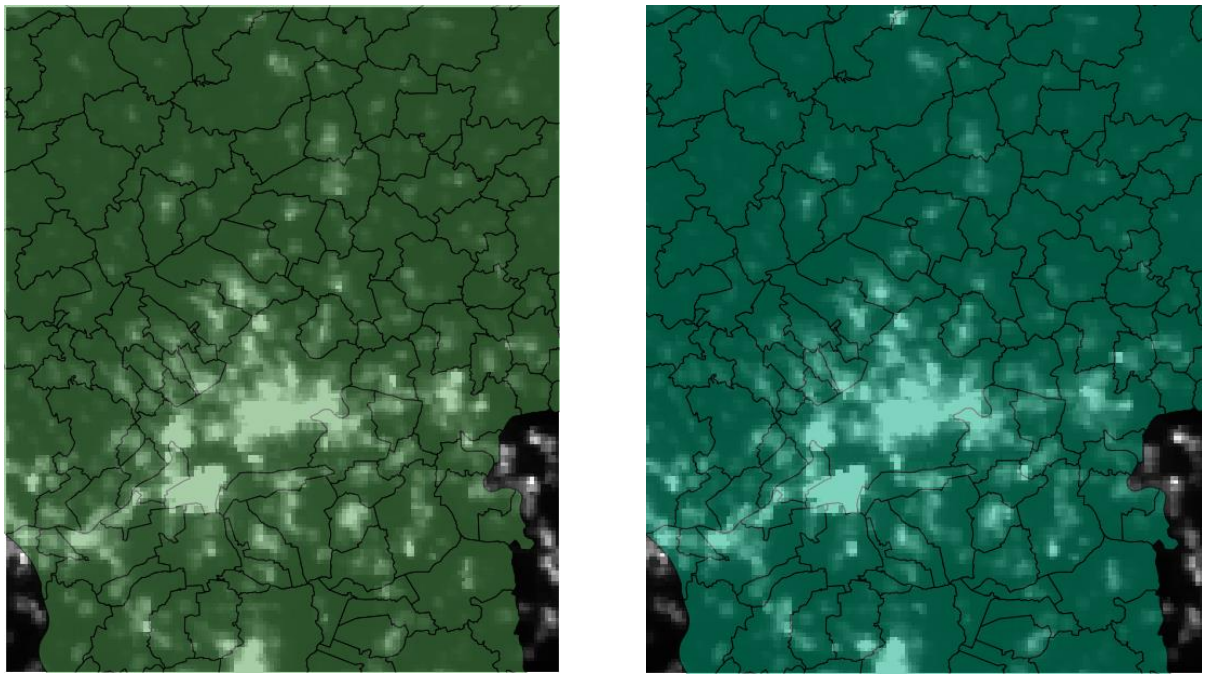
The total change of roughly 36 percent since 1996 is to be compared to the evolution at national level. During 2012 and 2017, the increase of average light emission of the RVFRM region is under-average (14 percent against 27 percent at national level). For 1996 until 2011 the change of light emission at national level is almost identical with 21.84 percent.[16]

How is radiance distributed within the administrative setting of the RVFRM? Within the local municipal boundaries, radiance is expressed as shown in the following maps. The images show that the variation of light is quite strong, even within urban boundaries (e.g. Frankfurt). But the true variation shown by the digital images – as pointed out earlier – is not detectable by visual inspection. There is no remarkable difference in urban extent when comparing the resulting maps of 2012 and 2017 as shown in the following map. But this visible variation is insufficient to determine a statistically determined precise and unambiguous threshold between urban and non-urban. In fact, Frankfurt city appears smaller than a statistical segmentation will show.

[15] The first radiance-calibrated composites were produced for the year 1996.

[16] Due to the completely different technologies, the different resolution levels, the large differences in straylight detection and the comparison of radiance ($nWcm^{-2}\ sr^{-1}$) with a radiance-calibrated unit-free scale makes this rough estimate rather unreliable. It is just presented as an auxiliary approach to monitor secular changes of light emission.

Figure 11: RVFRM: Light emission (radiance x 100) within administrative boundaries: 2012 and 2017



2012 *2017*

Source of data: NOAA

More interesting is thus to zoom-in the segmented area of the RVFRM and to compare the change of urban extent within this region (between 2012 and 2017). Therefore it is needed to focus on the respected patches extracted from the German national segmentation as these define local urban space confirmed at national level by Zipf's law. The extraction of the RVFRM area (2012 and 2017) is as follows:

Figure 12: RVFRM: Layers of segmented natural space: 2012 and 2017 compared

2017

2012

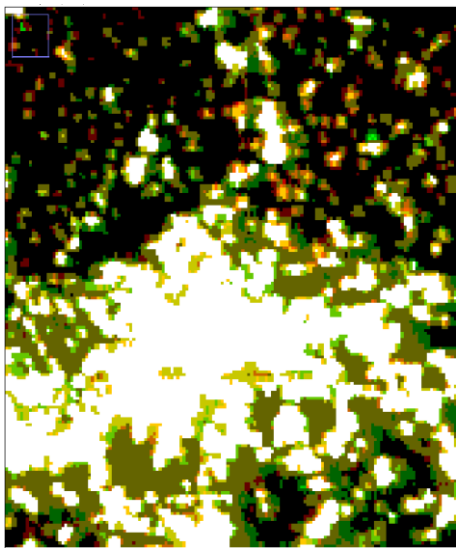
Source of data: NOAA

As expected for a time horizon of five years, Fig. 12 reveals only a minor change of natural urban size.

Also in terms of strength of light emission, distributed over the study area, differences appear minor rather than remarkable. This can be well depicted by the following image showing deviating and co-localized pixels. The yellow to green areas are those with increasing light emission between 2012 and 2017, while the few

brown/red areas are those with decreasing radiance[17]. Hence, few but not all parts of the northern conurbations of the city exhibit a decline of radiance.

Figure 13: RVFRM: Positive (green) and negative (red) change of light emission between 2012 and 2017



(White=saturated pixels in the 8-bit image; Source of data: NOAA)

The result is important (no major change) but not very spectacular.

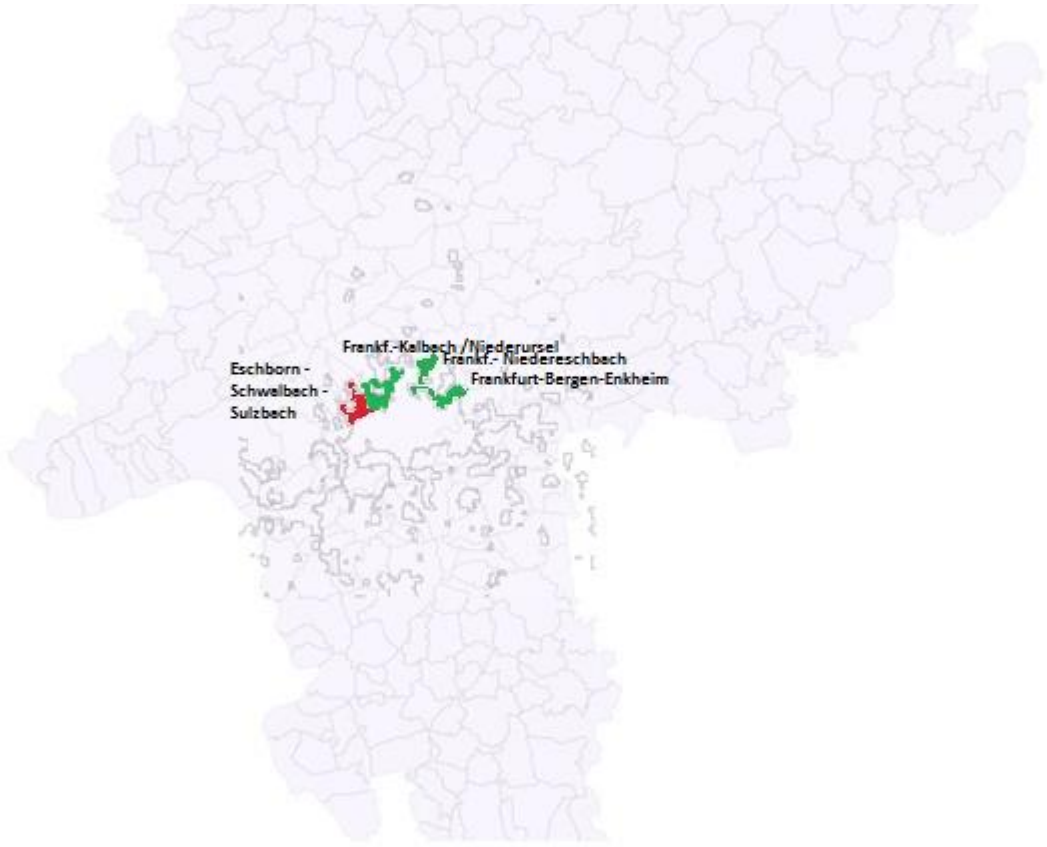
In the context of analyzing rural-urban synergies and to enhance information of true spatial differentiation it is more important to emphasize the analysis of the deviation of natural space from administrative boundaries. The following maps show the comparison between both classifications and years; they clearly reveal that administrative space and natural space differ and that there are both, natural rural areas within administrative urban space as well as *vice versa*. We clearly recognize a larger natural city as compared with Fig. 11 showing just light emission within administrative boundaries.

[17] The analysis is based on an image transformed into an 8-bit format. White areas comprise co-localised pixels (top-coded pixels with DN>255).

Figure 14: RVFRM: Layers of administrative and segmented natural space: 2012 and 2017

Source of data: NOAA

Figure 15: RVFRM: Natural versus administrative urban space 2017 zoomed in



Source of data: NOAA

Fig. 15 illustrates a selection of built environments at the northern urban fringe of Frankfurt. Non-urban/rural zones (green) belong to the administrative city of Frankfurt, peri-urban zones (red) belong to the segmented natural city of Frankfurt. On the one hand we recognize several larger zones that do not qualify as natural urban space. In fact, large parts of Frankfurt-Bergen-Enkheim, Frankfurt-Niederursel and Frankfurt-Niedereschbach consist of agricultural land and with a typical rural physiognomy. On the other hand, there are municipalities in the peri-urban belt that appear as a part of the natural city (e.g. Eschborn,

Schwalbach, Sulzbach, parts of Bad Soden). Furthermore, the cities of Frankfurt, Offenbach and Hanau are connected within the overall natural urban space.

There are several policy- and planning-relevant Implications from that finding:

The natural urban space (independent of administrative boundaries) is well visible by segmentation on light emission data; the change between 2012 and 2017 is very minor. This was to be expected because of the short time horizon. Deviations of natural urban space from administrative space for the core city are, however, remarkable; the natural city comprises a large part of conurbations and neighbor city regions stretching eastwards up to Hanau and south-westwards to Rüsselsheim/Mainz. However, on the one hand, the core city (Frankfurt) is not entirely urban (e.g. parts of Bergen-Enkheim, Niederursel and Niedererlenbach etc. are classified as non-urban). On the other hand, formerly rural and presently peri-urban areas are partly detected as urban, such as Eschborn, Schwalbach, Sulzbach and parts of Bad Soden appear to be directly connected with the city (thus belonging to the natural city).

Hence, the functional space differs considerably within administrative urban and rural areas. The information provided by the segmented images may help in the inter-municipal dialogue on policy and planning. With other words: Function and planning needs could be better defined when considering the differentiated and changing natural space instead of inflexible administrative boundaries.

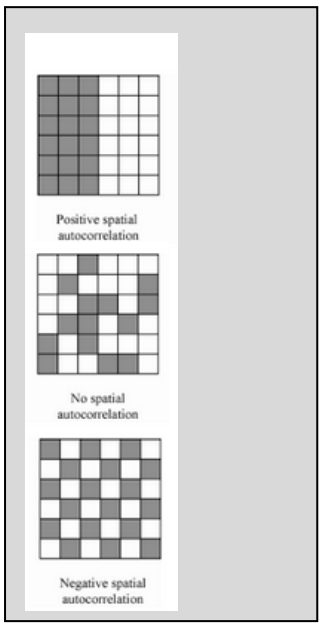
Even though this pattern described above can be also observed on a normal topographical map, the invisible variation of radiance gives a much more differentiated picture. This is particularly interesting in areas where there are settlements closer or further away from Frankfurt city limits; the extent of functional urban space takes course in the respective opposite direction. It is then important to link such results with issues of sub-optimum policy and planning coordination between urban and rural / peri-urban municipalities. A major problem is that municipal boundaries are the core parameter of such coordination (and thus also of local data). For Frankfurt Rhein-Main, but certainly for the other case studies as well, we might be faced with the well-known "Modifiable Areal Unit Problem" (MAUP). With the sat images we can shed light on that in addition to analyses of spatial dependence of light pollution and closely related variables like economic activity or population density.

5.3 Empirical results: Spatial dependence of light emission (Global Moran I analysis of the RVFRM region)

The first law of Geography (Tobler's law) says that "everything is related to everything else, but near things are more related than distant things." For the analysis of rural-urban interaction this spatial relation phenomenon is of major importance. Any influence or causal relationship within a spatial dimension is affected by decreasing spatial influence from contiguous to distant places. The statistical term for that is spatial autocorrelation. It is of major relevance in geo-statistics and spatial econometrics. For the analysis of spatial change (e.g. rural and urban evolution of a region), the analysis of spatial dependence can be useful and also relevant for planning and policy.

The database of radiance usually contains a strong level of spatial autocorrelation, simply due to the fact that bright and dark areas are not confined on single pixels but extend over larger sections of the area. It is impossible to find bright and dark pixels arranged like a chess board. Like inter-areal perturbation of waste gas emissions or emissions of noise, strong light emission from contiguous or neighboring areas will also result in higher light emission in the region regarded. It is to be noted that the higher the spatial resolution is (that is: the smaller the distances between areal units are) the higher will be the spatial autocorrelation.

Thus, the spatial issue of increasing light pollution can be directly monitored by the VIIRS images. Indirectly, when taking radiance as a proxy for environmental and socio-economic variables, it is then also possible to estimate spatial dependence for exhaust gas emission, power consumption, population density and economic activity.



A basic estimation of spatial autocorrelation can be done in viewing the radiance variable by the influence of contiguity for the whole area. It is assumed that mutual influence between two areas (in this case grid cells) is higher if they have a common border. It is possible to differentiate into "Rook" or "Queen" contiguity (in accordance with the possible chess maneuver of both chess pieces). In the following tables the Rook's case is regarded. Rook's case allows a wider range of spatial autocorrelation than the Queen's case.[18] Apart from contiguity, inverse distance is another parameter to be used for spatial autocorrelation. The explanatory power of the spatial autocorrelation coefficient is manifold[19]. In single or multiple fully homogenous areas (every grid cell has the same value), the coefficient is 1. A relatively high coefficient is also obtained for regions that are divided into two of few homogenous zones.

Spatial autocorrelation of many heterogeneous areas tend to be low, while spatial autocorrelation of a structure like a chessboard is perfectly negative (see box).

In fact, Moran's I depends on both, on size of the units and the size of space regarded. The Moran I for a region with a pixel resolution of the original VIIRS composites is much higher than Moran I for the same region with a resolution of 1 square kilometer (thus showing the average pixel value at one square kilometer). Spatial autocorrelation is thus not a constant to be used. It also neither contains a negative nor positive message, since both positive as well as negative neighborhood spill-overs are possible. When it comes to a functional differentiation of land use, e.g. the functional differentiation of von-Thunen-Rings around cities, Moran I is higher the more the separation is visible. Once urban activities more and more perturb into peripheral rural rings, spatial autocorrelation might slump to some extent because the functional space is not anymore clearly separated.

Change of spatial autocorrelation between 2012 and 2017 is marginal. This can be demonstrated by the following comparison:

Table 8: Area of the RV Frankfurt/Rhein-Main: Global Moran I 2012 (raw radiance data)

GRID	CONTIGUITY	MORAN_I
RVFRMLambertlit2012	Rook's case	0.895555

Estimated by QGIS (Saga command Moran I); data source: NOAA

[18] A chessboard is an example of autocorrelation of -1, if Rook's case is applied. For the Queen's case there would be still a positive influence due to the additional diagonal contiguity.

[19] Spatial autocorrelation is also to be considered in spatial econometric modelling such as the Spatial autoregressive Model (SAM), the Spatial Error Model (SEM) in order to avoid inefficient and inconsistent estimates implied by non-independently and non-identically distributed residuals in the presence of spatial autocorrelation. (Cf. Anselin and Florax (eds.) 1995)

Table 9: Area of the RV Frankfurt/Rhein-Main: Global Moran I 2017 (raw radiance data)

GRID	CONTIGUITY	MORAN_I
RVFRMLambertlit2017	Rook's case	0.905927

Estimated by QGIS (Saga command Moran I); data source: NOAA

There is no remarkable change of spatial dependence at the level of pixel size. At the level of one square kilometer the global spatial autocorrelation is lower because of the lower resolution, but the change is likewise negligible as the following comparison shows:

Table 10: Area of the RV Frankfurt/Rhein-Main: Global Moran I 2012 (mean radiance data at one square kilometer grid)

GRID	CONTIGUITY	MORAN_I
2012KM	Rook's case	0.782577

Estimated by QGIS (Saga command Moran I); data source: NOAA

Table 11: Area of the RV Frankfurt/Rhein-Main: Global Moran I 2017 (mean radiance data at one square kilometer grid)

GRID	CONTIGUITY	MORAN_I
2017KM	Rook's case	0.796854

Estimated by QGIS (Saga command Moran I); data source: NOAA

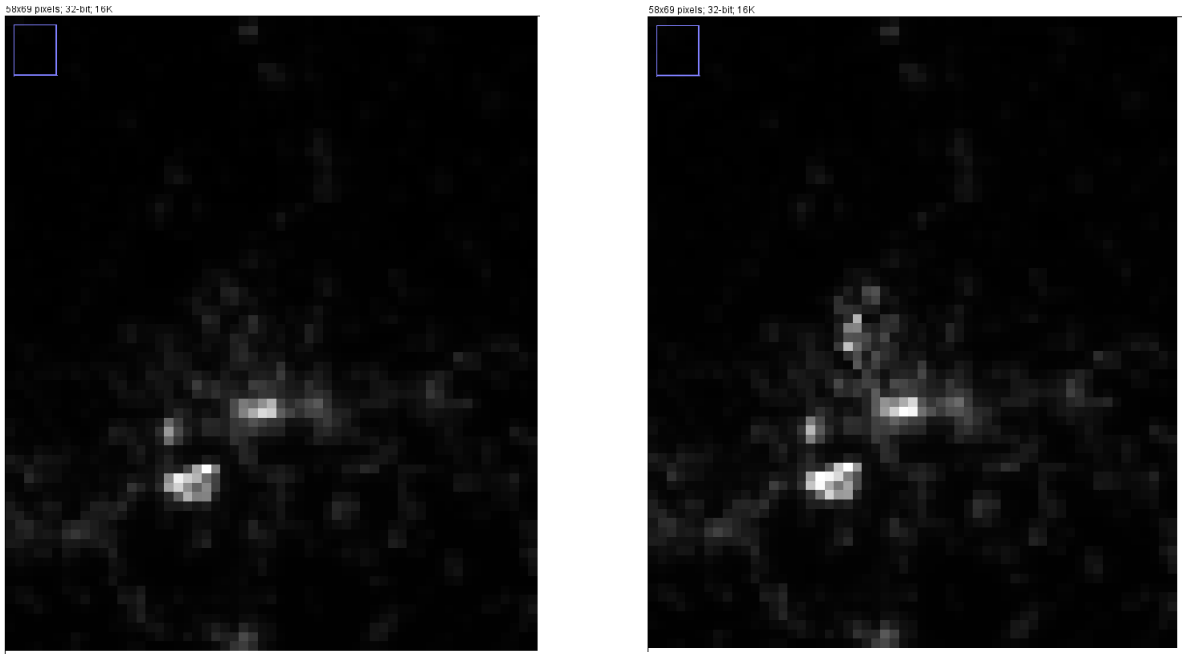
If we use light emission to segment functional space of a city region including its peri-urban conurbations and the rural periphery around (i.e. the ideal concentric segmentation of von Thunen rings), this segmentation should become visible by a relatively high autocorrelation coefficient (global Moran I). Hence, there are distinctly separated zones of different functions with a narrow transition band; regions are functionally defined and not converging. If cities are growing and conurbations and rural zones are further developed, the former functional division of land use will be more and more dissolved; the influence of urban forces in rural areas will increase and reduce spatial autocorrelation effects of rural specificity (e.g. low light pollution levels or low environmental degradation). Whether this process takes place uncontrolled (urban sprawl) or scheduled in the context of a land development plan, in both cases there will be essentially a modification of functional spaces and relationship between those. Hence spatial autocorrelation will decrease until boosting urbanization (development of new urban clusters merging around the city) will again increase spatial autocorrelation in the whole region. The evolution of spatial autocorrelation around growing cities over time takes thus a U-shape describing secular change of spatial functionality (cf. also: Yu and Wei 2006; Smętkowski M 2015, p.545)

Monitoring the evolution of spatial autocorrelation is of particular interest when forecasting scenarios of spatial development. Monitoring of spatial dependence might become highly important for urban-rural coordination in local policies and planning, especially in regions undergoing rapid growth processes. Therefore it is quite interesting to view spatial autocorrelation at the respective local levels and for certain urban development axes. The change of the global Moran I coefficient directly sheds light on spatial change with its social and environmental implications. Such an analysis can be done by image analysis along so called “concentric piece-of-cake” selections around cities. The following example should shed further light on that.

5.4 Simulation study new suburb along the urban fringe Frankfurt - Steinbach

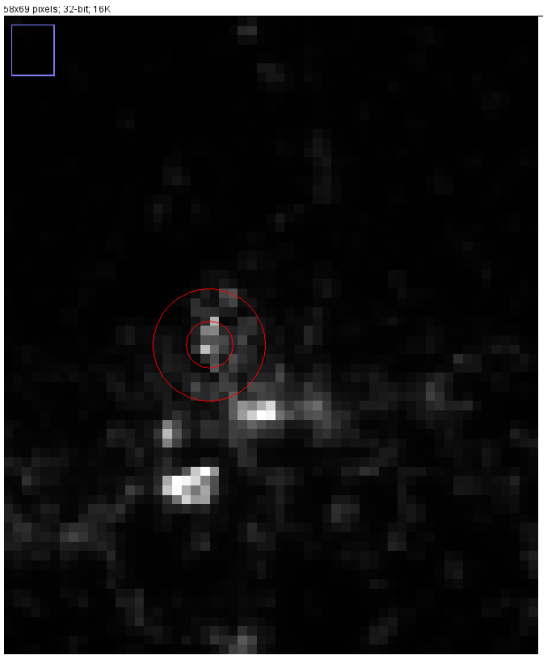
In order to relieve the enormous immigration pressure and further population growth there are major projects of Frankfurt city to develop open space within the city limits. One important project is the development zone at the north-west boundary to the municipalities of Steinbach and Oberursel. This area should evolve to a major suburb for dwellers and business. To estimate how spatial autocorrelation will change by that urban project, we simulate light emission for a time horizon of about ten years ("2027"), assuming a further mean increase of around 20 percent at national level, a further increase of a five-times higher light emission at the center of the selected development site and around two-times higher light emission at its periphery (from agricultural land use to residential and commercial land use). In Fig. 16 the positive change is shown by the different levels of the green spectrum, the white and black ones are co-localized pixels (with no or little relative change).The simulation does not necessarily represent the true future extent of new urban space and the corresponding light emission stemming from that. It is simply a statistical exercise of extrapolation of light emission for the next ten years in addition to a local enhancement of light emission effected by a change of land use in a certain area. Such a simulation forecast may help to illustrate the effect of developing further urban space of such a dimension.

Figure 16: Simulation of the spatial effect of a new suburb on urbanization (Frankfurt)

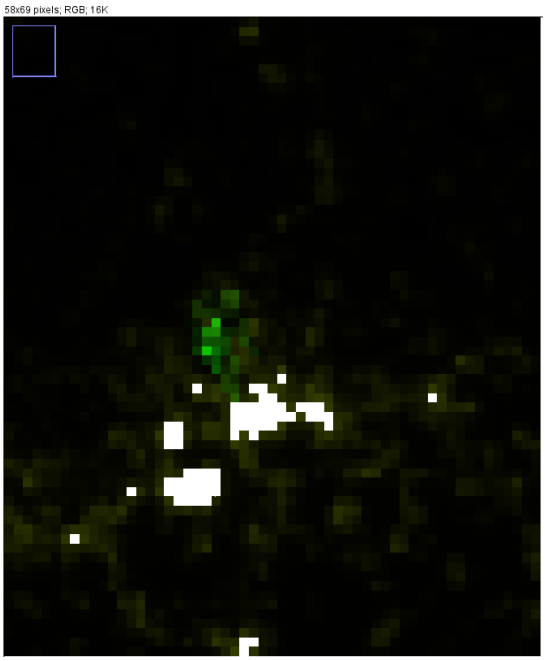


Light emission (1 px = 1 sq. kilometer) 2017

Simulation "2027"



Identification of the new suburb "2027" (center and periphery)



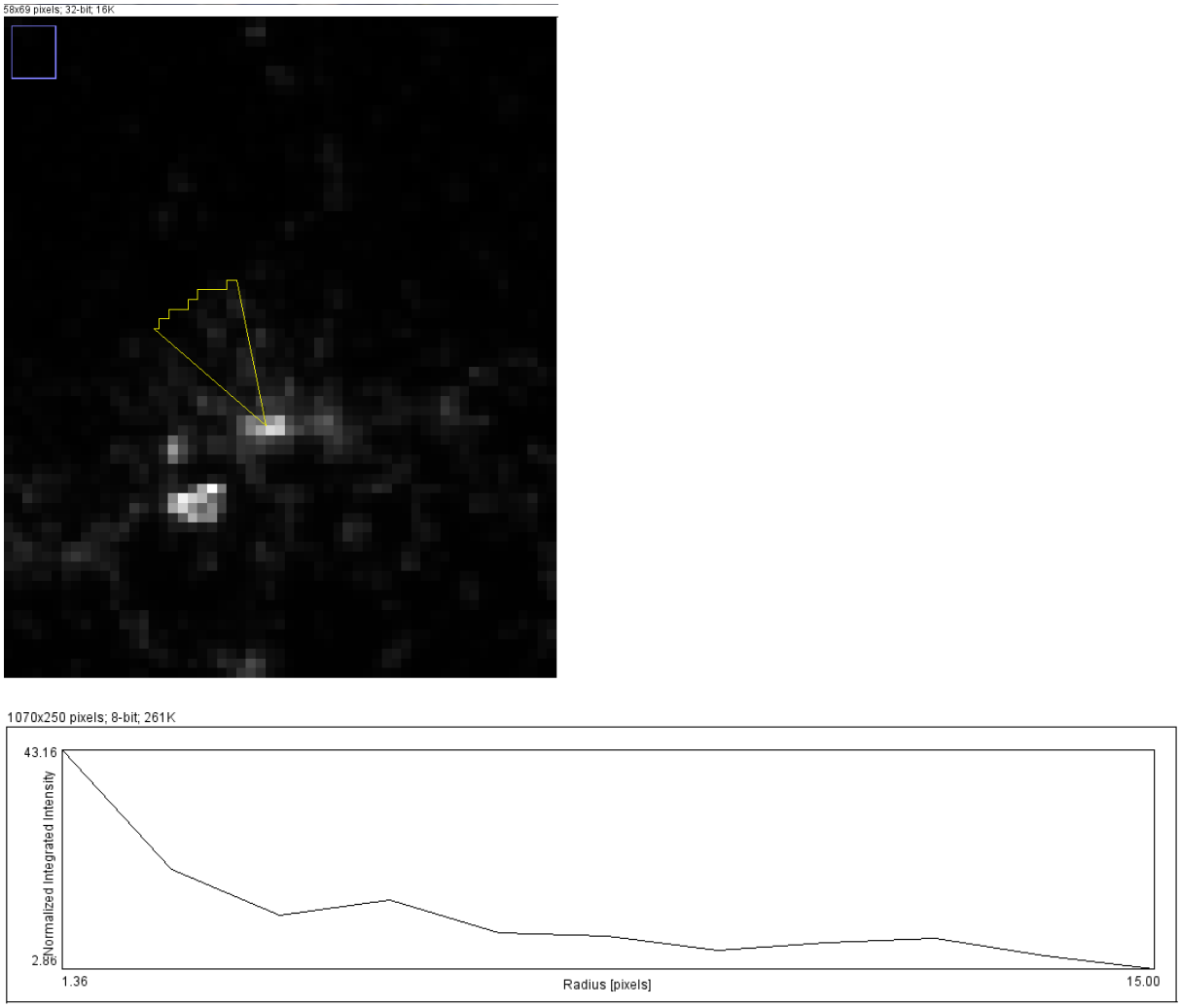
Co-localization and change analysis 2017/"2027"

(Sources of all figures above: NOAA; own calculations)

It can then be shown how the radial profile changes along the North-West development axis (piece-of-cake selection) from Frankfurt central city. For the modelled analysis (pixel=kilometer) we specified the city center at x=27 kilometer and y=42 kilometer of the following images. The radius for the analysis is 15 kilometer (=15 pixels), the starting angle is 120° and the integration angle is 20° to roughly capture the projected development zone.

The results for 2017 (normalised integrated intensity=pixel mean per concentric pixel circle) are as follows:

Figure 17: Frankfurt north-west axis: Radial profile results 2017



ImageJ OLS estimate:
Formula: y = a+bx
Sum of residuals squared: 460.27485
Standard deviation: 6.78436
R^2: 0.63543
Parameters:
a = 29.19985
b = -1.98040

(Sources of all figures above: NOAA; own calculations)

The change of spatial autocorrelation is negligible for the whole region (Table 10 and 11), but considerable for the selected city area (Table 12).

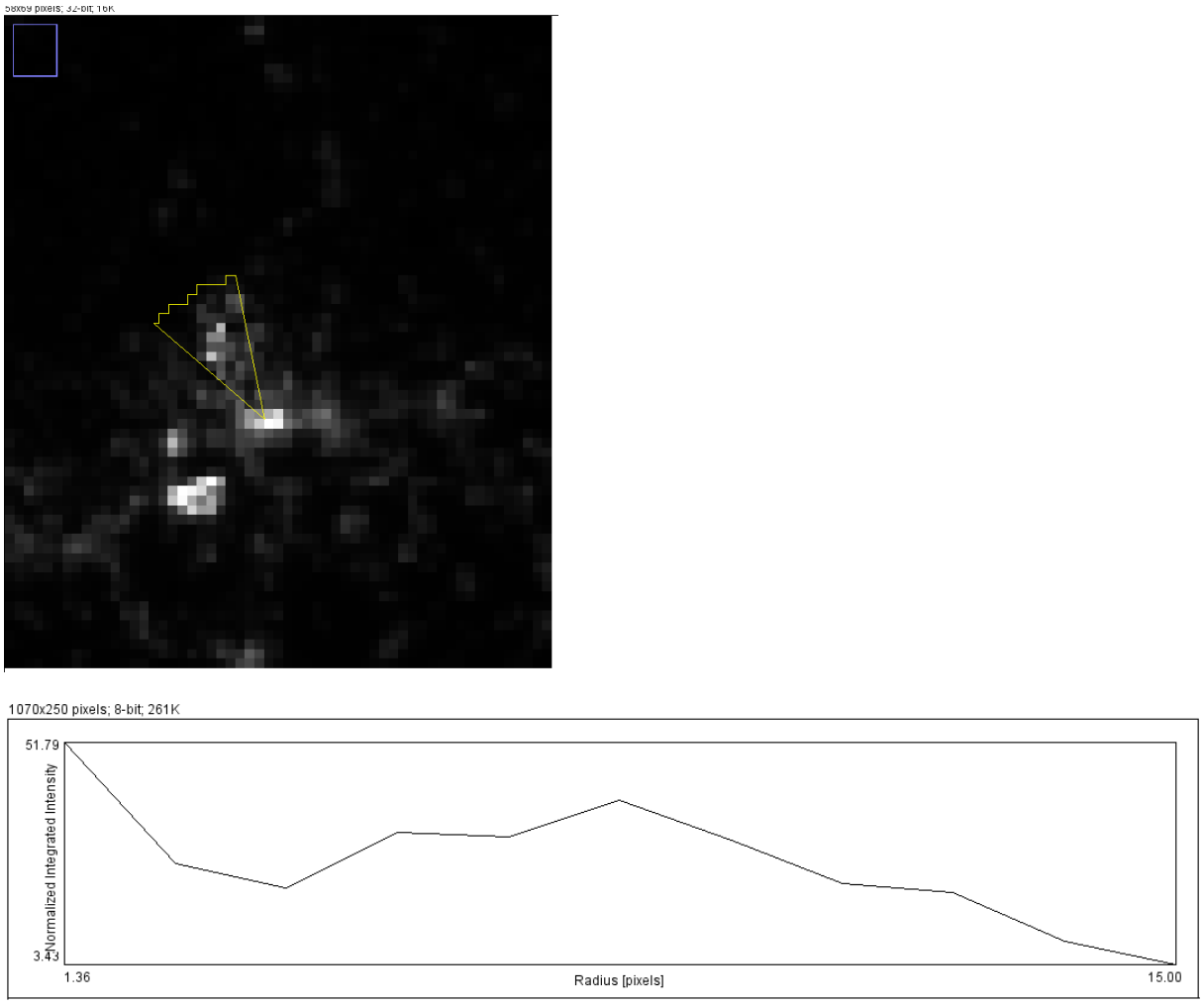
Table 12: Area of the RV Frankfurt/Rhein-Main: Global Moran's I for "2027" (mean radiance data at one square kilometer grid simulated)

GRID	CONTIGUITY	MORAN_I
2027KM	Rook's case	0.783983

Estimated by QGIS (Saga command Moran I; data sources: NOAA own calculation

The profile in the distance between five and twelve kilometers has markedly changed as shown by the radial profile in the next figure. [20]

Figure 18: Frankfurt north-west axis: Radial profile results "2027"



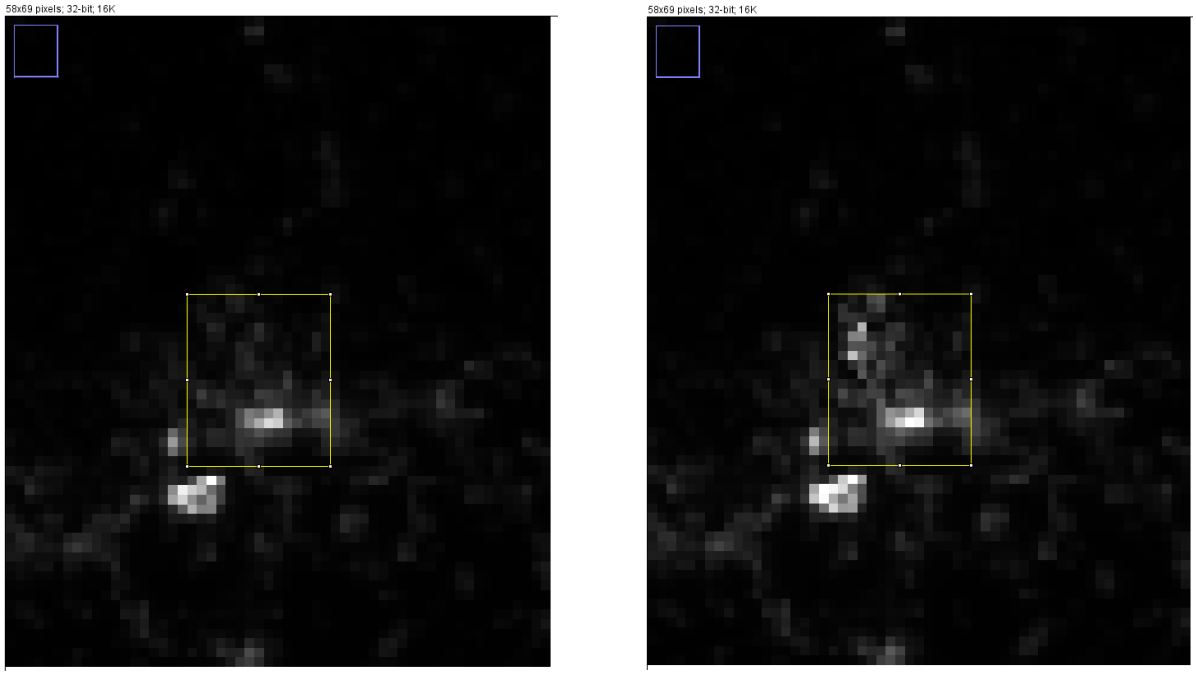
[20] The mean pixel values of the concentric circles can be then further used for a local Moran I (LISA) analysis in order to see how spatial dependence has changed at individual pixel level over that simulated time horizon.

ImageJ OLS estimate
Formula: y = a+bx
Sum of residuals squared: 825.84840
Standard deviation: 9.08762
R^2: 0.55268
Parameters:
a = 44.08254
b = -2.23347

(Sources of all figures above: NOAA; own calculations)

If we crop the map of the "Regionalverband" and zoom in a selection of Frankfurt (approximately defining the city) we see that spatial autocorrelation is decreasing when simulating the new town 2027. Space will become more heterogeneous with increasing distance from the center. If viewing the north-west development axis in 2017 there is still a continuously declining mean radiance, hence spatial autocorrelation is well determined by the existence of fairly homogenous urban and rural/peri-urban parts. In modelling "2027" the same axis becomes more heterogeneous and there is a spread of stronger illuminated space besides less illuminated one. This not only determines the lower R^2 in the linear regression but also the lower spatial autocorrelation estimate (0.69 versus 0.79). Hence, the urban-rural gradient in this corridor is disturbed by increased spatial variation of radiance, because the former rural or environmental function in this important urban – peri-urban corridor is reduced (trends of functional dissolution).

Figure 19: Frankfurt: a selected city region cropped for 2017 and "2027"



2017 *Simulation "2027"*

(Sources of all figures above: NOAA; own calculations)

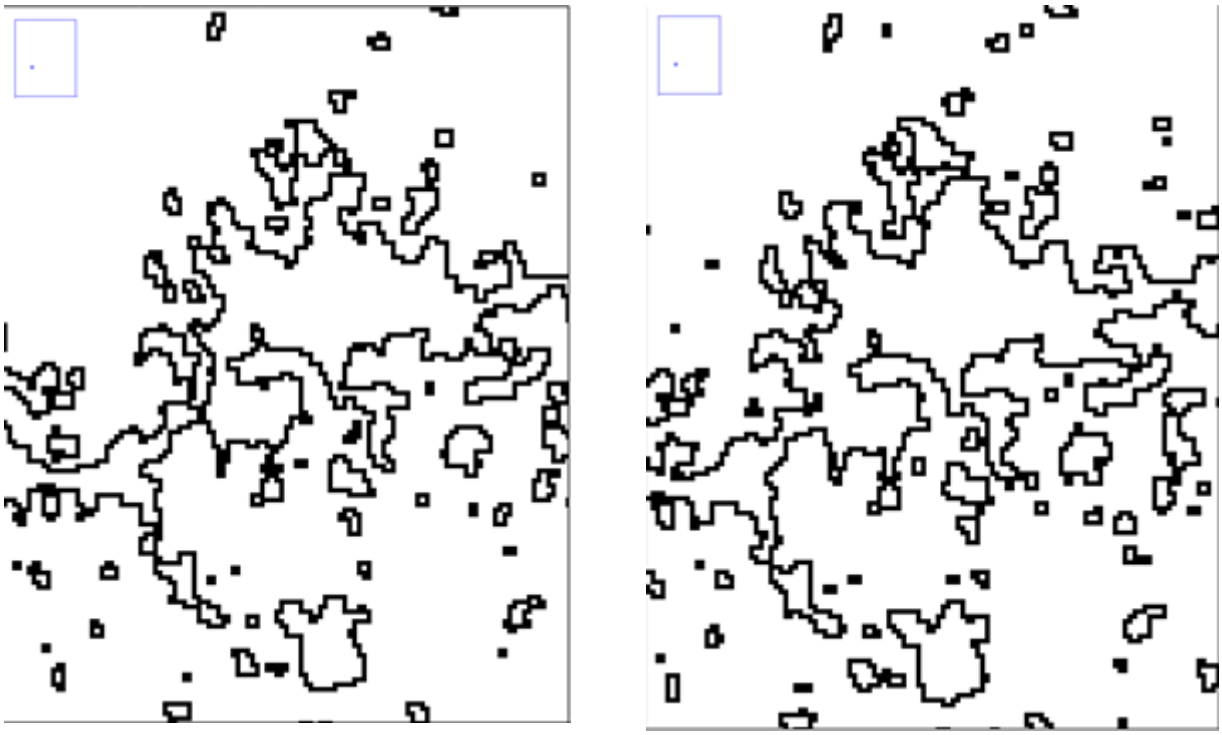
Table 13: Moran I for a selected city region of Frankfurt (2017 and 2027 simulated)

Year	Moran I
Moran's I (Frankfurt2017KM)	0.792200
Moran's I (Frankfurt2027KM)	0.690827

Estimated by QGIS (Saga command Moran I; data sources: NOAA, own calculation

The implication of change of spatial dependence can then be also made visible in the sphere of spatial heterogeneity (change of urban size). In this case we simulate the change at the level of pixels instead square kilometers using the 2017 threshold and an increased threshold based on a corresponding simulation for "2027". After segmenting the local urban space at a 20 percent higher threshold (based on the assumption of a corresponding increase of national light emission until 2027), the size and shape of the natural city will have changed considerably, as shown by the following figure:

Figure 20: Local segmentation: Frankfurt area 2017 and "2027"



2017 (original threshold: 810.82) *"2027" (new threshold: 972.98)*

Sources of the figures above: NOAA; own calculations

This simulation exercise demonstrates how a change of functional space of small sub-areas may have an over-proportional impact on natural city growth, because of progressively merging of natural urban space (such as the northern part of Frankfurt going to merge with the municipalities of Oberursel and Steinbach). In the above simulation model, a synergetic pattern of sustainable urban-rural evolution, with maintenance of functionality, is clearly absent in the respective zones of the urban fringe.

It is to be admitted that in theory a more precise segmentation can be only inferred from the national image in 2027, since here the right segmentation can be again tested by Zipf's law. In our exercise we simply

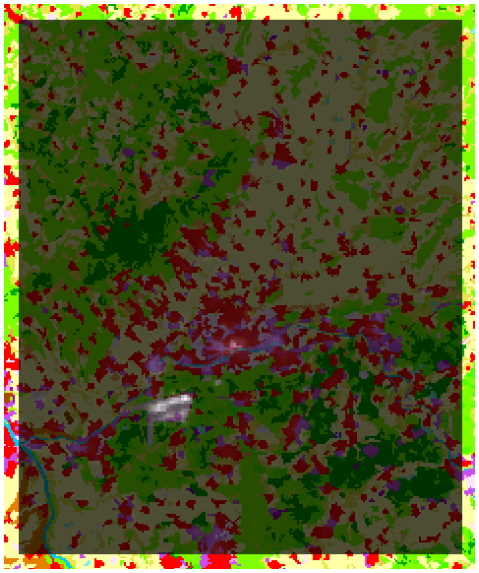
multiplied the pixel matrix by the factor 1.2 and manipulated the suburb region as for the grid map at square kilometer resolution. The problem is, however, that an isolated manipulation (such as artificially increasing local light emission for simulation purposes) would automatically distort the rank-size distribution of German cities as it directly violates Gibrat's law of city size growth being independent of its original size. With other words: It is unlikely that until "2027" light emission of the entire German space will have increased homogenously by *X* percent while just only for an isolated part of Frankfurt there is an additional increase (*X + x*) due to a recently constructed local suburb.

5.5 At what spatial level and for which variables can we use light emission as a proxy?

In addition to spatial segmentation and the analysis of spatial dependence, data from nocturnal satellite imagery can be alsso associated with other relevant grid data. We first tested to relate the distribution of light emission from the 2012 image to the 2012 CORINE land use map provided by the European Commission. On the following semi-transparent stack showing typically urban and rural land use types we can detect the following:

Light emission is particularly strong at the red delineated built-up urban areas in the core city stretching to Bockenheim and Rödelheim, some commercial areas in the East of the city and the industrial Park Frankfurt-Höchst. Most flashy is the area classified as transport infrastructure at Frankfurt airport. Hence, the spatial relationship between land use and light emission suggests a close association between light emission and economic activity. This relationship is recognizable by visual inspection.

Figure 21: RVFRM: Land use (CORINE) and light emission 2012



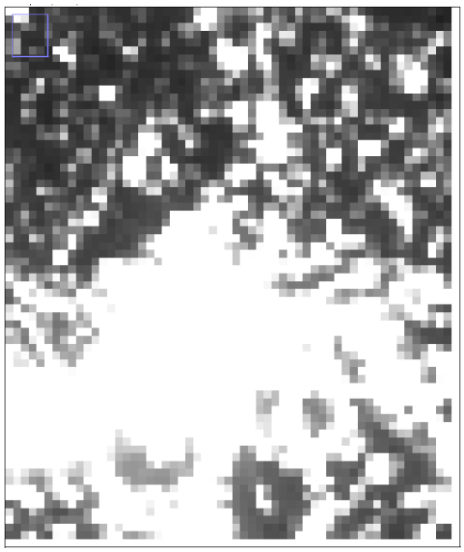
Source of data: NOAA, European Commission

But is this apparent association also confirmed by numerically transformed image data? We therefore also tested the relationship between light emission and more differentiated data of the socio-economy, land use and ecology. For that purpose we used the database of the IÖR-Monitor and transformed the original radiance images into maps with average radiance per square kilometer in accordance to the INSPIRE directive.

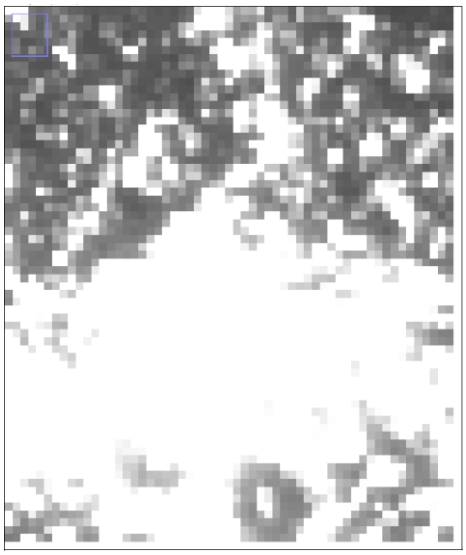
As shown earlier, there are high and highly significant correlation coefficients between light emission and certain socioeconomic or environmental variables at global level. At lower spatial levels this association is expected to be lower, at least for some variables. The reason is a decreasing variation of land use the smaller the regarded spatial area is. Hence if there is a high correlation between purchasing power and light emission at the level of a country, there could be a much lower association at the level of a city because people with a high income might live in wealthy quarters with moderate street lighting while poorer people perhaps live in more central and stronger illuminated parts of the city. It is therefore the intention to compare different types of sub-areas (entire region, rural, peri-urban, urban) to find out whether and where there appear systematically higher or lower correlations.

The two images 2012 and 2017 that show mean radiance (x 100) at one square kilometer grid level can be now directly related to other grid variables, such as Microm® (for Germany and Austria) and IÖR-Monitor.

Figure 22: RVFRM: Light emission at one square kilometer grid



2012 (radiance x 100)



2017 (radiance x 100)

Source of data: NOAA

The following table shows the image correlation at the resolution of a of one square kilometer grid with some relevant variables from the IÖR Monitor at the regional level of the Regionalverband:

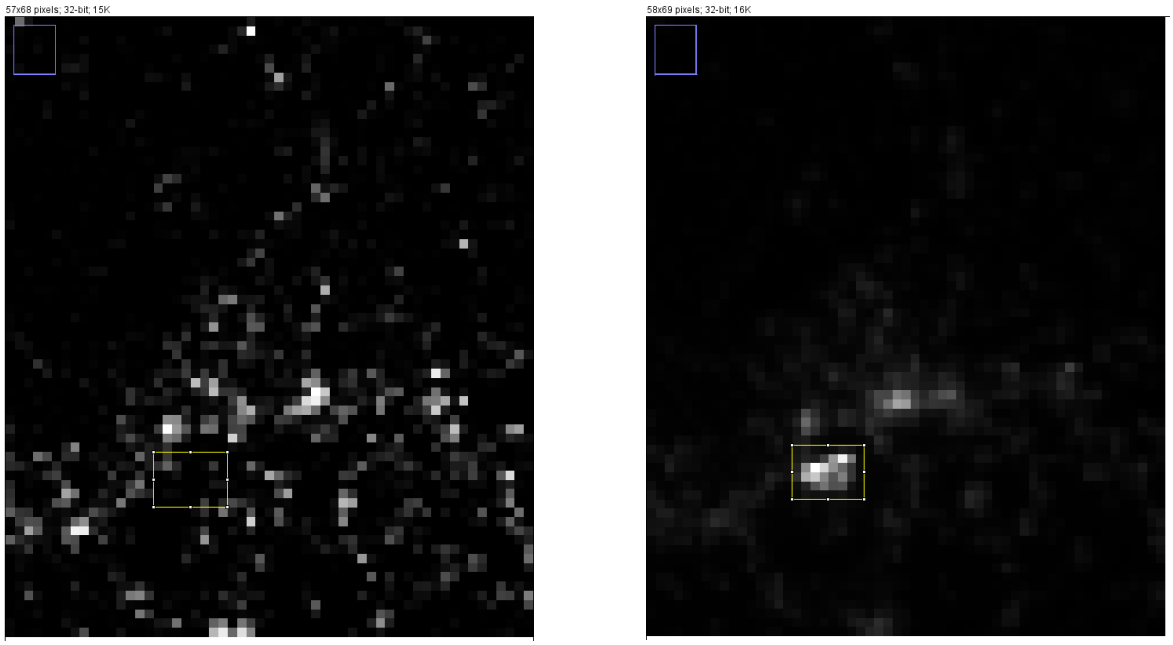
Table 14: RVFRM: Image correlation 2011/2012 and 2016/2017

Variable / Year	Image correlation: *Pearson r* (2011/2012)	Image correlation: *Pearson r* (2016/2017)
Areal share of Commercial estate	0.36***	0.40***
Motor traffic density	0.39***	0.42***
Areal share of sealed ground	0.60***	n.a.
Housing density	0.28**	n.a.

*Sources: NOAA, IÖR-Monitor (WCS layer); (**: $p \leq 0.05$, ***: $p \leq 0.01$)*

The level of association with certain land use variables appears quite lower than expected. Especially the share of commercial estate (thus an indicator closely related to economic activity) is only at 0.36 or 0.40 respectively. However, a closer look at the grid maps of the IÖR-Monitor reveals some peculiarities, such as the Rhein-Main airport not being largely displayed as a commercial estate (see earlier). While on the maps showing radiance, values around the airport are very high, the measured share of commercial estate appears not only low; the variation of both selected clippings suggests even a negative correlation as the two following images from 2011/2012 illustrate.

Figure 23: RVFRM: Comparison of light emission with share of commercial estate 2011/2012



Share of commercial estate 2011 *Radiance 2012*

Sources of data: NOAA, IÖR WCS layer; 1 px = 1 sq. kilometer

The only variable showing some closer association at the regional level of the Regionalverband is the share of sealed ground, with a *Pearson r* of about 0.60. The lower coefficient for housing density could have been expected. Usually, housing areas are less illuminated than areas with strong commercial activity, such as commercial estates or road traffic.

We also tested to merge the variables housing, commercial estate and transport infrastructure as different kinds of land-use representing economic activity (as a multiplicative composite variable) and to correlate it with night light emission. This test we executed for the entire region of the RVFRM, the city area of Frankfurt and a larger more rural area in the north-western part of the RVFRM:

Table 15: Image correlation between light emission and a composite land-use indicator for different types of area 2012

Area	Size (km²)	Image correlation 2011/2012 (*Pearson r*)
Entire area of the Regionalverband (gross size based on edge coordinates)	4,002	0.56***
Larger city area Frankfurt	494	0.51***
Selected area in the north-west of the RVFRM	735	0.77***

*Sources of data: IÖR, NOAA (***: $p \leq 0.01$)*

We first see that correlation between areal share of housing, transport and commercial estate (composite variable) and light emission is remarkably stronger (r = 0.56) than for the selected land-use variables alone. Furthermore, we see that for areas with stronger economic activity (i.e. the city region of Frankfurt) the association between both variables is slightly lower, while for areal parts with less economic activity, correlation is substantially stronger (r = 0.77).

We finally examined the strength of the composite land-use variable as a predictor of radiance in the context of a standard spatial econometric procedure. As a case for that we selected the larger city area of Frankfurt. The estimation consists of the effects of the predictor and that of the spatial distance, the latter either as an autoregressive effect (SAR):

$$y = \rho W y + X\beta + e,$$

or as part of the error term (SEM):

$$y = X\beta + u, \qquad where: u = \lambda W u + \varepsilon.$$

Y means the dependent variable, X represents the predictor to be estimated and W is a row-standardized weight matrix (inverse distance) with 494 cells of one square kilometer each. The error terms e and ε are assumed i.i.d. with mean 0 and finite variance ($\mu=0$, σ^2).

Table 166: Spatial econometric estimation of the composite variable (Larger city area Frankfurt 2011/2012)

Model	ML-SAR	ML-SEM
Composite (X)	0.34 (0.000)	0.54 (0.000)
Constant	8.63 (0.000)	32.55 (0.000)
ρ	0.65 (0.000)	
λ		0.67 (0.000)
Log likelihood	-1262.28	-1264.40
Wald test of ρ=0 / λ=0	156.266 (0.000)	154.101 (0.000)

Sources of data: NOAA; IÖR

Based on the weight matrix, both model types are suitable as confirmed by the Stata diagnostic tests for spatial dependence in OLS regression (spatdiag). In the two models, both, the spatial influence as well as the predictor are highly significant and confirm radiance to be well predicted by housing, commercial estate and transport infrastructure at the scale of Frankfurt city.

Another analysis of variable association has been executed for some of the Microm® variables. This analysis is subject to the second study dealing with socio-economic analysis of spatial change at grid level of one square kilometer.

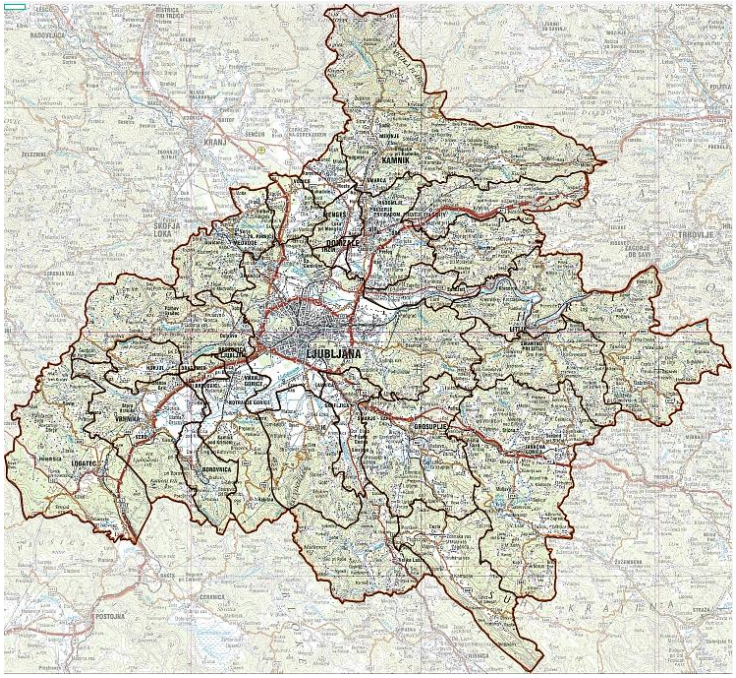
6 Case study “Ljubljana Urban Region” (LUR)

Ljubljana Urban Region (LUR) is the second case explored by the VIIRS images. LUR is growing region with major importance for a smaller EU member. But in contrast to the RVFRM is has not yet obtained the economic weight of a European growth pole.

The Slovene metropolitan region unites 26 municipalities and communities in the center of the country. It is the region with the most advanced potential in knowledge generation and creative potential. Key scientific, research, educational and cultural institutions of Slovenia are concentrated here. Furthermore, a large share of the private sector maintains its headquarters in this region, thus providing substantial employment opportunities and creating more than one third of Slovenia’s gross domestic product. Ljubljana Urban Region is thus the economic growth pole of the country. More than 500,000 people live in the region with the highest level of education and the highest added value per employee in Slovenia (http://www.rralur.si/en/regija/region).

The following map illustrates the region:

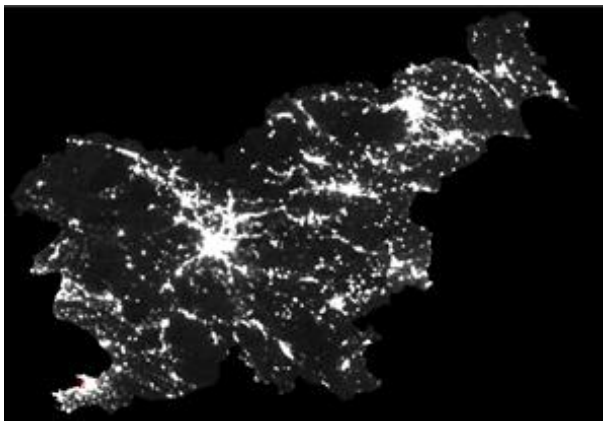
Figure 24: Map of Ljubljana Urban Region

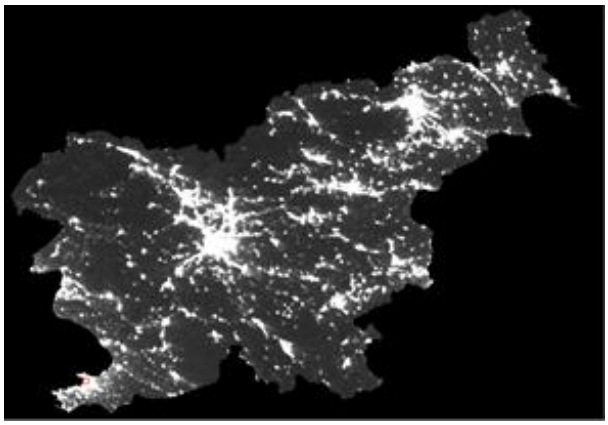


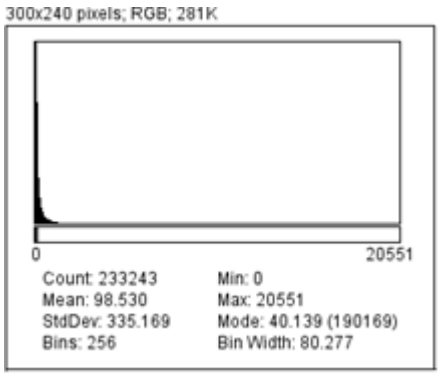
Source: www.rra-lur.si

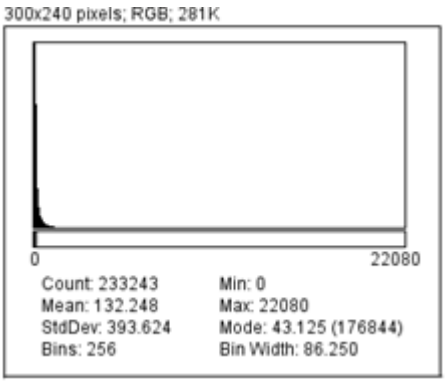
Before estimating the extent of natural cities in the LUR region, the national distribution of light emission needs to be viewed and classified. The following images, histograms and descriptive statistics on distribution, range, spread and means again show an extremely skewed power distribution.

Figure 25: Images and Basic moments of pixel distribution: Slovenia 2012 and 2017 (radiance x 100)









2012 *2017*

Source of data: NOAA

The brightest pixels are located close to the port area of Koper, thus the maximum radiance is found outside the study area.

For the national segmentation of urban versus non-urban space we have applied the same approach as for the RVRFM, i.e. pixel outlier removal followed by univariate k-means segmentation.

The following tables and figures show the cut-off points of outlier removal in accordance to Vandervieren and Huber (2004), the resulting distribution with basic moments and the corresponding segmentation of natural cities 2012 and 2017.

6.1 Empirical results: National segmentation of space – natural cities (Slovenia)

Segmentation of space has been established by k-means clustering after adequately trimming the pixel database (removal of outliers in a power distribution by the Vandervieren-Huber approach).

Table 177: Slovenia: Radiance outlier removal and cluster centroids(based on radiance x 100)

Year	Cut-off point	Percentage of removed pixels	Threshold	Cluster centroids
2012	2,087	0.50%	458.32	23.28 | 1,131.63
2017	2,501	0.50%	568.85	24.88 | 1,481.81

Source of data: NOAA; 0.5 percent of 233,243 px = 1166 ranks

The resulting thresholding of space with ImageJ (k-means clustering) has resulted in the following maps of Slovenia 2012 and 2017:

Figure 26: Natural cities: Slovenia 2012

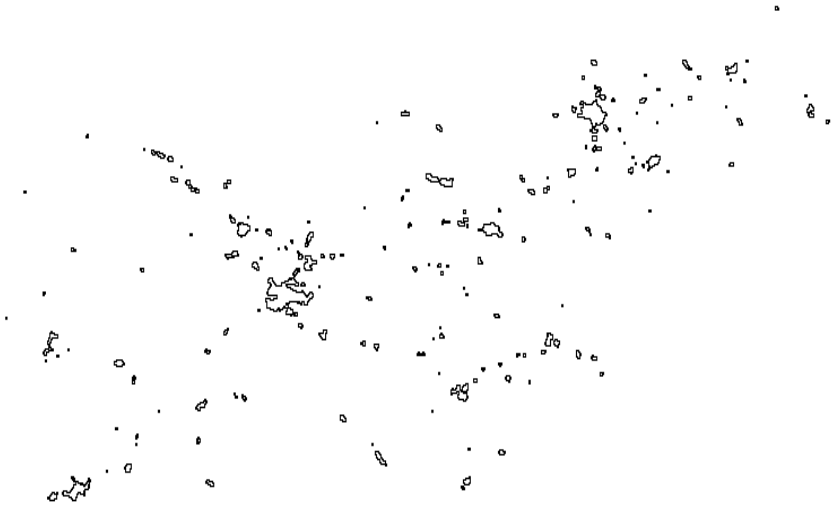
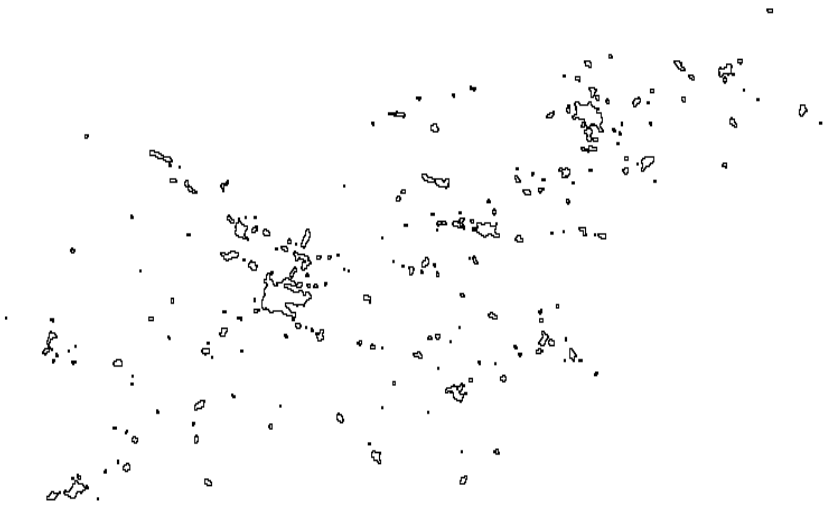


Figure 27: Natural cities: Slovenia 2017



Source of data for both figures above: NOAA

All patches have a different size in terms of pixels. Derived from the rank-size distribution of the segmented patches 2012 and 2017, a simple non-linear Pareto regression for cities larger than 50 pixels would support Zipf's law perfectly (2012: $y = 562.65x^{-0.979}$, $R^2 = 0.9631$; 2017: $y = 500.7x^{-0.965}$, $R^2 = 0.9495$). Even for cities larger than 25 pixels, Zipf's law would be still supported. However, the critical issue when dealing with a small number of observations is the bias implied by an underestimation of standard errors. This has been explained earlier. Therefore, the *Gabaix-Ibragimov* estimator is applied in this case. As we see, estimates are not that precise any more:

Table 188: Gabaix-Ibragimov-estimates of natural city rank-size distribution of Slovenia 2012 and 2017

Year / nat. city sizes	> 50 pixel	> 100 pixel
2012	y = -1.1932x + 7.1139 (0.0735) (0.3489) $R^2 = 0.9635$ Obs.: 12	y = -1.1678x + 6.9221 (0.1518) (0.8176) $R^2 = 0.9518$ Obs.: 5
2017	y = -1.1782x + 6.9184 (0.0750) (0.3535) $R^2 = 0.9648$ Obs.: 11	y = **-0.9779x** + 5.7591 (0.1733) (0.9460) $R^2 = 0.9409$ Obs.: 4

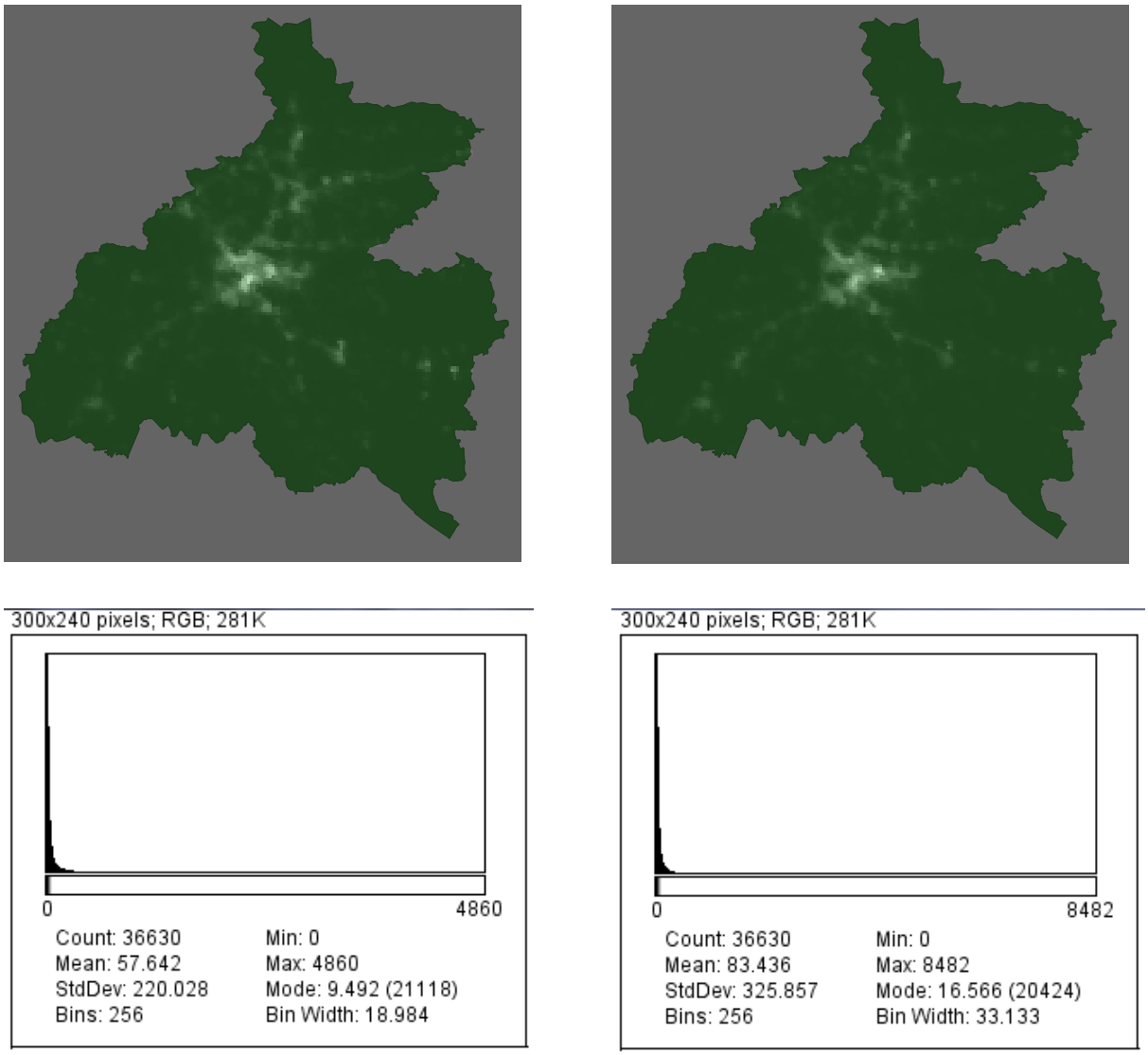
y means log of rank-0.5, x means log of size; standard errors in parentheses; For all estimates except 2017 with >100 px: p<0.001, (Source of data: NOAA)

For the rank-size distribution of cities larger than 100 pixels (there are only few natural cities in Slovenia of that size), only the 2017 data confirm Zipf's law, still at a significance level of $p<0.005$, despite the low number of degrees of freedom. For 2012 the estimate (-1.17) is slightly too low. The shape of extracted natural cities is thus not fully dependable any more. But this might probably stem from the very small number of observations in this special case. Anyway, the resulting segmentation is taken as the basis of Slovene naturally segmented space, but the risk of error is higher than in the case of RVFRM.[21]

The radiance detected within the boundary of LUR is illustrated by the next figure. By visual inspection there is no immediately detectable difference, but as in the case of Frankfurt/Rhein-Main, the basic moments differ considerably, as revealed by the histograms including minima, maxima and, means.

[21] It is less a problem of country size but rather the level of urbanization. While for Slovenia as a less urbanized EU country only very few patches > 50 pixel are detected by statistical segmentation the results for the Netherlands as a highly urbanized small country are very different (cf. Bergs 2018).

Figure 28: LUR: Basic moments for 2012 and 2017 (radiance x 100)



LUR 2012 *LUR 2017*

Source of data: NOAA

By using the national segmentation and zooming in the LUR region, it is possible to monitor the change of natural urban space.

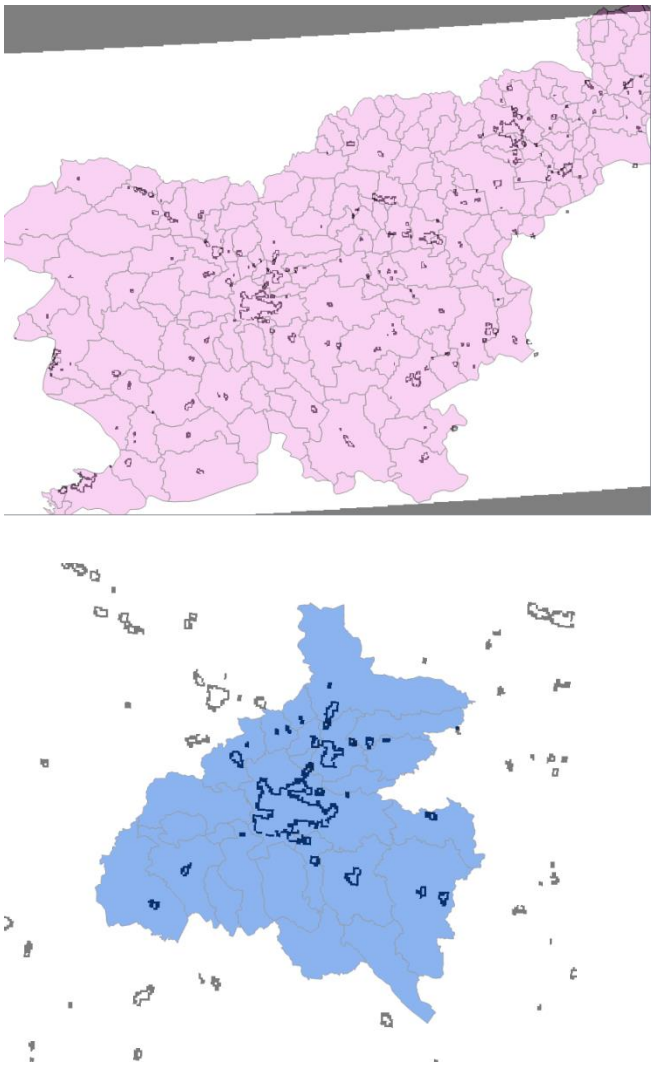
As Figures 28-30 show, the change of natural urban space between 2012 and 2017 is negligible. Basically, natural boundaries of 2012 coincide with those of 2017. In the periphery there are some municipalities with increased and some with decreased size.

Compared with the administrative boundaries, natural urban space appears substantially smaller than for the case study Frankfurt/Rhein-Main. The natural city of Ljubljana approximately covers a quarter of the city area delineated by administrative boundaries in 2012 and roughly one third in 2017. In contrast to Frankfurt there is no or only very little rural-urban intersection of at the urban fringe; thus a large part of Ljubljana is in fact non-urban and there is – at least for 2012 - no major peri-urban or rural area interfering into Ljubljana

natural city. For 2017 there has been some city growth of natural space in the south of Ljubljana municipality with some minor intersection.

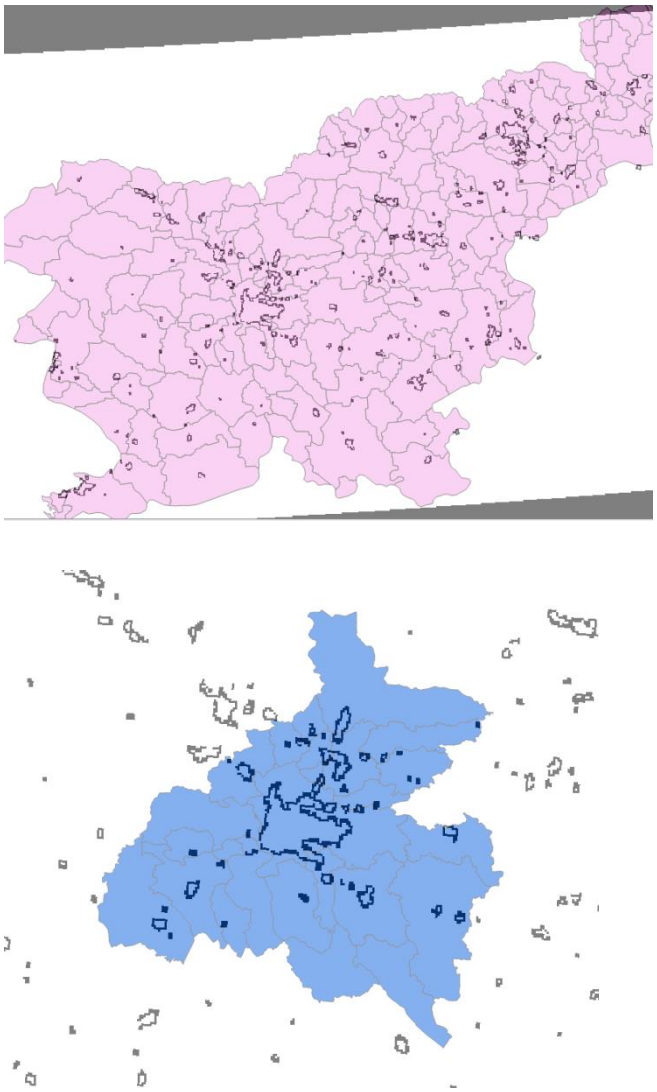
The following figures show natural urban space in relation to the administrative municipal delineation at different scales and a comparison layer to recognize the difference.

Figure 29: LUR: Natural versus administrative urban space 2012



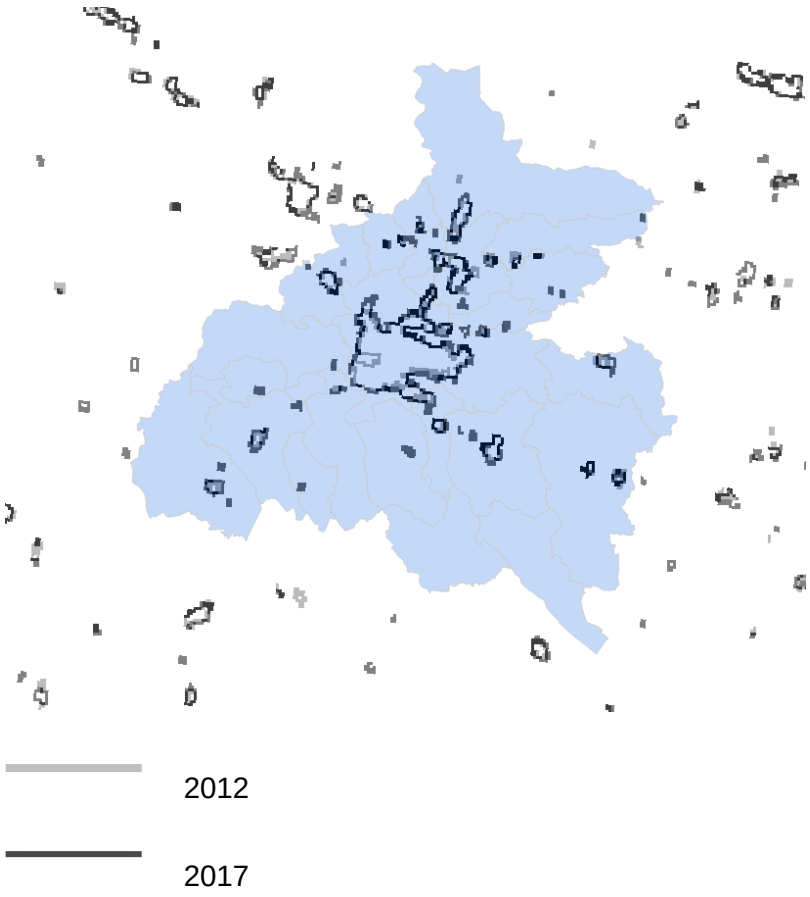
Source of data: NOAA

Figure 30: LUR: Natural versus administrative urban space 2017



Source of data: NOAA

Figure 31: LUR: Layers of segmented natural space: 2012 and 2017 compared



Source of data: NOAA

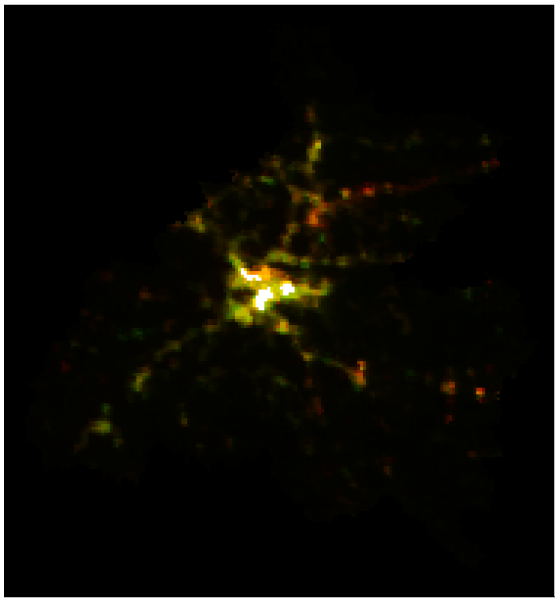
6.2 Change of the variation of luminosity over time

With image analysis it is also possible to monitor increasing or decreasing light emission over time. We can thus detect areas of growing and those of declining radiance, probably pointing to a change in local economic activity.

Figure 31 shows the change of light emission in a red-green spectrum. The green areas are those with increasing light emission during 2012 to 2017; the red ones are those with decreasing light emission. It is well visible that the city of Ljubljana in the north-east, north-west and south-west has gained in light emission while areas in the north of the city and municipalities in the periphery (north-east and south east) have lost. In addition to that there is an axis of increased light emission in the south-west (along the highway to Postojna/Trieste). The white areas in the city represent just co-localized space of top-coded pixels above DN 255.[22]

[22] For a so-called co-localisation analysis, usually applied for the analysis of medical images, raw imaged need to be transformed into 8-bit images. Hence, it is not possible to view changes above a DN of 255.

Figure 32: LUR: Change of light emission 2012-2017

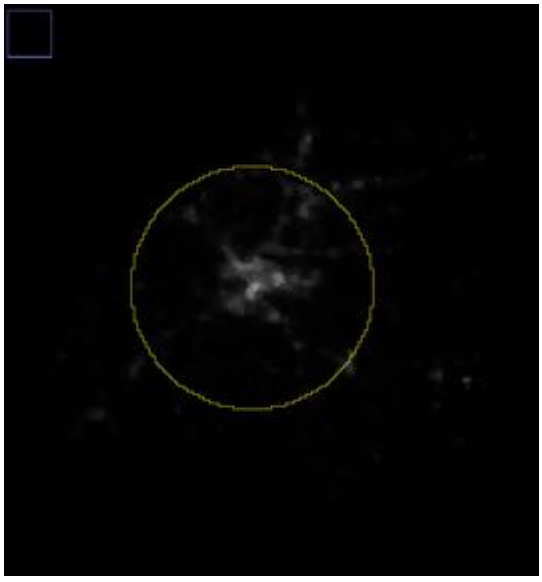


Source of data: NOAA

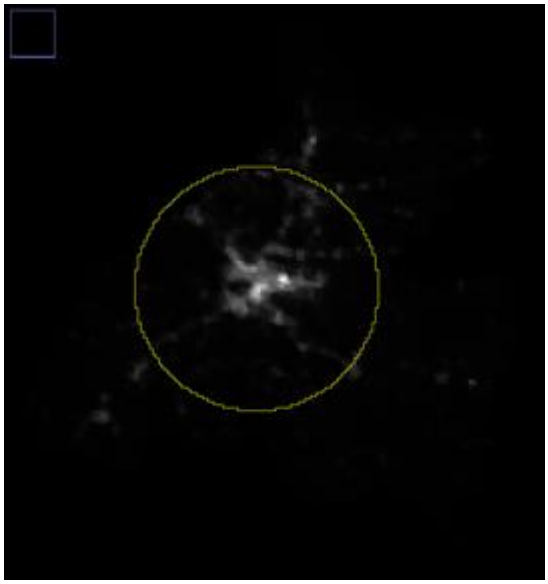
6.3 Empirical results: Spatial dependence of light emission (Global Moran I analysis of the LUR region)

In the following section we briefly look at change of the radial profile around the core city and spatial dependence.

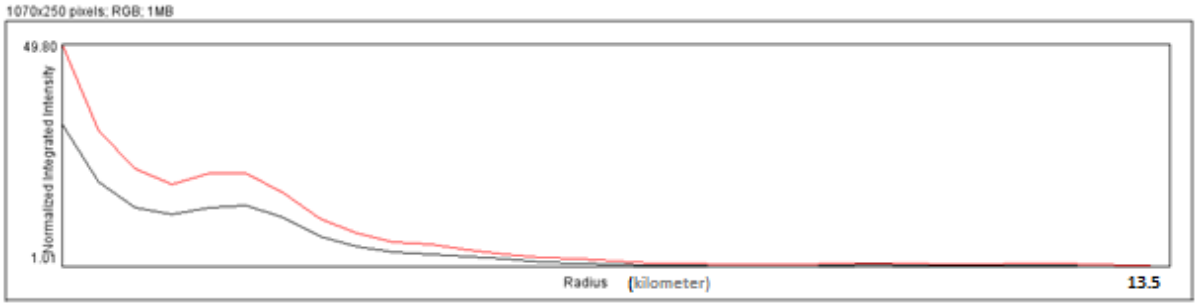
Figure 33: Radial profile around Ljubljana (2012: black; 2017: red)



2012



2017



Source of data: NOAA

Around the city (radius of some 13.5 kilometers), change is only in terms of light emission with a decreasing trend. In a distance of 13.5 kilometers from Ljubljana city center the difference becomes zero. Spatial autocorrelation has slightly increased between 2012 and 2017.

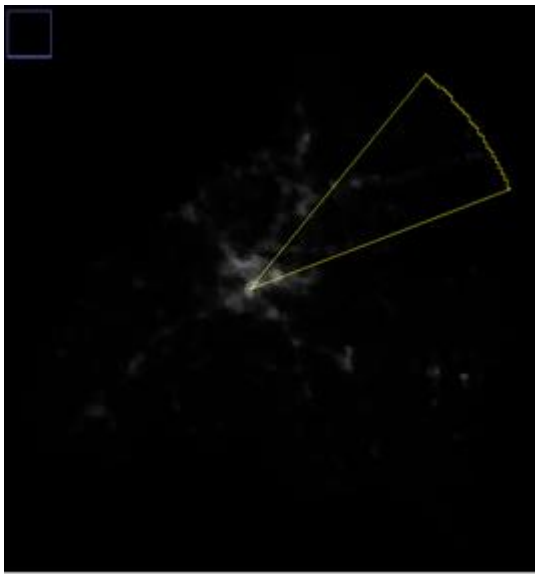
Table 199: Estimation of spatial autocorrelation 2012 and 2017

	CONTIGUITY	MORAN'S I
LUR2012Lambert	Rook's case	0.939406
LUR2017Lambert	Rook's case	0.945163

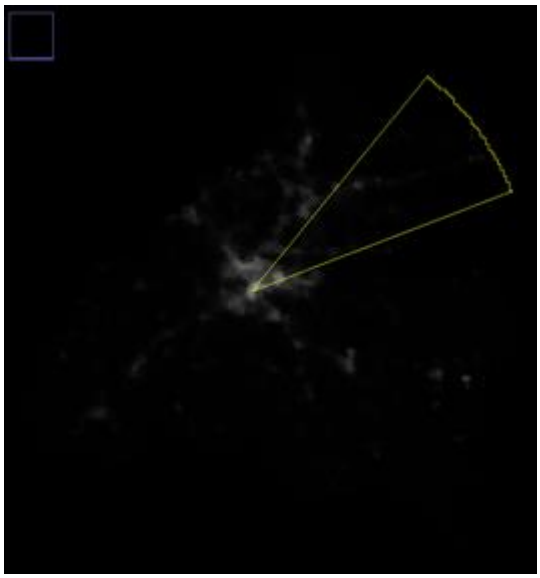
Estimated by QGIS (Saga command Moran I); source of data: NOAA

A similar trend is also observable when looking at a specific development axis (e.g. north-east). There are just two peaks, the inner city and the industrial zone BTC and a steep decrease at around four kilometers.

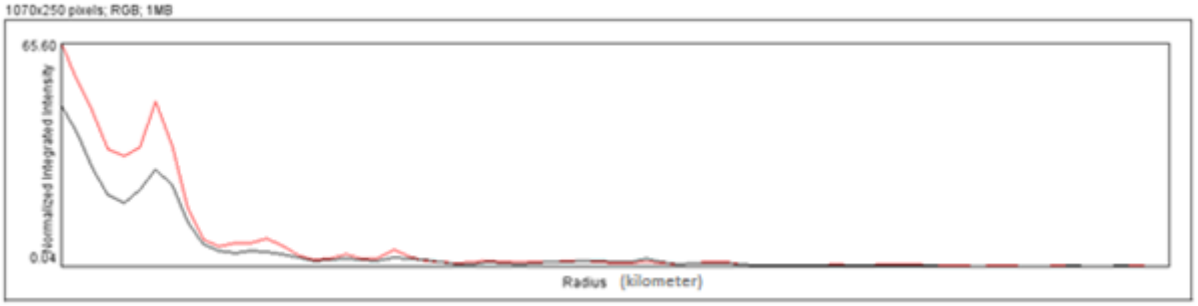
Figure 34: Radial profile of Ljubljana: North-east axis (2012: black; 2017: red)



2012



2017



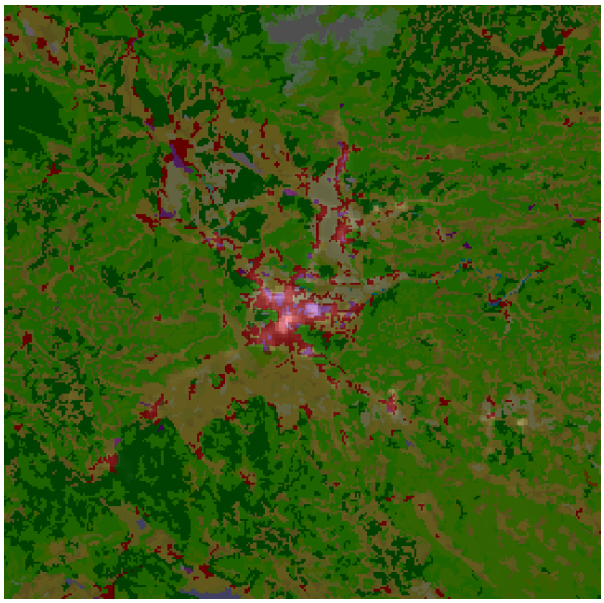
Source of data: NOAA

Compared with the Frankfurt/Rhein-Main area, spatial autocorrelation of light emission in the LUR area is slightly higher (based on the raw pixels). This may indicate a slightly higher functional separation of space for the LUR area. But the reasons for that are simply the different delineation of space and the distribution of pixels within. The more important message is that for both regions there has been a slight increase of spatial dependence between 2012 and 2017.

6.4 Association between light emission and land use

Apart from analyzing change of spatial heterogeneity and dependence we also explored the association between light emission and other variables as small spatial scale. Compared to the grid data available for the RVFRM it is only possible to look at the association between land use (CORINE) and the VIIRS image for 2012 by visual inspection.

Figure 35: LUR: Land-use (CORINE) and light emission 2012



Source of data: NOAA, European Commission

The core of this map is represented by the built-up city of Ljubljana and the surrounding commercial/industrial zones (red and violet areas). For some of them, the underlying light emission is well visible (lighter red and lighter violet). The smaller light red zone is just Ljubljana city center; the light violet zone in the northeast is the area of BTC (a former central warehousing and customs facility converted to shopping area – http://www.btc-city.com/lang/Eng) including industrial and logistics facilities along its southern edge. This is the largest shopping area in the country and probably the largest industrial zone in Ljubljana. The violet zone in the north appears to be the Dravlje-Stegne industrial zone – also a mixture of industry and logistics, but also services. Both areas have been very vibrant in terms of economic activity. The violet areas in southwest and southeast (Vič and Rudnik) are similar, but on a smaller scale. Rudnik is mostly a shopping area. (Information collected in a brief interview with Mojca Hrabar, Oikos).

We can conclude that, similar to RVFRM, light emission and economic activity (represented by central city areas and major industrial and commercial zones including important transport infrastructure like airports) are closely associated, not only at global level but also at the level of metropolitan and urban zones. It is rather a matter of definition of land use (e.g. transport infrastructure versus industrial/commercial space) as it has a major influence on the database to be compared with the satellite images.

7 Major questions revisited

In revisiting the above results, the following major questions have been at the centre of the research focus of the ROBUST project:

- **Can VIIRS images and the derived numerical data show the rural and the urban (and the continuum in- between)?**

Basically, these images show the whole continuum between urban and uninhabited space but no statistical sub-division. However, by applying Zipf's law on the rank-size distribution of the resulting patches (extracted from the cluster segmentation of light emission) those images can help to statistically separate the natural urban space (and its dynamic) from the non-urban one.

- **Can these images show the peri-urban?**

It is certainly possible to cluster space along the variation of light emission at the discretion of the researcher but the result always depends on the specification how to subdivide the respective continuum. The peri-urban is a discrete category not consistent with an assumed continuum. A peri-urban category (or sub-divided peri-urban categories) can be statistically clustered (e.g. K=3 or K=4), but the result cannot be statistically tested like the segmentation of the natural urban segment (i.e. with Zipf's law).

- **Can these images contribute to better urban rural synergies?**

The images allow insight into spatial heterogeneity and dependence and their dynamics. This may help to shed light on the evolution of natural urban and non-urban space and to enrich the urban-rural dialogue on spatial planning and local policies. More spatial information and a higher resolution (data at neighbourhood level) should facilitate a better outcome of coordinated policy and planning processes. In simulating spatial change implied by planned investment in the built environment, it is possible to detect risks of urban sprawl, disturbance of spatial functionality and environmental degradation.

- **Do these images show rural-urban synergies?**

Rural-urban synergies are largely the outcome of – on the one hand - market processes and – on the other hand - policy and planning activities. Nocturnal light emission cannot show those processes directly. But indirectly, these images may offer important information on synergetic functional urban and rural development. Notably the change of spatial autocorrelation around the urban fringe over time could be an important indicator of changing functionality of space to be directly related to the effectiveness of rural-urban coordination of spatial planning and the resulting synergies (e.g. active avoidance of sprawl).

8 Conclusion and discussion

Nocturnal satellite imagery has been widely used for the socio-economic analysis of space. Since 2012 the VIIRS composites promise to obtain even much more precise findings than the older DMSP-OLS images since 1992. The old right-censored 6-bit images were not a meaningful database for urban analysis because they could not capture the variance of light emission in cities. Any attempt of using such images to obtain a realistic functional segmentation of space would fail. The auxiliary radiance-calibrated DMSP products could solve that issue to some extent, even though it is not possible to precisely remove light blooming from the images. The 14-bit VIIRS images, used for the study at hand, are a much more powerful database, because the variance of pixel values in an extremely broad range helps to define functional space very precisely; furthermore cluster analysis finds a stable threshold after cleaning the database from outliers. Based on a properly prepared database after removal of true outliers it is possible to classify urban and non-urban space by univariate cluster analysis (k-means). The resulting patches represent natural urban space of a country. Related to the respective country we used Zipf's law as a test to confirm the segmentation result. A coefficient of $\alpha \approx 1$ would confirm a proper segmentation. This segmentation then indicates the natural urban and non-urban space in terms of its function. For Germany the segmentation result could be well confirmed by Zipf's law; for the Slovene case evidence is weaker, most probably implied by the low number of observations.

The most important added value of spatial segmentation based on the night satellite images is the proper delineation of natural space that can be compared with the administrative space of a city. It is thus possible to monitor rural-urban change along the urban fringe. Especially for inter-municipal land use planning such information (going beyond statistics depending on artificial boundaries) might become very helpful. Combined with analyses of changing spatial dependence, the status of functional areas within cities or rural areas can be monitored (cf. the simulation study for Frankfurt). It could be useful to use radiance as a variable to forecast urbanization under different prior assumptions. The simulation study for the planned new suburb in the north-west of Frankfurt shows how sensitive the size and shape of natural space might change by such investment in the built environment. Change of size of the natural city is not necessarily identical to the final size of the respective new urban zones but can be over-proportional. Even though there is a management of urban evolution in both case regions (RVFRM and LUR) and thus no uncontrolled urban sprawl it is not easy to estimate how investment into the built environment may change the true extent of urban space. Demographic pressure (local asymmetric shocks like the implications of Brexit for the banking sector or refugee migration etc.) might force planners and politicians to revise or adjust former land use plans disregarding limits of local resources of space and nature. The possible implication might not be uncontrolled sprawl, but a controlled one. Yet, the final impact could become likewise a negative one (Peiser

2001). Such evolutionary spatial scenarios can be rather well simulated by analysis of nocturnal satellite imagery.

Radiance might correlate quite significantly with various environmental and socio-economic variables at global level but not necessarily at regional and local levels. Thus, radiance is not always a good proxy for regions where micro-spatial data are unavailable. There is some stronger association between radiance versus sealed ground and radiance versus a composite indicator (housing + commercial estate + transport infrastructure), but this varies quite strongly along different types of space (rural or urban). For both case studies we could show, however, that areas exhibiting major light emission are also areas with a higher level of economic activity in terms of "people at work" (commercial shopping or industrial areas, central city areas). This could be demonstrated for both case studies by merging the respective VIIRS and the CORINE layers from 2012.

The purpose of this study was to examine a method for the analysis of spatial evolution of rural–peri-urban–urban settings over time and to open opportunities to base policy and planning decisions on a more precise empirical foundation. The digital datasets explored are useful to classify functional space and to separate it from municipalities in their administrative boundaries. If there are e.g. typically rural dwellers within city boundaries they are most likely represented by urban interests. This is often a source of conflict along the urban fringe. The possible improvement of urban-rural synergies is then improved by inter-municipal land use planning going beyond static historical boundaries and by recognizing functional space and its dynamic. Satellite images can help to shed light on that. A further purpose could be simulation of land use planning and its impact on urban growth and loss of valuable open space. The dataset thus helps to foster synergies by better insight, but to some extent they are able to even directly show synergies and dependencies. A secularly stable and well-differentiated functional space might suggest synergies, while a dynamic of urban sprawl shown by the data, might indicate a lack of urban-rural synergy.

In conclusion, the practical insight from this experimental exploration of nocturnal satellite imagery and its application in socio-economic and environmental monitoring urban and non-urban space can be summarized by the following chart (Figure 36):

Figure 36: Practical insight from the application of VIIRS nocturnal satellite imagery

9 List of references

Alexander C (1965) A City is not a Tree. Arch Forum 122(1): 58-62

Amaral S, Monteiro AVM, Camara G, Quintanilha JA (2006) DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. Int J Remote Sens 27(5): 855-870

Anselin L, Florax RJGM (1995) New Directions in Spatial Econometrics. Springer, Berlin

Bagan H, Yamagata Y (2015) Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. Gisci Remote Sens 52(6): 765-780

Bergs R (2018) The detection of natural cities in the Netherlands – Nocturnal satellite imagery and Zipf's law. Rev Reg Res , Vol. 38 (2)

Brakman S, Garretsen H, van Marrewijk C (2009), The New Introduction to Geographical Economics. Cambridge University Press, Cambridge

Brakman S, Garretsen H, van Marrewijk C, van den Berg M (1999) The Return of Zipf: Understanding the Rank-size Distribution. J Regional Sci 39(1): 183-213

Budde R, Neumann U (2016) Gravitation and dispersion - a disaggregate view on urban agglomeration and sprawl in Germany. Paper, Scorus Conference, Lisbon 2016

Caffyn A, Dahlstrom M (2005), Urban-rural interdependencies: joining up policy in practice, Reg Stud Vol. 39 (3)

Cauwels P, Pestalozzi N, Sornette D (2014) Dynamics of spatial distribution of global nighttime lights. EPJ Data Science 2014 3(2)

Chen X (2015) Explaining Subnational Infant Mortality and Poverty Rates: What Can We Learn from Night-Time Lights? Spat Demogr 3(1) 27-53

Chen X, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. PNAS 108(21), 8589-8594

Duranton G (2007) Urban Evolutions: The Fast, the Slow, and the Still. Am Econ Rev 97(1): 197-221

Eeckhout J (2004) Gibrat's Law for (All) Cities. Am Econ Rev 94(5): 1429-1451

Egger PH, Lumeau G, Püschel N (2017) Natural City Growth in the People's Republic of China. Asian Dev Rev 34(2): 51-85

Elvidge CD, Safran J, Nelson IL, Tuttle BT, Hobson VR, Baugh KE, Dietz JB, Erwin EH (2004) Area and position accuracy of DMSP nighttime lights data. In: Lunetta RS and Lyon JG (eds) Remote Sensing and GIS Accuracy Assessment. CRC Press. Boca Raton: 281-291

Eurostat (2017) Urban-rural typology (http://ec.europa.eu/eurostat/statistics-explained/index.php/Urban-rural_typology#Database), accessed 30 May, 2017

Fujita M, Krugman P, Venables AJ (1999) The Spatial Economy: Cities Regions and International Trade. MIT Press, Cambridge (Mass.)

Gabaix X (1999a) Zipf's Law and the Growth of Cities. Am Econ Rev 89(2): 129-132

Gabaix X (1999b) Zipf Law for Cities: An Explanation. Q J Econ 114(3): 739-767

Gabaix X, Ibragimov R (2011), Rank-1/2: A simple way to improve the OLS estimation of tail exponents. J Bus Econ Stat 29(1): 24-39

Gan L, Li D, Song S (2006) Is the Zipf law spurious in explaining city-size distributions? Econ Lett 92(2): 256-262

Gehlke CE, Biehl K (1934) Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material. J Am Stat Assoc 29(185A): 169-170

Ghosh T, Powell RL, Elvidge CD, Baugh KE, Sutton PC, Anderson S (2010) Shedding Light on the Global Distribution of Economic Activity. The Open Geography Journal 3: 147-160

Ghosh T, Anderson SJ, Elvidge CD, Sutton PC (2013) Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. Sustainability 2013(5), 4988-5019

Henderson V, Storeygard A, Weil DN (2011) A Bright Idea for Measuring Economic Growth. Am Econ Rev 101(3): 194-199

Hsu FC, Baugh KE, Ghosh T, Zhizhin M, Elvidge CD (2015) DMSP-OLS Radiance Calibrated Nighttime Lights Time Series with Intercalibration. Remote Sens 7(2): 1855-1876

Jiang B (2015) Geospatial Analysis Requires a Different Way of Thinking: The Problem of Spatial Heterogeneity. GeoJournal 80(1): 1-13

Jiang B, Jia T (2011) Zipf's law for all the natural cities in the United States: a geospatial perspective. Int J Geogr Inf Sci 25(8)

Jiang B, Miao Y (2014) The Evolution of Natural Cities from the Perspective of Location-Based Social Media. Prof Geogr 67(2): 295-306

Jiang B, Yin J, Liu Q (2015) Zipf's Law for all the Natural Cities around the World. Int J Geogr Inf Sci 29(3): 498-522

Jiang B, Yin J, Liu Q (2015) Zipf's Law for all the Natural Cities around the World. Int J Geogr Inf Sci 29(3): 498-522

Krugman P (1996) The Self Organizing Economy. Blackwell, Cambridge (Mass.)

Krugman P (1997) Development, Geography and Economic Theory. MIT Press, Cambridge (Mass.)

Kyba C, Kuester T, Sanchez de Miguel A, Baugh K, Jechow A, Hölker F, Bennie J, Elvidge C, Gaston K Guanter L (2017) Artificially lit surface of earth at night increasing in radiance and extent. Sci Adv 2017(3)

Li X, Zhou Y (2017) Urban mapping using DMSP/OLS stable night-time light: a review. Int Rev Remote Sens 2017

Liu Q (2014) A Case Study on the Extraction of the Natural Cities from Nightlight Image of the United States of America. Master thesis, University of Gävle

Mellander C, Stolarick K, Matheson Z, Lobo J (2013) Night-Time Light Data: A Good Proxy Measure for Economic Activity? Martin Prosperity Institute, Toronto

Miller S, Mills S, Elvidge C, Lindsay D, Lee T, Hawkins J (2012) Suomi satellite brings to light a unique frontier of nighttime environmental sensing capabilities. PNAS 109 (39)

Min B, Gaba KM, Darr OF, Agalassou A (2013) Detection of rural electrification in Africa using DMSP-OLS night lights imagery, Int J Remote Sens Vol 34 (22)

Morrish M (2015) Pressures and conflicts in Cambridge's rural-urban fringe, GeoActive 27 (1)

Nordhaus W, Chen X (2015) A sharper image? Estimates of the precision on nighttime lights as a proxy for economic statistics. J Econ Geogr 15: 217-246

OECD (2013) Definition of Functional Urban Areas (FUA) for the OECD metropolitan database, OECD, Paris

Pandya M, Baxi A, Potdar MB. Kalubarme MH, Agarwal B (2013) Comparison of Various Classification Techniques for Satellite Data. Int J Sci Eng Res 4(1): 1-6

Peiser R (2001) Decomposing urban sprawl. Town Plan Rev 72(3): 275-298

Pinkovskiy ML (2013) Economic discontinuities at borders: Evidence from satellite data on lights at night. MIT, Cambridge (Mass.)

Pritchard SB (2011) The trouble with darkness: NASA's Suomi satellite images of earth at night, Environ Hist Durh N C 22(2): 312-330

Proville J, Zavata-Araiza D, Wagner G (2016) Night-time lights: A global long-term look at links to socio-economic trends. PLoS One 12 (3)

Reggiani A, Nijkamp P (2015) Did Zipf anticipate Socio-economic Spatial Networks? Environ Plan B 42(3): 468-489

Repp A et al. (2012) Urban-rurale Verflechtungen: Analytische Zugänge und Governance-Diskurs (Urban-rural linkages: Analytic access and governance discourse), Müncheberg: Leibniz-Zentrum für Agrarlandforschung

Rosen KT, Resnick M (1980) The Size and Distribution of Cities: An Examination of Pareto Law and Primacy. J Urban Econ 8(2): 165-186

Roychowdhury K, Taubenböck H, & Jones S (2011) Delineating urban, suburban and rural areas using Landsat and DMSP-OLS night-time images Case Study of Hyderabad, India. In Stilla U, Gamba P, Juergens C, Maktav D (Eds) JURSE 2011 - Joint Urban Remote Sensing Event, Munich

Small C, Pozzi F, Elvidge CD (2005) Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sens Environ 96(3-4): 277-291

Smętkowski M 2015, Spatial scope of regional economic growth in Central and Eastern European Countries. Geographia Polonica, Vol 88(4)

Soo KT (2005) Zipf's Law for Cities: A Cross Country Investigation. Reg Sci Urban Econ 35(3): 239-263

Starrett D (1978) Market allocations of location choice in a model with free mobility. J Econ Theory 17(1): 21-37

Stefanovic IL, Scharper SB eds. (2012) The Natural City: Re-Envisioning the Built Environment. University of Toronto Press, Toronto

Suedekum J, Giesen K (2011) Zipf's law for cities in the regions and the country. J Econ Geogr 11(4): 667-686

Sutton P (1998) Modelling Population density with Night-Time Satellite Imagery and GIS. Comput Environ Urban 21(3-4): 227-244

Vandervieren E , Huber M (2004) An adjusted box-plot for skewed distributions. In: Antoch J (ed.) Compstat 2004 Symposium, Physica/Springer, Heidelberg, pp 1933 -1940

Woods M, Haley L (2017) Conceptualisation of Rural-Urban Relations and Synergies, University of Aberystwyth (ROBUST deliverable 1.1)

Wu S (2015) Zipf's Law for Natural Cities Extracted from Location-Based Social Media Data. Master thesis, University of Gävle

Wu W, Zhao H, Jiang S (2018) A Zipf's Law-Based Method for Mapping Urban Areas Using NPP-VIIRS Nighttime Light Data. Remote Sens 2018 10, 130

Yang W, Hou K, Liu B, Yu F, Lin L (2017) Two-Stage Clustering Technique Based on the Neighboring Union Histogram for Hyperspectral Remote Sensing Images. IEEE Access 5: 5640-5647

Yu D and Wei YD 2006, Spatial data analysis of regional development in Greater Beijing, China, in a GIS environment. Pap Reg Sci, Vol. 78 (1)

Zipf GK (1949) Human Behaviour and the Principles of Least Effort. Addison Wesley, New York