

Article

Spatial Analysis of Settlement Structures to Identify Pattern Formation Mechanisms in Inter-Urban Systems

Katharina Henn, John Friesen , Jakob Hartig and Peter F. Pelz 

Chair of Fluid Systems, Technische Universität Darmstadt, Otto-Berndt-Straße 2, D-64287 Darmstadt, Germany; katharina.henn@stud.tu-darmstadt.de (K.H.); jakob.hartig@fst.tu-darmstadt.de (J.H.); peter.pelz@fst.tu-darmstadt.de (P.F.P.)

* Correspondence: john.friesen@fst.tu-darmstadt.de; Tel.: +49-6151-16-27100

Received: 12 August 2020; Accepted: 7 September 2020; Published: 8 September 2020



Abstract: Dissipative structures known from non-equilibrium thermodynamics can form patterns. Cities are regarded as open, dissipative structures due to their self-organisation and thus in theory are also capable of pattern formation. In a first step to understand similarities between nonlinear pattern formation and inter-urban systems, we investigate how inter-urban structures are arranged. We use data from the Global Urban Footprint to identify spatial regularities in seven regions (Argentina, China, Egypt, France, India, Ghana and USA) and to quantitatively describe settlement patterns by number of objects and density. We find that small areas of the examined data sets show a regular arrangement, the density and number of settlements differ widely between the different regions and the portion of regular areas within this regions strongly correlates with these two parameters. The results can be used to develop mathematical models that describe inter-urban pattern formation on the one hand and to investigate to what extent the respective settlement patterns are related to infrastructural, economic or political boundary conditions on the other.

Keywords: spatial analysis; pattern formation; urban modelling; settlement pattern; remote sensing

1. Introduction

Urbanisation coupled with major demographic transformations is, along with climate change, one of the most striking phenomena of the early 21st century [1] and has increasingly moved into the focus of scientific work in recent decades [2]. This development has led to a “science of cities” [3], where with the help of quantitative mathematical models, various attempts are made to gain access to the different processes that shape and change today’s cities [4,5].

A major focus is the investigation of urban morphology [6] and land use within cities and their surroundings. To describe and predict changes in land use, cellular automata are a commonly used tool [7]. In the last decades, these are extended by mobile agents, resulting in agent-based models describing the interaction between the actors and the respective environment they are operating in [8].

However, due to the integration of ever larger amounts of data [9], more complex algorithms [10], especially machine learning and deep learning, the models are becoming more and more complex [11]. The “black-box” structure of these models, coupled with an increasing number of nuances considered in the models, leads to a loss of comprehensibility [12]. A contrary approach is taken when certain phenomena are mapped with comparatively simple mathematical models to describe core processes of urbanisation.

A very famous model to describe urban structures was made by Walter Christaller in the early 1930s, when he developed the concept of the central place theory (CPT) [13]. This theory is based on the assumption that from a locally homogeneous distribution of small settlements, urban hierarchies develop over time in which larger towns take over certain infrastructural tasks for their neighbours.

Although the theory has been criticised (see, e.g., in [14]), both the hierarchy in settlement structures for different regions [15] and the regularity in the arrangement of rural settlements [16] can be shown empirically.

While many models based on CPT focused on the development from regular to hierarchical structures [14], less attention was paid to the development of the initial regular structures. The only approach known to us following this direction for inter-urban system deals with the regular spatial arrangement of industrial plants. Paul Krugman [17] proposed that the regular spatial arrangement of these plants can be described by a diffusion-driven instability.

Besides diffusion-driven instability, other open thermodynamic systems can be found in nature that form regular structures such as the Taylor–Couette flow, Rayleigh–Benard convection and Benard–Maragoni convection. The order far from thermodynamic equilibrium arises by instabilities that grow, stimulated by small disturbances or imperfections [18]. As urban systems are also open systems [19], the question arises whether they can be described by nonlinear mathematical models of this type. In recent years, there have been increasing attempts to transfer these physical nonlinear models to intra-urban patterns or structures [20–22].

However, we are concerned with patterns in inter-urban systems. Qualitative different pattern of urban structures can be linked to the growth mechanisms urban sprawl [23,24] and sprinkling [25,26], which have social and ecological effects [23]. A deeper understanding of these patterns, processes and modelling possibilities [25,27] is therefore of great importance.

In a first step, we examine the nature of the structures formed by inter-urban systems on the basis of different examples in order to examine the applicability of these nonlinear models in future work with the knowledge that such an approach is only an indication of pattern formation mechanism [28]. Our research question is as follows.

How are settlements arranged in different predominantly rural regions of the world?

Although there are studies that examine the regularity of settlement arrangements [15,16], these are often not globally comparative. Therefore, we compare settlement patterns determined by satellite data in seven different regions of the world, using the following framework.

2. Conceptual Framework

Against the background of the strong spread of machine learning, a wide variety of scientific domains are concerned with patterns. In our analysis, we investigate two-dimensional binary patterns.

To answer the research questions stated above, consistent methods are in need. The properties of the settlement patterns have to be quantified in a correct yet condensed manner, to be able to compare settlement structures in different regions. Therefore, we use the following three parameters (Figure 1),

- (i) the number of objects investigated;
- (ii) the density of objects within the investigated area; and
- (iii) the regularity, which we quantify with a special characteristic value.

While the number and density represent continuous values from “low” to “high”, the regularity is divided into three classes. These are clustered, random and regular. In this procedure we neglect the shape of the individual objects. With these methods we quantify settlement patterns, using the framework in Figure 2.

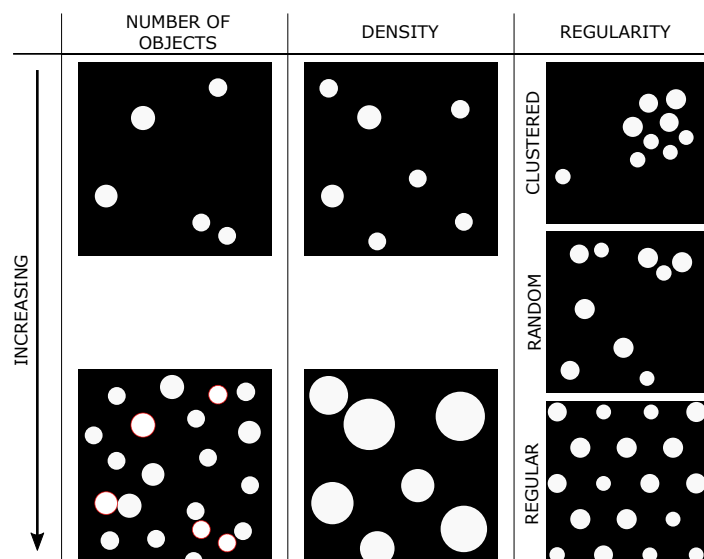


Figure 1. Types of pattern classified by number of objects, density and regularity.

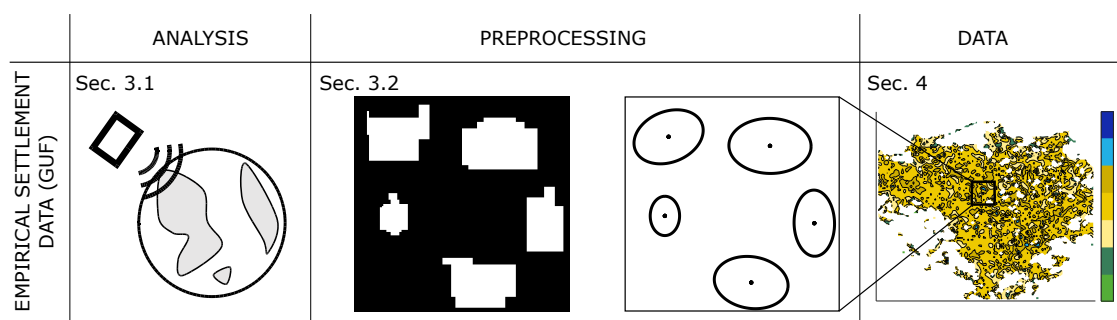


Figure 2. Workflow of this paper starting from the empirical data of the Global Urban Footprint (GUF), a binary settlement map derived from satellite data (Section 3.1) [29,30]. Preprocessing of the data as follows (Section 3.2) by identifying related objects and determine their centre of gravity and analysis of the spatial distribution of the number of objects, density and regularity (Section 4).

We first present and briefly explain in Section 3 the empirical data and analysis methods used before we present the results in Section 4 and discuss them afterwards in Section 5.

3. Data and Methods

3.1. Data

As empirical data the GUF is used, a data set by the German Aerospace Center (DLR), which is based on data from the TanDEM-x mission. The data of the GUF are from the years 2011 and 2012. The radar data collected by the two satellites are evaluated fully automatically in three steps by the Urban Footprint Processor and a binary settlement map with a resolution of approximately 12 m is produced [29–31]. The binary settlement map thus consists of pixels of the resolution just mentioned, which indicate whether the respective area is settled or not.

In our analysis, we investigated sections from seven primarily agricultural regions on five continents, namely, regions in Argentina, China, Egypt, France, Ghana, India and USA, which all are equally sized with 200 km by 200 km (Figure 3). The coordinates of the data sets are given in the Table A1. Figure 4 shows examples of the GUF examined in China and Egypt. In this analysis, we focus on agricultural dominated landscapes as different settlement theories like the CPT [13] or others [32,33] are based on regular distribution of non-urban settlements as initial state. In selecting the regions, we have tried to obtain as comprehensive a picture as possible by examining countries with different economic development (developing and developed). In addition, as few metropolitan regions

as possible should be present. Cairo (Egypt) is the exception, as the Nile Delta is a very interesting and otherwise a homogeneous settlement area north of the Sahara.



Figure 3. Schematic geographical position of the seven analysed areas in Argentina, Egypt, China, France, Ghana, India and USA.

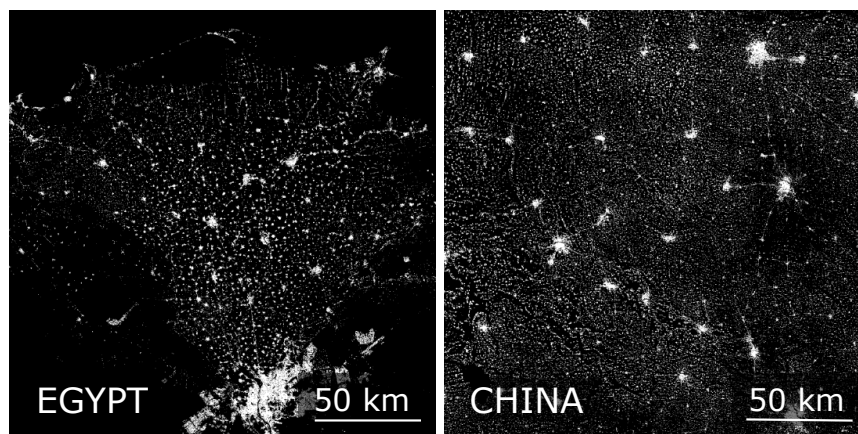


Figure 4. Parts of the Global Urban Footprint (GUF). The data set Egypt includes the Nile Delta (**left**), the investigated region in China is located in den province Henan and Anhui (**right**).

3.2. Method

As introduced in Section 2 the settlements patterns resulting from the GUF are examined with regard to the three dimensionless parameters, the number of objects N , density ϱ and the Average-Nearest Neighbour Index (ANN) as a measurement for regularity.

We define density as the quotient of the populated area $A_{\text{build-up}}$ and investigated area A :

$$\varrho := \frac{A_{\text{build-up}}}{A}. \quad (1)$$

The ANN [34] is commonly used to measure the distribution of settlements in geographical investigations [15,16]. It distinguishes clustered, random and regular patterns, represented by ANN values near 0, 1 or 2.15, respectively. When calculating the ANN , the arithmetic mean \bar{r}_A of the distance r_i between every data point $i \in [1, N]$ and its nearest neighbour in the investigated area A is used. This distance is compared to the distance $\bar{r}_E = 0.5 * \sqrt{A/N}$ of data points in a random distribution [34]:

$$ANN = \frac{\bar{r}_A}{\bar{r}_E} = \frac{2 \sum_{i=1}^N r_i}{\sqrt{N \cdot A}} \quad (2)$$

Similar to the work in [16], we differentiate between clustered ($0 \leq ANN \leq 0.6$), random ($0.6 < ANN \leq 1.5$) and regular ($ANN > 1.5$) distribution.

In order to quantify the reliability of the classification of the pattern as deviating from a regular distribution, the z-value for the *ANN* is calculated. It indicates the probability with which the hypothesis of a random point distribution can be rejected correctly. A z-value < 1.96 or > 1.96 forms the 90% confidence interval and the value is defined as

$$z = \frac{\bar{r}_A - \bar{r}_E}{0.26136} \cdot \frac{N}{\sqrt{A}} \quad (3)$$

We first use the parameters introduced in Section 2, namely, number of objects *N*, density *q* and *ANN* to describe the pattern of settlement arrangement. Calculating the *ANN* by considering all settlement objects in the area consisting of at least five pixels, we get a *ANN* < 1 for all data sets.

As this does not allow detailed analyses, the number of objects *N*, the density *q* and the *ANN* are evaluated in sample squares spreading out on all seven investigated regions. For each analysis, the equally sized sample squares are equidistantly arranged in vertical and horizontal direction. The distance between adjacent sample squares is considerably less than their length of the edge. The distance is 0.5 km for sample square length equal to 2 km. 1 km for sample square length 3 km and larger sample squares have a distance of 2 km. Because squares crossing the border of the investigated area would result in misleading results, the sample squares are not covering the data set's edges. The sample square's result for *N*, *q* and *ANN* are assigned to its centre to depict the results of each sample square for the entire data set.

4. Analysis

4.1. Settlement Arrangement

In Figure 5, the parameters *N*, *q* and *ANN* for sample squares with a length of 8 km are shown.

It should be noted that the abstraction of the data in GUF to the centre of gravity and surface area of the object can lead to densities greater than 100%. This is the case when a large number of objects with a large size have their centre of gravity in the sample square, but their dimensions considerably exceed the sample square's border. This results in a settlement area that is larger than the area of the sample square. We limit the value of density to 100%.

Figure 5 allows us to characterise each data set: Data set Argentina shows equidistant spots with higher levels of density, which belong to larger cities in the observed region. The southeastern corner in the data set China is characterised by high numbers of objects and low density, therefore representing small settlement objects. The settlements in the data set Egypt clearly follow the Nile Delta and the agglomeration of Cairo can be identified easily. The vein-shaped areas with low porosities that run through the data set France are characterised by rivers on whose banks a higher number of settlement objects can be found. In the data set Ghana, the clustering of settlement objects around the million city of Kumansi is clearly visible. Clusters of dense settlement structures are located in the north of the data set India, while the southern part contains very few settlement objects. The data set USA includes only a small number of objects, which have a small expansion. Therefore, the chosen sample square length is not suitable for *ANN* evaluation in this data set.

In every regions the clear majority of sample squares is characterised as random distribution by the *ANN*. Nevertheless, the individual data sets do have differences: In the data set Argentina, the majority of sample squares is represented by an *ANN* ranging from 0.6 to 1.2; in the data sets China and Egypt, most of the sample squares have an *ANN* greater 0.9 and smaller 1.2; in France, Ghana, India and USA, the dominating range is $0.6 < ANN \leq 0.9$.

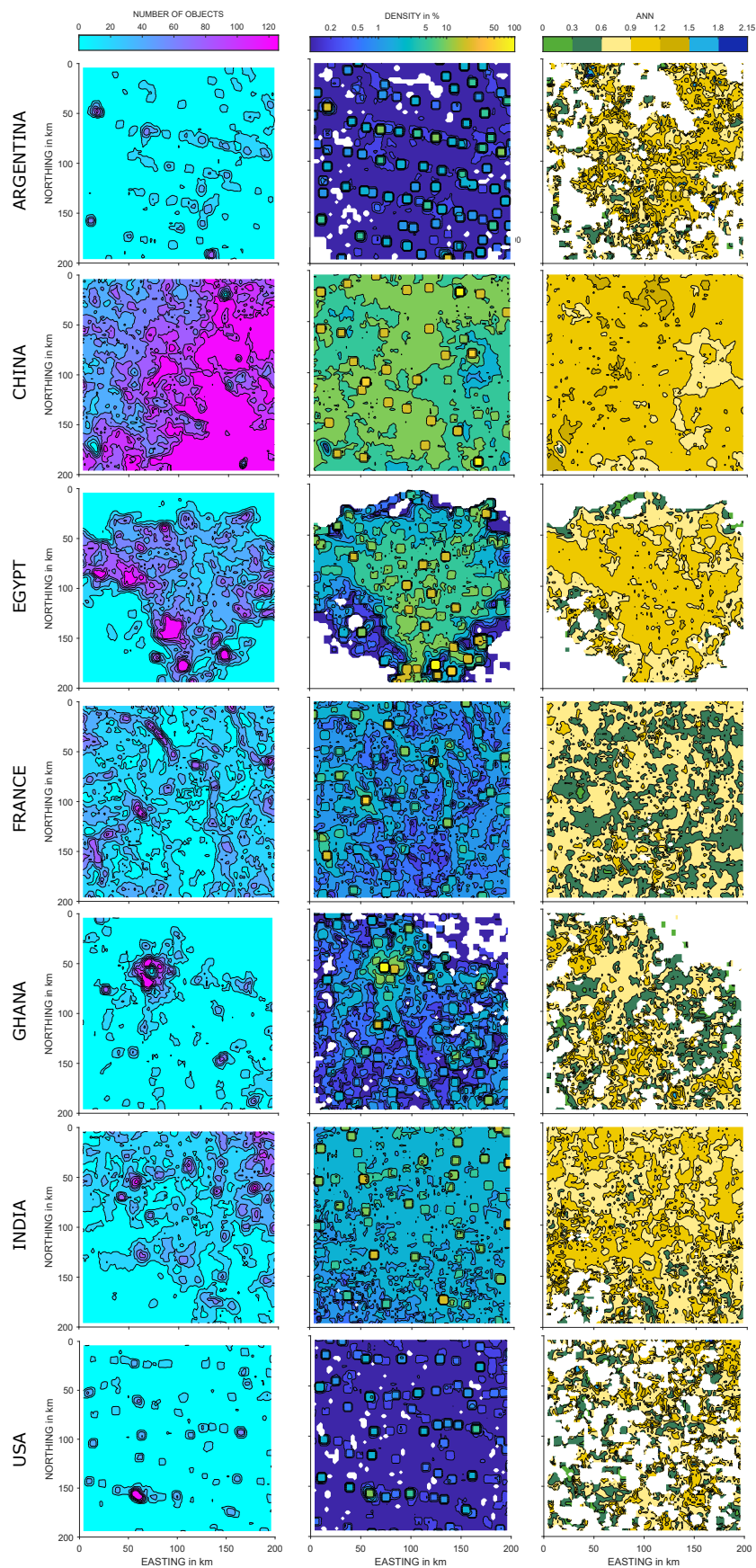


Figure 5. Number of objects, density and ANN for sliding window analysis for windows of sample size of 8 km.

Nevertheless, in some regions a small percentage of squares, e.g., 2.7% in data set Argentina and 2.6% in data set USA, contain, according to the ANN, regular structures (see Table A2).

In order to investigate the arrangement of settlement objects in more detail, the described analysis with sample squares was repeated for squares with different edge lengths. The averaged results are shown in Figure 6. The effect of the sample squares' size on the ANN becomes visible: With increasing edge length the ANN decreases, whereas in sparsely populated areas an increase to medium sample square edge lengths is detected before the decrease. It also becomes clear that the settlement objects in the data sets China and Egypt are more regularly distributed than the settlements in the other data sets.

However, sample squares with regular settlement distribution can be found for all data sets. For the share of regular sample squares in the each data refer to Table A2. While for a sample window size of $2 \times 2 \text{ km}^2$ the proportion of windows with a $ANN > 1.5$ is between 0.8% (USA) and ~14% (China), for $3 \times 3 \text{ km}^2$ between 0.95% (France) and 4% (India), for larger sample windows ($15 \times 15 \text{ km}^2$) the values are between 0 % for China and 0.23% for Argentina (Table A2). Areas of regular settlement structures can thus be observed above all in small settlement sections.

All sample squares listed with $ANN > 1.5$ result in a z-value above 1.96. In this case, the ANN delivers the reliable result of a non-random distribution.

The investigated regions show noticeable differences in terms of number of objects, density and portion of regular spots in the settlement structures. Nevertheless, similarities in the data sets Argentina and USA, and Egypt and India can be seen.

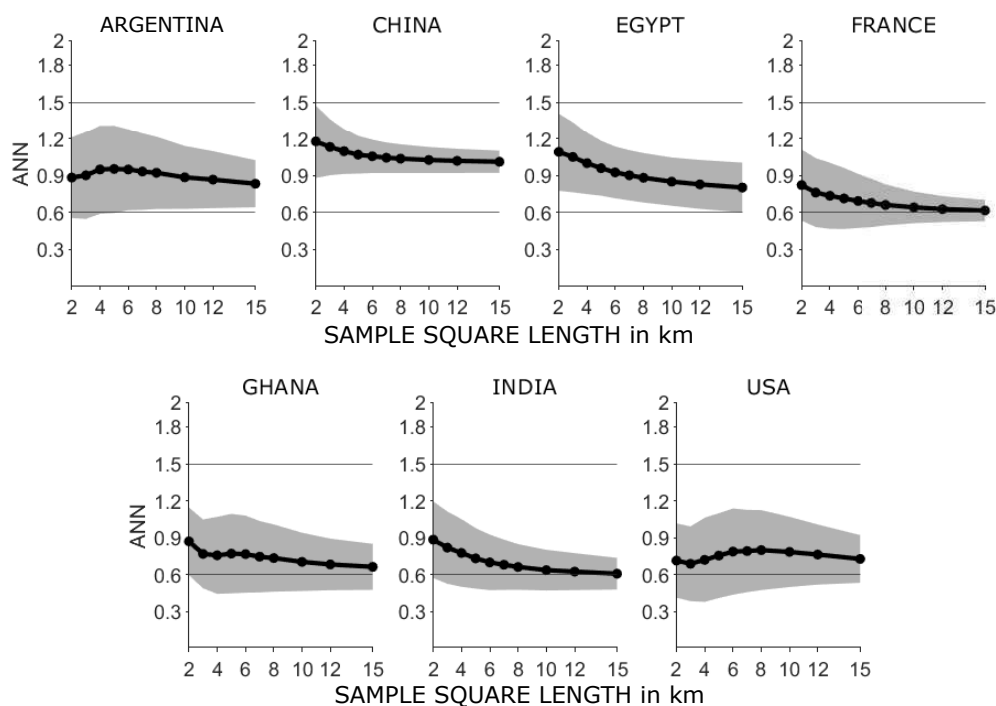


Figure 6. Results for the mean ANN calculated in sample squares of different edge length in all data sets. The grey area represents the standard deviation of the calculated values.

4.2. Size Distribution

To examine some of the points from above more closely, in addition to the three parameters N , ρ and ANN, the size distribution of the expansion of the settlement objects in all data sets was examined, see Figure 7. It is found that the median of all seven data sets is in a similar range of 10^4 m^2 , as well as the geometric mean S_0 given in Table 1. The characteristic size of the inter-urban structures investigated here is thus in the same range as the intra-urban structures of morphological slums investigated in [35,36].

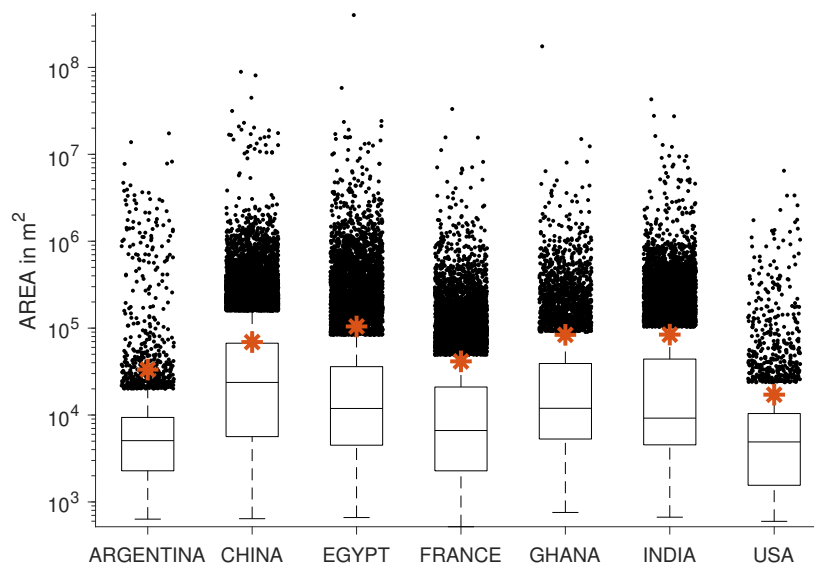


Figure 7. Box plot showing the size distribution of settlement objects in all seven data sets. The arithmetic mean is marked with a red star for each data set.

Table 1. Table containing the number N_0 of objects in every data set, and the geometric mean S_0 of the objects' area. The data sets' average characteristic size is $\bar{S}_0 = 0.0112 \pm 0.0051 \text{ km}^2$. All values are defined as in [35].

Data Set	N_0	S_0 in m^2	σ	$l = \sqrt{S_0}$ in m	Density in %
Argentina	7119	4995	1.2045	70.7	0.59
China	65,736	19,179	1.6694	138.5	11.4
Egypt	22,744	13,512	1.6810	116.2	5.94
France	20,771	7621	1.6670	87.3	2.15
Ghana	9147	14,293	1.5846	119.6	1.91
India	17,270	14,266	1.7161	119.4	3.63
USA	6605	4471	1.2792	66.9	0.28

It also becomes clear that there are few settlement objects to be found in the regions investigated in Argentina and the USA. In addition, these two data sets do not contain as many large settlement objects as the other investigated regions, and are very similar in this aspect. This fact does not seem surprising, as the data sets show a part of the Argentinean Pampas and the American Great Plains. In the following, these two areas are referred to as sparsely populated. The data set China contains by far the most settlement objects, which also have the largest extension. The data sets India and Egypt contain a similar number of objects and have a characteristic size of the objects, which differs only slightly. Both number and characteristic length range between the extremes of the sparsely populated data sets and the extremely dense populated data set China.

The data sets Ghana and France cannot be easily classified into the categories of sparsely, medium and densely populated areas. France has a similarly large number of objects as India and Egypt, but contains objects of much smaller dimensions, whereas in Ghana, the extent of the objects is comparable with the extent of the objects in the data set India and the number of objects is close to the number of sparsely populated data sets.

4.3. Classes

The small percentage of regular structures in Section 4.1 seems surprising when visually comparing the results to the GUF data, e.g., for Egypt (see Figure 4). The larger cities in the Nile Delta

appear to be equidistant to each other, like it was shown for the cities in the Nile Valley [15]. Taking that into account a second analysis regarding only settlement objects larger in size, is carried out.

The data is divided into three logarithmic bins, ranging from the smallest object's area to the largest object's area. The entirety of objects, contained in the three bins, has been analysed in the previous subsections. In the following, this data set is referred to as *Class 1*. The middle and the last bin containing medium and large objects are denoted as *Class 2*, while the large objects are representing *Class 3*. For a visualisation of the data sets' division into classes, please refer to Figure A1. The different classes in the data set Egypt are depicted in Figure 8.

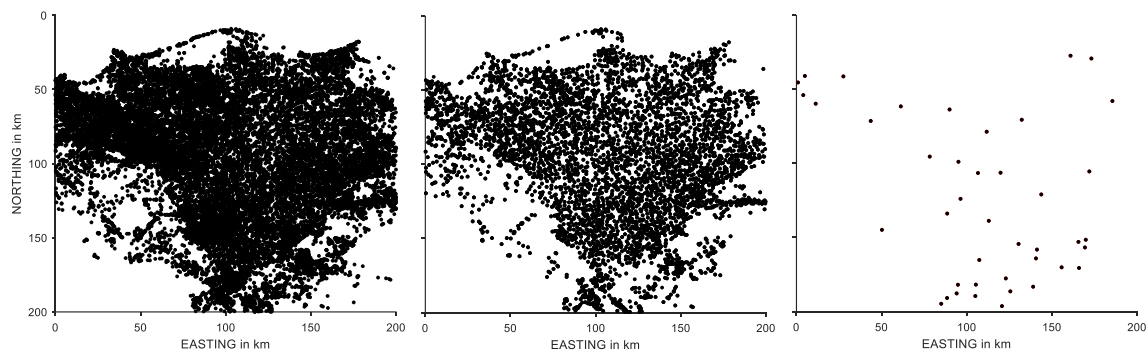


Figure 8. Data points for *Class 1*, *Class 2*, *Class 3* (from left to right).

Repeating the described analysis with regarding only *Class 2*, the mean ANN in the data set Egypt increases, see Figure 9. In the data set Egypt there are only few objects in *Class 3*, that is why the sample square length must be increased significantly to allow a evaluation of the ANN. Therefore, the analysis results are not easily comparable with those for *Class 1* and *Class 2*. To get an impression of the results for *Class 3*, Figure 9 gives the average ANN for sample squares with edge length equal to 75 km, taking into account only the objects of *Class 3*.

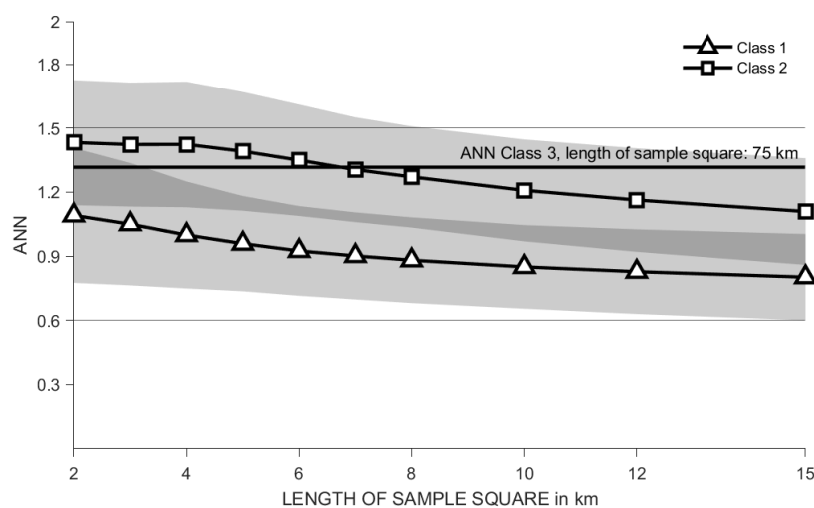


Figure 9. Average ANN in sample squares of different length for *Class 1* and *Class 2* in the data set Egypt.

Similar results can be obtained by analysing *Class 2* or *Class 3* in the other data sets. Some data sets show higher proportions of regular windows in *Class 3* than in *Class 1*, while others show higher proportions of regular windows in *Class 2* than in *Class 1*. This leads to the hypothesis that settlement objects larger in size are arranged in a more regular pattern than the entirety of settlement objects.

When not just identifying the ANN in the sample squares but also the number of settlement objects regarded to calculate the index, it becomes clear that the sample squares with higher ANN usually include less objects (see Table A2). When analysing *Class 2* and *Class 3*, the number of settlement objects decreases while the sample square's edge length is kept constant. One may argue that the ANN in *Class 2* and *Class 3* therefore only increases because decreasing number objects in the sample squares. Regarding Table 2 this is not the case. Only for small sample squares with high ANN the data set *Class 1* includes more data points than in *Class 2*. For larger sample squares the number of objects increases when moving from *Class 1* to *Class 2*.

Table 2. For *Class 1* and *Class 2* in the data set Egypt, percentage of sample squares (s.s.) with ANN > 1.5 are listed; *N* is the average number of data points in these sample squares.

s.s. Length		3 km	5 km	7 km	10 km	12 km	15 km
<i>Class 1</i>	% of sample squares	5.77	0.73	0.31	0.16	0.00	0.03
	<i>N</i>	6.79	7.89	7.00	6.75	-	10.00
<i>Class 2</i>	% of sample squares	38.91	35.26	20.20	6.43	2.60	0.65
	<i>N</i>	5.41	6.51	9.44	16.09	20.08	15.34

4.4. Regular Settlement Structures

None of the seven regions studied consists mainly of regular settlement structures. However, regular structures can be identified for all seven data sets—see Figure 10. Table 3 shows three such sections with ANN > 1.5 from the GUF and their parameters. It should be noted that the sections of the data sets China and Egypt for *Class 2* do not appear to be distributed as regularly as the high ANN suggests. This is due to the fact that in *Class 2* the small objects, which are, in this case, clustered around the larger ones, are not taken into account.

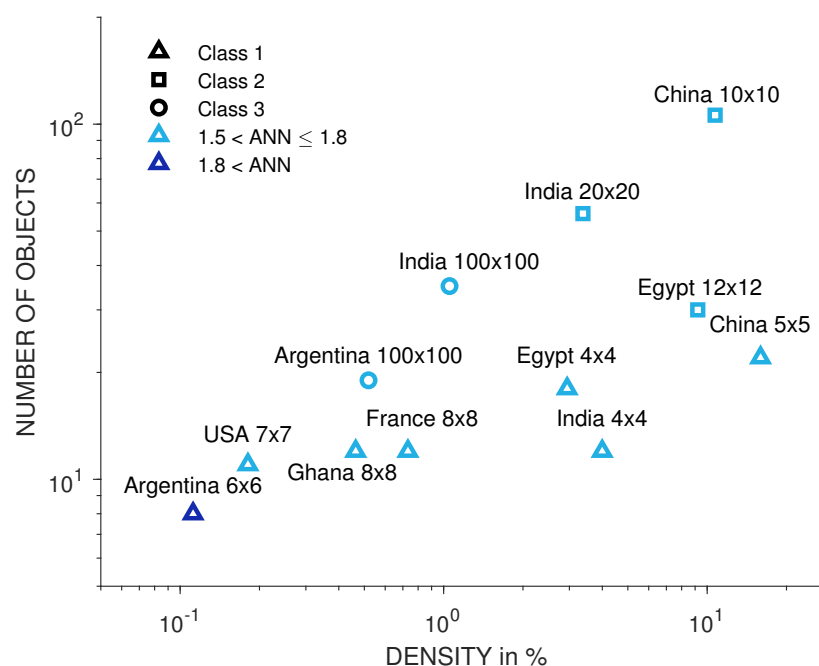
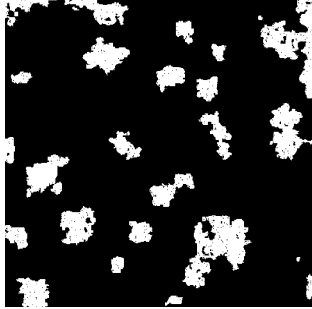
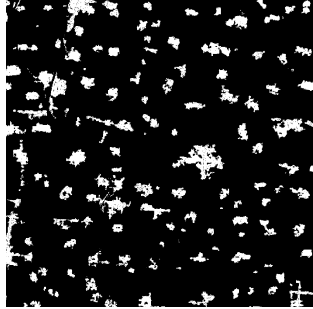
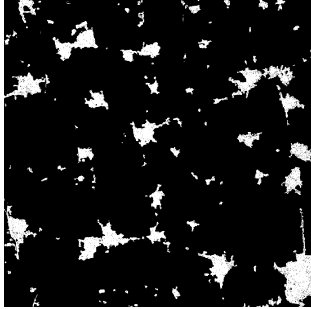


Figure 10. The marker represent the parameters of the chosen sample squares. The numbers following the data set's name indicate the edge length of the sample square.

Table 3. Examples for regular settlement structures in the data sets China and Egypt.

	CHINA Class 1	CHINA Class 2	EGYPT Class 3
s.s. Length	5 km	10 km	12 km
parameters			
N	22	106	30
ρ in %	15.95	10.72	9.21
ANN	1.69	1.50	1.57
z	6.22	9.87	5.98
original GUF data			

5. Discussion

The regions investigated in this study, especially rural areas, show different settlement patterns. The GUF is analysed with three parameters (number, density and regularity) that define a two dimensional pattern without considering the shape of the different objects. However, the analysis carried out is limited to a specific data set at a defined point in time. When assessing the results, it must be taken into account that the data may contain artefacts that are not discussed further here. For an evaluation of the original GUF data, please refer to the corresponding publications [37].

Looking at the size distribution of the different settlements and their mean value, it is striking that the values of $\sim 0.01 \text{ km}^2$ found here are similar to those from intra-urban analyses [35,36]. This size seems to be an universal quantity of settlement structures, as it is also similar to the size of city blocks in American and Australian Cities [38].

We use the ANN as a method to quantify the regularities of patterns in awareness of the limitations of this approach: The influences of the area and number of objects taken into account for the calculation cannot be neglected (for a more detailed discussion on pattern indices we refer to the work in [39]). The disadvantages are understandable, as the spatial arrangement of a structure is reduced to a single parameter leading to an information loss. As the ANN still gives a first indication of the regularity of the pattern, it is an appropriate metric to answer the research questions formulated above.

Looking at the analysed regions from a macroscopic view, settlement clusters can be found, especially towards the coastlines. This is due to the fact that the process of settlement formation does not take place homogeneously over an area, but the proximity to resources and infrastructure (fishing, trade and economy) seems to play a role in the choice of location. The area boundaries also have a great influence on the result: while south of the region studied in Ghana the Atlantic Ocean begins directly and north of it the rainforest, in China the region is part of a very large inter-urban structure with high regularities [16].

The empirical results show a correlation between high density and a high number of objects (Table A2). At the same time, the density and the number of objects correlate strongly ($R = 0.9327$) with the percentage of areas of high regularity (high ANN), especially when using small investigation windows ($2 \times 2 \text{ km}^2$). With larger windows, the number of objects in areas with high ANN values decreases, so that it is uncertain how valid the results on the regularity are. It is also qualitatively shown in Table 3 that although density and ANN can be mapped, the size distribution in the empirical data fluctuates more than a theoretically regular pattern. As indicated by an analysis for large cities in

the Nile Valley [15], the analyses can be continued in different classes. In further studies, the regularity of the larger cities introduced in Section 4.3 should be examined in detail.

At the beginning of this paper we asked the question whether urban structures (or parts of them) are arranged in a regular pattern and thus exhibit a typical property of non-equilibrium pattern formation in dissipative, open systems. As one might expect, regular structures cannot be observed in the entire study areas. However, we have been able to identify regular settlement patterns, especially in areas with a high number and density of settlement structures.

Based on these findings, a further step is to investigate which models can be used to depict the pattern formation of urban structures. For example, it can be analysed whether reaction–diffusion models, the pattern formation through the interaction of different morphogenes, are suitable to describe the structure formation. For this purpose, fundamental work by Rudy et al. [40] or Zhao et al. [41] can be referred to, who have identified the underlying mechanisms of pattern formation by analysing image sequences. Their proposed methods can be used to analyse time-resolved settlement maps and quantify their changes by using concepts from physics.

Since regular structures were identified mainly in small evaluation windows, the nonlinear pattern formation models to be developed should also focus primarily on small scales.

6. Conclusions

The results in this paper aid to answer whether settlements are arranged regularly in different regions of the world. The motivation for posing this question is whether cities as thermodynamic open systems share similarities in pattern with dissipative structures.

The settlement patterns of seven regions in five continents were investigated. While the region in China had a high number of settlement objects and a high settlement density, and the regions in the USA and Argentina had a low number of objects and settlement density, the other regions studied (Ghana, France, India and Egypt) were in between. The settlement density and the number of settlements correlate with each other, which can be explained by the relatively similar size distributions of the settlements with a geometric mean of nearly 0.01 km^2 . The number of settlements as well as the density correlates with the proportion of sample windows with a high regularity $ANN > 1.5$. The analyses with the highest resolution (sample window of $2 \times 2 \text{ km}^2$) showed that between approximately 1% (USA) and 14% (China) of the investigated area show regular settlement structures. This means that the more densely populated a region is, the more likely it is that regular settlement structures can be found in it.

These results can be used to develop mathematical models that can describe the pattern formation of urban structures by models of non-equilibrium thermodynamics. Furthermore, the results could be expanded in future work and brought together with other research fields: For example, (i) in infrastructural studies it could be analysed whether the temporal development of transport infrastructures had an influence on the inter-urban pattern formation. Furthermore, (ii) it could be investigated to what extent political systems (politics and policies) are correlated with pattern formation. These aspects could then be explicitly considered in the mathematical models. Finally, (iii) it can be investigated to what extent the economic development of a region is related to or influences the arrangement of settlements. All these insights are useful for a better understanding of urbanisation processes and can, for example, help urban planners to develop sustainable concepts to meet these challenges.

Author Contributions: Conceptualisation, John Friesen and Jakob Hartig; Methodology, Katharina Henn, John Friesen and Jakob Hartig; Formal analysis, Katharina Henn; Investigation, Katharina Henn; Writing—original draft preparation, Katharina Henn, John Friesen and Jakob Hartig; Writing—review and editing, Katharina Henn, John Friesen and Jakob Hartig; Visualisation, Katharina Henn; Supervision, Peter F. Pelz; Project administration, Peter F. Pelz. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank the German Aerospace Center for providing the data of the Global Urban Footprint and acknowledge support by the DFG and the Open Access Publishing Fund of Technical University of Darmstadt for publishing.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript.

ANN Average-Nearest Neighbour Index
 CPT central place theory
 DLR German Aerospace Center
 GUF Global Urban Footprint
 s.s. sample square

Appendix A. Further Data

Table A1. Coordinates of all seven considered datasets. The reference system is WGS 84, the coordinates are given in decimal degrees.

Data Set	West	South	East	North
Argentina	−63.424	−33.800	−61.265	−31.996
China	114.065	32.800	116.200	34.603
Egypt	30.014	29.876	32.084	31.680
France	3.350	47.401	6.000	49.200
Ghana	−2.278	5.419	0.473	7.228
India	74.400	28.896	76.451	30.700
USA	−100.000	38.500	97.707	40.301

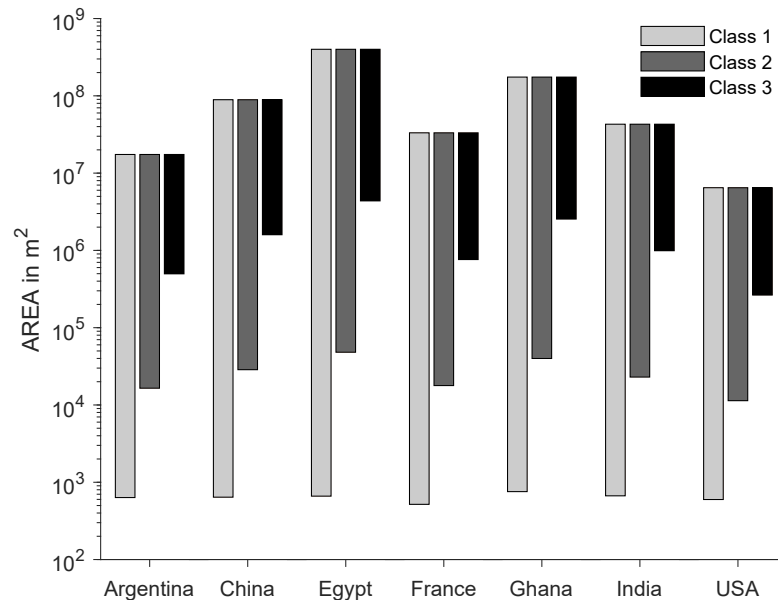


Figure A1. The columns represent the range of the settlement objects' area for the different classes in each data set.

Table A2. Percentage of sample squares (s.s.) with $ANN > 1.5$ for each data set and the according number of data points N considered to evaluate the $ANN > 1.5$ in these s.s. Additionally, the percentage of sample squares with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$, and the average number of objects N in these sample squares.

	2 km	4 km	6 km	8 km	10 km	12 km	15 km
Argentina							
% of s.s. with $ANN > 1.5$	4.28	7.45	5.05	2.69	1.25	0.94	0.23
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	68.18	71.53	79.42	85.03	81.63	87.50	75.00
average N in s.s. with $ANN > 1.5$	5.44	5.69	6.77	7.20	7.10	7.65	6.20
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	5.58	5.85	7.16	7.49	7.49	8.00	6.33
China							
% of s.s. with $ANN > 1.5$	13.97	1.75	0.07	0.00	0.00	0.00	0.00
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	82.96	94.71	100.00	0.00	0.00	0.00	0.00
average N in s.s. with $ANN > 1.5$	6.56	12.24	22.86	-	-	-	-
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	6.79	12.60	22.86	-	-	-	-
Egypt							
% of s.s. with $ANN > 1.5$	9.58	2.30	0.50	0.18	0.16	0.00	0.03
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	77.03	84.73	75.76	76.92	58.33	0.00	100.00
average N in s.s. with $ANN > 1.5$	6.08	7.54	6.42	7.31	6.75	-	10
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	6.31	7.91	6.68	8.00	7.57	-	10
France							
% of s.s. with $ANN > 1.5$	1.87	0.95	0.31	0.07	0.00	0.00	0.00
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	70.6	71.43	67.86	71.43	0.00	0.00	0.00
average N in s.s. with $ANN > 1.5$	5.78	5.89	5.86	7.14	-	-	-
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	6.00	6.09	6.16	7.60	-	-	-
Ghana							
% of s.s. with $ANN > 1.5$	1.62	2.86	2.03	0.87	0.28	0.06	0.19
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	74.80	70.83	79.28	82.54	86.36	100.00	68.75
average N in s.s. with $ANN > 1.5$	6.31	5.53	6.68	7.24	8.95	5.80	5.63
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	6.62	5.61	6.97	7.52	9.42	5.80	5.64
India							
% of s.s. with $ANN > 1.5$	7.30	4.00	1.56	0.65	0.31	0.20	0.08
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	70.27	75.23	76.74	67.80	85.71	94.44	100.00
average N in s.s. with $ANN > 1.5$	5.82	6.34	6.30	6.59	6.32	7.39	11.29
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	6.03	6.69	6.54	7.23	6.54	7.41	11.29
USA							
% of s.s. with $ANN > 1.5$	0.80	2.73	3.51	2.60	1.73	0.80	0.21
% of s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	63.64	58.54	77.30	73.99	76.43	85.51	94.44
average N in s.s. with $ANN > 1.5$	5.67	5.37	5.91	6.09	6.41	6.86	7.28
average N in s.s. with $ANN > 1.5$ and $z < -2.58$ or $z > 2.58$	5.86	5.42	6.09	6.32	6.72	7.05	7.41

References

- Retief, F.; Bond, A.; Pope, J.; Morrison-Saunders, A.; King, N. Global megatrends and their implications for environmental assessment practice. *Environ. Impact Assess. Rev.* **2016**, *61*, 52–60. [\[CrossRef\]](#)
- Zhu, Z.; Zhou, Y.; Seto, K.C.; Stokes, E.C.; Deng, C.; Pickett, S.T.; Taubenböck, H. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote. Sens. Environ.* **2019**, *228*, 164–182. [\[CrossRef\]](#)
- Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013.
- Batty, M. Building a science of cities. *Cities* **2012**, *29*, S9–S16. [\[CrossRef\]](#)
- Portugali, J. *Complexity, Cognition and the City*; Springer complexity, Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2011.
- Taubenböck, H.; Debray, H.; Qiu, C.; Schmitt, M.; Wang, Y.; Zhu, X. Seven city types representing morphologic configurations of cities across the globe. *Cities* **2020**, *105*, 102814. [\[CrossRef\]](#)
- Benenson, I.; Torrens, P.M. *Geosimulation: Automata-Based Modelling of Urban Phenomena*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
- Liu, Y.; Batty, M.; Wang, S.; Corcoran, J. Modelling urban change with cellular automata: Contemporary issues and future research directions. *Prog. Hum. Geogr.* **2019**. [\[CrossRef\]](#)
- Rathore, M.M.; Ahmad, A.; Paul, A.; Rho, S. Urban planning and building smart cities based on the internet of things using big data analytics. *Comput. Netw.* **2016**, *101*, 63–80. [\[CrossRef\]](#)
- Wu, N.; Silva, E.A. Artificial intelligence solutions for urban land dynamics: A review. *J. Plan. Lit.* **2010**, *24*, 246–265.
- Zhao, K.; Tarkoma, S.; Liu, S.; Vo, H. Urban human mobility data mining: An overview. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1911–1920.
- Healy, K. Fuck Nuance. *Sociol. Theory* **2017**, *35*, 118–127. [\[CrossRef\]](#)
- Christaller, W. *Die Zentralen Orte in Süddeutschland: Eine Ökonomisch-Geographische Untersuchung Über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit Städtischen Funktionen*; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1933.
- Allen, P.M.; Sanglier, M. A dynamic model of growth in a central place system. *Geogr. Anal.* **1979**, *11*, 256–272. [\[CrossRef\]](#)
- AbouKorin, A.A. Spatial analysis of the urban system in the Nile Valley of Egypt. *Ain Shams Eng. J.* **2018**, *9*, 1819–1829. [\[CrossRef\]](#)
- Yang, R.; Xu, Q.; Long, H. Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanisation. *J. Rural Stud.* **2016**, *47*, 413–424. [\[CrossRef\]](#)
- Krugman, P. *The Self-Organizing Economy*; Number 338.9 KRU 1996; CIMMYT: Mexico City, Mexico, 1996.
- Cross, M.C.; Hohenberg, P.C. Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **1993**, *65*, 851. [\[CrossRef\]](#)
- Purvis, B.; Mao, Y.; Robinson, D. Entropy and its application to urban systems. *Entropy* **2019**, *21*, 56. [\[CrossRef\]](#)
- Levashova, N.; Sidorova, A.; Semina, A.; Ni, M. A Spatio-Temporal Autowave Model of Shanghai Territory Development. *Sustainability* **2019**, *11*, 3658. [\[CrossRef\]](#)
- Medda, F.; Nijkamp, P.; Rietveld, P. A Morphogenetic Perspective on Spatial Complexity. In *Complexity and Spatial Networks*; Reggiani, A., Nijkamp, P., Eds.; Springer: Berlin/ Heidelberg, Germany, 2009; pp. 51–60.
- Pelz, P.F.; Friesen, J.; Hartig, J. Similar size of slums caused by a Turing instability of migration behavior. *Phys. Rev.* **2019**, *99*. [\[CrossRef\]](#)
- Barrington-Leigh, C.; Millard-Ball, A. A century of sprawl in the United States. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8244–8249. [\[CrossRef\]](#)
- Urbietta, P.; Fernandez, E.; Ramos, L.; Méndez Martínez, G.; Bento, R. A land-cover based urban dispersion indicator suitable for highly dispersed, discontinuously artificialized territories: The case of continental Portugal. *Land Use Policy* **2019**, *85*, 92–103. [\[CrossRef\]](#)
- Romano, B.; Zullo, F.; Fiorini, L.; Ciabò, S.; Marucci, A. Sprinkling: An Approach to Describe Urbanization Dynamics in Italy. *Sustainability* **2017**, *9*, 97. [\[CrossRef\]](#)

26. Saganeiti, L.; Favale, A.; Pilogallo, A.; Scorza, F.; Murgante, B. Assessing Urban Fragmentation at Regional Scale Using Sprinkling Indexes. *Sustainability* **2018**, *10*, 3274. [\[CrossRef\]](#)
27. Ewing, R.H. Characteristics, Causes, and Effects of Sprawl: A Literature Review. In *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*; Marzluff, J.M., Shulenberger, E., Endlicher, W., Alberti, M., Bradley, G., Ryan, C., Simon, U., ZumBrunnen, C., Eds.; Springer: Boston, MA, USA, 2008; pp. 519–535.
28. Kondo, S.; Miura, T. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* **2010**, *329*, 1616–1620. [\[CrossRef\]](#)
29. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *Isprs J. Photogramm. Remote. Sens.* **2017**, *134*, 30–42. [\[CrossRef\]](#)
30. Esch, T.; Bachofer, F.; Heldens, W.; Hirner, A.; Marconcini, M.; Palacios-Lopez, D.; Roth, A.; Üreyen, S.; Zeidler, J.; Dech, S.; et al. Where we live—A summary of the achievements and planned evolution of the global urban footprint. *Remote Sens.* **2018**, *10*, 895. [\[CrossRef\]](#)
31. Esch, T.; Schenk, A.; Ullmann, T.; Thiel, M.; Roth, A.; Dech, S. Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information. *IEEE Trans. Geosci. Remote. Sens.* **2011**, *49*, 1911–1925. [\[CrossRef\]](#)
32. Hudson, J.C. A Location Theory for Rural Settlement. *Ann. Assoc. Am. Geogr.* **1969**, *59*, 365–381. [\[CrossRef\]](#)
33. Friesen, J.; Hartig, J.; Henn, K.; Pelz, P.F. Diffusion-driven demographics—Turing model as a concept for the emergence of sedentism. *arXiv* **2020**, arXiv:2005.05107.
34. Clark, J.P.; Evans, C.F. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology* **1954**, *35*, 445–453. [\[CrossRef\]](#)
35. Friesen, J.; Taubenböck, H.; Wurm, M.; Pelz, P.F. The similar size of slums. *Habitat Int.* **2018**, *73*, 79–88. [\[CrossRef\]](#)
36. Friesen, J.; Taubenböck, H.; Wurm, M.; Pelz, P.F. Size distributions of slums across the globe using different data and classification methods. *Eur. J. Remote Sens.* **2019**, *52*, 99–111. [\[CrossRef\]](#)
37. Klotz, M.; Kemper, T.; Geiß, C.; Esch, T.; Taubenböck, H. How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sens. Environ.* **2016**, *178*, 191–212. [\[CrossRef\]](#)
38. Siksna, A. The effects of block size and form in North American and Australian city centres. *Urban Morphol.* **1997**, *1*, 19–33.
39. Goodall, D.W.; West, N.E. A comparison of techniques for assessing dispersion patterns. *Vegetatio* **1979**, *40*, 15–27. [\[CrossRef\]](#)
40. Rudy, S.H.; Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Data-driven discovery of partial differential equations. *Sci. Adv.* **2017**, *3*, e1602614. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Zhao, H.; Storey, B.D.; Braatz, R.D.; Bazant, M.Z. Learning the physics of pattern formation from images. *Phys. Rev. Lett.* **2020**, *124*, 060201. [\[CrossRef\]](#) [\[PubMed\]](#)



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).