Scale-dependent complexity in administrative units and implications for data-driven decision-making models

Søder, Peter Højrup

Published in:
Planning Theory

DOI:
10.1177/14730952231203151

Publication date:
2024

Document Version
Peer reviewed version

Citation for published version (APA):
Søder, P. H. (2024). Scale-dependent complexity in administrative units and implications for data-driven decision-making models. *Planning Theory*, *23*(2), 131-156. https://doi.org/10.1177/14730952231203151

Scale-dependent complexity in administrative units and implications for data-driven decision-making models

Planning Theory

Abstract

Through analysis and discussion of basic systemic properties of a rural municipality, this paper explores how aggregating data in planning and land use modeling can potentially obscure intricate real-world behavior. Complexity theory is applied as a theoretical framework for explaining this hypothesis. Thus, the aim of this study is to address the author's desire to understand systemic complexity when designing a data-driven decision-making model for rural planning. The novelty of this approach is two-fold: one, most studies on scalability issues in planning addresses spatial complexity, more so than systemic complexity within the complex system that the very act of planning strives to dictate. Two, although delimited to the scope of the study, the accessibility to and use of complete and valid socio-demographic data enables a rarely demonstrated accurate representation of an entire population. It is ultimately observed that on the disaggregated municipal level, systemic dispersion increases parallelly with population size, a correlation that is significantly influenced by gender ratio in any given parish – a characteristic that was not visible at the aggregated municipal level. In addition to advancing the understanding and placement of complexity science within spatial data science, these insights will make it easier to assess the generalizability of any given administrative unit by quantifying basic complexity attributes; in this case based on the correlation dispersion caused by the fragmentation of a municipality into its comprising parishes.

Keywords: Planning; Spatial data science; DDDM; Scale; MAUP; Complexity theory; Regression analysis

1. Introduction

Planning has traditionally been associated with a reductionist perspective of cause and effect (Byrne, 2003; Chettiparamb, 2014) – a perspective that essentially delimits itself to an understanding of mechanical complicatedness, more so than social complexity. However, aggregating and simplifying data across spatial scales causes issues with oversimplification and loss of statistical generalizability (Jelinski and Wu, 1996; Kar and Hodgson, 2012; Dapena et al., 2016; Garreton and Sánchez, 2016; Stillwell et al., 2018; Xiao, 2021). This discrepancy is the entry point to the research question of this paper, which concerns a basic premise of planning: i.e., how abstract can a political decision-making model be without compromising its generalizability? The aim of this paper is to address complex systemic properties of planning, and their potential implications for a data-driven decision-making (DDDM) model for planning. This is achieved through deliberately simple and abductive, pragmatic analysis of the complexity embedded within the administrative units on which planning is stratified, and by how aggregating basic statistical data can potentially disguise complex behavioral patterns. Thus, this paper is an analysis of model scalability in the context of rural planning that utilizes a predominantly quantitative approach, albeit qualitatively informed by the theoretical framework of complex systems.

Complexity theory has been widely applied to a plethora of different research areas (Chettiparamb, 2006; Walton, 2014), including the social sciences (Ratter, 2006; Timmermans, 2015; Ekman, 2018; Chettiparamb, 2019; Kwon & Silva, 2020; Fredslund et al., 2022) and the development of modern planning theory (Portugali, 2012; Eräranta, 2020; Portugali, 2021a; Portugali 2021b; Dhamo, 2021). A recent example of application of complexity theory in planning is Totry-Fakhouri and Alfasi (2017), who identifies towns and cities in Israel as adaptive, self-organizing urban developments that are guided by spontaneous emergence and social order, defined as the 'urban code' of the settlements. Moreover, fractal development structures are observed across singular plots, neighborhoods, and entire villages, which is an inherent characteristic of complex systems (Mandelbrot, 1982; Gleick, 1987). According to Chettiparamb (2005), planning takes places in a fractal system, as decisions are largely consistent going across vertically different administrative levels – i.e., national planning strategies are reflected in local planning strategies. However, as the abstraction level decreases along the vertical levels, the system acquires unpredictable properties at each separate horizontal level, thus reflecting scale-dependency; a fundamental property of complex systems (Mitleton-Kelly, 2004). A well-known example of fractal properties in spatial complexity, defined as the complexity of surfaces and spatial objects (Papadimitriou, 2020), is the coastline paradox (Mandelbrot, 1982), which demonstrates the fractal properties of coastlines. In this paradox, smaller units of measure cause larger measurements, making the seemingly simple task of measuring the coastal length quite complex. In the context of planning and related sciences, this property can cause issues when comparing or compiling different datasets, especially if the datasets represent different scales or levels of abstraction (Jacobs-Crisioni, 2014). More recently, Salvati (2022) observes nonlinear properties in urban growth patterns on the municipal scale in Greece by analyzing official data on building activity. Conclusively, the author calls for future research that enriches the proposed methodological approach, while emphasizing identification of target indicators. The present paper provides a response to this call by analyzing complexity target indicators across multiple levels in a Danish rural municipality. By adopting a theoretical framework based on complex social systems, the analysis presented in this paper contributes to positioning complexity science within the still maturing field of spatial data science, as well as provides insights for assessing the generalizability of administrative units.

The use case for this paper is Guldborgsund Municipality (GM), a rural municipality located in the southern part of Region Zealand. The municipality have experienced a population decline of 4% since 2012 (Statistics Denmark, 2022a) and is along with Lolland Municipality one of the most socioeconomically disadvantaged areas of Denmark (Jepsen et al., 2018; 2020). Roughly half the population of GM lives in its rural districts, which are characterized by the effects of emigration and centralization (Guldborgsund Municipality, 2018). With demographic changes in the rural districts and increasing diversity in living arrangements like multi-family housing, communal farming, ecological villages and minor settlements, rural planning is quickly starting to become a complex task in the municipality. For this reason, the municipality has decided to employ a DDDM model for political planning purposes, to provide an informed basis for the development of future planning strategies.

2. Theory

Complexity science is concerned with the fundamental logical properties of the behavior of nonlinear and network feedback systems, no matter where they are found (Stacey, 1995). Although there is no official and unambiguous definition of theoretical complexity, it can be defined as the degree of interrelatedness between systemic attributes and interfaces, and their consequential impact on predictability and functionality (Kermanshachi et al., 2016). Thus, to reduce the complexity of any given system, its potential points of failure must be reduced. Deriving from complexity theories of the natural and social sciences, ten principles of complexity are identified by Mitleton-Key (2004) under the unifying term 'complex evolving systems'. Among these theories are complex adaptive systems (Kaufmann and Macready, 1995), autopoiesis (Maturana and Valera, 1992) and chaos theory (Gleick, 1987), all of which possesses the overall characteristic of creating new order through self-organization, emergence, connectivity, interdependence, feedback, historicity and time, and path-dependence. Summarily, the creation of new order through iterations and feedback loops is identified as a distinguishing characteristic for complex systems.

Specifically for urban design, Boeing (2018) identifies five complexity dimensions: temporal, visual, spatial, scaling, and connectivity. For the purposes of this paper, I will be focusing on the scaling dimension, which inscribes self-similarity across multiple structures, and fractal patterns: *"Consider the example of an urban street network. At the largest scale, the city has a few major arterial roads and boulevards that serve as the key routes for system-wide traffic circulation. But if we zoom into this picture, a larger number of mid-sized collector streets appear, branching off from these few large arteries. As we zoom in further to a fine scale, a denser mesh of local streets appears, branching off from these collector streets. Certain distributions within a complex system may produce greater efficiency when they follow a power law rather than, say, an even distribution. For example, it is not ideal for a neighborhood to have the same number of arterial roads, collector streets, and local streets. Rather, there might be a small number of large arterial roads, a medium number of midsized collector streets, and a large number of capillary local streets. Such a system resembles a fractal"* Boeing (2018: p. 10). At this point, addressing a potential distinction between 'urban planning' and 'rural planning' is arguably due, but beside the point for this paper, as it merely addresses general characteristics of complexity in planning in a rural context. As a sightline, however, the definition of 'rural' is understood as *"of or relating to the country, country people or life, or agriculture"* (Merriam-Webster, 2022), while the definition of a rural municipality is based on accessibility to jobs and the number of inhabitants in the largest city in the municipality (Statistics Denmark, 2018). Finally, Boeing (2018: p. 3) provides an interpretation of complexity as a function of disorder, where complexity increases with disorder, in which complexity is highest *"[…] when objects are scrambled-up with the greatest variety and diversity."* This interpretation of complexity is represented in this paper by observing standard deviation (SD) as an indication of how dispersed a given dataset is – i.e., more dispersion equals less order (Shiner et al., 1999). It should be noted, however, that I merely apply SD as a *comparative metric* for complexity, more so than an absolute measure of complexity itself. This enables me to observe potential changes in the demographic data across different spatial scales.

Issues relating to scalability in planning are commonly referred to as the *modifiable areal unit problem* (MAUP), which describes how data tabulated for different spatial scale levels or zonal systems will not provide consistent analytical results, despite representing the same region (Wong, 2009). Simply put, the MAUP expresses the sensitivity of analytical results to the definition of units for which data are collected (Fotheringham and Wong, 1991). Figure 1 illustrates different levels of scale in a fractal system for planning on two axes. A vertical axis through which information is distributed across the different levels, and a horizontal axis through which the system acquires disorder on each level. This way the global and the local gets defined simultaneously, and *"the onward progression along the coupled multiple levels can thus allow for variety and order to co-exist at the same time"* (Chettiparamb, 2005: p. 330). Specifically for this paper, these levels represent the administrative units wherein demographic data is aggregated. It is thus hypothesized, that self-similarity in complex system characteristics can be observed across parishes, municipalities, and regions.

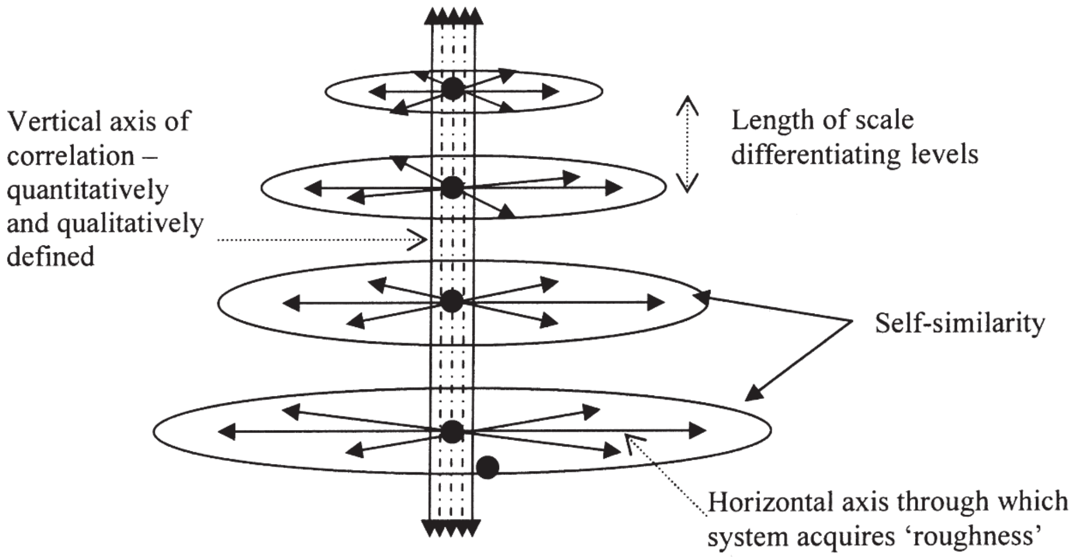

Figure 1: The four guiding parameters of a fractal system for planning (Chettiparamb, 2005).

In summation, the theoretical framework for this paper observes complexity as the degree of interrelatedness between systemic attributes and interfaces (Kermanshachi et al., 2016); the practice of planning as taking place in a social system in which complexity is a function of disorder (Boeing, 2018); and, finally, planning itself as a fractal system where self-similarity can be observed across its hierarchical levels (Chettiparamb, 2005; 2019).

3. Methodology

This paper is a predominantly quantitative analysis, albeit qualitatively informed by a theoretical framework of complexity. Hence, a mixed methods approach is applied (Frederiksen et al., 2014), which utilizes a combination of statistical regression analysis of demographic registry data, and a discussion of the results from this analysis that is inferred through complexity theory in a social science context. The methodology section features two tables for providing an overview of the applied theoretical terms and concepts (table 1), and methodology for data aggregation analysis across spatial scales (table 2), in the subsequent analysis. Note however, that these descriptions and examples of use are general and will be elaborated throughout the paper as needed.

Context	Term/concept	Description	Use in paper
Complex systems	Complexity	Iterative and nonlinear real-world behavior	Theoretical framework
	Interrelatedness	Having mutual or reciprocal relations	Defining attribute of complexity
	Numerosity	Consisting of many components	Defining attribute of complexity
	Disorder	State prone to unpredictable behavior	Comparative complexity measure
	Fractals	Self-similar patterns across different scales	Framing of administrative units
Statistical analysis	Standard deviation (SD)	Measure of how dispersed a dataset is	Comparative metric for dispersion
	Standard error (SE)	SD relative to sample size (n)	Proxy for numerosity when needed
	Correlation (R^2)	Dependence between two variables	Modeling relationship strength
	Residuals	Difference between observation and prediction	Identification of outliers
	Significance (p-value)	How likely it is a result happens by chance	Test for statistical significance

Table 1: Overview of general theoretical terms and concepts used in this paper.

The Danish municipal unit corresponds to a local administrative unit (LAU) in Eurostat's 'Nomenclature of Territorial Units for Statistics' (NUTS) and is generally used in Danish planning. To identify and analyze the hypothesized statistical changes between administrative units, a hierarchical deconstruction of the municipal administrative unit of Denmark, in which a group of parishes constitutes a municipality (LAU in NUTS), is conducted. In Denmark, an ecclesiastical parish is a geographically delimited area with at least one church.

Each deconstruction consists of two datasets, referred to as 'method A' and 'method B', respectively. As illustrated in table 2, on the cross-regional level, the dataset for method A consists of one column for each municipality that features the number of people who are the same age – i.e., aggregated data. For instance, 258 people of age 0, 279 people of age 1, 343 people of age 2 and so on up to 173 people of age 82. In this example for method A, this results in 83 observations – one for each age increment – and an SD of 84. In method B, these aggregations are deconstructed as a matrix across the parishes that constitute the given municipality – i.e., disaggregated data. For instance, the sum of 258 for age 0 in method B is spread across five parishes, and so is the sum for age 1, age 2 and on so on up to age 82. Following the same example, this results in 415 observations – one for each age increment multiplied by the number of parishes – and an SD of 40.

	Cross-regional level analysis						Municipal level analysis				
	Method A	**Method B**					**Method A**				
Age	Municipality	Municipality					Municipality				
		Parish	Parish	Parish	Parish	Parish	Parish	Parish	Parish	Parish	Parish
0	258	26	26	59	95	64	26	26	59	95	64
1	279	28	19	60	105	67	28	19	60	105	67
2	343	36	30	79	119	75	36	30	79	119	75
…	…	…	…	…	…	…	…	…	…	…	…
82	173	11	20	14	92	32	11	20	14	92	32
n	*83*	*415*					*83*	*83*	*83*	*83*	*83*
SD	*84*	*40*					*13*	*9*	*16*	*31*	*27*

Table 2: Methodology for comparing aggregated data to disaggregated data.

Two variables are then isolated for each dataset for comparison. These are the sum of inhabitants, and SD of age distribution. For the purposes of this paper, I utilize SD as a basic, comparative measure of dispersion, which I in turn interpret as a measure of complexity by increase of disorder (Boeing, 2018). This approach is inspired by Lee and Rogers (2019), who applies the coefficient of variance (SD divided by the mean) as a dispersion measure for geographic distribution. The correlations (R^2) between the two variables were then calculated by linear regression to compare the strength of the relationship between the sum of inhabitants, and SD of age distribution across different spatial scales. The parish level (also a LAU in NUTS) represents the lowest abstraction level of the analysis and for this reason only method A can be applied to this administrative level (table 2), seeing that a parish is the smallest administrative unit accessible in the dataset. This means an SD is calculated for each parish on the municipal level, the same way an SD is calculated for each municipality on the cross-regional level. Herein lies the comparison, namely how the analysis and resulting outcomes are affected by the two different methods regarding the number of observations, SD, and correlation dispersion. Because this is a comparison between two different treatments of the same population, the analytical focus lies on the differences in the results between the two methods and how these can be interpreted as complexity indicators.

The results from this analysis will inform my working hypothesis, which states that statistical outliers based on aggregated data differs considerably from results based on disaggregated data, even though both datasets are intended to represent the same population. All the empirical data is extracted from Statistics Denmark, the central authority on Danish demographic statistics. In the following analysis, I will thus be focusing on age as an independent variable, i.e., the number of people between age 0 and 82 – 82 years, which is the average combined lifespan for Danish men and women (Statistics Denmark, 2021). This variable is selected because it is simple, universally applicable, and its representative data is easily obtained. In addition to this delimitation, Statistics Denmark's definition of a rural municipality (Statistics Denmark, 2018) will be applied in the analysis. This definition states that a rural municipality cannot contain any cities with more than 30,000 inhabitants and must have a job availability below 40,000 – i.e., the number of jobs that are accessible within the municipal border. This means that 67 municipalities from a total of 98 are excluded from the analysis, as these all contain relatively large cities with high job availability, which opposes the rural scope of the paper. The remaining 31 municipalities represents a total population sample of 1,121,813 (19.7%) of the country's entire population within the age group 0 to 82, and 806 (37.8%) of all the country's parishes.

4. Analysis and results

4.1 Cross-regional deconstruction of rural municipalities

As a proof-of-concept for the basic logic that is applied in the analysis, a cross-regional analysis of all the 31 rural municipalities is conducted by exploring the relationship between population size and SD by its residuals. Simply put, residual values represent the distance between dot and trendline in a scatterplot, commonly referred to as the difference between predicted and observed value. A value of zero is a perfect fit to the trendline, whereas a negative number puts the dot beneath the line, while a positive number puts it above it. The larger the number, negative or positive, the weaker the relationship between the dot and the trendline. Moreover, this analysis serves to demonstrate the effects on correlation dispersion in aggregated data versus disaggregated data; specifically, as a demonstration of the importance of spatial data scientists recognizing the potentially concealing effects statistical treatment can have on DDDM in planning, as well as to demonstrate how such practices can benefit from adopting a qualitative understanding of the real-world phenomena they strive to model, namely through complexity theory.

The regions of Denmark (NUTS 2) represent the highest abstraction level of the analysis with all five regions represented. Table 3 ranks all 31 rural municipalities according to their residual values from top to bottom and compares the results in method A to the results in method B, represented by two major columns. Each column is divided into intervals with dotted lines that indicates a range of residuals with similar values. The grey area indicates residuals that are within the range -5,000 to 5,000, which are the municipalities that are closest to zero, thus isolating the largest deviations at the far top and bottom of the table. This range was chosen solely to capture enough differences between the two methods to make a comparison, while staying relatively close to zero. Note how for method A, the R^2 (model strength) is 0.81 and the residuals range from -18,021 to 13,333, while for method B the R^2 is 0.41 and the residuals range from -28,854 to 24,835. Thus, the data in method B is more dispersed than the data in method A.

For instance, with a residual value of 13,333, Svendborg Municipality is the largest positive outlier for method A, while Odsherred Municipality is the largest negative outlier, with a residual value of -18,021. For each method in table 3, 13 municipalities have been denoted with one or two asterisks. These asterisks denote considerable changes in rank between the two methods relative to the residual range closest to zero (the grey area). One asterisk denotes that a municipality considered an outlier in method A falls within the low residual range in method A, and vice versa. In method A these municipalities are Varde (9,454), Frederikshavn (-7,113), Vordingborg (-10,303), Bornholm (-14,765), and Odsherred (-18,021). The SD for the whole dataset in method A ranges from 12 to 166 with an average of 105. This places Varde close to the average with an SD of 101, and Frederikshavn, Vordingborg, Bornholm and Odsherred well above with SDs of 166, 141, 138, and 130, respectively. However, this order changes in method B.

Method A R^2 = 0.81 | N = 2,573						Method B R^2 = 0.41 | N = 66,898					
Municipality	**Population**	**Parishes**	**n**	**SD**	**Residual**	**Municipality**	**Population**	**Parishes**	**n**	**SD**	**Residual**
Svendborg	56,944	28	83	115	13,333	Ringkøbing-Skjern	53,549	3	3,818	18	24,835
Sønderborg	70,409	25	83	156	10,773	Sønderborg	69,273	45	2,075	32	21,736
Haderslev	53,059	27	83	115	9,468	Svendborg	56,782	20	2,324	23	21,428
Varde*	47,644	28	83	101	9,454	Hjørring	61,153	13	3,320	28	18,895
Ringkøbing-Skjern	53,993	46	83	119	8,685	Kalundborg**	46,536	25	2,490	21	14,870
Hjørring	61,188	40	83	141	7,576	Lolland	38,050	27	3,984	17	10,953
Aabenraa	56,490	20	83	130	7,249	Skive	43,456	23	3,735	21	10,845
Brønderslev	34,864	20	83	76	6,432	Thisted	41,437	1	4,399	20	10,742
Thisted	41,450	53	83	94	6,122	Guldborgsund**	57,597	28	3,569	32	10,569
Billund	25,741	11	83	54	5,835	Tønder**	35,268	28	2,490	19	6,105
Vesthimmerlands	34,794	38	83	78	5,474	Vesthimmerlands	34,885	26	3,154	19	5,708
Skive	43,193	45	83	100	5,384	Haderslev	52,982	11	2,241	32	5,524
Jammerbugt	36,817	28	83	88	3,818	Jammerbugt	36,835	48	2,324	21	4,663
Mariagerfjord	40,098	29	83	102	1,393	Mariagerfjord	39,899	21	2,407	24	3,172
Morsø	19,159	32	83	51	313	Frederikshavn	56,162	22	1,826	37	2,856
Kalundborg	46,660	30	83	126	-1,142	Vordingborg	43,431	46	2,158	29	-169
Norddjurs	35,407	35	83	97	-1,371	Lemvig	18,518	35	1,909	12	-980
Tønder	35,230	30	83	97	-1,531	Varde	47,216	32	2,324	33	-1,206
Guldborgsund	57,696	43	83	155	-1,533	Odsherred	31,446	28	1,079	22	-2,050
Struer	19,907	16	83	59	-1,906	Norddjurs	35,406	43	2,905	26	-3,890
Læsø	1,656	3	83	12	-1,907	Bornholm	37,648	40	1,743	29	-4,903
Fanø	3,278	1	83	19	-3,139	Langeland	11,543	20	1,494	11	-7,163
Lemvig	18,509	23	83	58	-3,150	Læsø**	1,644	18	249	5	-8,179
Samsø	3,503	1	83	21	-3,669	Aabenraa	55,199	30	1,660	44	-8,603
Ærø	5,652	6	83	31	-5,443	Morsø**	19,141	1	2,656	18	-9,354
Frederikshavn*	56,141	22	83	166	-7,113	Brønderslev	34,750	6	1,660	33	-14,124
Vordingborg*	43,493	26	83	141	-10,303	Struer**	19,903	30	1,328	22	-14,135
Langeland	11,555	18	83	60	-10,803	Ærø	5,642	53	498	12	-14,202
Bornholm*	37,721	21	83	138	-14,765	Fanø**	3,274	29	83	19	-26,515
Lolland	38,153	48	83	141	-15,511	Samsø**	3,505	16	83	21	-28,575
Odsherred*	31,409	13	83	130	-18,021	Billund	25,607	38	913	37	-28,854

* Denotes a municipality with a residual range that is outside the range -5,000 to 5,000 (grey area) in method A, but is inside the range in method B.
** Denotes a municipality with a residual range that is outside the range -5,000 to 5.000 (grey area) in method B, but is inside the range in method A.

Pop. = Population in each municipality
n = Number of observations for each municipality

Table 3: Cross-regional analysis of data aggregation based on the 31 rural municipalities of Denmark.

With a considerably lower correlation of R^2 0.41, method B is more dispersed than method A, as previously illustrated. The SD for the whole dataset in method B ranges from 5 to 44 with an average of 25, calculated on sample sizes that ranges from 83 to 4,399. Here, only nine municipalities fall within the grey area in table 3, compared to 12 in method A. Moreover, eight municipalities that were within the grey area in method A, are placed outside the grey area in method B, with four of them at the far ends of the residual range. In other words – one third of the municipalities that were most correlated in method A, are among the least correlated in method B. The outliers in method B shares some common characteristics. Samsø, Fanø, and Læsø are all island municipalities with populations below 5,000 – a characteristic that contradictorily appeared to place their residual values closer to zero in method A. Interestingly, some of the municipalities with large populations have remained extreme outliers in both methods, e.g., Svendborg, Sønderborg and Ringkøbing-Skjern.

Given that the only difference between the two methods is based on the number of parishes within each municipality, and thus their fragmentation, this is a plausible part of the explanation. To test this notion, SD was replaced with standard error (SE) for a second trial. Method A is based on an aggregated dataset with equal sample sizes (there are 83 observations in each municipality) and is therefore unaffected by replacing SD with SE, observed with an identical R^2 of 0.81. Contrarily, method B displayed a virtually non-existing correlation between the two variables SE and population size, observed with an R^2 of 0.09. Considering SE as a function of SD and sample size, this observation supports the notion of the model's sensitivity to numerosity and fragmentation – i.e., on the municipal level, there are 26 times more samples in method B than method A. As observed by the reduced correlation between SD and population size from method A to method B, the model's sensitivity to numerosity and aggregation becomes even more apparent when observing SE instead of SD.

The preliminary results from this cross-regional analysis indicates noticeable differences between using aggregated data (method A) or disaggregated data (method B) for analyzing variance within populations. Overall, the analytical model demonstrated here is sensitive to scale, as quantifiable differences in SD and SE can be observed when parishes factor in on the population samples, versus when it does not. Figuratively, in the coastline metaphor mentioned earlier on (Mendelbrot, 1982), this compares to choosing one unit of measurement over the other: large units produce simple measurements but at the same time they potentially obscure the underlying nuances embedded within them. Following this proof-of-concept, which by a clear dispersion of correlation indicated an increase in complexity in the datasets used between the two methods, the next step is to home in on the paper's use case and apply this analytical framework to identify complexity attributes within the municipality, and potentially reveal otherwise hidden statistical patterns within the aggregated data.

4.2 Deconstruction of Guldborgsund Municipality

The municipal level represents the basic abstraction level of the analysis, defined as the level on which rural planning is administered in Denmark. There are 57,696 inhabitants distributed across 43 different parishes in GM, ranging from 88 inhabitants in Sønder Alslev parish to 16,163 inhabitants in Nykøbing Falster parish (Statistics Denmark, 2022b). A parish will be identified by its parish code and name, for instance '7584 Sønder Alslev' or '7578 Nykøbing Falster'.

The municipality is located withing Region Zealand. A region that represents some of the poorest rural municipalities of Denmark, which includes GM, despite having a relative proximity to the Capital Region and Copenhagen (Jepsen, et al., 2020). According to Jepsen et al. (2018), the population of GM and Lolland Municipality (collectively 'Lolland-Falster') has a lower life expectancy, is less educated than the rest of Denmark, has a lower income, has more people outside the workforce and, especially on Lolland, more people move out of the area than into it.

Similar to the cross-regional comparison, figure 2 illustrates the correlation between SD and population size, while figure 3 illustrates the correlation between SE and population size for GM. With an R^2 of 0.81, there is a strong correlation between increase in population and increase in SD, which is in line with the findings made on the cross-regional level of the analysis. The SD for the whole dataset ranges from 1 to 34 with an average of 9. The correlation of 0.29 for SE indicates a high sensitivity to numerosity on the municipal level, also similar to the cross-regional level.

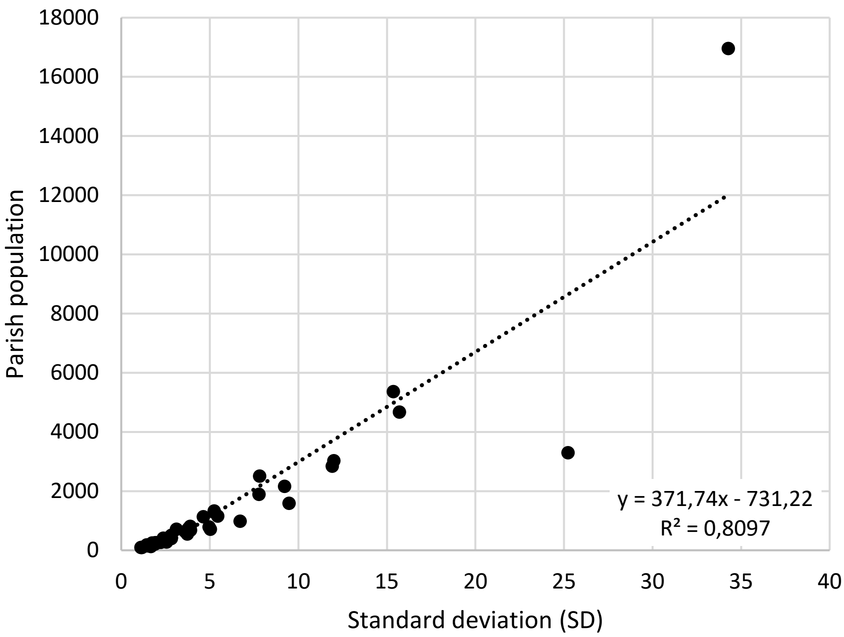

Figure 2: Parish population vs. SD in GM

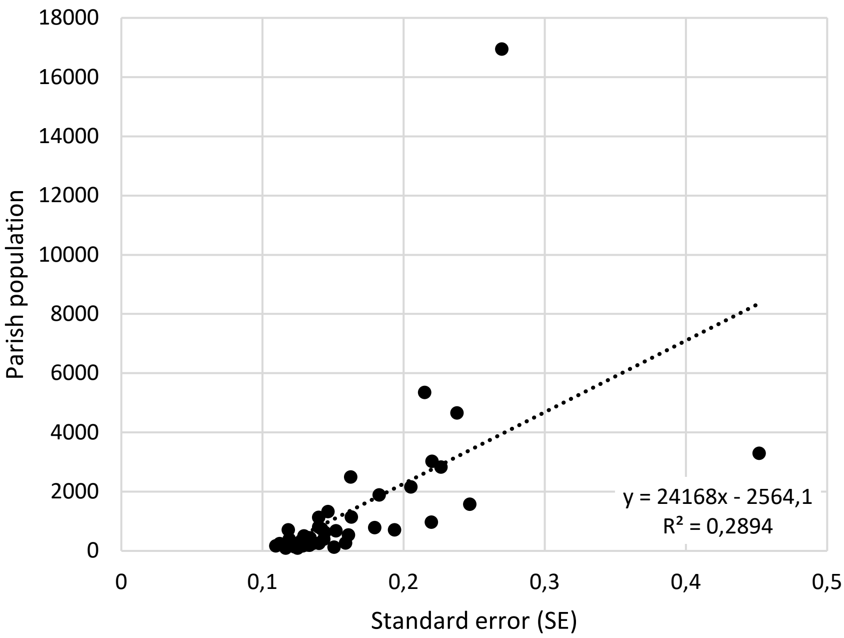

Figure 3: Parish population vs. SE in GM

Two considerable outliers can be identified in both charts as they do not fit the trendline. These outliers are 7580 Væggerløse and 7578 Nykøbing-Falster. The parish 7580 Væggerløse is characterized by a disproportionate amount of summerhouses compared to the rest of GM, while 7578 Nykøbing-Falster accounts for roughly 25% of the total population of GM. The next part of the analysis further deconstructs the parishes within GM for the purpose of isolating complexity factors other than numerosity within the parishes, and outline some basic demographic characteristics across GM. The results from this deconstruction are presented in table 4, with each parish ranked by parish code.

Code	Test	Area[a]	Population[b]	Density[c]	Avg. age[d]	Gender[e]	Education[f]	Employment[g]	Commuters[h]	Ethnicity[i]	Church[j]
	Jarque-Bera	0.49	0.21	0.00	0.22	0.58	0.10	0.25	0.42	0.23	0.26
	VIF	4	6,902	8	2	2	202	906	101	58,560	42,536
	Coefficient β	0.1665	-	-	-0.009	-19.2275	-	-	-	-	-
	P-value	0.0034	-	-	0.9667	0.0058	-	-	-	-	-
Code	**Parish**										
7575	Ønslev	15	681	47	45.3	1.1	20	352	286	642	562
7576	Eskilstrup	15	1,331	91	44.6	1.0	39	614	474	1,241	1,091
7577	Tingsted	32	3,026	95	40.6	1.0	138	1,605	1,368	2,875	2,536
7578	Nykøbing Falster	19	16,954	889	44.8	0.9	743	7,249	3,200	14,558	12,582
7579	Idestrup	33	1,889	58	48.5	1.0	54	866	685	1,781	1,562
7580	Væggerløse	45	3,297	73	56.2	1.0	117	1,257	890	3,123	2,771
7581	Skelby	19	263	14	49.3	1.0	11	107	86	244	202
7582	Gedesby	9	276	31	55.0	1.1	14	104	79	248	195
7583	Sønder Kirkeby	10	250	25	41.1	1.0	11	137	117	224	180
7584	Sønder Alslev	11	93	8	59.3	0.8	10	50	36	89	69
7585	Karleby	9	257	30	46.2	1.1	13	127	107	239	205
7586	Horreby	11	505	44	45.5	1.0	7	255	205	485	432
7587	Nørre Ørslev	12	447	37	40.9	1.2	27	246	209	425	370
7588	Systofte	12	713	59	45.6	1.1	19	366	319	658	593
7589	Horbelev	30	1,133	38	45.9	1.1	40	521	379	1,058	895
7590	Falkerslev	21	247	12	44.2	1.1	9	124	97	225	207
7591	Aastrup	37	789	21	51.1	1.1	40	367	261	733	627
7592	Tårs	23	453	20	43.2	1.1	15	253	200	442	385
7593	Vigsnæs	12	127	11	53.8	1.0	10	78	55	124	105
7594	Majbølle	23	975	42	51.0	1.1	28	446	336	914	784
7595	Radsted	36	744	21	45.2	1.2	14	381	318	693	625
7596	Toreby	71	5,359	76	45.7	1.0	242	2,450	1,809	4,983	4,351
7597	Slemminge	19	444	23	45.8	1.2	13	198	160	412	372
7598	Fjelde	7	178	24	46.6	1.2	2	75	55	160	149
7599	Døllefjelde	8	172	23	45.8	1.1	8	86	70	165	142
7600	Musse	9	198	23	48.1	1.4	4	92	79	188	170
7601	Herritslev	16	400	25	49.9	1.1	11	188	140	378	338
7605	Kettinge	30	1,148	38	46.1	1.2	27	581	450	1,065	934
7606	Bregninge	10	146	14	41.2	1.3	8	72	54	142	123
7607	Stubbekøbing	4	2,163	562	50.6	0.9	72	760	504	1,991	1,734
7608	Maglebrænde	11	288	26	47.6	1.2	9	149	121	265	224
7609	Torkilstrup	17	406	24	46.0	1.1	18	196	143	364	312
7610	Lillebrænde	8	264	32	46.1	1.3	9	128	109	230	199
7611	Gundslev	23	544	24	47.4	1.2	20	277	225	520	446
7613	Nørre Alslev	14	2,500	176	48.2	0.9	55	1,040	742	2,376	2,139
7614	Nørre Kirkeby	9	211	24	42.4	1.3	4	116	87	202	182
7642	Sakskøbing	23	4,667	207	49.7	1.0	115	1,771	1,197	4,343	3,878
7644	Våbensted	22	630	28	44.7	1.1	15	333	257	590	520
7645	Engestofte	11	94	9	47.9	1.5	2	39	37	86	78
9249	Gedser	8	711	90	53.7	1.2	14	250	169	652	586
9318	Nordvestfalster	81	2,836	35	46.8	1.1	123	1,393	1,299	2,646	2,326
9319	Nysted-Vantore*	29	1,582	54	53.9	0.9	59	575	0	1,479	1,275
9330	Krumsø**	39	807	20	45.2	1.3	0	0	0	764	680

[a] Km^2, based on vector data from Danmarks Administrative Geografiske Inddeling (DAGI), extracted on November 14th, 2022
[b] Sum of males and females, Statistics Denmark, KM5: Population 1. January by parish, sex, age and member of the National Church, released February 17th, 2022
[c] Population per km^2, calculated using data from SOGN1 and DAGI
[d] Average age regardless of age, Statistics Denmark, KMGALDER: Average age 1. January by parish and sex, released February 17th, 2022
[e] Male to female ratio as 1:X, Statistics Denmark, KM5: Population 1. January by parish, sex, age and member of the National Church, released February 17th, 2022
[f] People with an education level equal to or above a bachelor's degree, Statistics Denmark, KMST007A: Population 1. October (15 years+) by parish, member of the National Church and highest education completed, released March 21st, 2022
[g] Employment rate, Statistics Denmark, KMSTA005: Population 1. January (15 years+) by parish, socioeconomic status and member of the National Church, released February 17th, 2022
[h] Commuters to other parishes, Statistics Denmark, KMSTA009: Population 1. January (15 years+) by parish, member of the National Church and commuting, released December 8th, 2020
[i] Ethnic Danes, Statistics Denmark, KMSTA001: Population 1. January by parish, ancestry and member of the National Church, released February 17th, 2022
[j] National Church membership, Statistics Denmark, KM5: Population 1. January by parish, sex, age and member of the National Church, released February 17th, 2022

* Contains null values for commuting and is therefore ignored in this analysis.
** Contains null values for education, employment, and commuting, and is therefore ignored in this analysis.

Table 4: Multiple regression analysis variables

A total of ten different independent variables were tested by multiple regression analysis against SD as an independent variable, which was previously identified as a preliminary indication of complexity within an administrative unit – i.e., a parish. These values were selected solely on availability of official demographic data from Statistics Denmark and applicability to the parish level. A multiple regression analysis including all variables resulted in a high R^2 of 0.93, with population size as the only significant result with p-values below 0.05. Nine of the original ten variables were normally distributed, calculated by applying the Jarque-Bera test to the dataset for each separate variable – a test for normal distribution of a population sample, which in turn enables further statistical testing. A value above 0.05 indicates that a population sample is normally distributed. Seven of these variables displayed a variance inflation factor (VIF) above five and was then excluded from the analysis, as these would otherwise artificially inflate the variance and overall correlation of the model. The R^2 for the remaining three variables was 0.36. Two of these variables scored a p-value below 0.05 and were thus deemed statistically significant. These variables were the parishes' area in km^2 and gender ratio. Conclusively, gender ratio provides the strongest explanation for an increase in SD, followed by the physical size of the parish. Thus, uneven gender distributions (illustrated by highlighted parishes within GM in figure 4) combined with large parishes, seemingly cause an increase in SD.

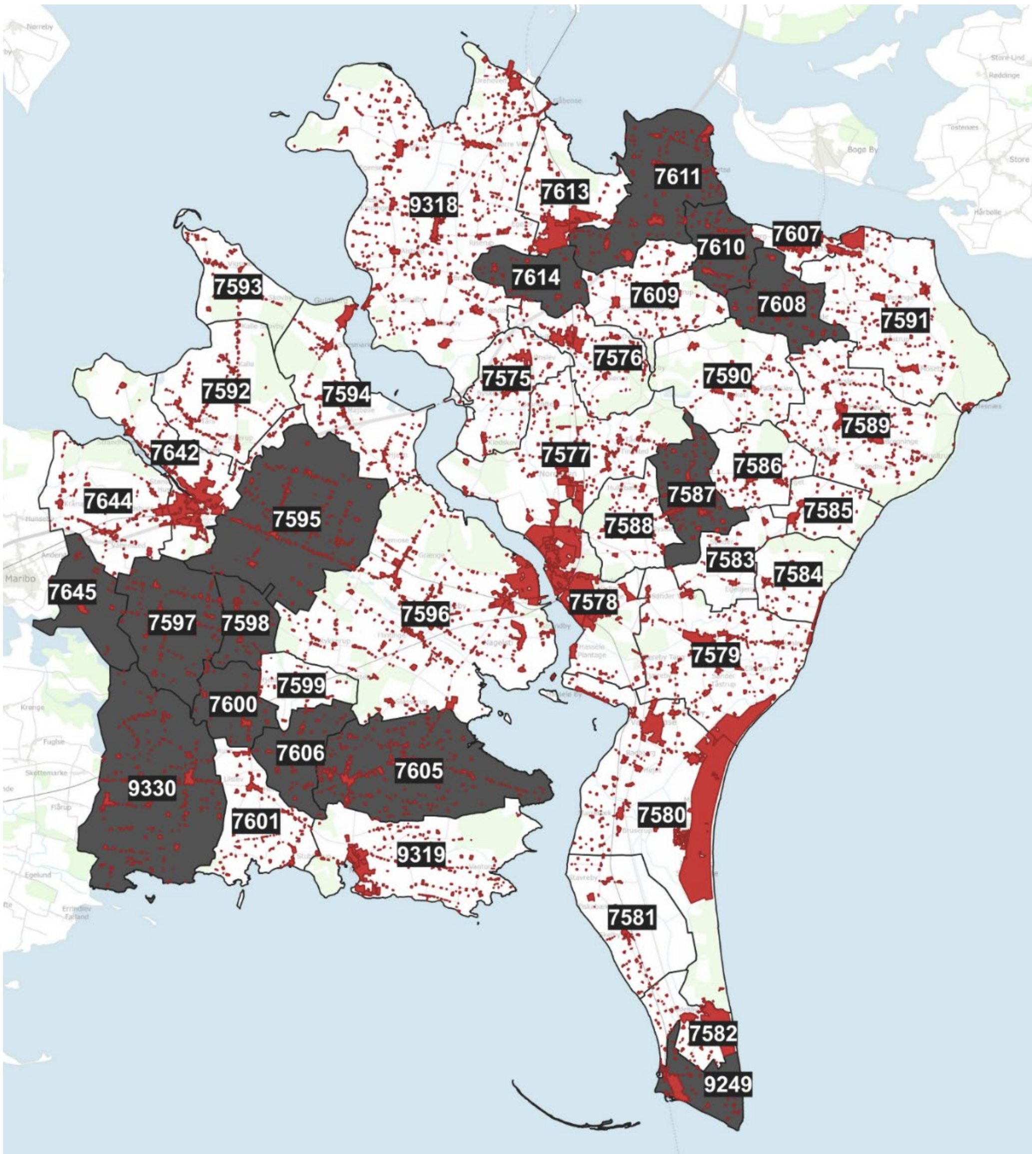

Figure 4: Gender ratio distribution of GM vs. built-up areas. Numbers = parish codes; light areas = low gender ratios; grey areas = high gender ratios; red areas = buildings. Source: author's own work based on data from the analysis in this paper and vector data from GeoDanmark.

There are two notable clusters in figure 4, namely on either side of the municipality (grey areas). Each cluster represents a gender ratio between 1.15:1 and 1.5:1. The cluster on the far left is the largest of the two and consist of eight parishes with an average SD of 3: 7595, 7597, 7598, 7600, 7605, 7606, 7645 and 9330. The top-right cluster consist of four parishes, also with an average SD of 3: 7608, 7610, 7611, and 7614. The two remaining parishes, 7587 and 9249, have an SD of 3 and 5, respectively, placing all four SDs well below the overall average SD of 9 for the dataset. This observation is arguably counterintuitive considering the prior observation that a disproportionate gender ratio contributes to an increase in SD, albeit explainable by the overall model strength of R^2 0.36. However, as figure 4 also depicts, the clusters generally cover sparsely populated areas with relatively scattered settlement patterns (red areas). This raises another question, namely as to why these parishes are disproportionately populated by males, compared to other parishes with similar settlement patterns within GM. A possible explanation to this phenomenon is clustering, i.e., an emergent pattern that has been formed by and within the municipality (Jeon and Jung, 2019). For instance, based on figure 4, one might suspect that men are attracted to the grey areas, whereas an alternative explanation could be that they are in fact figuratively pushed into or retained in them instead, by the properties of the surrounding areas.

Following this analysis, the upcoming discussion addresses three different themes: one, qualifying the theoretical framework for observing and analyzing municipalities as complex systems; two, the paper's use of statistical methods for comparative analysis, and model strength; and three, the results' potential implications on a DDDM planning model for GM and subsequent consequences.

5. Discussion

5.1 Municipalities as complex systems

The premise of informing planning theory through complexity theory is widely accepted (e.g., Portugali, 2012; Hartman and Roo, 2013; Samet, 2013; Rydin and Tate, 2016; Totry-Fakhouri and Alfasi, 2017; Boeing, 2018; Chettiparamb, 2019). According to Chettiparamb (2006), complexity theory has been applied as a theoretical metaphor to many different research areas, including the natural sciences (thermodynamics, physics, chemistry, biology, computer science, information technology, etc.) and social sciences (economics, political science, management science, etc.). Chettiparamb (2006) identifies two general streams within this body of work concerning complexity theory in planning: a quantitative stream concerned with modeling-related issues, and a non-quantitative stream concerned more with qualitative aspects. Both streams deal with similar concepts within complexity theory, but the theory has received criticism for the arguably vague manner it has been applied. For instance, in analyzing current literature and planning theory, Cameli (2021) typifies cities as cyborg systems, referring to a clear distinction between the human beings living in them and their social structures, and the buildings and infrastructure that constitutes its physical environment. Although the distinction between the built environment and the humans who occupies it is logically sound from an analytical standpoint, such metaphors arguably support the argument of somewhat vague applications of complexity theory in research areas outside of the natural sciences, including the social sciences. Mitleton-Key (2004) soberly emphasizes how complexity theory derives from different research disciplines that shares the common notion of creation of new order – i.e., emergent properties. Byrne (2003: p. 176) preemptively addresses this notion regarding complexity theory in planning and modeling: *"The crucial thing complexity offers in terms of modeling is the idea of alternatives – not a limitless range of alternatives but a set of more than one future with action determining which future is possible. There is a really substantial problem of method here. The traditional linear models of statistical reasoning are absolutely not isomorphic with processes of social transformation, precisely because they are incremental and cannot deal with changes of kind."* In line with Byrne's (2003) reasoning in terms of modeling, this methodological problem calls for a pragmatic take on how 'real world' phenomena are abstracted for decision-making modeling purposes.

There is an interesting catch-22 in the field of spatial data science, as demonstrated in the paper, namely in the context of political decision-making. As a political organization, Guldborgsund Municipality intend to base future planning strategies on the results of this type of research, which effectively creates a potential causal relationship between the research and the future physical layout of the municipality – and, arguably more importantly, its citizens. Consequently, the research will inevitably impact the subject matter that is being studied. This entails an ethical responsibly in terms of how the results are communicated to the decision-makers. They must understand the context in which the analysis has been conducted, and how the apparently objective data produced through e.g., statistical modeling and probability analysis is entirely dependent on how the reality they represent is perceived. Summarily, I argue that for spatial data science to make sense in a real-world setting, any research design based hereon should combine inductive and deductive reasoning. Inductive in terms of how we observe the world we live in (i.e., ontological logic), and deductive in terms of how we can detect the large-scale emergent patterns that each of us contributes to as individual systemic components (i.e., epistemological logic). Utilizing the hypothetico-deductive model or abductive reasoning (Upmeier zu Belzen et al., 2021), thus combining 'the best of both worlds' is arguably a sensible middle ground here.

First, by using metaphors to both articulate and empirically analyze complex dynamics in physical planning (Chettiparamb, 2006), the theoretical framework of this paper constitutes an important role as a communicator between the research and the municipalities' practitioners. Emphatically heavily reduced, complexity theory characterizes a complex system as dynamic and ever evolving, perpetually creating new order (Mitleton-Kelly, 2004). Check, municipalities change over time, demographically as well as physically. Thus, the theoretical framework serves as a metaphorical mediator to emphasize what exactly it is we ontologically, qualitatively mean by 'complex', when addressing the complications of designing a data-driven decision-making model for planning. Moreover, as complex behavior is a measurable phenomenon, the framework also serves as an epistemological, quantitative methodology.

A complicated system differs from a complex system in that complicatedness is a linear dynamic and thus predictable, unlike complexity, which is a nonlinear dynamic – a fundamental trait in deterministically chaotic and complex systems (Gleick, 1987). Following this reasoning, is it then reasonable to expect that a DDDM model can be based on a complex system? This distinction is difficult to make, seeing that a fractal system is arguably predictable in the sense that it is self-replicating across its hierarchical levels (Chettiparamb, 2005), thus providing a pattern. As self-contradictory this may seem, these patterns are the result of emergent behavior, which itself is unpredictable (Stacey, 1995). However, when it comes to modeling 'real world' phenomena, this distinction is arguably not necessary to make: *"Models do not necessarily have to be very detailed in order to provide an understanding of the strategic changes that may occur, and this is where one can distinguish between a 'complicated model', for example, with a great deal of detail and needing perhaps a very large computer, and a 'complex systems' model, that may generate changing, complex patterns of behavior and of organization, from relatively simple interacting equations"* (Allen, 1997: p. 2). Applying this distinction to a practical modeling setting inevitably fosters a high level of abstraction, thus effectively defining the subject matter as a highly aggregated entity. In other words, because Guldborgsund Municipality as the subject matter encompasses more than 60,000 people and 85,000 buildings spread across an area of over 90,000 hectares, a high level of abstraction is a prerequisite more so than a methodological choice.

5.2 Statistical analysis and model strength

As the research question of this paper is on model generalizability, the strength of the statistical model is discussed prior to discussing potential implications of a DDDM model for rural planning in GM. The input data for the statistical model stems from Statistics Denmark, an official institution that provides free and highly valid and almost complete registers of demographic data of the Danish population (Petersson et al., 2011; Baadsgaard and Quitzau, 2011; Jensen and Rasmussen, 2011), a service which is accessible in only a few other countries apart from Denmark (Jepsen et al., 2020). This service allowed for the model's input data to be extracted as aggregated (method A) and disaggregated (method B) data sets with a high level of accuracy, thus enabling the analysis of the paper to identify differences in correlation dispersion across different scales. The data was then applied to multivariate regression analysis, which produced a model strength of $R^2 = 0.36$. Intuitively, this indicates a weak statistical model.

However, on the discussion of the importance of R^2 in social sciences, Moksony (1999) argues that a low R^2 simply indicates that the dependent variable is affected by other factors than those included in the model, and that the R^2 in general is overvalued and misused as a measure for model reliability and validity. In the case of GM, this means that on a finer aggregation level that includes parishes, the variables 'area size' and 'gender ratio' are only part of the explanation as to why age dispersion increases parallelly to population size. This notion is supported by Ozili (2022), who argues that an R^2 lower than 0.50 is acceptable in social science research when most of the explanatory variables are statistically significant. There are three variables included in this model, of which two are statistically significant, meaning the model meets these criteria. A similar view is presented by Tian et al. (2014) in a study of rural settlement and policy implications, in which an R^2 between 0.20 and 0.50 is considered correlated, while a value above 0.60 is considered significantly correlated. This metric generally fits the current literature on planning in social science that utilizes regression analysis (e.g., Villarraga et al., 2014; Chen and Lui, 2016; Li et al., 2016; Jeon and Jung, 2019; Huang et al., 2021; Qi et al., 2022). With model strengths ranging between R^2 values of 0.51 and 0.75 with p-values below 0.01, Zhang et al. (2017: p. 12) attributes complexity properties by linking the sprawl of built-up land to systemic disorder: *"Fractal dimension values increased significantly during the ten years, which means that urban growth brought a more complex, scattered and disordered distribution of built-up land patches in Wen-Tai region. If this trend continues, complex and fragmented landscapes will increase rapidly with urbanization, which might lead to the inefficient usage of built-up land resources. Accordingly, the authors suggest that local government implement reasonable built-up land plans by balancing economic growth with the construction of settlements and industrial land in order to guide the city toward sustainable development."* This argument is supported by Xu et al., (2020: p. 163), whose model produces an R^2

value of 0.37 based on ordinary least squares: *"The authors suggest that in the process of resettlement, the distance between settlements should be kept less than [5,000 meters] if possible. The government needs to further improve and implement preferential land use policies, with appropriate increases in the area of built-up land."* The results presented in this paper are in line with the current literature on the general topic of planning and regression analysis, namely wherein disorder can be interpreted as a notion of an increase in complexity. Referring to the basic property of numerosity and interrelatedness in complex systems (Mitleton-Kelly, 2004), this notion entails a discussion of aggregated vs. disaggregated datasets and potential implications on a DDDM model for planning in GM.

5.3 Implications on a data-driven decision-making model

Due to the macro-organizational nature of their work, politicians on all hierarchal levels deals with abstract issues. Prime ministers deal with national issues such as international relations, while mayors and city councilors deal with local issues for whichever city they represent. Issues can arise, however, regarding the human factor, when the people in charge makes decisions out of ignorance that are misaligned with practical concerns and feasibility (Laursen et al., 2021). For this reason, the appropriate scale of analysis depends on what is being analyzed (Qi et al., 2019). Thus, if one cannot simply go out and modify a real-world system like a municipality to reduce its complexity for the sake of improving the generalizability of a DDDM model, only the complicatedness of the DDDM model representing a complex system can be reduced – e.g., by data aggregation.

Employing a model-based planning practice arguably fosters a reductionist approach to qualitative properties. However, there are arguably some potential issues regarding generalizability and reductionism when applying quantitative decision-making models to qualitative research fields, and vice-versa. According to McGreevy (2018), modern planners evoke a top-down planning approach that is belligerently set on reducing complexity, though also fascinated by the bottom-up self-organizing capabilities of complex adaptive systems such as cities. Moreover than addressing local issues, planning is also a global concern, as evident by the United Nations' (UN) 11th Sustainable Development Goal (SDG), specifically target 11.a: *"[…] support positive economic, social and environmental links between urban, peri-urban and rural areas by strengthening national and regional development planning"* (United Nations, 2015). The UN provides a quantitative indicator for this target (11.a.1), which is based on whether a country has a national urban policy or regional development plan or not. This arguably provides an issue regarding oversimplification, as the target is paired with this one simple indicator alone (Berisha, Caprioli and Cotella, 2022). Raising concerns with reducing the planning strategies and approaches of an entire country to one single indicator is perfectly reasonable, although the point in doing so is arguably a pragmatic necessity. However, one must consider how abstract a scale-dependent decision-making model can be without compromising its generalizability (Costanza et al., 1993). For instance, the European Union (EU) operates with four administrative levels (NUTS), which are defined by population thresholds. Starting at the national level, the NUTS-system abstracts from populations of 7M to 150k inhabitants. In the case of Denmark, with close to 6M citizens, this effectively puts the entire population within one abstraction level of populations ranging from 800k to 3M, divided into five regions, thus not considering e.g., municipalities, parishes, cities, or towns.

The effects of aggregating spatial units are well-known and well-established: *"Aggregating smaller areal units into regions filters out the harmonics whose wavelengths are smaller than the size of the regions, [and, if two or more] space series have harmonics which are filtered out by a given aggregation, the correlation and regression coefficients of the series before the aggregation will differ from the coefficient obtained after the aggregation"* (Casetti, 1966; Tobler, 1989). A notion that is supported by Jelinski and Wu (1996), who argues that because aggregating spatial data has a filtering effect, the impact of MAUP is far more complex and unpredictable on multivariate analysis, than it is on traditional statistical analysis, e.g., correlation analysis and linear regression. In a more recent study, Jacobs-Crisioni (2014: p. 53) observes this phenomenon where the model strength decreases as the data aggregation level increases: *"[...] aggregating by averaging is a sound strategy for limiting scale dependencies; on the other hand […] data in very coarse resolutions are not fit for assessing the impacts of factors that are important on the micro level"*. Hennerdal and Nielsen (2017) supports this position and argues that if one does not vary the geographical extent of the area of reference to test how the results differ with different delineations of the area of reference, then the results of such an analysis can be rendered arbitrary. Utilizing a disaggregated dataset (NUTS-4) on rural development domains/variables in Poland and Slovakia, a rural development index (RDI) is developed by Michalek and Zarnekow (2012) for the purpose of measuring the overall level of rural development and quality of life in individual rural regions. Comparable to the methodology in this paper, the authors emphasize that in addition to the need for simplicity to encourage practical application, the model should be based on cheap and freely available data. Lee and Rogers (2019) demonstrate a similar methodology in their analysis of geographic distributions in a political context and stresses that *"[…] the best way to manage concerns with the measure of geographic distribution lies in the articulation of a clear theory linking the unit of geography to a particular characteristic of interest."* I.e., political practice in planning is unavailing and potentially aimless if not informed by scientific insight and theorization.

Now, as we are approaching the conclusion of the discussion, and of this paper, the underpinning MAUP-issue of this paper's methodology and analysis is duly addressed. Naturally, there is a prevailing focus on areal and spatial units in studies that deals with the MAUP. Hennerdal and Nielsen (2017: p. 572) provides part of the explanation for this notion: *"It is vital to acknowledge the bias due to the part of the MAUP related to the area of reference because the outcomes of clustering and segregation studies influence policy. Therefore, methodologies that cannot fully cope with the MAUP, including the part of the problem related to the area of reference, should be avoided."* This argument entails a direct implication for any DDDM planning model, namely its requirement to be able to manage the MAUP. Because not just the areal units themselves are susceptible to the MAUP, but also the variability of the descriptive statistics, such as SD and correlation coefficients, are scale-dependent (Gomes and Cunto, 2020). So why even bother designing a DDDM planning model, if the MAUP will inevitably distort its predictions? Ironically, part of the answer may lie within the nature of complexity itself. Because although sensitive to control parameters and attractors, complex systems are robust, for two reasons (Byrne, 2003): one is that they remain complex systems, regardless of changes; two is that they can change radically without losing their systemic integrity. Moreover, the identification of MAUP as a problem can be misleading, as it simply reflects the nature of the real hierarchically structured systems (Jelinski and Wu, 1996). In terms of modeling, complexity theory offers the idea of alternative futures determined by social actions: *"Despite these qualities, bottom-up and unplanned, self-organized spaces are not easily adapted to the rapid change in the needs of a modern, urbanized society. In the face of modernization, unexpected conflicts arise that require planning intervention. Although self-organizing systems, urban areas and traditional societies included, are most flexible and adjusting, the current lifestyle demands external intervention consistent with local values and existing circumstances"* (Totry-Fakhouri and Alfasi, 2017: p. 37). Accordingly, I argue, that recognizing systemic complexity in planning, including the MAUP and its effects on data aggregation and generalization etc., is a prerequisite – not necessarily a limitation – for designing any model intended to represent the real world, as defined by those who intend to plan it.

6. Conclusions

With a practice-oriented purpose in mind, I aimed to obtain an understanding of the intrinsically complex nature of planning through complexity theory with this paper. This practical purpose consisted of laying out the foundation for designing a DDDM planning model for GM that is both quantitively accurate and qualitatively realistic, all the while striving for as much generalizability as possible. Perhaps most noticeably, the analytical results demonstrated three things: one, scale-dependency by sensitivity to numerosity, as there are considerable differences in correlation coefficients between aggregated and disaggregated data; two, fractal properties, as dispersion increases with fragmentation both cross-regionally and municipally; three, emergence by clustering in gender ratios across the municipality. Having demonstrated these systemic properties, I conclude that GM can justifiably be identified and analyzed as a complex system. These insights make for a rather straightforward answer to the paper's research question on how abstract a political decision-making model can be without compromising its generalizability: the model should inform the user, not the other way around. For instance, instead of relying on administrative units for spatial and demographic analysis, a finer grid would presumably produce more accurate results and reveal sub-level behavior otherwise hidden on even the parish level. Thus, the administrative units ought to be removed from the equation of the DDDM planning model to reduce the effects of the MAUP – albeit the MAUP merely reflects the natural behavior of the real world and therefore only presents an issue to humans who insist on quantifying reality. In addition to this implication, a DDDM planning model should be scenario-based, more so than simply a passive data collection platform. Case in point, changes in one parish may affect the dynamics of another, which an optimal DDDM planning model should be able to predict.

By applying a theoretical framework rooted in complex social systems, the analysis in this paper provides a novel contribution to the integration of complexity science into the field of spatial data science, which is arguably still in its formative stages. Moreover, this study offers valuable insights into the evaluation of the applicability and transferability of administrative units. The adoption of such a framework not only enhances our understanding of the intricate dynamics of social systems, but also enriches the methodology and approach employed in spatial data science. These insights have the potential to refine our comprehension of the complexities inherent in administrative units and enable more accurate assessments of their generalizability. Overall, this research sheds light on the practical implications of complexity science within the evolving field of spatial data science.

Further studies could benefit from employing a grid-based methodology on a low abstraction level smaller than parishes and exploring alternative metrics for complexity other than SD or SE, as well as conducting multivariate regression analysis or similarly on a finer level. This could identify more explanatory variables with stronger correlations and reveal a more comprehensive depiction of any given rural municipality, or other administrative unit, which in turn could provide for a more informed planning strategy.

7. References

Allen, P. M., 1997. *Cities and Regions as Self-organizing Systems: Models of Complexity.* Bedford, United Kingdom: International Ecotechnology Research Centre, Cranfield University.

Bandana, K. & Hodgson, M. E., 2012. A Process Oriented Areal Interpolation Technique: A Coastal County Example. *Cartography and Geographic Information Science,* 39(1), pp. 3-16.

Berisha, E., Caprioli, C. & Cotella, G., 2022. Unpacking SDG target 11.a: What is it about and how to measure its progress?. *City and Environment Interactions,* Volume 14, p. 100080.

Boeing, G., 2018. Measuring the Complexity of Urban Form and Design. *Urban Design International,* Volume 23, pp. 281-292.

Byrne, D., 2003. Complexity theory and planning theory: A necessary encounter. *Planning Theory,* 2(3), pp. 171-178.

Baadsgaard, M. & Quitzau, J., 2011. Danish registers on personal income and transfer payments. *Scandinavian Journal of Public Health,* 39(7), pp. 103-105.

Cameli, S. A., 2021. Natural or artificial? A reflection on a complex ontology. *Planning Theory,* 20(3), pp. 191-210.

Casetti, E., 1966. Analysis of Spatial Association by Trigonometric Polynomials. *Canadian Geographer,* Volume 10, pp. 199-204.

Chen, S. & Liu, Z., 2016. What determines the settlement intention of rural migrants in China? Economic incentives versus sociocultural conditions. *Habitat International,* Volume 58, pp. 42-50.

Chettiparamb, A., 2005. Fractal spaces for planning and governance. *The Town Planning Review,* 76(3), pp. 317-340.

Chettiparamb, A., 2006. Metaphors in complexity theory and planning. *Planning Theory,* 5(1), pp. 71-91.

Chettiparamb, A., 2014. Complexity theory and planning: Examining 'fractals' for organising policy domains in planning practice. *Planning Theory,* 13(1), pp. 5-25.

Chettiparamb, A., 2019. Responding to a complex world: Explorations in spatial planning. *Planning Theory,* 18(4), pp. 429-447.

Costanza, R., Wainger, L., Folke, C. & Maler, K.-G., 1993. Modeling Complex Ecological Economic Systems. *BioScience,* 43(8), pp. 545-555.

Dapena, A. D., Vázquez, E. F. & Morollón, F. R., 2016. The role of spatial scale in regional convergence: The effect of MAUP in the estimation of β-convergence equations. *The Annals of Regional Science,* Volume 56, pp. 473-489.

Dhamo, S., 2021. *Understanding Emergent Urbanism: The Case of Tirana, Albania.* Cham, Switzerland: Springer.

Ekman, U., 2018. Smart City Planning: Complexity. *International Journal of E-Planning Research,* 7(3), pp. 1-21.

Eräranta, S., 2020. Social complexities in collaborative planning processes. In: *Handbook on planning and complexity.* Cheltenham, United Kingdom: Edward Elgar Publishing, pp. 171-185.

Fotheringham, A. S. & Wong, D. W. S., 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A,* Volume 23, pp. 1025-1044.

Frederiksen, M., Gundelach, P. & Nielsen, R. S., 2014. *Mixed methods-forskning.* Copenhagen: Hans Reitzels Forlag.

Fredslund, L., Søder, P. H., Hansen, E. J. d. P. & Koch, C., 2022. An Institutional Analysis of How the Construction Industry's Complexity Affects Waste: An Alternative Response to Future Market Demands for Sustainability. *Proceedings of the 38th ARCOM Conference,* pp. 612-621.

Garreton, M. & Sánchez, R., 2016. Identifying an optimal analysis level in multiscalar regionalization: A study case of social distress in Greater Santiago. *Computers, Environment and Urban Systems,* Volume 56, pp. 14-24.

Gleick, J., 1987. *Chaos: Making a new science.* New York, United States: Viking.

Gomes, M. J. T. L. & Cunto, F. J. C., 2020. Influence of Different Spatial Aggregations on Variables Implemented in Macroscopic Road-Safety Modeling. *Journal of Transportation Engineering, Part A: Systems,* 146(4), p. 04020014.

Guangjin, T., Qiao, Z. & Xiaolu, G., 2014. Rural settlement land dynamic modes and policy implications in Beijing metropolitan region, China. *Habitat International,* Volume 44, pp. 237-245.

Guldborgsund Kommune, 2018. *Det rige hverdagsliv i en kommune som rækker hånden ud: Planstrategi 2018.* Nykøbing-Falster: Guldborgsund Kommune.

Hartman, S. & Roo, G. D., 2013. Towards managing nonlinear regional development trajectories. *Environment and Planning C: Government and Policy,* Volume 31, pp. 556-570.

Hennerdal, P. & Nielsen, M. M., 2017. A Multiscalar Approach for Identifying Clusters and Segregation Patterns That Avoids the Modifiable Areal Unit Problem. *Annals of the American Association of Geographers,* 107(3), pp. 555-574.

Huang, X., Weesep, J. V. & Tang, S., 2021. To move or not to move? Residential mobility of rural migrants in a medium-sized Chinese city: the case of Yangzhou. *Housing Studies,* 36(2), pp. 278-301.

Jacobs-Crisioni, C., Rietveld, P. & Koomen, E., 2014. The impact of spatial aggregation on urban development analyses. *Applied Geography,* Volume 47, pp. 46-56.

Jelinski, D. E. & Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology,* 11(3), pp. 129-140.

Jensen, V. M. & Rasmussen, A. W., 2011. Danish education registers. *Scandinavian Journal of Public Health,* 39(7), pp. 91-94.

Jeon, G. & Jung, S., 2019. Comparative analysis of residential patterns of foreigners from rich and poor countries in Seoul. *Sustainable Cities and Society,* Volume 47, p. 101470.

Jepsen, R. et al., 2018. Lolland-Falster Health Study: Study protocol for a household-based prospective cohort study. *Scandinavian Journal of Public Health,* pp. 1-9.

Jepsen, R. et al., 2020. Socio-economic determinants of participation in the Lolland-Falster health study. *Journal of Public Health: From Theory to Practice,* Volume 28, pp. 657-664.

Kaufmann, S. & Macready, W., 1995. Technological Evolution and Adaptive Organizations: Ideas from biology may find applications in economics. *Complexity,* pp. 26-43.

Kermanshachi, S., Dao, B., Shane, J. & Anderson, S., 2016. Project Complexity Indicators and Management Strategies: A Delphi Study. *Procedia Engineering,* Volume 145, pp. 587-594.

Kwon, H. R. & Silva, E. A., 2020. Mapping the Landscape of Behavioral Theories: Systematic Literature Review. *Journal of Planning Literature,* 35(2), pp. 161-179.

Laursen, K. G. et al., 2021. *Uvildig evaluering af Roskilde Badet: Forsinkelser og fordyrelser i byggeprojekter. BUILD 2021:17,* Copenhagen: Institute of the Built Environment, Aalborg University.

Lee, D. W. & Rogers, M., 2019. Measuring Geographic Distribution for Political Research. *Political Analysis,* Volume 27, pp. 263-280.

Li, D. et al., 2016. The Effects of Urban Sprawl on the Spatial Evolution of Rural Settlements: A Case Study in Changchun, China. *Sustainability,* 8(8), p. 736.

Mandelbrot, B. B., 1982. *The fractal geometry of nature.* San Francisco, United States: W. H. Freeman.

Maturana, H. R. & Varela, F. J., 1992. *The Tree of Knowledge: The Biological Roots of Human Understanding.* Boston, United States: Shambhala.

McGreevy, M. P., 2018. Complexity as the telos of postmodern planning and design: Designing better cities from the bottom-up. *Planning Theory,* 17(3), pp. 355-374.

Merriam-Webster, 2022. *Rural.* Springfield, United States: Merriam-Webster.

Michalek, J. & Zarnekow, N., 2012. Application of the Rural Development Index to Analysis of Rural Regions in Poland and Slovakia. *Social Indicators Research,* Volume 105, pp. 1-37.

Mitleton-Kelly, E., 2004. The Information Systems Professional as a Hermit: Of plural rationalities, information rejection and complexity. *Innovation,* 17(4), pp. 289-323.

Moksony, F., 1999. Small Is Beautiful: The Use and Interpretation of R2 in Social Research. *Szociologiai Szemle,* pp. 130-138.

Ozili, P. K., 2022. *The acceptable R-square in empirical modelling for social science research.* s.l.:s.n.

Papadimitriou, F., 2020. *Spatial Complexity: Theory, Mathematical Methods and Applications.* Cham, Switzerland: Springer.

Petersson, F., Baadsgaard, M. & Thygesen, L. C., 2011. Danish registers on personal labour market affiliation. *Scandinavian Journal of Public Health,* 39(7), pp. 95-98.

Portugali, J., 2012. Complexity Theories of Cities: First, Second or Third Culture of Planning?. In: *Complexity and Planning: Systems, Assemblages and Simulations.* Routledge: London, United Kingdom, pp. 117-140.

Portugali, J., 2021a. Introduction to the Handbook on Cities and Complexity. In: *Handbook on Cities and Complexity.* Cheltenham, United Kingdom: Edward Elgar Publishing, pp. 1-11.

Portugali, J., 2021b. Cities, complexity and beyond. In: *Handbook on Cities and Complexity.* Cheltenham, United Kingdom: Edward Elgar Publishing, pp. 13-27.

Qi, J., Liu, H., Liu, X. & Zhang, Y., 2019. Spatiotemporal evolution analysis of time-series land use change using selforganizing map to examine the zoning and scale effects. *Computers, Environment and Urban Systems,* Volume 76, pp. 11-23.

Qi, W., Deng, Y. & Fu, B., 2022. Rural attraction: The spatial pattern and driving factors of China’s rural in-migration. *Journal of Rural Studies,* Volume 93, pp. 461-470.

Ratter, B. M. W., 2006. Complexity theory and geography: a contribution to the discussion on an alternative perspective on systems. *Mitteilungen der Österreichischen Geographischen Gesellschaft,* Volume 148, pp. 109-124.

Rydin, Y. & Tate, L. E., 2016. *Actor networks of planning: exploring the influence of actor network theory.* London, United Kingdom: Routledge.

Salvati, L., 2022. A Long Way to Complexity: Nonlinear “Growth Stages” and Spatially Uncoordinated Settlement Expansion in a Compact City (Athens, Greece). *Geographical Analysis.*

Samet, R. H., 2013. Complexity, the science of cities and long-range futures. *Futures,* Volume 47, pp. 49-58.

Shiner, J. S., Davison, M. & Landsberg, P. T., 1999. Simple measure for complexity. *Physical Review E,* 59(2), pp. 1459-1464.

Stacey, R. D., 1995. The Science of Complexity: An Alternative Perspective for Strategic Change Processes. *Strategic Management Journal,* Volume 16, pp. 477-495.

Statistics Denmark, 2018. *Municipality Groups v1:2018-.* Copenhagen: Statistics Denmark.

Statistics Denmark, 2021. *HISB7: Life expectancy for new born babies by sex.* Copenhagen: Statistics Denmark.

Statistics Denmark, 2022a. *BY2: Population 1. January by municipality, size of the city, age and sex.* Copenhagen: Statistics Denmark.

Statistics Denmark, 2022b. *FOLK1AM: Population at the first day of the month by region, sex and age.* Copenhagen: Statistics Denmark.

Statistics Denmark, 2022c. *BYGB12: Buildings by region, ownership, use and areal intervals.* Copenhagen: Statistics Denmark.

Stillwell, J., Daras, K. & Bell, M., 2018. Spatial Aggregation Methods for Investigating the MAUP Effects in Migration Analysis. *Applied Spatial Analysis and Policy,* Volume 11, pp. 693-711.

Timmermans, W., López, F. Ó., Tubio, J. M. & Maseda, R. C., 2015. A Complexity Perspective on Institutional Change: Dealing with Land Fragmentation in Galicia. *Social Evolution & History,* 14(2), pp. 77-107.

Tobler, W. R., 1989. Frame independent spatial analysis. In: *The Accuracy Of Spatial Databases.* London, United Kingdom: Taylor and Francis, pp. 115-122.

Totry-Fakhoury, M. & Alfasi, N., 2017. From abstract principles to specific urban order: Applying complexity theory for analyzing Arab-Palestinian towns in Israel. *Cities,* Volume 62, pp. 28-40.

Upmeier zu Belzen , A., Engelschalt, P. & Krüger, D., 2021. Modeling as Scientific Reasoning - The Role of Abductive Reasoning for Modeling Competence. *Education Sciences, 11(9),* p. 495.

Villarraga, H. G., Sabater, A. & Módenes, J. A., 2014. Modelling the Spatial Nature of Household Residential Mobility within Municipalities in Colombia. *Applied Spatial Analysis,* Volume 7, pp. 203-223.

Walton, M., 2014. Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning,* Volume 45, pp. 119-126.

Wong, D., 2009. The Modifiable Areal Unit Problem. In: *The SAGE Handbook of Spatial Analysis.* London: SAGE Publications, pp. 105-124.

Xiao, J., 2021. Spatial Aggregation Entropy: A Heterogeneity and Uncertainty Metric of Spatial Aggregation. *Annals of the American Association of Geographers,* 111(4), pp. 1236-1252.

Xu, J. et al., 2020. Do settlements isolation and land use changes affect poverty? Evidence from a mountainous province of China. *Journal of Rural Studies,* Volume 76, pp. 163-172.

Zhang, L. & Qian, Q., 2017. How mediated power affects opportunism in owner-contractor relationships: The role of risk perceptions. *International Journal of Project Management, 35(3),* 35(3), pp. 516-529.

Zhang, Z. et al., 2017. Characterizing Factors Associated with Built-Up Land Expansion in Urban and Non-Urban Areas from a Morphological Perspective. *Sustainability,* 9(8), p. 1411.