JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 26 (2023), pp. 1–25 doi:10.5311/JOSIS.2023.26.212

RESEARCH ARTICLE

A geographical perspective on Simpson's paradox

Mehak Sachdeva and A. Stewart Fotheringham

Spatial Analysis Research Center, Arizona State University, United States

Received: February 27, 2022; returned: August 19, 2022; revised: September 16, 2022; accepted: December 16, 2022.

Abstract: The concept of scale is inherent to, and consequential for, the modeling of geographical processes. However, scale also causes huge problems because the results of many types of spatial analysis appear to be dependent on the scale of the units for which data are reported (measurement scale). With the advent of local models and the fundamental difference in their scale of application compared to global models, this issue is exacerbated in unexpected ways. For example, a global model and local model calibrated using data measured at the same aggregation scale can also result in different and sometimes contradictory inferences (the classic Simpson's Paradox). Here we provide a geographical perspective on why and how contrasting inferences might result from the calibration of a local and global model using the same data. Further, we examine the viability of such an occurrence using a synthetic experiment and two empirical examples. Finally, we discuss how such a perspective might inform the analyst's conundrum: when the respective inferences run counter to one another, do we believe the local or global model results?

Keywords: local models, Simpson's paradox, global models, process scale, multiscale geographically weighted regression (MGWR)

---

1 Introduction

A fundamental aspect of geographical research is to understand processes operating between people, objects and events through examination of their observed spatial patterns [37]. Of prominent interest in investigating spatial processes is the scale at which they occur. Though many statistical techniques have been created to explore various properties of spatial patterns, inferring underlying spatial processes from these remains a persistent problem. This gap has often been attributed to the "scale problem" [37]. The issue is intrinsic to all geographical research and has been an active area of inquiry for many decades [37, 46, 59, 70, 71]. Since different processes contribute to spatial patterns at unique scales of observation (often termed observation or measurement scale), it is difficult to make accurate inferences about them by studying patterns at any one scale. For example, where some processes affecting house prices, such as economic performance, may be operational at a national or even global scale, others such as proximity to public transportation or population density may operate at the scale of local neighborhoods.

A long-standing problem related to scale is that the inferences we make based on an analysis of spatial data might vary according to the definition of the spatial units for which the data are measured. This is often referred to as the Modifiable Areal Unit Problem (MAUP) which has long beset spatial analysis [25, 53, 54]. Here we describe and discuss another aspect of scale, Simpson's paradox, which is perhaps less well-known in spatial analysis and occurs when the results obtained from the same data aggregated to different scales (or to different groups in the context of aspatial data) yields contradictory conclusions [10, 11, 20, 68, 77]. A popular aspatial example of Simpson's paradox was observed in a study on possible gender bias in the admission process at University of California, Berkeley in 1973 [10]. In this study, the pattern of graduate admissions across the university as a whole indicated a bias against female applicants, whereas data on individual departments suggested female applicants were generally favored over male applicants [10]. Since female applicants tended to apply to more competitive departments with lower rates of admission and the opposite was true for male applicants, the trends at the aggregate level were seen as contradicting those at the disaggregated department level [10]. Another well-known example of the paradox from the aspatial statistics literature concerns the batting averages of players in professional baseball [62]. Ross [62] demonstrated how between two high ranking players, one player had a better annual batting average for 1995 and 1996 while the other had a higher batting average when the statistics for the two years were combined. Although much has been written about this phenomenon in aspatial analysis, it is rarely recognized in the analysis of spatial data. Here, we demonstrate through the lens of local modeling: (i) that this is a fundamental problem about which geographers should be aware; (ii) how and why this paradox might commonly plague inferences from the analysis of spatial data using local and global models (where 'global' refers to the use of all data points in a predefined study area in the model calibration); and (iii) that by refocusing the problem in terms of processes, we contribute significantly to a greater understanding of the scale problem in spatial modeling.

To date, our understanding of Simpson's paradox in spatial analysis is limited. It has rarely been identified and, consequently, there is little understanding of its cause or potential implications, especially for policy oriented research. While in a recent paper, Fotheringham and Sachdeva [24] briefly showcase an empirical example of its occurrence in spatial analysis, it has not yet been explored in detail. However, as we show below, the increasing popularity of local models means that the spatial variant of the paradox is likely to be encountered with increasing regularity and therefore needs to be better understood. Local models are calibrated with subsets of the data and yield results which are specific to each location thereby allowing the modeling of processes which vary across space. When the calibrations of local and global models applied to the same data are compared, Simpson's paradox can arise in the following way. Suppose a global model is calibrated and a parameter estimate associated with covariate x is found to be significantly negative. In the calibration of a local model on the same data, the expectation is that local estimates of the parameter associated with covariate x will also be significantly negative. However, the local parameter estimates can be insignificant or, in rarer cases, they can be significantly positive. This is essentially a process scale issue—at one spatial scale we infer that y and x are negatively related; at a different spatial scale we infer they are positively related. As with aspatial examples of Simpson's paradox, both of these statements can be true, which can be confusing to the analyst. What do we do if the results of calibrating the same model locally and globally produce contrasting inferences—the classic Simpson's paradox? Should we believe the results of the local model or the results of the global model?

In what follows, we introduce Simpson's paradox, provide a geographical perspective on Simpson's paradox, explain how it might be identified and understood, give an example of how local modeling can be employed to investigate the paradox using simulated data and then provide two empirical instances of the occurrence of Simpson's paradox in real-world spatial data. Finally, we discuss the implications of the paradox in the spatial sciences and future directions for research.

2 Simpson's paradox

In a recent comparative study analyzing Covid-19 Case Fatality Rates (CFR) in Italy and China, Kugelen et al. [73] found that for each age group using analysis from age-stratified data, CFR in Italy were lower than in China. However, on analyzing aggregated CFR for both countries (i.e., not disaggregated by age), fatality rates in Italy were found to be higher than in China. How could fatality rates for all individual age groups in Italy be lower while the aggregated rate was higher?

This conundrum is a classic example of a well-known statistical phenomenon referred to as Simpson's paradox or Simpson's reversal and it arises when the analysis of aggregated data reveals an opposite trend from that observed when the data are disaggregated according to some criterion (for example, age in the above example). The phenomenon was first technically pointed out by Simpson [68] using a hypothetical example where a treatment was found to be ineffective for the total population but effective for both men and women when analyzed using gender-stratified data. Similar effects, however, were reported fifty years earlier [58] although these studies described the disappearance or diminishing of a trend through aggregation of data rather than a complete reversal. Cohen and Nagel [20] appear to be the first to observe a complete reversal of trend during data aggregation and coined the phenomenon a 'paradox'. Subsequently, instances of the paradox or reversal have been reported frequently in the statistical literature, especially in the fields of epidemiology and the social sciences. While the paradox may be seen as a statistical curiosity from one perspective, its occurrence in policy oriented fields causes consternation and uncertainty because it would seem that completely contradictory courses of action can be validated using the same data.

Simpson's paradox has in fact been observed in a variety of forms. First coined by Samuels [66], the most intriguing occurrence and the one that is most frequently investigated is known as Association Reversal, [7, 11, 12, 16, 22, 47]. This occurs when the trends in all subpopulations (the original population disaggregated based on some criterion such as gender or race) are reversed when the entire population is analyzed. A weaker case of the Association Reversal, where the trend observed in the overall population disappears in the analysis of disaggregated data, is called Yule's Association Paradox [78]. Such an occurrence is typically observed when associations between variables have a common but omitted cause. For example, the probability of having skin cancer is typically observed to be positively correlated with exercising in the general population but on disaggregating the population into subgroups, based on average exposure to sunlight, the association disappears [76]. Another scenario where the individual subpopulation trends add up to a smaller or larger trend than the one observed for the total population is called the Amalgamation Paradox [31] and is perhaps the most common version of Simpson's paradox. These types are visualized in Figure 1.

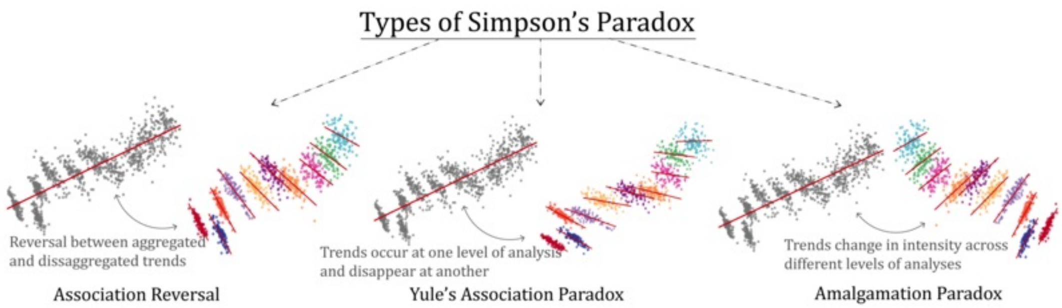


Figure 1: Types of Simpson's Paradox

The Covid-19 CFR example stated previously exhibits the Association Reversal type of the paradox and the reversal in trends between the age-stratified groups and the overall populations for the two countries can be explained in part by focusing on each of their demographic profiles. On observing the confirmed cases for the two countries, Kugelen et al. [73] found that while most cases in China fell within the 30-59 age group, the majority of cases reported in Italy were for people aged 60 and above. Since older people are generally at a higher risk of succumbing to the disease than the younger population, the 60+ subgroup rates drive the overall population rates to be higher for Italy over China while lower CFR are recorded for each individual age group. Additionally, since the population in Italy (median age 45.4) is generally older than in China (median age 38.4) and there may well be social behavioral differences between the two age groups, these factors could also cause the observed reversal between overall and age-stratified results [73].

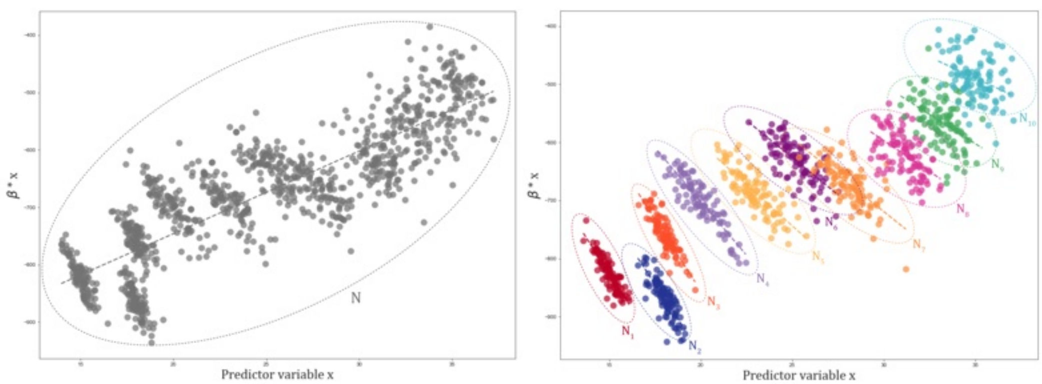
Several other examples of Simpson's Paradox, often using aspatial data, have been observed and investigated in the literature [17, 60, 62, 66, 74]. While these aspatial examples are well-known, the presence of Simpson's Paradox in spatial analysis has largely gone unnoticed. Understanding the cause of its occurrence in the aspatial literature provides clues on how the paradox might occur in spatial analysis. Mathematically, Simpson's paradox is observed when the partitioning criterion used to stratify the population into groups is correlated with both the predictor and the response variables. While some generic versions of Simpson's paradox might still occur despite the lack of correlation, we will focus on the more common and standard kinds of the paradox in this paper. Hence, for an Association Reversal to occur, the stratifying criterion must be correlated with x and y in a model [45, 49]. Re-examining the exercise example from above: generally, people who exercise (x) more appear to be at a higher risk of developing skin cancer (y). However, when the population is stratified based on the amount of average sun exposure (*stratifying* *criterion*) a person receives, the association disappears. Since people who exercise more also tend to get higher sun exposure, the stratifying criterion is correlated with x. Similarly, the stratifying criterion is also correlated with y. Hence, for two people A and B who have the same average exposure to sunlight, frequency of physical activity does not affect their risk of developing skin cancer. This occurrence is an example of a fundamentally aspatial statistical study. It is however not difficult to imagine the possibility of observing the paradox in geographical analysis with location-specific data. As a simplistic example, if we were to expand the scenario of the study and stratify the patients based on the city in which they lived (spatial stratifying criterion), rather than on the basis of exposure to sunlight, would we still observe the stratified trends between skin cancer and physical activity to disappear? We would argue, yes. Cities with higher average sunlight incidence tend to have people who exercise more outside and gain more sun exposure than those who live in colder and greyer places. Albeit the overly simplistic nature of the example presented here, it is easy to imagine that spatial instances of Simpson's paradox might occur and possibly be at least as frequently encountered as in aspatial analysis.

3 A geographical perspective on Simpson's paradox

Research within the field of spatial statistical analysis typically pursues a blend of two lines of enquiry—one is the description of spatial patterns exhibited by the data we measure, and the other is the investigation of the spatial processes that produced those data, which we usually cannot observe and need to infer. Spatial processes are the associations or conditioned relationships between observed phenomena and factors or variables that are hypothesized to affect those phenomena. These relationships are often measured using models within the well-known regression framework. While studies exploring spatial patterns are invaluable in describing how data are distributed over space [2, 4, 13, 19, 29, 30, 48, 50, 51, 55, 57, 61, 75], various spatial and aspatial models of relationships enable investigation of the why questions related to spatial data [2, 5, 6, 9, 18, 21, 34–36, 38, 42, 72]. In traditional models within the regression framework such as Ordinary Least Squares (OLS) regression, a single parameter estimate is calculated to represent the trend between the predictor and a response variable. The inherent assumption in these models is that processes or conditioned associations are constant or stationary across space. For example, in an OLS regression model investigating house prices as a function of number of bedrooms in a house, the assumption would be that no matter where the house is located within the study area, the marginal value of a bedroom is the same.

However, during the last two decades it has become increasingly recognized that some processes, particularly those related to human preferences, decision-making and actions, might vary over space. Consequently, the expectation is that in some neighborhoods of a city the marginal cost of a bedroom might be higher than in others. If processes do vary over space, traditional global models of behavior will be misspecified and their calibration will hide interesting spatial variations in the way in which covariates affect the spatial distribution of a variable of interest. This has led to the development of various local spatial modeling frameworks which can be employed to investigate possible variations in processes over space, examples of which include those of Geographically Weighted Regression [14, 15, 23] and the more recent multiscale version, MGWR [26, 56]; Bayesian Spatially Varying Coefficients Models [8, 27]; Spatial Filtering methods [33, 52]; and Spatially Clustered Coefficient models [43]. Recognizing the innate difference between the spatial and aspatial schools of thought and the contrast between local and global models, leads to some interesting insights into Simpson's paradox.

As a corollary to the expectation that associations might vary across space, a response variable could be assumed to be a function of both the predictor variables and space. Variables that are generally used as predictors in studies related to human-behavior and preferences are also expected to vary as a function of space, that is, exhibit spatial heterogeneity [3]. Consequently, it can be imagined that space is correlated with both the response and predictor variables, and so omitting space could lead to a case of Simpson's paradox similar to that discussed above for the aspatial context. For example, suppose a global model is calibrated using all the data in a study area and a parameter estimate associated with covariate x is found to be significantly positive. It is possible that in the calibration of a model using subsections of the same data partitioned across space, the associations are insignificant or they may be significantly negative. Using simulated data as in Figure 2(a) for instance, the expectation would be that an OLS model calibration would result in a positive significant beta estimate representing a positive correlation. However, calibrating individual models using disaggregated data according to spatial neighborhoods from $\{N_1, N_2, ..., N_{10}\}$ as depicted in Figure 2(b), would result in individual negative significant beta estimates representing a negative correlation between the response and predictor variables. This simplistic example shows how global and local models can reveal seemingly contradictory processes.



(a) (b)

Figure 2: Simulated data example of spatial Simpson's Paradox

To highlight the issues Simpson's paradox might raise in real-world spatial analysis, a more realistic example is now described. Suppose an analyst has a statewide set of geocoded residential real estate prices along with data on various determinants used in hedonic models, one of which is population density. Suppose further that on calibrating a global model, the relationship between house prices and population density is significantly positive suggesting that across the state, cities with greater population density have higher average house prices perhaps owing to access to jobs, amenities and resources. However, the calibration of a local model might reveal the relationship to be significantly negative in many local neighborhoods suggesting that people prefer neighborhoods which are more spacious and less densely populated. This then poses the question: which level of analysis is correct—are house prices positively or negatively related to population density? Simpson's paradox is therefore a scale issue—at one spatial scale we infer that y and x are positively related; at a different spatial scale we infer they are negatively related. As with aspatial examples of Simpson's paradox, both of these statements can be true which poses a substantial challenge to drawing robust inferences from the data.

While Wilson [77] describes one occurrence of Simpson's paradox using a global model and Sachdeva et al. [64] and Fotheringham and Sachdeva [24] more recently show a limited empirical instance of Simpson's paradox using a local model, the potential prevalence of the problem in spatial analysis is unknown. The empirical instances of Simpson's paradox such as those described above could simply be attributed to data anomalies or highly unusual examples. Here, we not only describe three further instances of the paradox across two real world examples, but we also provide a more general argument as to why occurrences of Simpson's paradox when comparing local and global model results might be the norm rather than the exception. The aim of this paper is therefore to illustrate that: (i) Simpson's paradox might be more common than previously thought in spatial information science; (ii) when the paradox is observed in spatial analytic research, it is not a symptom of a problem with the analysis but is simply a manifestation of different processes operating at different spatial scales; and (iii) when the paradox is observed through the calibration of local and global models, this gives us the opportunity to identify the spatial scale at which processes change.

These objectives are addressed in sections 4 and 5 where we examine whether a spatial variant of Simpson's paradox can occur when the results of local and global models are compared using a general simulation experiment and two distinct empirical examples from the field of housing market analysis. Next, we first provide a brief overview of local modeling with MGWR and lead onto the simulation experiment exemplifying its use in detecting an occurrence of Simpson's Paradox in spatial analysis.

4 A demonstration of spatial Simpson's paradox in simulated data

The varying spatial scales at which local and global models are calibrated is demonstrated in Figure 3. Whereas a global model is calibrated using all the data available in a study area (Figure 3a), a local model such as MGWR uses subsets of the data to perform multiple local calibrations with data weighted as inverse functions of distance from the location of each local calibration (Figure 3b).

In what follows, we use Multiscale Geographically Weighted Regression (MGWR) as an example of a local modeling technique [26] and OLS regression as an example of an equivalent global model. For the calibration of the MGWR model, the mgwr Python package [56] with computation improvements from Li and Fotheringham [44] is employed. For the calibration of the global OLS model, the ols function within the `statmodel` package for python is used [67]. In the MGWR calibrations, an adaptive bisquare kernel is employed and GWR bandwidth initialization is used. Ten groups of 100 points each (1000 points in total) are constructed such that x and y are negatively correlated in each of the individual

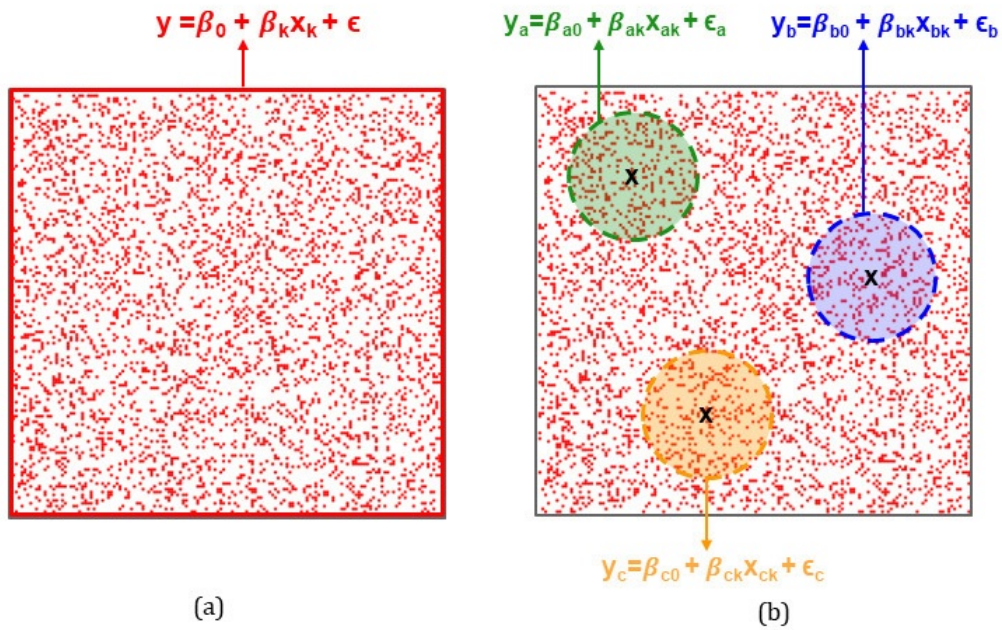


Figure 3: The scale of data used in (a). Global linear regression (left) and (b). MGWR (right)

groups while the whole dataset of 1000 points displays a positive correlation. The covariates $\{x_1, x_2, ..., x_{10}\}$ where x_m represents the m^{th} group of 100 points, are simulated as random normal distributions with a positive mean ranging between 15 and 35 with standard deviation = 1 and the subsequent $\{\beta_1, \beta_2, ..., \beta_{10}\}$ are constructed as random normal distributions with a negative mean ranging between -15 and -55 with standard deviation = 1. For example, for group 1, x_1 is defined as a vector $(n = 100)$ following a random normal distribution with mean = 25 and standard deviation = 1 and β_1 is defined as a vector $(n = 100)$ following a random normal distribution with mean = -25 and standard deviation = 1. The range chosen for the means of the distributions of x_m and β_m are deliberately spread out such that the association for each disaggregated group is negative but when aggregated for all the groups it is positive. The dependent variables for each of the groups $\{y_1, y_2, \ldots, y_{10}\}$ are then constructed from the x_m and β_m of the corresponding groups and a random normal error (ϵ) with mean = 0 and standard deviation = 1, using the following equation:

$$y_{i,m} = \beta_{i,m} x_{i,m} + \epsilon_{i,m} \tag{1}$$

The ten groups of data are then distributed randomly across space as shown in Figure 4. On calibrating the model in equation (1) by MGWR and OLS, the parameter estimates from the OLS calibration are constant and significantly positive at the 95% confidence limit (Figure 4d), whereas those estimated using MGWR (Figure 4e) are all significantly negative at the 95% CL (adjusted for dependent multiple hypothesis tests) and closely replicate the

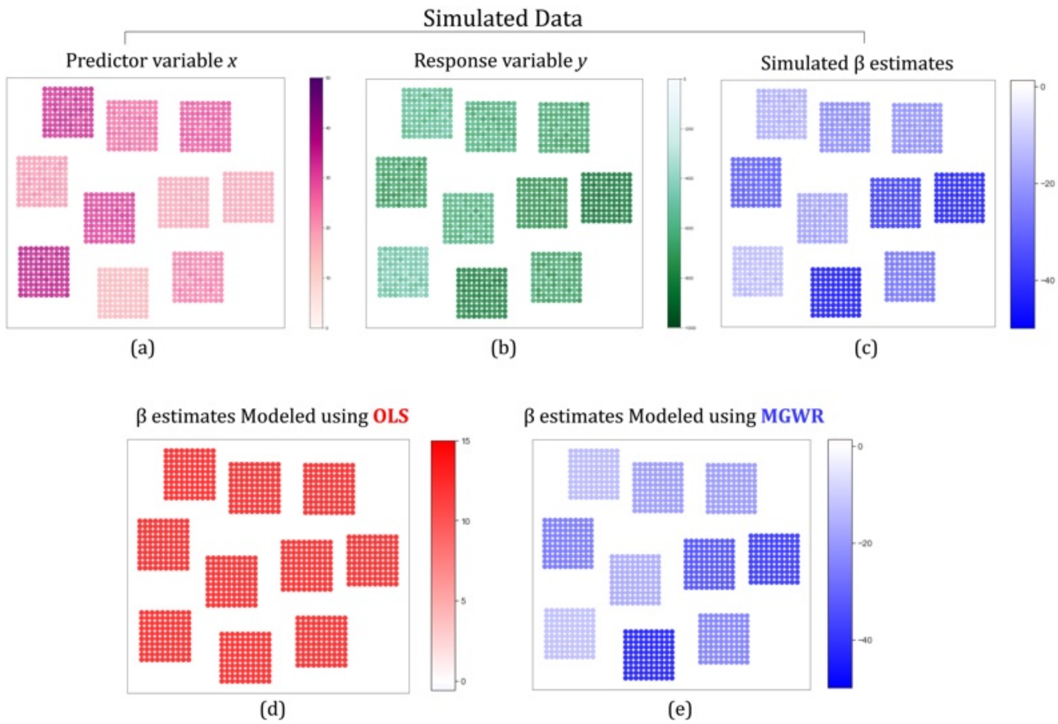


Figure 4: (a). Simulated covariate, (b). Simulated response variable, (c). Simulated process, (d). Parameters estimated using OLS, (c). Parameters estimated using MGWR

'real' beta estimate values (Figure 4c). Consequently, calibration of a global model such as employing the OLS technique results in a positive association between y and x since it considers the whole dataset at once. In contrast, a local model such as MGWR is able to detect the local neighborhoods of negative association and results in a contradictory inference regarding the relationship between y and x. Figure 5 amplifies the relationships estimated using OLS and MGWR in comparison to the simulated associations. The blue clouds of local relationships which reflect the estimates from MGWR are representative of the neighborhood-level associations (gray), whereas the single red trend estimated using OLS describes the global relationship between y and x.

In the above, we describe a simple, simulated example in which a spatial variant of Simpson's paradox is demonstrated to convey the important difference in the scale of the questions being answered when global and local models are calibrated. We now employ these frameworks on real-world data representing two distinct housing markets to detect and discuss occurrences of Simpson's paradox in spatial data analysis.

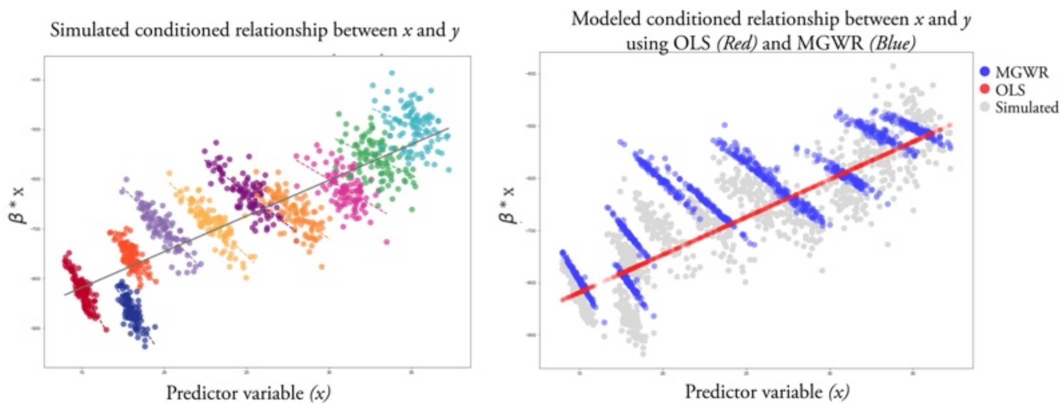


Figure 5: Simulated vs modeled relationships between y and x using MGWR and OLS

5 Two empirical examples of spatial Simpson's paradox

The simulation experiment in Section 4 describes how a spatial variant of Simpson's paradox can occur in theory. We now demonstrate its presence in two real data sets. In both cases a hedonic price model is calibrated: in the first a global and local hedonic price model is calibrated for individual house sales in King Co., Washington; in the second, a different form of hedonic price model is calibrated for individual house sales in Los Angeles Co., CA.

5.1 House prices in King Co. WA

In this example, we draw our model and data from Sachdeva et al. [64] where house sales in the form of 19,832 georeferenced properties in King County, WA sold between May, 2014 and May, 2015, along with the various covariates such as living area, access to waterfront, unemployment rate etc. are used to calibrate a hedonic model. The data were obtained from the online data science competition platform Kaggle [1] and consist of the sales price of each house and various structural attributes of each property. The formulation and other details of the model can be found in Sachdeva et al. [64]. Further, resulting local associations have been tested for potential non-linearities in relationships using the diagnostic tool as described in Sachdeva et al. [63], and no indication of related biases are found. In their paper, Sachdeva et al. [64] report one instance of Simpson's paradox observed between the conditioned relationship of age of a property and its price. Fotheringham and Sachdeva [24] further discuss that instance and comment briefly on its implications for inference. Here, we elaborate on this instance of Simpson's paradox, but go beyond it to report the results of calibrating the local model with different spatial aggregations of the data to investigate if the results vary when the data are aggregated (a novel version of the MAUP) and also to see at what scale the results flip signs. House prices are regressed using the following log-log hedonic price equation:

$$\begin{aligned} ln(house \quad price \quad value_i) = \beta_{i,0} + \beta_{i,1} \quad ln(sqft \quad living_i) + \beta_{i,2} \quad ln(age_i) \\ + \beta_{i,3} \quad basement \quad present_i + \beta_{i,4} \quad ln(dist \quad to \quad waterfront_i) \\ + \beta_{i,5} \quad ln(tech \quad jobs_i) + \beta_{i,6} \quad ln(unemployment \quad rate_i) + \beta_{i,7} \quad ln(index_i) + \epsilon_i \end{aligned} \tag{2}$$

In the calibration of the model in equation (2) by both a global and local model, three instances of Simpson's paradox are uncovered out of the seven slope parameters. The first and most clear instance of the association reversal paradox is observed in the parameter estimate measuring the association of the age of housing on house prices. In an OLS model applied to all the data points in the study area, the estimate of this parameter is 0.01 and significant at the 95% CL suggesting, perhaps counter-intuitively, that across this region older houses are worth more than newer houses, ceteris paribus, and house prices rise by 0.1% for every 10% increase in the age of the housing units. When the model is calibrated locally, on the other hand, most local estimates are significantly negative suggesting that as the age of a house increases, the value declines. Figure 6 depicts these contrasting results at the local model calibration level (a), at the global level (b) and by comparison of the associations at both levels (c and d). Figure 6(c) shows the significant parameter estimates resulting from MGWR model calibration (in blue) and the average, negative, trend line depicting the association between the age of a property and its price. The red dotted line is the association as estimated by OLS and is positive and more representative of the insignificant beta estimates resulting from MGWR (those in red). In Figure 6(d) we show two randomly selected local calibrations (for neighborhoods N_1 and N_2) of the 19,832 in MGWR which are both significantly negative at the 95% CL.

The explanation for the seemingly contradictory results lies in realizing that the questions being answered by the two calibrations are different because of the difference in the spatial scale of the two applications. When all the data are analyzed in a global model, the parameter estimates are based on a comparison of house sales across the whole of King County, WA, and at this scale, areas with older houses, such as in parts of Seattle, are more desirable than other areas containing newer houses: older areas of housing may be seen as more stable and/or as having a cachet associated with age. This generates the significant positive relationship seen in the global model and the local models using highly aggregated data. However, at the local scale the model parameter estimates are not based on a comparison of house prices across the whole county but only within local neighborhoods around each property and where, presumably, the age of properties is more uniform. At this scale, within neighborhoods of similar housing, newer houses are worth more than older ones, ceteris paribus, leading to locally significant negative relationships. We return to this result later in the section for a discussion on how the inferences from such an instance relate to ecological fallacy.

A second milder version of the paradox is seen in the parameter estimate measuring the association of unemployment rates on house prices. In the global model the estimate of this parameter is -0.128 suggesting that across this region lower house prices are associated with higher unemployment rates. When the model is calibrated locally for each origin, as shown in the large panel in Figure 7, while most local estimates are significantly negative reaffirming the conclusions from the global model estimate, there are some positive estimates near Lake Sammamish to the East. This is the more common kind of Simpson's paradox where the trends are not completely reversed for all the subpopulation groups. While even the negative local parameter estimates provide more information about the processes being modeled than the single global estimate, the results do not contradict each other for the majority of the region. To explore this further, we can examine at what scale the inference changes by calibrating the local model for the 11 different scales of aggregated data using grids ranging from 400m by 400m to a larger scale of 4km by 4km and these results are shown in the smaller panels of Figure 7.

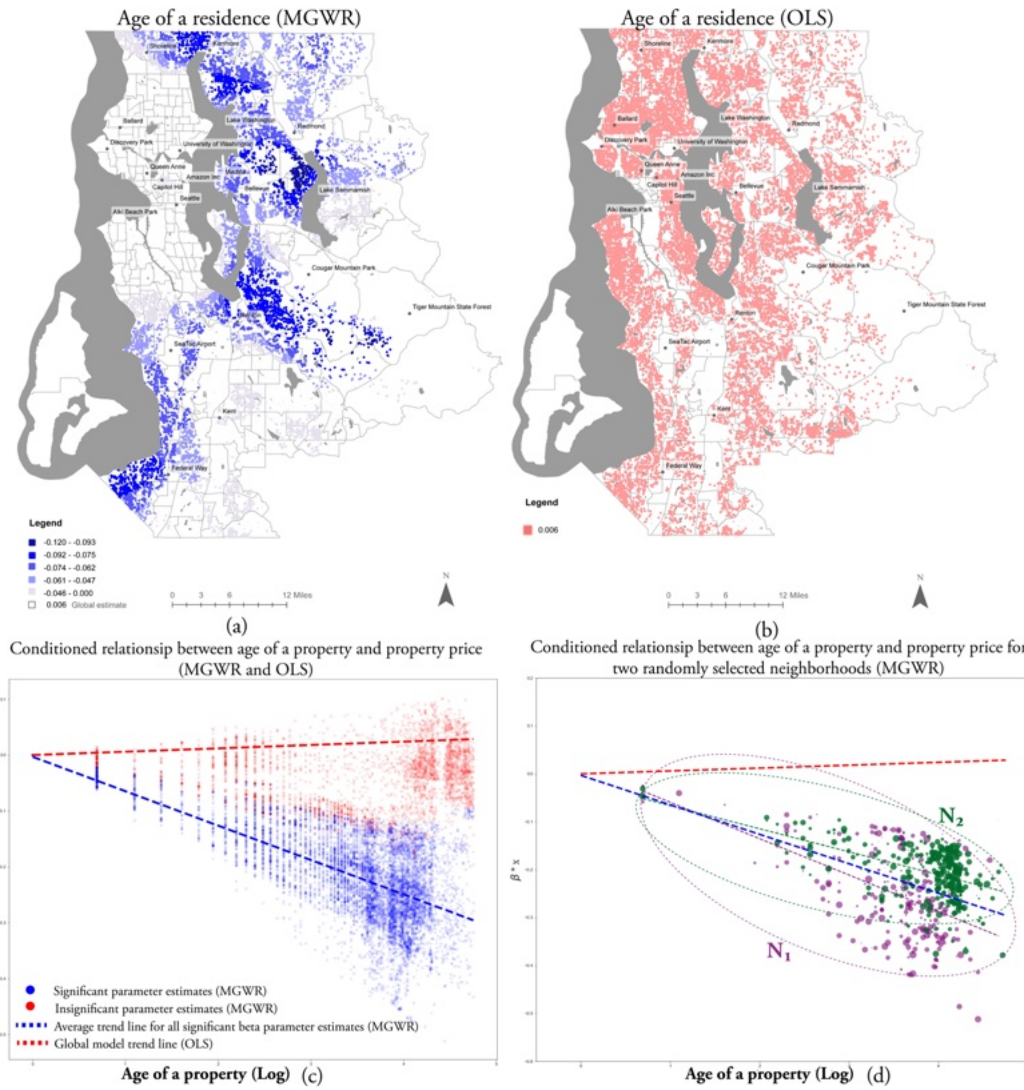


Figure 6: Significant global and local associations between age of a property and its price

In spatial analyses, inference and interpretation of such results becomes all the more important due to the added complexity of the geographical scale of measurement, analysis and interpretation. Contradictory results for some regions at the finer scales (individual

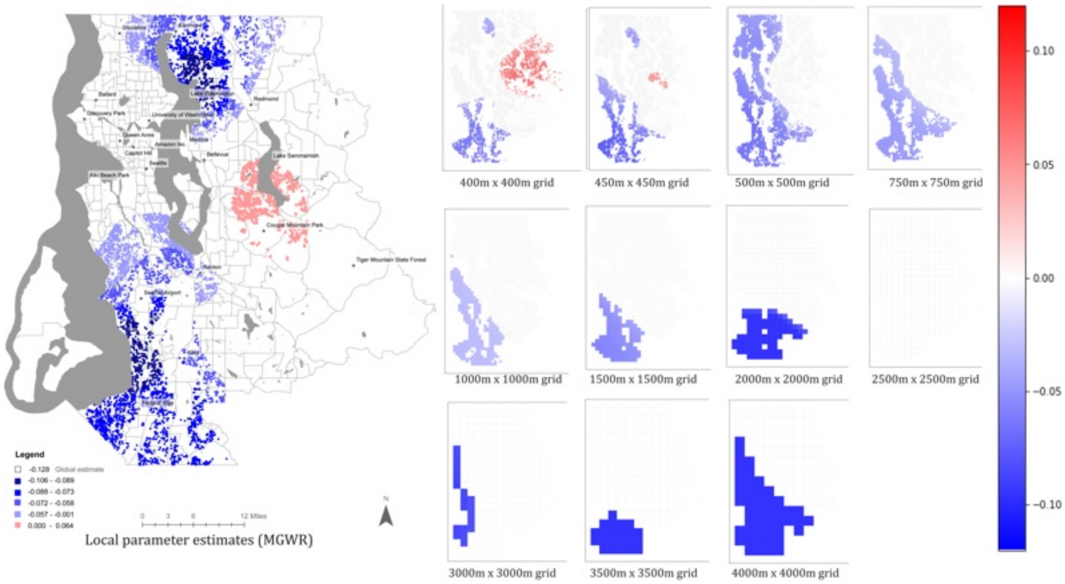


Figure 7: Significant (95% CL adjusted for dependent multiple hypothesis tests) local parameter estimates for the association of rate of unemployment on house price at 12 different spatial scales

point level, 400m and 450m in Figure 7) that disappear in other coarser scale estimates, may again seem spurious, indicating an issue with model specification. However, as observed with the age example, the more likely explanation is that we are inferring different trends at different scales since we are asking different questions from these distinct models. At the global scale, when all the data are used for the model calibration, areas with high unemployment rates, such as South of Seattle (Federal Way, Des Moines), are also generally less desirable than areas with low unemployment rates around Seattle, such as Kirkland and Queen Anne, and are hence less preferred residential neighborhoods. At finer scales, inference drawn from comparisons within smaller housing clusters for areas in North and South of Seattle, exhibit a similar trend. The neighborhoods near Lake Sammamish, however, at finer scales produce a trend which is contradictory to that at the global scale.

Interestingly, the areas around Sammamish exhibit high unemployment rates (ACS 2014 data) as well as high house prices. The tech industry around Seattle was booming at that time when industry giants such as Facebook, Google and Uber were setting up huge offices in the area. The explosion of jobs from these new campuses, in addition to the existing expanding Microsoft headquarters, led to an increase in employed, skilled people over the following years in the area. The local trends around Sammamish are therefore potentially signals caught at the start of when the housing market in the area began to change. People with high-paying jobs moving into the area, as well as real estate companies that were projecting a rising demand of housing in the area, were potentially acquiring properties in previously impoverished suburbs with close access to the new office communities. This process, revealed using local modeling, is reaffirmed by the following data. The population in Sammamish increased by about 29% and unemployment rate decreased by about 40% from 2014 to 2018 yet the city has one of the costliest housing markets in the country. Perhaps the lag between high house prices for properties sold between 2014 and 2015 and the influx of employed and stable population (which would be reflected in community survey data in the years after 2014) was captured by the local model estimates. Since houses with high house prices and high unemployment rates are compared amongst each other at finer scales in the local model calibration, higher unemployment rates are seen as affecting house prices in a contradictory manner at those scales.

Finally, the third instance of Simpson's Paradox is observed in the waterfront view parameter estimates in the model. In addition to accounting for the effects of access to a waterfront, a composite measure including waterfront access and the elevation of a residence was also included in the model to measure the influence of sea or lake views on price. The index, ranging between 0 and 1, has a higher value when houses have a high elevation and are close to a waterfront and a lower value when houses are at low elevations farther from a water body. The global parameter estimate is significant but negative, which is counter-intuitive. The significant local estimates range from -5.8 to 8.6 and are displayed in Figure 8. The positive estimates in areas around Queen Anne, Capitol Hill and Bellevue, reflect an increase in housing price with an increase in the index, ceteris paribus which is intuitive as in these areas increased elevation would lead to a better view of a bay or lake. In some of the areas with negative local parameter estimates the index appears to be acting as an inverse proxy for access to parks and golf courses. Figure 8 also reveals that there are relatively few locations with significant parameter estimates despite the global estimate being significant. This is also a feature of both local modeling and Simpson's paradox.

5.2 House prices in Greater Los Angeles

A second real-world study based on house-price determinants in greater Los Angeles is used to support the inferences drawn from the previous example. Here, a subset of a 1990 California census dataset often referred to in the literature [28, 41, 65] is employed to calibrate a local (MGWR) and global (OLS) hedonic price model. The dataset consists of median house price values and some summary attributes such as median age of housing, median income of residents, average proximity of the block group to the coastline, etc, for all census block groups in California. Here we use median house price in each block group as the dependent variable. A log-log hedonic house price model where both the independent and dependent variables are log transformed, is formulated as follows:

$$\begin{aligned} ln(median \quad house \quad price_i) = \beta_{i,0} + \beta_{i,1} \quad ln(median \quad income_i) \\ +\beta_{i,2} \quad ln(median \quad age \quad of \quad housing_i) \\ +\beta_{i,3} \quad ln(no. \quad of \quad beds \quad per \quad pop_i) + \beta_{i,4} \quad ln(no. \quad of \quad households_i) \\ +\beta_{i,5} \quad ln(dist. \quad from \quad coastline_i) + \epsilon_i \end{aligned} \tag{3}$$

An adaptive bisquare kernel using nearest neighbors is used to calibrate the MGWR model. A local model is calibrated using the original 7,126 block group data and then on aggregations of the data to eight spatial scales as shown in Figure 9. The r-squared value for the individual level base MGWR model is 0.91.

A similar instance of Simpson's Paradox to the one observed for King County WA is observed between median house prices in LA and the predictor measuring median age. In

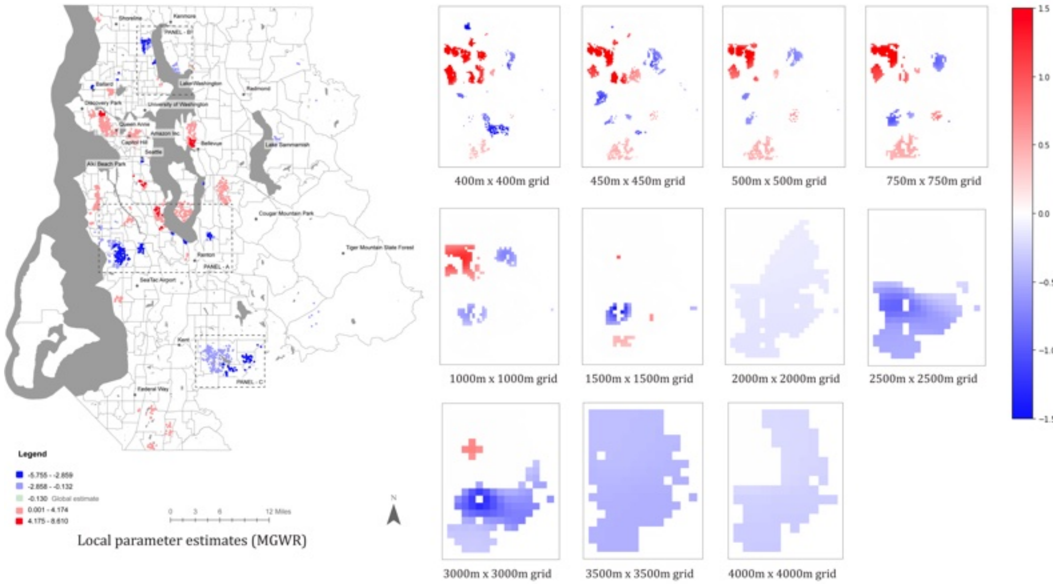


Figure 8: Significant (95% CL adjusted for dependent multiple hypothesis tests) local parameter estimates for the influence of waterfront view on house price at 12 different spatial scales

the global model applied to all 7,126 block groups, the estimate of this parameter is 0.065 with a SE of 0.008 suggesting a preference for older housing neighborhoods at the global level. The map of the significant local parameter estimates in Figure 10 displays a mix of negative and positive parameter estimates. For example, positive estimates are observed around Beverly Hills, where, presumably, block groups containing older and more expensive housing units are preferred to block groups with newer and cheaper house values. On the other hand, block groups around areas such as Inglewood display a preference for newer housing communities. At the most coarse scale of the analysis (10km x 10km grid), the local parameter estimates are all positive and significant and are concentrated in the richer, more preferred part of the city, similar to that observed in the context of the King County housing market. This example reaffirms that the different and contradictory results obtained using the King County dataset were not peculiar to that data set and may well be a common feature of spatial data analysis.

6 Spatial processes and their scales of transition

6.1 Simpson's paradox and the ecological fallacy

The observation of Simpson's Paradox in spatial analysis is not just central to the measurement and operational scales of analyses in spatial sciences but also to the scale at which interpretations are made. Studies using global models of analysis and spatially aggregated

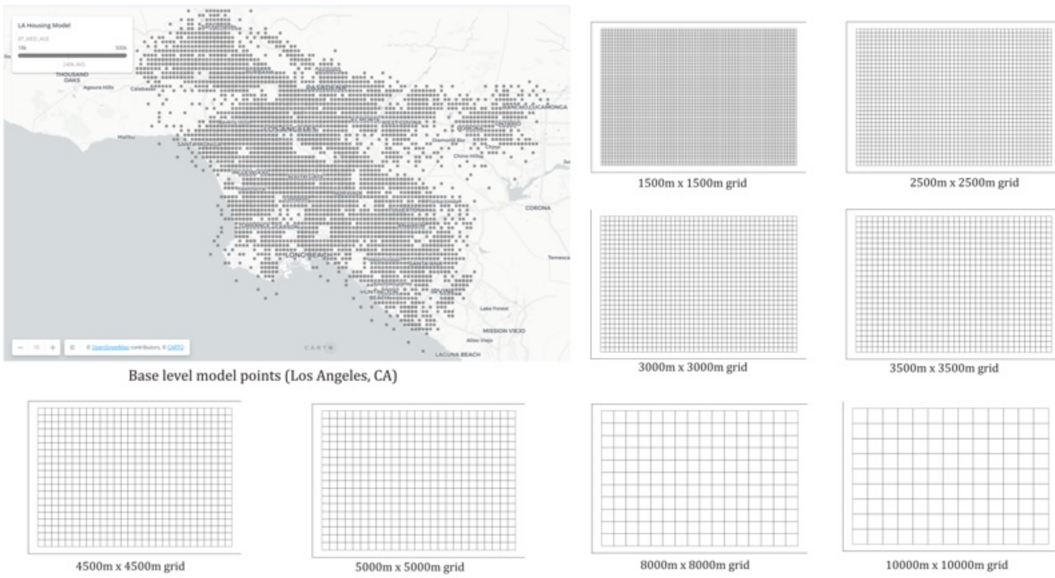


Figure 9: Original LA housing block group centroids and grid scales used for aggregation

data are often used to infer behavior or properties relating to individual units within the study area. When observed in the context of the scale at which the data are measured, this is referred to as the *ecological fallacy*. Another type of ecological fallacy can occur when interpreting the results from local and global models. As observed in the examples of Simpson's paradox above, local and global models are operationally different and provide evidence on fundamentally different processes. In the presence of Simpson's paradox, inferences from global and local models are therefore not as contradictory as they appear. For example, consider the effect of the age of a housing unit on house prices in the King Co. example. As shown in Figure 11, if the data on individual housing units are aggregated to different administrative units (block groups, census tracts and zip codes), a reversal in trend between the age of housing and house prices is observed which is similar to that observed when a comparison between the local and global model calibration is made. At the property level, a local model is estimated using only a weighted subset of the data and essentially the model's parameter estimates report on the choice process of individuals comparing houses within neighborhoods with newer houses within a neighborhood of largely similar houses being preferred (Figure 12 panel a). At the zip code level however, large areas of housing are being compared with older neighborhoods being preferred to newer ones (Figure 12 panel b). These are fundamentally different processes that answer different questions about the observed phenomena. While the marginal cost of a newer housing unit is higher than its older counterpart, zip codes or areas with older, more stable housing markets are preferred over newer, less well-established, neighborhoods.

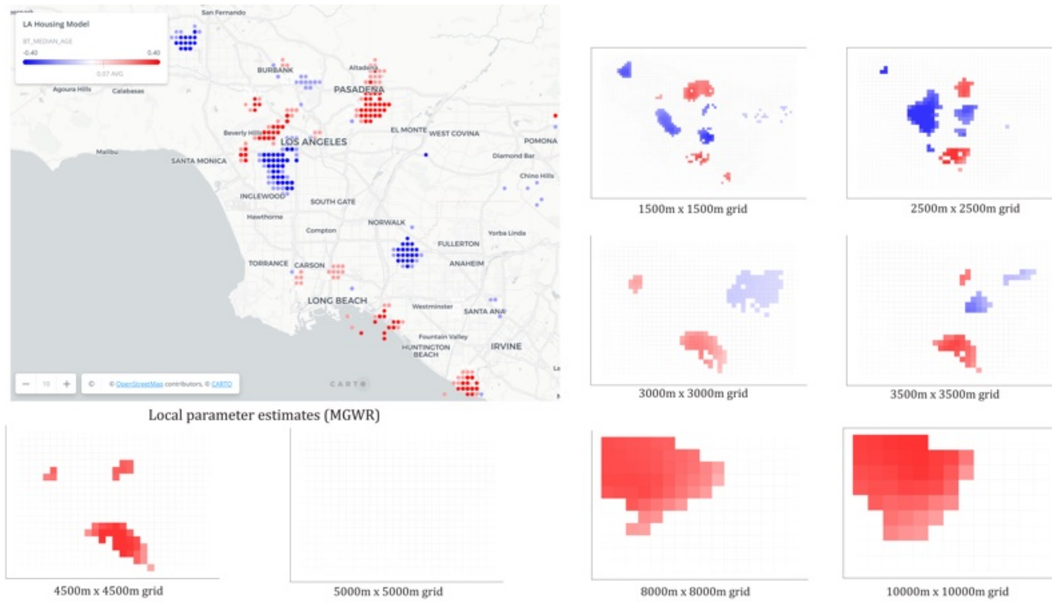


Figure 10: Significant (95% CL adjusted for dependent multiple hypothesis tests) local parameter estimates for the association between median age of housing on house prices at 8 different spatial scales.

6.2 Spatial scales at which processes transition

An advantage of calibrating global and local models at different aggregations of data is the potential to identify the critical spatial scale at which conditioned relationships change sign For instance, in the King Co. empirical example presented in section 5.1, the aggregation at which the paradox is observed is different for different conditioned relationships, as shown in Figure 13. The transition from mixed conditioned associations between unemployment rate and house price for different neighborhoods to completely negative associations occurs at a local scale of around 450 m whereas the transition of the waterfront view and housing age conditioned associations with house price both occur at about 1.5k . This variation in transition scales may shed light on different aspects of what constitutes a relatively homogeneous neighborhood for different facets of house price determinants. In this case, for example, the scale at which unemployment rates affects house prices appears to be more local than the scale at which the conditioned relationship between house price and age becomes more homogeneous suggesting larger neighborhoods with relatively homogenous construction ages. Neighborhoods larger than about 1.5 square kms are compared with one another, the average age of housing in such units exerts a greater distinction between older and wealthier neighborhoods, such as historic districts, and newer, less wealthy suburban housing tracts.

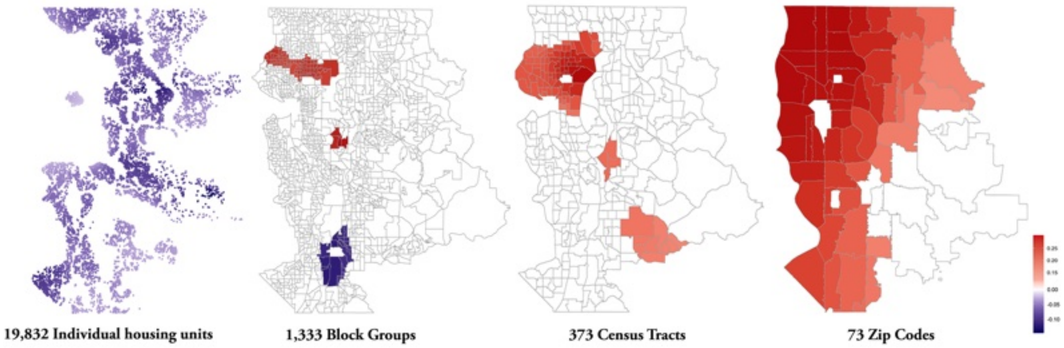


Figure 11: Significant (95% CL adjusted for dependent multiple hypothesis tests) local parameter estimates for the influence of the age of a residence on house price at four different spatial scales—King Co., WA.

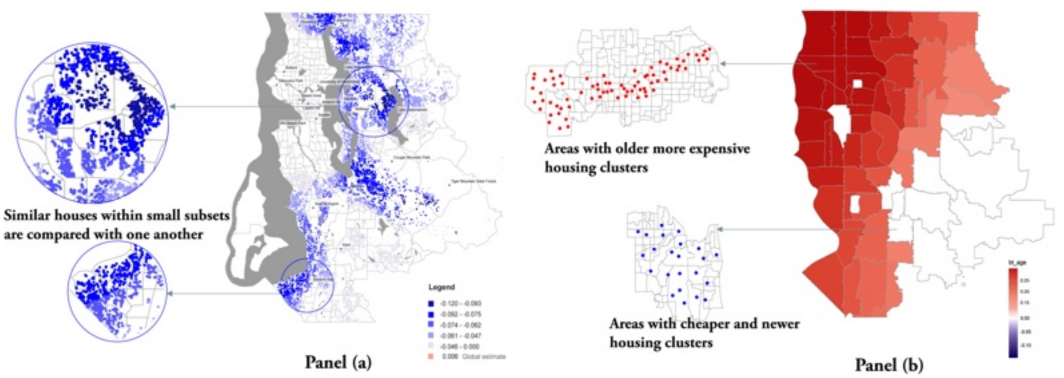


Figure 12: Detailed panels representing how different spatial processes are estimated at different scales.

7 Discussion

The geographical scale of analysis is an inherent component of any form of spatial analysis yet policy implications and interpretations from spatial analysis research are beset with issues around the appropriate scale at which phenomena are measured, analyzed and interpreted. Here we demonstrate how this is an even greater problem than previously thought given the growing popularity of local statistical modeling and the ever-increasing likelihood of encountering a spatial version of Simpson's paradox. Until now, instances of Simpson's paradox have rarely been identified and explored in spatial analysis so our understanding of the phenomenon and its implications, especially for policy oriented research, is limited. Here, we demonstrate through the lens of local modeling that: (i) this is a fundamental problem about which spatial analysts should be aware; (ii) by refocusing

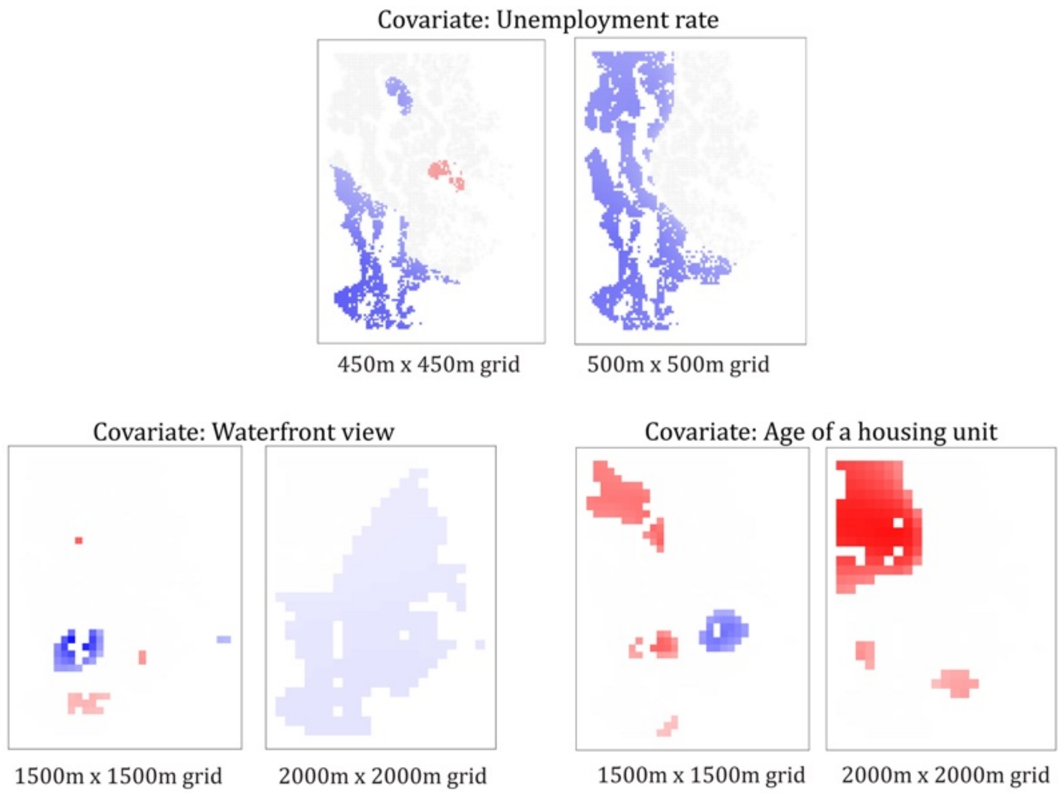


Figure 13: The spatial scales at which processes change in the King Co. housing empirical example

the problem in terms of processes, we contribute significantly to a greater understanding of the paradox as it occurs in local modeling; and (iii) an occurrence of the paradox in studies comparing local and global models is not a problem and is to be expected. We examine the role of scale and space in the occurrence of Simpson's paradox in research using results from both local and global models. The raison d'etre of local models is that a global scale (where 'global' simply refers to all locations within a predefined area of interest) might be the incorrect scale at which to undertake any analysis of spatial processes; the alternative being a local scale (where 'local' refers to individual locations). Using results from local and global models, we discuss and exemplify the extreme differences that can result when calibrating global and local models with the same data and how Simpson's paradox can arise in this context. We highlight through this study that where scale refers to the geographical entity for which a model is calibrated, the results from neither the local nor global models are incorrect—these just answer different spatial questions related to the phenomena.

Further, through this study we show that parameters estimated in the calibration of global models are not necessarily 'averages' of their equivalent local estimates. In some instances, a spatial variant of Simpson's paradox can arise whereby a relationship is significantly positive (negative) at the global level but is significantly negative (positive) at the local level. There is nothing wrong in this and there is no reason to expect global estimates to summarize local estimates, although in some cases, they might. It bears repeating that while an interpretation of a spatial process at one scale might contradict that at another, one interpretation is not necessarily more correct than the other. Models aggregated to different spatial scales essentially answer different spatial questions. Global parameter estimates inform on the conditioned relationship between y and x across the whole study area; local parameter estimates inform on the same conditioned relationship but around a single location. These relationships might be very different, as exemplified by the conditioned relationship between the value of a property and its age shown above in two empirical examples.

With the growing number of applications that use local modeling techniques, these findings are important in their own right but are also crucial in understanding the issue of reproducibility and replicability in geographical analysis [32,39,40,69]. A well-conditioned model (i.e., one which is properly specified and contains no serious statistical flaws) could still generate different parameter estimates when applied to the same data set aggregated to different sets of zones. Further, if the same well-conditioned model were calibrated globally and locally, we should not expect to see similar results because we are asking different questions at the two scales. Consequently, the issues of reproducibility and replicability in geographical analysis are much more nuanced than in some areas of science and the occurrence of Simpson's paradox is not a strange, inexplicable property that invalidates geographical analyses. Rather, variations in estimated parameters from the same model are to be expected and are a natural reflection of spatially varying processes and because we ask different questions at different spatial scales.

References

[1] House Sales in King County, USA | Kaggle.

[2] ANSELIN, L. *Spatial econometrics: methods and models*. Kluwer Academic Publishers, 1988.

[3] ANSELIN, L. What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis (89-4). https://escholarship.org/uc/item/3ph5k0d4, 1989.

[4] ANSELIN, L. Local Indicators of Spatial Association–LISA. *Geographical Analysis 27*, 2 (1995), 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.

[5] ANSELIN, L. Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics 27*, 3 (2002), 247–267. doi:10.1111/j.1574-0862.2002.tb00120.x.

[6] ANSELIN, L. *Spatial Regression. The SAGE Handbook of Spatial Analysis*. SAGE Publications, Ltd, 2009.

[7] BANDYOAPDHYAY, P., NELSON, D., GREENWOOD, M., BRITTAN, G., AND BERWALD, J. The logic of Simpson's paradox. *Synthese 181*, 2 (2011), 185–208. doi:10.1007/s11229-010-9797-0.

[8] BANERJEE, S., CARLIN, B., AND GELFAND, A. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2014.

[9] BANERJEE, S., GELFAND, A., AND CARLIN, B. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2003.

[10] BICKEL, P. J., HAMMEL, E. A., AND O'CONNELL, J. W. Sex bias in graduate admissions: Data from berkeley: Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation. *Science 187*, 4175 (1975), 398–404. doi:10.1126/science.187.4175.398.

[11] BLYTH, C. On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association 67*, 338 (1972), 364–366. doi:10.1080/01621459.1972.10482387.

[12] BLYTH, C. Simpson's Paradox and Mutually Favorable Events. *Journal of the American Statistical Association 68*, 343 (1973), 746–746. doi:10.1080/01621459.1973.10481419.

[13] BOOTS, B., AND GETIS, A. *Point Pattern Analysis*. SAGE Publications, 1988.

[14] BRUNSDON, C., FOTHERINGHAM, A., AND CHARLTON, M. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis 28*, 4 (1996), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x.

[15] BRUNSDON, C., FOTHERINGHAM, S., AND CHARLTON, M. Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician 47*, 3 (1998), 431–443. doi:10.1111/1467-9884.00145.

[16] CARTWRIGHT, N. Causal Laws and Effective Strategies. *Noûs 13*, 4 (1979), 419–437. doi:10.2307/2215337.

[17] CHARIG, C. R., WEBB, D. R., PAYNE, S. R., AND WICKHAM, J. E. Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *Br Med J (Clin Res Ed) 292*, 6524 (1986), 879–882.

[18] CHUMNEY, E., AND SIMPSON, K. *Methods and Designs for Outcomes Research*. ASHP, 2006.

[19] CLIFF, A., AND ORD, J. *Spatial Processes Models and Applications*. Pion Ltd, 1981.

[20] COHEN, M., AND NAGEL, E. *An Introduction to Logic and Scientific Method*. Harcourt, Brace and Company, 1934.

[21] DEMPSTER, A., SCHATZOFF, M., AND WERMUTH, N. A Simulation Study of Alternatives to Ordinary Least Squares. *Journal of the American Statistical Association 72* (1977), 357. doi:10.1080/01621459.1977.10479910.

[22] EELLS, E. *Probabilistic Causality*. Cambridge University Press, 1991.

[23] FOTHERINGHAM, A., BRUNSDON, C., AND CHARLTON, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, 2003.

[24] FOTHERINGHAM, A., AND SACHDEVA, M. Scale and local modeling: new perspectives on the modifiable areal unit problem and Simpson's paradox. *Journal of Geographical Systems* (2022). doi:10.1007/s10109-021-00371-5.

[25] FOTHERINGHAM, A., AND WONG, D. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A: Economy and Space 23*, 7 (1991), 1025–1044. doi:10.1068/a231025.

[26] FOTHERINGHAM, A., YANG, W., AND KANG, W. Multiscale Geographically Weighted Regression (MGWR. *Annals of the American Association of Geographers 107*, 6 (2017), 1247–1265. doi:10.1080/24694452.2017.1352480.

[27] GELFAND, A., KIM, H.-J., SIRMANS, C., AND BANERJEE, S. Spatial Modeling With Spatially Varying Coefficient Processes. *Journal of the American Statistical Association 98* (2003), 462. doi:10.1198/016214503000170.

[28] GÉRON, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2017.

[29] GETIS, A. Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach. *Environment and Planning A: Economy and Space 23*, 9 (1991), 1269–1277. doi:10.1068/a231269.

[30] GETIS, A., AND ORD, J. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis 24*, 3 (1992), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x.

[31] GOOD, I., AND MITTAL, Y. The Amalgamation and Geometry of Two-by-Two Contingency Tables. *The Annals of Statistics 15*, 2 (1987), 694–711. doi:10.1214/aos/1176350369.

[32] GOODCHILD, M. F., FOTHERINGHAM, A. S., KEDRON, P., AND LI, W. Introduction: Forum on reproducibility and replicability in geography. *Annals of the American Association of Geographers 111*, 5 (2021), 1271–1274.

[33] GRIFFITH, D. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Springer Science & Business Media, 2003.

[34] GRIFFITH, D., AND CSILLAG, F. Exploring relationships between semi-variogram and spatial autoregressive models. *Papers in Regional Science 72*, 3 (1993), 283–295. doi:10.1007/BF01434277.

[35] HAINING, R. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, 1993.

[36] HAINING, R. *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, 2003.

[37] HARVEY, D. Pattern, Process, and the Scale Problem in Geographical Research. *Transactions of the Institute of British Geographers 45* (1967), 71–78. doi:10.2307/621393.

[38] HUTCHESON, G. *The Multivariate Social Scientist*. SAGE Publications, Ltd, 1999.

[39] KEDRON, P., FRAZIER, A. E., TRGOVAC, A. B., NELSON, T., AND FOTHERINGHAM, A. S. Reproducibility and replicability in geographical analysis. *Geographical Analysis 53*, 1 (2021), 135–147.

[40] KEDRON, P., LI, W., FOTHERINGHAM, S., AND GOODCHILD, M. Reproducibility and replicability: opportunities and challenges for geospatial research. *International Journal of Geographical Information Science 35*, 3 (2021), 427–445.

[41] KELLEY PACE, R., AND BARRY, R. Sparse spatial autoregressions. *Statistics & Probability Letters 33*, 3 (1997), 291–297. doi:10.1016/S0167-7152(96)00140-X.

[42] LESAGE, J., AND PACE, R. *Spatial Econometric Models. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer, 2010.

[43] LI, F., AND SANG, H. Spatial Homogeneity Pursuit of Regression Coefficients for Large Datasets. *Journal of the American Statistical Association 114*, 527 (2019), 1050–1062. doi:10.1080/01621459.2018.1529595.

[44] LI, Z., FOTHERINGHAM, A., LI, W., AND OSHAN, T. Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science 33*, 1 (2019), 155–175. doi:10.1080/13658816.2018.1521523.

[45] LINDLEY, D., AND NOVICK, M. The Role of Exchangeability in Inference. *The Annals of Statistics 9*, 1 (1981), 45–58. doi:10.1214/aos/1176345331.

[46] LLOYD, C. *Exploring Spatial Scale in Geography*. Wiley, 2014.

[47] MALINAS, G. Simpson's Paradox: A Logically Benign, Empirically Treacherous Hydra. *The Monist 84*, 2 (2001), 265–283.

[48] MATHERON, G. Principles of geostatistics. *Economic Geology 58*, 8 (1963), 1246–1266. doi:10.2113/gsecongeo.58.8.1246.

[49] MITTAL, Y. Homogeneity of Subpopulations and Simpson's Paradox. *Journal of the American Statistical Association 86* (1991), 413. doi:10.1080/01621459.1991.10475016.

[50] MORAN, P. Notes on Continuous Stochastic Phenomena. *Biometrika 1*, 2 (1950), 17–23. doi:10.2307/2332142.

[51] MORAN, P. A. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological) 10*, 2 (1948), 243–251.

[52] MURAKAMI, D., AND GRIFFITH, D. Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems 17*, 4 (2015), 311–331. doi:10.1007/s10109-015-0213-7.

[53] OPENSHAW, S. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the institute of British geographers* (1977), 459–472. doi:10.2307/622300.

[54] OPENSHOW, S. A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical applications in the spatial science* (1979), 127–144.

[55] ORD, J., AND GETIS, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis 27*, 4 (1995), 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x.

[56] OSHAN, T., LI, Z., KANG, W., WOLF, L., AND FOTHERINGHAM, A. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS International Journal of Geo-Information 8*, 6 (2019), 269. doi:10.3390/ijgi8060269.

[57] PAELINCK, J., AND KLAASSEN, L. *Spatial Econometrics*. Gower, 1979.

[58] PEARSON, K. On the Theory of Genetic (Reproductive) Selection. Philosophical Transaction of the Royal Society. *Series A 192* (1899), 260–278.

[59] QUATTROCHI, D., AND GOODCHILD, M. *Scale in remote sensing and GIS*. Lewis Publishers, 1997.

[60] RADELET, M. Racial Characteristics and the Imposition of the Death Penalty. *American Sociological Review 46*, 6 (1981), 918–27.

[61] RIPLEY, B. Spectral Analysis and the Analysis of Pattern in Plant Communities. *The Journal of Ecology 66*, 3 (1978), 965. doi:10.2307/2259308.

[62] ROSS, K. *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*. Pi Press, 2004.

[63] SACHDEVA, M., FOTHERINGHAM, A. S., LI, Z., AND YU, H. Are we modelling spatially varying processes or non-linear relationships? *Geographical Analysis 54*, 4 (2022), 715–738. doi:10.1111/gean.12297.

[64] SACHDEVA, M., FOTHERINGHAM, S., AND LI, Z. Do places have value?: Quantifying the intrinsic value of housing neighborhoods using mgwr. *Journal of Housing Research 31*, 1 (2022), 24–52.

[65] SACHDEVA, M., OSHAN, T., AND FOTHERINGTON, A. S. Local modeling in a regression framework. In *Handbook of Spatial Analysis in the Social Sciences*. Edward Elgar Publishing, 2022, pp. 123–141.

[66] SAMUELS, M. Simpson's Paradox and Related Phenomena. *Journal of the American Statistical Association 88* (1993), 421. doi:10.1080/01621459.1993.10594297.

[67] SEABOLD, S., AND P., J. statsmodels: Econometric and statistical modeling with python, 2010.

[68] SIMPSON, E. H. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological) 13*, 2 (1951), 238–241. doi:10.1111/j.2517-6161.1951.tb00088.x.

[69] SUI, D., AND KEDRON, P. Reproducibility and replicability in the context of the contested identities of geography. *Annals of the American Association of Geographers 111*, 5 (2021), 1275–1283. doi:10.1080/24694452.2020.1806024.

[70] TATE, N., AND ATKINSON, P. *Modelling scale in geographical information science*. Wiley, 2001.

[71] THOMAS, M. Statistical Geography. Problems in analyzing areal data. *Science 136*, 3515 (1962), 191–196. doi:10.1126/science.136.3515.510. Place: New York.

[72] TIEFELSDORF, M. *Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I*. Springer-Verlag, 2000.

[73] VON KÜGELGEN, J., GRESELE, L., AND SCHÖLKOPF, B. Simpson's paradox in covid-19 case fatality rates: a mediation analysis of age-related causal effects. *IEEE Transactions on Artificial Intelligence 2*, 1 (2021), 18–27.

[74] WAGNER, C. Simpson's Paradox in Real Life. *The American Statistician 36*, 1 (1982), 46–48. doi:10.1080/00031305.1982.10482778.

[75] WHITTLE, P. On Stationary Processes in the Plane. *Biometrika 3*, 4 (1954), 434–449. doi:10.2307/2332724.

[76] WILLIAMS MERTEN, J., KING, J., WALSH-CHILDERS, K., VILARO, M., AND POMERANZ, J. Skin Cancer Risk and Other Health Risk Behaviors: A Scoping Review. *American Journal of Lifestyle Medicine 11*, 2 (2017), 182–196. doi:10.1177/1559827615594350.

[77] WILSON, R. Changing Geographic Units and the Analytical Consequences: An Example of Simpson's Paradox. *Cityscape 15*, 2 (2013), 289–304.

[78] YULE, G. Notes on the Theory of Association of Attributes in Statistics. *Biometrika 2*, 2 (1903), 121–134. doi:10.1093/biomet/2.2.121.