

ARTICLE

The Modifiable Areal Unit Problem in Political Science

Dong Wook Lee¹ , Melissa Rogers²  and Hillel David Soifer³ 

¹Department of Political Science and International Relations, Adelphi University, Garden City, NY 11530, USA;

²Department of Politics and Policy, Claremont Graduate University, Claremont, CA 91711, USA; ³Department of Political Science, University of California, Berkeley, CA, USA.

Corresponding author: Melissa Rogers; Email: melissa.rogers@cgu.edu

(Received 26 September 2023; revised 30 January 2024; accepted 21 February 2024)

Abstract

Building on the availability of geospatial data, improvements in mapping software, and innovations in spatial statistics, political scientists are increasingly taking geography seriously. As we adopt the tools of geographers, we must also consider the methodological challenges they have identified. We focus on the modifiable areal unit problem (MAUP)—the idea that the size of aggregate spatial units and the location of their borders affect the empirical results we obtain. We first describe the logic of the MAUP, and then demonstrate the MAUP through simulations, showing MAUP-related inconsistency in regression results in randomly generated and real-world data. We identify MAUP concerns, and best practices, in top journals in political science. We conclude by suggesting how scholars may respond in theoretical and empirical terms to concerns about validity and reliability that arise from the MAUP.

Keywords: geospatial data; modifiable areal unit problem; measurement

Edited by: Jeff Gill

1. Introduction

Whether we study the dynamics of collective processes, or how individual actions are shaped by the context in which they occur, we are engaged in the study of spatial aggregate units (Darmofal 2015). Yet scholars can often choose from a range of plausible options as they operationalize those spatial units. These are inherently theoretical choices that are often not explicitly considered in political science, despite several decades of scholarship in geography on the modifiable areal unit problem (hereafter, modifiable areal unit problem [MAUP]) that show their importance. We argue that much is at stake in the choice of spatial unit. Our paper explains the logic underlying the MAUP, shows it is common in political science research, and demonstrates through simulations that when we conduct the same analysis with individual data points aggregated to different spatial units, we may find entirely different relationships between variables, and therefore different descriptive findings and regression results. Our central claim is that the most effective response to the threats to inference posed by the MAUP is a theoretical one; a clear justification of the choice of the spatial unit of analysis. This fundamental though too often neglected element of research design deserves more explicit attention because it provides the most powerful response to the MAUP.

The remainder of the paper is organized into three parts. We first describe the MAUP by demonstrating it with a simple example drawn from voting data from American politics. Second, we identify inferential threats from the MAUP through simulation using generated and real-world data. We find changing the scale and border locations of units may impact correlations and regression output

in unpredictable ways. Next we describe our re-analysis of influential recent scholarship with data aggregated to plausible alternative spatial units, finding that results change across spatial mappings. We also code five years of articles in two flagship journals to show that the MAUP threatens inference in a wide range of political science scholarship, and highlight best practices from that research. Before concluding, we present potential ways to manage the inferential threats of the MAUP, emphasizing validity as the most powerful response and drawing examples from scholarship.

Existing recommendations for responding to the MAUP center on demonstrations of reliability of findings across spatial mappings. Based on our assessment of the reliability of results from our simulated data, however, we show that reliability is more limited than scholars might have hoped, and thus cast doubt on the effectiveness of demonstrations of reliability at a single alternative spatial mapping as a persuasive response to the MAUP.

We argue that scholars may instead get bigger payoffs from justifying the validity of the spatial units they choose. This justification, which is fundamentally a theoretical rather than a methodological response to the MAUP, is the most effective way to stem potential threats to inference that it poses. For scholars working in settings where they are unable to justify a single spatial unit as unique and valid, or where data availability is a constraint, reliability is therefore a second-best but potentially useful response. We offer an intuitive way to assess the reliability of results across multiple spatial mappings.

2. The MAUP

Geographers have long been concerned that studies of aggregate phenomena using data at different levels of aggregation produce different descriptive patterns, and underpin divergent interpretations of the causal claims they evaluate. Fotheringham and Wong (1991), Amrhein (1995), Openshaw (1984) and others have shown that the results of statistical analyses using both linear and logit models change in wildly unpredictable ways across spatial mappings. They show that we cannot characterize the underlying distribution from which the results of a given set of border mappings are drawn, and that there is no way to predict how a given change in mapping will affect the covariation among variables. This is the problem geographers have called the MAUP. Political scientists, while devoting a great deal of energy to careful data analysis, have not sufficiently considered the implications of the MAUP despite the fact that many of our analyses examine spatially-aggregated phenomena.

2.1. The Logic of the MAUP

The MAUP refers to the fact that the same individual-level data aggregated into different spatial units will produce distinct inferences about relationships among characteristics of those units. Both the values of spatial aggregate units on each attribute and the covariation among attributes across cases can potentially change in unpredictable, substantively large, and statistically significant ways as the units of analysis change.

The MAUP arises when researchers are interested in studying phenomena at the aggregate level, but they cannot identify a single uniquely valid aggregate unit for operationalizing those processes. If we aggregate individuals into groups in different ways, the covariation across characteristics for that group of individuals changes in unpredictable ways as group composition changes, even if the mean scores for the group on each characteristic do not change (Amrhein 1995). Moreover, a change in the unit of analysis for any variable—whether independent, dependent, or control variable—sets off logically equivalent effects on the results of regression analysis.¹ As shown in Amrhein (1995), the most substantively important effects of the MAUP are seen not in the parameters of individual variables (mean, variance) but in the relationship among them.

¹ A change to any element of the variance–covariance matrix of the fitted model’s coefficient estimates (the vector of $\hat{\beta}$) due to changes in where unit borders are drawn will affect every element of $\hat{\beta}$ so long as the variable affected by the MAUP is correlated with any other variable in the regression analysis.

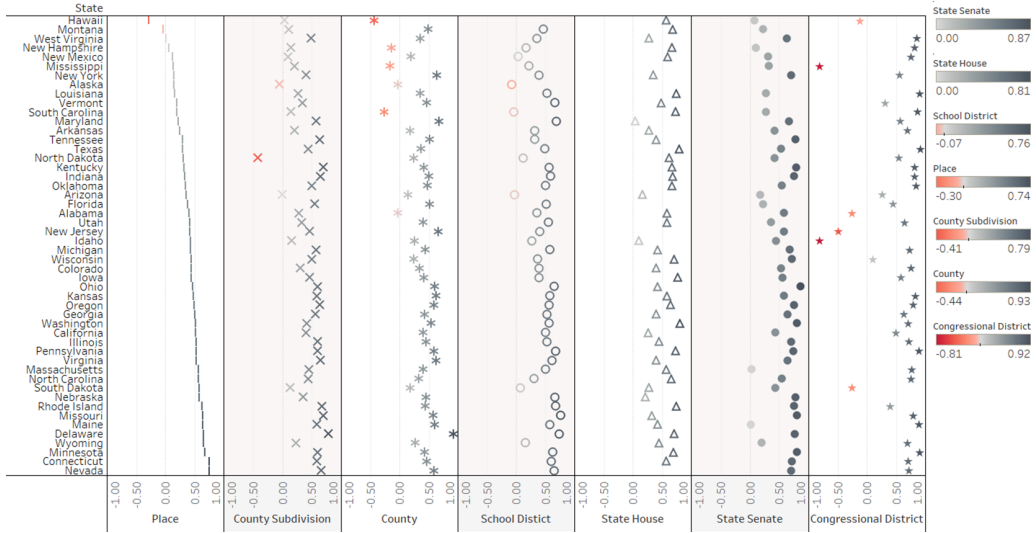


Figure 1. Modified areal units and changes in Pearson correlation.

Notes: Correlation between population density and 2020 Democratic presidential vote share across different administrative and political units. Data from Warshaw and Tausanovitch (2022).

The aggregate spatial units we use can vary in two ways—scale (or unit size) and zoning (or border location). Geographers therefore refer to two sub-problems of the MAUP: the scale sub-problem and the zoning sub-problem. Both cause the data generated from each aggregate unit on any variable of interest to change as different border mappings are used.² We explore below how both of these sub-problems affect scholarship in political science.

We can see the substantive effects of the MAUP in an example from American politics in Figure 1, which illustrates the correlation between population density and Democratic presidential vote share in 2020 across multiple spatial units within each state, using data from Warshaw and Tausanovitch (2022). Existing research on partisan polarization has emphasized the importance of the urban-rural divide, with population-dense locations more likely to support the Democratic party, and more Republican support in places with low population density (McKee and Teigen 2009); Tam Cho, Gimpel, and Hui 2013). But it is unclear how “location” should be operationalized to assess this relationship. We explore in Figure 1 various possibilities, using multiple spatial units available in the data set, including Census “place” (unincorporated places within counties, such as cities, towns, or villages), county subdivisions (sub-county governmental or administrative units, typically towns or townships), counties, school districts, state upper and lower legislative chambers, and Congressional districts. The first crucial observation is that while the relationship between population density and Democratic vote share is generally positive, its strength and consistency across states varies depending on the spatial unit chosen. Thus, the inferences we draw about this relationship, and patterns across states, would depend heavily on the unit chosen.

The relationships run the full spectrum from close to perfect negative correlation in the case of Idaho’s Congressional districts, to a near-perfect positive correlation in Minnesota or Louisiana’s districts. While some may argue these extreme results emerge from the gerrymandering of the borders of these

²The MAUP is distinct from the uncertain geographic context problem (Kwan 2012) which refers to the idea that simple mapping does not reflect people’s exposure to contextual influences which vary across time and space. It is also distinct from the ecological inference (EI) problem, which arises when scholars wish to explore a relationship at a particular spatial scale, but lack sufficiently disaggregated data to do so. While the EI problem is fundamentally a data problem, the MAUP is a theoretical one.

electoral districts, we also see dramatic cross-state variation in units such as counties that are long-standing and whose borders are not (currently) manipulated by politicians.

The most important implication of the values shown in Figure 1, however, is that we would get very different findings from a regression explaining vote choice if we include a population density variable operationalized at different scales. Crucially, our findings for individual-level predictors of vote choice may change depending on the spatial unit we choose to operationalize population density. Without a clear justification for why we should study the effect of density at one of the scales shown in Figure 1 and not the others, we would not know what inferences to draw about either this variable or about other determinants of vote choice included in our regression. More generally, the implication of the MAUP is that the spatial units we choose affect the findings we get for any study in which one or more variables capture a characteristic of a spatial unit that can plausibly be operationalized in multiple ways, so long as the inference is based on the co-variation among variables. In other words, the MAUP can impact both descriptive findings about relationships among variables like the one explored in Figure 1, and regression analyses.

3. Monte Carlo Simulations of the MAUP

In this section, we use simulations to show the core features of the MAUP and to explore its effects on regression. We simulate results on gridded data, which allow us to scale and zone the observations in standardized ways. Since gridded data are an increasingly common data structure, particularly in studying international relations (Tollefsen, Strand, and Buhaug 2012), these simulations reflect actual practice. We manipulate both the size of grids, to explore the scale sub-problem of the MAUP, and the location of their borders, to explore the zoning sub-problem.³

Figure 2 illustrates a simplified structure of the simulation with a 10×10 grid. We can think of each grid-cell as a “location” with values for several variables, which we label as “first variable,” “second variable,” and “third variable” respectively because their designations as the independent and dependent variables are arbitrary within the simulation. Each variable at each location can take any of three values {no event = 0, event occurred = 1, no observation = -}, thus the combination of the three variables that characterize a particular location can take one of 27 possible forms.⁴ We include non-observations to simulate a realistic data environment in which we do not have observations evenly distributed over space.

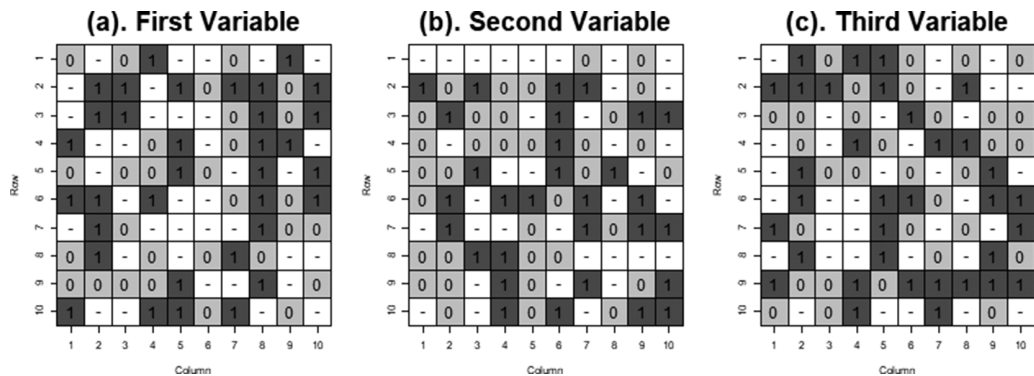


Figure 2. The distribution of random draws for three binary values.
Notes: The figure represents the same locations, with the value of their three variables (First Variable = DV, Second Variable = IV1, Third Variable = IV2).

³Note, however, that any change to unit size also implies a change in zoning, as the borders move with changes in scale.
⁴See the full list of possible mappings that cover the full grid in Section 1 of the Supplementary Material.

With this structure, expanded to a 100×100 grid, we draw 10,000 observations from a random sample, and then we aggregate the locations into areal units of between 4 and 100 cells (2×2 grid boxes to 10×10 grid boxes, respectively). Having this structure allows us to examine the impact of scale as we increase the number of locations within each unit from 4 to 16 to 25 to 100 grid cells and of zoning by shifting the areal units one grid-cell over to create a different zoning structure.⁵ We adjust the zoning by shifting our aggregated units eastward or southward starting from the northwest corner of our “map” of grid-cells. We treat our grid map as a “globe,” allowing for aggregations that extend beyond the right and bottom sides of the grid map to circle back to the left and top sides of the map. Summary statistics in Tables A.1–A.3 in the Supplementary Material show how the standard deviation and distribution of values are altered as we change unit size and border location.

We draw values (0, 1, -) from a random shuffle, using both a uniform probability distribution across these values and a non-uniform unknown probability distribution so that we can investigate whether the MAUP can be observed with both uniform data and more realistically distributed data. We (arbitrarily) set the first variable as our outcome of interest (dependent variable), and the other two as independent variables (IV1 and IV2). By design, we should find no or little correlation between our variables at the level of the unit location. But (as shown in Section 1.2 of the Supplementary Material) we find that each bivariate correlation varies in unsystematic but substantial ways across the different aggregate units.

In Figure 3, we show how shifts in zoning and scale affect multivariate logistic regression estimates from the simulated data. We plot the association between the two variables that we arbitrarily designated as IV1 and IV2, and the DV, shown as the coefficient estimate and the surrounding bar representing the standard error at the 95% confidence interval level. The association between IV1 and the DV is shown on the top figures and the association between IV2 and the DV is shown in the bottom figures. We divide our simulation into uniform (left column) and non-uniform (right column) sampling probabilities and show the results at increasing levels of aggregation and shifts in zoning to the east and the south, as indicated in the figure legend.

These results reinforce the conclusion that changes in scaling and zoning impact inference, and do so in unpredictable ways. In the uniform probability models, the initial coefficient estimate is zero, with some error, as expected for a random distribution draw with equal sampling probability. Yet there is some variation across mappings in the results. But we see more troubling results for the estimates based on a non-uniform distribution of data at the individual location level, shown on the right. For IV1, we see associations that vary in sign and significance as the scale and zoning of aggregate units change. For IV2, most aggregate mappings show a positive relationship significant at the 95% level, even as the association at the individual level is negative and significant. Notably, the aggregation effect cannot account for the findings we observe; by comparing 4(+1E Shift) to 100(+1E Shift) as indicated in the figure, for example, we can see that moving to a larger scale brings increased significance on our coefficient rather than the attenuated effect we would expect from aggregation.

3.1. Additional Simulations with Irregular Lattice Data

Our simulations above employ regular lattice data in which the units (grid-cells) are of uniform size and shape. This is common practice in simulation exercises examining the MAUP (Amrhein 1995; Lee *et al.* 2019). These findings suggest that the impact of the MAUP is widespread and significant. Yet there is reason to think that regular lattice data may understate its impact: Arbia (2012) shows that the effects of the MAUP are minimized if units are identical in terms of size, shape, and neighboring structure, and spatially independent of one another.

Thus, we should also examine the impact of the MAUP in simulations that better approximate common research settings in political science in which spatial units are irregular and may exhibit spatial dependence. Explorations using irregularly-shaped administrative and electoral unit boundaries

⁵We choose these grid sizes because they use all locations and thus rule out the possibility that any effects are driven by excluding locations in a particular mapping.

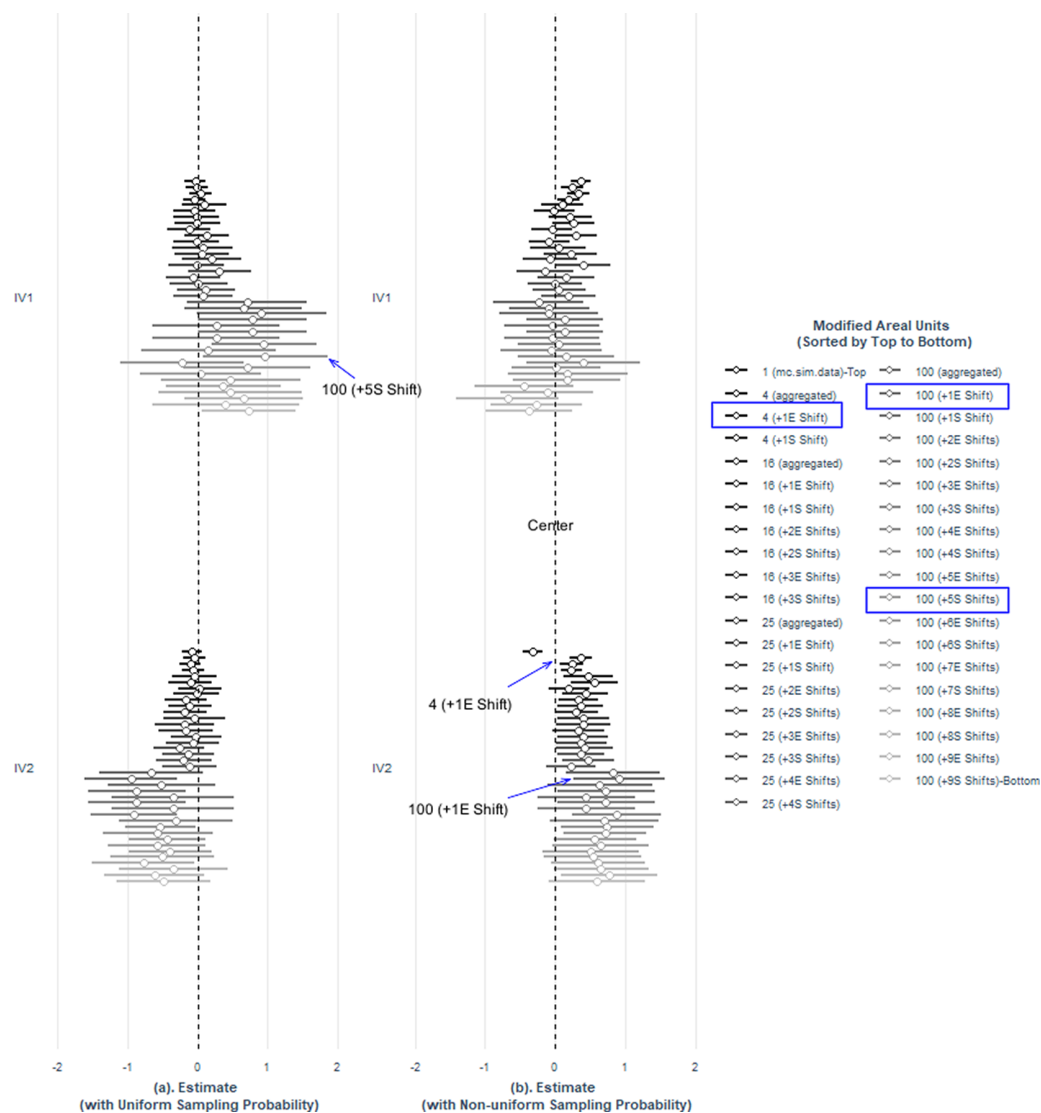


Figure 3. Coefficient estimates using fractional logistic regressions.
Notes: Each block of horizontal lines should be read from the top to the bottom to see the impact of increased scale and changed zoning. Hollow circles are average coefficient estimates. Horizontal lines are 95% confidence intervals. All regression estimates are based on Monte Carlo simulations with 10,000 observations in comparison to the smallest unit (1).

suggest that the consequences of the MAUP are important (Bisbee and Zilinsky 2023; Briant, Combes, and Lafourcade 2010; Hipp 2007).

In Section 2 of the Supplementary Material, we conduct simulations that demonstrate the MAUP in irregular lattice data, comparing regression analyses of predictors of Democratic presidential vote share in 2016 conducted at the county level to those conducted at the Congressional district level in California. We visualize results in the “baseline scenario” of OLS bivariate regression in Figure A.3 in the Supplementary Material, which shows that changing the boundaries of the units we use (from counties to Congressional districts) can have large effects on correlation and regression results.

Another important concern with the MAUP is that it may be exacerbated by characteristics of spatial data known to impact statistical analysis, such as spatial dependence (Arbia 2012). Dungan *et al.*

(2002, 631) show that spatial statistics such as the Moran's I are subject to the MAUP - their values change across spatial mappings. Lee *et al.* (2019) conduct a simulation of the effects of the MAUP as underlying spatial autocorrelation in the data varies, and find that as spatial autocorrelation increases, the variance of aggregate-level regression is subject to effects of the zonation sub-problem of the MAUP.

One might wonder, then, whether accounting for spatial properties of data in regression may mitigate the effects of the MAUP. To explore this, we build on the "baseline scenario" in Figure A.3 in the Supplementary Material by running multivariate regression models for spatial lag and spatial error dependence. We first conduct 100 simulations, drawing a random sample from real-world data measuring Democratic presidential vote share as a function of median household income and population share working from home across California's 58 counties and 53 Congressional districts (see notes from Figures A.2 and A.3 in the Supplementary Material for data sources). We then analyze these simulated data using spatial lag models with k-nearest, distance-decay, and queen contiguity weights (Figure A.4 in the Supplementary Material) and spatial error models with those three neighboring connectivity weights (Figure A.5 in the Supplementary Material). The results are stark: even after accounting for spatial properties of the underlying data, the impact of the MAUP remains significant. Results at the county and Congressional district levels, calculated with the same underlying individual data, are significantly different from each other in many cases. We suggest two implications from these analyses. First, they provide further backing for our suggestion that the real-world impact of the MAUP may be even greater than what we showed in our simplified, grid-based simulations. Second, and dovetailing with other scholars' claims, they also suggest that the MAUP is not simply an artifact of spatial properties of the data.

3.2. Replication and Reanalysis

To supplement our simulation results, we replicate and reanalyze two recent studies at alternative scales and zoning in Section 3 of the Supplementary Material. Our replications show that results are altered in important ways when reanalyzed using an alternative, plausibly valid, spatial mapping to operationalize key variables.

In Section 3.1 of the Supplementary Material, we show that changing the spatial unit to an alternative that is commonly employed in related studies produces different interpretations of the determinants of support for the Brexit referendum than those found by Colantone and Stanig (2018). Support for Brexit is explained as a function of regional economic hardship and increased exposure to immigration, which are both characteristics of spatial aggregate units. Existing studies employ several different spatial units to capture this phenomenon, most commonly the "NUTS" (classification of territorial units for statistics) level 2 and level 3 units, neither of which align with political constituencies in the United Kingdom (Schraff, Vergioglou, and Demirci 2023). When we use the commonly-employed NUTS level 2 data (Arnorsson and Zoega 2018; Los *et al.* 2017) rather than NUTS level 3 as employed by Colantone and Stanig (2018), we find that in some models the import share (their proxy for economic hardship) is not significantly associated with support for Brexit and immigrant arrivals are negatively and significantly associated with Brexit support. In Section 3.2 of the Supplementary Material, we show that changing the spatial unit for a moderator or contextual variable can impact results using a study from American politics on neighborhood effects (Anoll 2018).

4. How Widespread is the MAUP in Political Science Research?

If scholars provide incomplete justifications for their choice of spatial units, the MAUP may be a widespread threat to inference.⁶ This threat is only fully eliminated in research settings where there

⁶This is not to downplay the difficulty of accessing data at theoretically appropriate spatial units in political science. Data challenges are pervasive, particularly where political units change over time, or where there are multiple, overlapping but non-nested units. Our recommendation is for scholars to be explicit about its role in constraining their research designs.

are clear theoretical reasons to prefer one particular unit of analysis to all alternative possibilities, and the associated data are available. Under those conditions, scholars can rule out as irrelevant the results of analyses conducted at other spatial mappings, and set aside the MAUP as a threat to inference. More commonly the choice of a spatial unit is based on data availability, which may not have a theoretical foundation. We should therefore expect that the MAUP is a common potential threat to inference in empirical scholarship.

To demonstrate the breadth of research that is potentially affected, we reviewed all articles published in the *APSR* (2016–2020) and *AJPS* (2016–2020). Of the 325 papers contained in these volumes of *APSR*, we focus on the 221 that contained empirical analysis, setting aside those that are purely theoretical, normative, or methodological. Of these 221 papers, we found that 124 (56 percent) utilized an aggregate spatial unit in their analyses. Of these 124, we can identify a plausible alternative spatial unit for 53 (43 percent), meaning there is a substantial share of articles that is potentially vulnerable to the MAUP.⁷ Table A.8 in the Supplementary Material offers descriptions of the MAUP concerns we identify with the 53 articles and shows that only a small share of these papers provide any responses. The results were similar to our review of *AJPS*. We found that 114 out of 251 (45 percent) of articles used an aggregate spatial unit. Of these 114, we identified a potential alternative spatial unit in 59 articles (52 percent).

In Section 4 of the Supplementary Material, we also highlight examples of best practices from these articles to provide scholars with tools to address MAUP concerns in their research. Below, we identify three grounds on which validity can be justified: (1) the logic of the theory in terms of the spatial mapping at which treatment is assigned, (2) empirical evidence from the research setting about that spatial mapping, and (3) empirical or theoretical evidence about the spatial mapping associated with a particular theorized causal mechanism. We provide descriptions of articles that take each of these approaches.

5. Addressing the MAUP: Theoretical Precision and Empirical Strategies

A few responses to the MAUP have been proposed in existing political science scholarship. Darmofal (2015) points out that researchers with appropriate questions and data might use a spatial filtering approach to produce smoothed maps that reduce noise from the aggregation process. Yet as Openshaw (1981) and Darmofal (2015) point out, this approach requires the researcher to choose an areal unit, which may re-introduce MAUP concerns. It may, however, aid visualization and help to diagnose whether results depend on the areal unit selected.

Darmofal also proposes that scholars with point-level data construct “bespoke neighborhoods” that build out from individuals to include, say, the 50, 100, or 500 nearest households (MacAllister *et al.* 2001). This approach is related to one developed by Wong *et al.* (2012) that allows individuals to define the boundaries of their neighborhoods and thus accounts for the fact that people perceive and use space differently. These approaches are promising when scholars have individual-level data, but (as discussed throughout) scholars often lack the disaggregated data they require. Others suggest spatial transformation methods such as overlays, interpolation, and kriging (Gill 2021). Zhukov *et al.* (2023) provide a dataset and approach that matches data across misaligned spatial units.

In some research settings, scholars might follow Lee and Rogers (2019) and develop scale-invariant measures of quantities of interest. For example, scholars using administrative units such as US counties, which vary wildly in terms of land area, population, and population density, can scale them based on those features to better capture the intended concept. This approach, however, is often data intensive and depends on the availability of information needed to weight sub-units with theoretically important variables. Most fundamentally, this approach is only appropriate where a scale-invariant measure is of interest to the researcher.

⁷To provide a conservative estimate, we do not consider how errors are clustered (i.e., municipality clustered errors), or controls for geographic characteristics (i.e., rural designations for survey respondents), although these could have MAUP implications (Nemerever and Rogers 2021).

These ways of responding to the MAUP apply in certain settings but more practical and broadly applicable guidance is needed. We suggest that researchers can attenuate the concerns raised by the MAUP if they are attentive to the distinction between two types of threats to inference that it poses: challenges to validity and reliability of findings. In the remainder of this section, we advocate for validity as the most fundamental threat to inference raised by the MAUP and point to the role of theoretical and conceptual precision about spatial units as the most powerful response researchers have in the face of these concerns. We then discuss the utility of demonstrations that results are reliable across spatial mappings, which is a common response to the MAUP.

5.1. Validity

As demonstrated above, the concern at the heart of the MAUP is that for any given set of findings, results conducted using an alternative spatial mapping could differ in unpredictable ways, leading to different inferences. In this section, we argue that the only way to completely rule out this concern is to identify a unique spatial mapping at which analysis could be validly conducted. If no other spatial mapping is valid, it logically follows that scholars can set aside any results obtained from other mappings as irrelevant. Justifying the claim that a particular spatial mapping is valid for a given analysis, and that alternatives are not is fundamentally a theoretical exercise that rests on how the concepts in a given study are appropriately operationalized.

As we show in this section, it is not always possible to rule out all alternative spatial mappings for a given analysis. But the more tightly scholars can draw the bounds on this set of spatial mappings—the less modifiable their spatial units are—the more effectively concerns about the MAUP can be addressed. Where alternative spatial mappings are valid and cannot be ruled out as irrelevant, scholars can turn to strategies for demonstrating reliability that are discussed in the next section.

The case for which spatial mappings are and are not valid for a given analysis rests fundamentally on theoretical and conceptual grounds. In theoretical terms, as Arjona (2019, 218) writes, scholars should focus on “the spatial units that...we [as researchers] consider to be the relevant context in which a decision is made.” In some research settings, a sufficiently precise version of our theory would point to a particular spatial aggregate unit as the one at which a given causal variable is assigned.

In some instances, the concepts on which our theory rest are associated with particular spatial mappings. For example, an analysis of the effects of electoral institutions on vote choice would assign individuals to electoral districts. Alternatively, where theory identifies a unit of analysis with particular properties as the one at which a cause is assigned, or the one through which a theorized causal mechanism operates, scholars can work to identify the unit with those properties in their empirical setting. Finally, scholars can collect empirical information about which spatial units are appropriate for analysis in a given research setting.

In other instances, the concepts and theory central to a researcher’s investigation can go a long way to help rule out threats to inference posed by the MAUP. To see this, we can compare explanations for subnational patterns of violence in Mexico’s contemporary drug conflict, and observe that some can more effectively address this concern than others. Since homicide is what Harbers and Ingram (2019) call an “unbounded” process that could plausibly be studied using either administrative units or a range of other spatial mappings, in principle there are many plausible spatial units to which we might aggregate homicide data and explore variation. This section briefly explores how several prominent currents of scholarship use the concepts, causes, and mechanisms contained in their theories to justify the validity of their findings and rule out concerns related to the MAUP.

A first set of studies in this research area: Durán-Martínez (2017), Dell (2015), and Trejo and Ley (2020) link violence to patterns of municipal political control, pointing to various reasons why partisan alignment between municipal and state-level governments, among other factors, is associated with lower homicide rates. Since this variable is assigned at the municipal level, the only appropriate way to aggregate individual-level homicide data into spatial aggregates for analysis to test this hypothesis is to do so at the municipal level. Here, then, the logic of the theory helps us to determine which spatial units are valid and which are not.

Things are somewhat less straightforward for a second set of studies which focus on competition among criminal actors over territory (Calderón *et al.* 2015; Durán-Martínez 2015; Phillips and Ríos 2020). Here, the concept of criminal competition over territory does not provide sufficient information to identify the appropriate spatial mappings for analyses, or to rule out others. Scholars can instead rely on empirical information drawn from the Mexican case about this proposed cause. To the extent that they can make a convincing claim that influence over local-level policing is crucial to competition over shipping routes and drug-selling locations, they can justify the municipal level as the appropriate spatial scale at which to conduct their analyses.

Finally, in addition to empirical and conceptual information about independent variables, scholars can also draw on information about causal mechanisms to identify spatial units. We can see this by looking at studies that explore the impact of income inequality on violence (Enamorado *et al.* 2016; Phillips 2017). Since income inequality could plausibly be measured across any spatial aggregate that contained more than one individual or household, the concept does not help us to isolate a particular spatial mapping as valid, or to rule out alternatives. How, then, can these studies address concerns raised by the MAUP? Here, information about the causal mechanism linking inequality to violence can be informative. For example, Phillips (2017) argues that one mechanism linking income inequality to the formation of militias to combat drug cartels is its impact on the social control of landowners over municipal police and legal institutions. This mechanism, then, is appropriately tested by measuring inequality at the municipal level and exploring its impact on militia formation. By contrast, Enamorado *et al.* (2016) link inequality to homicide rates without specifying a clear causal mechanism; thus, neither concepts nor theory provides sufficient information to justify their findings from a municipal-level analysis as valid in the face of the threat to inference posed by the MAUP.

As the examples provided in the previous paragraphs suggest, relying on concepts and theory to justify a particular spatial mapping as valid and to rule out alternatives is a more effective response to the MAUP in some research settings than others. Where scholars can identify a particular mapping as uniquely valid and rule out alternatives, they can fully address the concerns raised by the MAUP. More commonly, our theories and the concepts they deploy can limit the set of plausibly valid spatial mappings, ruling out many possibilities while leaving others. This partial but not fully sufficient response to concerns about validity is, as we suggest below, precisely the setting in which information about the reliability of results may be informative.

Our instinct might be that smaller spatial units are in and of themselves more valid. Yet this is not the case: instead, going smaller artificially introduces variation and spatial autocorrelation, and going below the “true” spatial scale of assignment is analogous to ignoring clustering in treatment assignment, making us more likely to find an effect even when no “true” effect exists. Finally, as Cook and Weidmann (2022) point out, analyses using smaller aggregate units are more vulnerable to measurement error in the location of individuals.

Before turning to reliability, however, it is worth pointing out a possible situation researchers may confront, in which concepts and theory provide no guidance on how spatial mappings are to be chosen. Here, scholarship that relies on geographic grid cells (Ahmed and Stasavage 2020; Hager, Krakowski, and Schaub 2019; Harris and Posner 2019) poses a particularly interesting case. This choice of spatial unit is justified precisely because grid cells are exogenous to any political or social process; i.e., because they are arbitrary in scale and zoning. However, because there are no theoretical grounds for the choice of a particular grid mapping, it follows that the MAUP may pose serious concerns to the validity of the findings that these studies produce. As the simulation results above showed limited reliability across grid mappings, we suggest that there is reason to believe that the MAUP poses a threat to what we can learn from analyses using these spatial units.

5.2. Reliability

The previous discussion argued that theories and the concepts they build on can provide validity-based grounds for using particular spatial mappings in analyses and ruling out alternatives. But in many

Table 1. Robustness checks (Effects of IV2 on DV).

Robustness check	Initial Result			
	Uniform sampling		Non-uniform sampling	
	Not significant	Significant	Not significant	Significant
<i>Scale</i>				
Significant	0	0	4	9
Not significant	11	4	2	0
<i>Zoning</i>				
Significant	4	8	14	36
Not significant	64	4	16	14

Note: All regression outputs are based on the spatial units listed in Tables A.9 in the Supplementary Material (15 exercises for scaling) and A.10 (80 exercises for zoning). The number in each cell indicates the number of pairs of analyses. Bolded numbers denote the off-diagonal cells in which we observe that results are not reliable across alternative spatial units. All instances of unreliable results reflect variation in significance.

research settings, scholars cannot identify a uniquely valid mapping; more than one plausibly valid set of spatial units can be identified. Under these circumstances, scholars may turn to evidence about the reliability of their findings by relying on robustness tests showing that results from a given analysis hold at an alternative, plausibly valid spatial mapping (Jelinski and Wu 1996). In our review of papers in the *APSR* and *AJPS*, 14 papers took such an approach. For example, Hassell, Holbein, and Baldwin (2020) explore the effects of school shootings on electoral mobilization and outcomes at various spatial scales, confirming their initial findings that these events have no discernable electoral impact. We can think of aggregating our data to a new spatial unit and seeing whether we get consistent results as a robustness test that is focused on seeing whether “potentially arbitrary” assumptions in the specification of the baseline model could be replaced with “alternative plausible assumptions” (Neumayer and Plümper 2017).

Still, we know little about what to expect from this exercise: how powerful is a demonstration of reliability in response to the threat to inference posed by the MAUP? How likely are we to find results reliable across multiple spatial scales, or, conversely, that results are not consistent? To further explore these issues, we return to the grid-cell simulation exercise and conduct a series of reliability checks. In each one, we arbitrarily choose a particular spatial unit as the one at which our primary analysis is conducted, and our initial results generated, and compare the findings from that baseline to those at a larger scale or a different zoning. The details of this procedure are described in Sections 5.1 and 5.2 of the Supplementary Material. Table 1 summarizes the results of these reliability checks, showing the results for the independent variable that we arbitrarily labeled IV2.

Our main finding is that regression results based on our simulated data are quite often inconsistent across scale and zoning. Moving across scale, across zoning, and from uniform to non-uniform sampling has no systematic effect on reliability. This can be seen by comparing across the quadrants of Table 1. The diagonal connecting the bottom-left and top-right cells in each quadrant contains those instances where results were reliable across two spatial mappings; thus, the larger the share of overall instances within those cells, the more reliable our inferences across any given pair of spatial aggregates. In all four quadrants, we observe a sizable number of instances in which our findings are unreliable across a second spatial mapping. Indeed, we find that the extent of reliability is unpredictable even across variables within a single simulated dataset; this can be seen by comparing Table 1 (estimation results for IV2) to Table A.13 (estimation results for IV1). Moreover, the rate of reliability overall is fairly low: for example, as the off-diagonal entries in the lower-right quadrant of Table 1 shows, nearly 1/3 of all pairs of spatial mappings produced inconsistent results, and even the best-performing setting (in the lower-left quadrant of Table A.13) finds unreliability in more than 10% of cases.

As the set of plausible spatial mappings gets larger, the likelihood that results will remain consistent across all of them declines. It logically follows, then, that in research settings where only a small number of spatial mappings are plausibly valid, a small number of robustness checks with different mappings will provide a great deal of information about the reliability of findings. By contrast, where theory and concepts provide little guidance about the spatial units to be used, we may learn less about reliability from conducting a single re-analysis at an alternative scale. The implication is that these robustness checks are more informative where scholars can also argue that the set of plausible alternative units is small. Once the set of plausible alternative mappings can be shown to be small, demonstrations of reliability can indeed be an effective response to concerns about the MAUP.

6. Conclusions

We began our discussion by suggesting that despite the ubiquity of spatial units in political science scholarship, little attention has been paid to the MAUP, and virtually none outside the study of racial context in American politics (Tam Cho and Baer 2011). Through simulations and re-analysis of prominent scholarship, we show that results are altered in important ways when the boundaries of spatial units are re-drawn. Moreover, we find that these changes can be substantively and theoretically important: core findings may be specific to the spatial units used and they are not always replicated when spatial borders are changed. We show that these effects can be observed not only when core variables are spatial, but also where moderating variables are characteristics of spatial aggregates. Thus, our paper makes the case that a wide range of studies in political science are potentially subject to the inferential threat posed by the MAUP. More work remains to be done to develop ways to mitigate its effects.

We have proposed that scholars should mitigate the impact of the MAUP on the validity on their findings where possible by connecting their choices of spatial units to theory and concepts in well-considered and explicit ways. Where there are multiple theoretically plausible spatial units, and scholars cannot therefore rule out the threat to inference the MAUP poses on the grounds of validity of spatial units, scholars can instead ameliorate concerns about the reliability of results across spatial units by conducting robustness checks using alternative plausible spatial aggregates. These suggestions are both practical and broadly applicable across research settings, and provide some ways for scholars to move forward and address the concerns raised by the MAUP.

Data Availability Statement. Replication code and data for this article has been published in Dataverse at <https://doi.org/10.7910/DVN/L18QRY> (Rogers, Lee, and Soifer 2024).

Supplementary Material. For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2025.2>.

References

- Ahmed, A. T., and D. Stasavage. 2020. "Origins of Early Democracy." *American Political Science Review* 114 (2): 502–518.
- Amrhein, C. 1995. "Searching for the Elusive Aggregation Effect." *Environment and Planning A* 27 (1): 105–119.
- Anoll, A. P. 2018. "What Makes a Good Neighbor?" *American Political Science Review* 112 (3): 494–508.
- Arbia, G. 2012. *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, vol. 14, Berlin: Springer Science & Business Media.
- Arjona, A. 2019. "Subnational Units, the Locus of Choice, and Concept Formation." *Inside Countries: Subnational Research in Comparative Politics* 214: 214–242.
- Arnorsson, A., and G. Zoega. 2018. "On the Causes of Brexit." *European Journal of Political Economy* 55: 301–323.
- Bisbee, J., and J. Zilinsky. 2023. "Geographic Boundaries and Local Economic Conditions Matter for Views of the Economy." *Political Analysis* 31 (2): 288–294.
- Briant, A., P.-P. Combes, and M. Lafourcade. 2010. "Dots to Boxes." *Journal of Urban Economics* 67 (3): 287–302.
- Calderón, G., G. Robles, A. Díaz-Cayeros, and B. Magaloni. 2015. "The Beheading of Criminal Organizations and the Dynamics of Violence in Mexico." *Journal of Conflict Resolution* 59 (8): 1455–1485.
- Colantone, I., and P. Stanig. 2018. "Global Competition and Brexit." *American Political Science Review* 112 (2): 201–218.
- Cook, S. J., and N. B. Weidmann. 2022. "Race to the Bottom: Spatial Aggregation and Event Data." *International Interactions* 48 (3): 471–491.

- Darmofal, D. 2015. *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press.
- Dell, M. 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105 (6): 1738–1779.
- Dungan, J. L., et al. 2002. "A Balanced View of Scale in Spatial Statistical Analysis." *Ecography* 25 (5): 626–640.
- Durán-Martínez, A. 2015. "Drugs Around the Corner." *Latin American Politics and Society* 57 (3): 122–146.
- Durán-Martínez, A. 2017. *The Politics of Drug Violence*. New York: Oxford University Press.
- Enamorado, T., L. F. López-Calva, C. Rodríguez-Castelán, and H. Winkler. 2016. "Income Inequality and Violent Crime." *Journal of Development Economics* 120: 128–143.
- Fotheringham, A. S., and D. W. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." *Environment and Planning A* 23 (7): 1025–1044.
- Gill, J. 2021. "Measuring Constituency Ideology using Bayesian Universal Kriging." *State Politics & Policy Quarterly* 21 (1): 80–107.
- Hager, A., K. Krakowski, and M. Schaub. 2019. "Ethnic Riots and Prosocial Behavior." *American Political Science Review* 113 (4): 1029–1044.
- Harbers, I., and M. C. Ingram. 2019. "Politics in Space." In *Inside Countries*, edited by A. Giraudy, E. Moncada, and R. Snyder. New York: Cambridge University Press.
- Harris, J. A., and D. N. Posner. 2019. "(Under What Conditions) Do Politicians Reward Their Supporters?." *American Political Science Review* 113 (1): 123–139.
- Hassell, H., J. Holbein, and M. Baldwin. 2020. "Mobilize for Our Lives?." *American Political Science Review* 114 (4): 1375–1385.
- Hipp, J. R. 2007. "Block, Tract, and Levels of Aggregation." *American Sociological Review* 72 (5): 659–680.
- Jelinski, D. E., and J. Wu. 1996. "The Modifiable Areal Unit Problem and Implications for Landscape Ecology." *Landscape Ecology* 11 (3): 129–140.
- Kwan, M.-P. 2012. "The Uncertain Geographic Context Problem." *Annals of the Association of American Geographers* 102 (5): 958–968.
- Lee, D. W., and M. Rogers. 2019. "Measuring Geographic Distribution for Political Research." *Political Analysis* 27 (3): 263–280.
- Lee, S.-I., M. Lee, Y. Chun, and D. A. Griffith. 2019. "Uncertainty in the Effects of the Modifiable Areal Unit Problem under Different Levels of Spatial Autocorrelation." *International Journal of Geographical Information Science* 33 (6): 1135–1154.
- Los, B., P. McCann, J. Springford, and M. Thissen. 2017. "The Mismatch Between Local Voting and the Local Economic Consequences of Brexit." *Regional Studies* 51 (5): 786–799.
- MacAllister, I., R. Johnston, C. Pattie, H. Tunstall, D. Dorling, and D. Rossiter. 2001. "Class Dealignment and the Neighbourhood Effect." *British Journal of Political Science* 31 (1): 41–59.
- McKee, S., and J. Teigen. 2009. "Probing the Reds and Blues." *Political Geography* 28 (8): 484–495.
- Nemerever, Z., and M. Rogers. 2021. "Measuring the Rural Continuum in Political Science." *Political Analysis* 29 (3): 267–286.
- Neumayer, E., and T. Plümper. 2017. *Robustness Tests for Quantitative Research*. Cambridge: Cambridge University Press.
- Openshaw, S. 1981. "The Modifiable Areal Unit Problem." In *Quantitative Geography: A British View*, edited by N. Wrigley and R. J. Bennett, 60–69. Boston, MA: Routledge and Kegan Paul.
- Openshaw, S. 1984. "Ecological Fallacies and the Analysis of Areal Census Data." *Environment and Planning A* 16 (1): 17–31.
- Phillips, B. J. 2017. "Inequality and the Emergence of Vigilante Organizations." *Comparative Political Studies* 50 (10): 1358–1389.
- Phillips, B. J., and V. Ríos. 2020. "Narco-Messages." *Latin American Politics and Society* 62 (1): 1–24.
- Rogers, M., D. W. Lee, and H. Soifer. 2024. "Replication Data for: Lee, Rogers, Soifer. 'The Modifiable Areal Unit Problem in Political Science'." <https://doi.org/10.7910/DVN/L18QRY>. Harvard Dataverse, V1.
- Schraff, D., I. Vergioglou, and B. B. Demirci. 2023. "The European NUTS-level Election Dataset." *Party Politics* 29 (3): 570–579.
- Tam Cho, W., J. Gimpel, and I. S. Hui. 2013. "Voter Migration and the Geographic Sorting of the American Electorate." *Annals of the Association of American Geographers* 103 (4): 856–870.
- Tam Cho, W. K., and N. Baer. 2011. "Environmental Determinants of Racial Attitudes Redux." *American Politics Research* 39 (2): 414–436.
- Tollefsen, A. F., H. Strand, and H. Buhaug. 2012. "PRIO-GRID." *Journal of Peace Research* 49 (2): 363–374.
- Trejo, G., and S. Ley. 2020. *Votes, Drugs, and Violence: The Political Logic of Criminal Wars in Mexico*. New York: Cambridge University Press.
- Warsaw, C., and C. Tausanovitch. 2022. "Subnational Ideology and Presidential Vote Estimates (v2022)." Harvard Dataverse, V1. <https://doi.org/10.7910/DVN/BQKU4M>.
- Wong, C., J. Bowers, T. Williams, and K. D. Simmons. 2012. "Bringing the Person Back In." *The Journal of Politics* 74 (4): 1153–1170.
- Zhukov, Y. M., J. S. Byers, M. A. Davidson, and K. Kollman. 2023. "Integrating Data Across Misaligned Spatial Units." *Political Analysis* 32: 17–33.