Uncertainty in grid data: a theory and comprehensive robustness test

Akisato Suzuki

Accepted: 4 October 2022 / Published online: 8 November 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
This methodological note makes two novel contributions to spatial political and conflict research using grid data. First, it develops a methodological theory of how uncertainty specific to grid data affects inference. Second, it introduces a comprehensive robustness test on sensitivity to this uncertainty, implemented in R. The uncertainty stems from (1) establishing the correct size of grid cells, (2) deciding the correct locations where the dividing lines of grid data are drawn, and (3) a greater effect of measurement errors due to finer grid cells. The proposed test diversifies grid cell sizes, by aggregating original grid cells into a multiple of these grid cells. The test also varies the locations of the diving lines, by using different starting points of grid cell aggregation (e.g., starting the aggregation from the corner of the entire map or one grid cell of the original size away from the corner). I apply the test to Theisen et al. (Int. Secur. 36(3):79–106, 2011), who utilize the PRIO-GRID data (Tollefsen et al., J. Peace Res. 49(2):363–374, 2012), to substantiate its use.

Keywords Geocoding · Geo-referenced · Spatial data · Grid · Uncertainty · R package

1 Introduction

Political and conflict research have been developing geo-referenced data and analytical techniques for such data (e.g., Lee and Rogers 2019; Kikuta 2022; Pickering 2016; Schutte and Donnay 2014; Shaver et al. 2019; Sundberg and Melander 2013; Tollefsen et al. 2012). Geo-referenced data enable disaggregating macro-level units into smaller ones. This is useful when it is expected that a causal relationship is local. For example, drought in one location might have an impact on the probability of political violence in that particular location, but not on the probability of political violence in other locations, even within the same country (Theisen et al. 2011; Von Uexkull et al. 2016).

One of the universal formats in geo-referenced data is grid cells (for details on grid data, see Pickering 2016; Tollefsen et al. 2012). In grid data, the entire map of interest (e.g., the African continent) is divided into artificial grid cells of a certain size. The result
is an artificial, apolitical division of the map. Each observation corresponds to a grid cell. As grid cells are typically created in a smaller size than the entire territory of the average country, grid data allow researchers to focus on the local relationship between a causal factor and an outcome (for examples of applied research, see Buhag et al. 2011; Linke et al. 2017; O’Loughlin and Witmer 2012; Ruggeri et al. 2017; Schutte and Weidmann 2011; Theisen et al. 2011; Wood and Sullivan 2015).

However, using grid data comes with a cost. It brings additional sources of uncertainty in statistical and causal inference, because of the necessity to specify grid cells. I identify three such sources.

The first source is ambiguity in the correct size of grid cells. It is often difficult to theoretically pinpoint the only correct size (if any). Is the correct size 3000 km², 12,000 km², or 28,000 km²? What should we do, if there is more than one plausible grid cell size?

The second is ambiguity in the correct locations where the dividing lines of grid data are drawn. In theory, it is possible to draw the first dividing line from any point of latitude and longitude. What are the correct locations of the dividing lines that will make each grid cell correctly include causally related variables and separate causally unrelated variables?

The third is the impact of measurement errors as a function of grid cell sizes. All else equal, finer grid cells are more prone to measurement errors; an event recorded in a grid cell might have actually taken place in a nearby grid cell. This is because it is often difficult to identify the exact zone of the phenomenon of interest, practically or conceptually (Kikuta 2022). An illustration of a practical challenge is identifying the exact location of political violence in a remote area that is problematic for researchers or journalists to reach. On a conceptual level, the difficulty is exemplified by the challenges of drawing a clear line between a conflict area and a non-conflict area within a country.

As known, a measurement error affects statistical and causal inference (Pearl 2010), while the first two sources of uncertainty specific to grid data are related to the so-called Modifiable Areal Unit Problem (MAUP): the choice of a geographic unit affects inference (Lee and Rogers 2019; Lee et al. 2020; Openshaw 1983; Soifer 2019). For political geographic units, this problem can sometimes be addressed based on a theoretical justification; for example, the norm of national sovereignty may be a good justification for why nation-states are a plausible unit for the study of foreign policy. In grid data, on the other hand, it is difficult to find the same kind of theoretical justification (Soifer 2019, 107–108). This is because grid cells are by definition apolitical geographic units, i.e., “fixed in time as well as space and are insensitive to political boundaries and developments” (Tollefsen et al. 2012, 365).

This methodological note makes two novel contributions. First, it develops a methodological theory of exactly how each of the three sources of uncertainty specific to grid data affects inference. While the existing literature sometimes explores empirically how changing the size of grid cells affects statistical inference (e.g., Ito and Hinkkainen Elliott 2020; Lee et al. 2020; Linke et al. 2017; Schutte and Weidmann 2011), this methodological note provides a theoretical foundation for why it is necessary to use different grid cell specifications for robustness tests. A specific choice of grid cell specification affects statistical and causal inference (Soifer 2019, 95). Thus, results from a single grid cell specification must not be taken for granted (unless that specification is the only theoretically conceivable way, which may be unlikely in social science), without examining the sensitivity to different specifications.

Second, this methodological note proposes a comprehensive robustness test to accommodate the three sources of uncertainty specific to grid data (for a general principle of robustness tests, see Neumayer and Plüumper 2017). The proposed test diversifies grid cell
sizes, by aggregating original grid cells into a multiple of these grid cells.¹ In the process of aggregation, the proposed test also allows for different starting points of grid cell aggregation, to vary the locations where the dividing lines of aggregated grid data are drawn. For example, if the sides of grid cells are to be doubled, the starting point of the aggregation can be either from the corner of the entire map or one grid cell away from the corner. This creates a shift in the dividing lines between the grid cells of the aggregated size, by one grid cell of the original size. Finally, the proposed test can address measurement errors, in that if the results are robust regardless of plausible aggregations, it suggests the measurement errors are not as significant as to threaten a substantive interpretation.

This methodological note is also accompanied by a new R package (co-developed with Johan A. Dornschneider-Elkink) to make the robustness test easy to implement. While Pickering (2016) introduces software to generate the dataset of grid cells of any size, which is useful to create new grid data based on one’s research purpose, our R package can be used for any existing grid data as long as the row and column numbers of the grid cells are available. This is useful particularly when the original source of data is unavailable and only a grid data format is available (e.g., in the case of replication analysis). The R package also enables streamlining the robustness test efficiently, without the necessity of creating multiple datasets of different grid cell sizes in advance.

To exemplify the use of the robustness test, I apply it to a previous study that, using the PRIO-GRID dataset (Tollefsen et al. 2012), finds “little evidence” that drought increases the likelihood of the onset of civil armed conflict (Theisen et al. 2011, 81). Varying the size of grid cells and the starting point of grid cell aggregation, I find greater uncertainty in the estimates than the original study showed.

In the rest of the note, I first discuss further how the three sources of uncertainty in grid data affect inference. Second, I explain how aggregating grid cells can serve as a comprehensive robustness test on possible sensitivity in statistical estimation using grid data. Third, I present the applied example of the robustness test. I conclude with final remarks.

2 A methodological theory on how the uncertainty specific to grid data affects inference

In this section, I explicate the three sources of uncertainty specific to grid data and how they affect inference, in greater detail. The first is the uncertainty over what size of grid cells we should use to divide the entire map. Practical limitations often define the smallest possible grid cell size. But there are, in theory, numerous ways to aggregate the smallest grid cells because distance is a continuous measure. Should one side of the cell be 50, 51, 100, or 125 km? Practically, we may think in a more discrete way, such as 50, 100, 150 km, and so on. No matter how we aggregate grid cells, the point is that the size of grid cells directly affects statistical and causal inference.

Figure 1 is a crude but illustrative example—an approach similar to the explanation by the canonical MAUP study (Openshaw 1983), but fine-tuned for grid data in particular. In the left panel, the entire map is divided into four grid cells, while in the right panel, it is divided into sixteen grid cells. The locations indicated by “x” and “y” represent a value of

¹ Depending on the setting, after the aggregation a few grid cells might remain unaggregated or partially aggregated because the aggregated size of grid cells might not evenly divide the entire map. This point is discussed further in the next section.
one for the binary treatment and outcome variables $X$ and $Y$ (e.g., the presence of drought and that of political violence); the locations not marked “$x$” or “$y$” mean a value of zero for these variables.\footnote{In theory, both “$x$” and “$y$” can be observed exactly in the same location, but such a case is not included here for legibility.} Let us assume that $X$ has a positive causal effect on the probability of $Y$; thus, in data, a positive statistical correlation between them should be observed. Let us also assume that the data record a value of one for these binary variables per grid cell, if the presence of the treatment and that of the outcome are observed in any location within a grid cell (e.g., in the left panel of Fig. 1, the top left cell has $X = 0, Y = 1$). The empty grid cells are therefore where neither the treatment nor the outcome is present in any location within the grid cell, i.e., $X = 0, Y = 0$ (e.g., in the right panel of Fig. 1, there are twelve such grid cells). The shaded cells mean either $X = 1, Y = 1$ or $X = 0, Y = 0$. If we compute the correlation coefficient between $X$ and $Y$ based on these grid data, they are negatively correlated in the left panel ($-0.33$) with a greater p-value ($0.67$), while positively correlated in the right panel ($0.59$) with a smaller p-value ($0.02$). The difference is, of course, an artifact of grid cell specifications, because the entire map is the same in both panels. In other words, the size of grid cells used to divide the entire map influences statistical and causal inference.

Causal theories in political and conflict research are often not sufficiently specific to pinpoint the correct size of grid cells. Instead, they might suggest the range of plausible grid cell sizes. The implications of this uncertainty are as follows. Let us assume that there is the latent causal mechanism that justifies a particular size of grid cells, given how far a causal effect can go to affect an outcome. Sizing grid cells larger than such a mechanism warrants would risk an ecological fallacy, creating a statistical correlation between the treatment and outcome variables in the areas where the actual causal effect does not exist. Alternatively, grid cells that are smaller than warranted by the causal mechanism would risk the opposite: causing an absence of the statistical correlation between the treatment and outcome variables in the areas where the effect actually does exist.

The above issue is further complicated by the second source of uncertainty specific to grid data. Inference could also be affected by the ambiguity in the correct locations for

### Fig. 1
How different grid cell sizes affect statistical and causal inference. $x$: a binary treatment variable $X$ taking a value of one; $y$: a binary outcome variable $Y$ taking a value of one. Locations not marked “$x$” or “$y$”: both binary variables taking a value of zero. Shaded cells: either $X = 1, Y = 1$ or $X = 0, Y = 0$. 

---

\footnote{In theory, both “$x$” and “$y$” can be observed exactly in the same location, but such a case is not included here for legibility.}
drawing the dividing lines between the grid cells; this would be the case even if the size of grid cells could be justified theoretically. This ambiguity is impossible to solve theoretically, if grid cells are supposed to be apolitical and, therefore, atheoretical units. In other words, any dividing lines that are drawn by grid cells must not have any political or theoretical meaning attached, unlike national borders, which represent the political/theoretical meaning of sovereignty.\(^{3}\)

In terms of Fig. 1, let us assume that we could theorize that the treatment variable can have a causal effect only on the immediate neighboring area, and also that an area of $3 \times 3$ units is the correct areal size from this theoretical point of view, provided that an area marked by “x” represents a measurement unit of one. Even with these assumptions, we could not create grid data like those in Fig. 2. This is because the dividing lines are defined by the theoretical causal effect of the treatment and, therefore, no longer apolitical. Defining the locations of dividing lines theoretically would also require determining theoretically how to divide the map in such a way that the \textit{counterfactual} states are correctly specified. In the running example, a number of areas sized $3 \times 3$ without “x” must be defined such that these areas correctly capture the causal relationship between \(X\) and \(Y\) if “x” were observed.\(^{4}\) Of course, spatial data created in the above way might be plausible to use to examine the spatial effect of the treatment on the outcome. But they are not what grid data are supposed to be.

How are statistical and causal inference affected by a possible variation in the locations of the dividing lines? Let us keep assuming that the treatment variable can have a causal effect only on the immediate neighboring area. Dividing the current example map by this theoretically correct size of grid cells results in two difficulties.

First, there are grid cells in the periphery that cannot have the same size as the other grid cells, if the theoretically correct size of grid cells ($3 \times 3$, assuming that a location marked by “x” represents a measurement unit of one) divides the entire map of the current example. Figure 3 shows four possible ways to divide the map. If the dividing lines of grid

\(^{3}\) Note the difference between the correct (range of) grid cell size, which should be justified theoretically, and the locations where the dividing lines of grid data are drawn, which cannot be justified theoretically.

\(^{4}\) This requirement is in addition to the general requirement for the identification of the average treatment effect: both areas with “x” and those without “x” are (conditionally) exchangeable as treatment/control groups. See, for example, Hernán and Robins (2020).
data are supposed to be apolitical and atheoretical, we cannot decide which way is correct from a theoretical point of view. It would indeed be unlikely that grid cells sized in a theoretically correct way can perfectly divide a real map, which is much more complex than the square used here.

Second, depending on how lines are drawn to place grid cells of the correct size, the statistical correlation between $X$ and $Y$ can change. For example, there are two cases where the presence of $X$ is causally related to the presence of $Y$ from a theoretical point of view (one on the top right corner and the other on the middle right corner). Only the left top and bottom panels of Fig. 3 correctly capture both causal relationships, given this specific configuration of the locations of the variables. However, again, if the dividing lines of grid data are supposed to be apolitical and atheoretical, we cannot decide the correct location of each grid cell, based on the location of the phenomenon of interest. In short, the requirement

![Fig. 3](image)

Fig. 3 How the variation in the locations of the dividing lines of grid cells affects statistical and causal inference. $x$: a binary treatment variable $X$ taking a value of one; $y$: a binary outcome variable $Y$ taking a value of one. Locations not marked “$x$” or “$y$”: both binary variables taking a value of zero. Shaded cells: either $X = 1, Y = 1$ or $X = 0, Y = 0$.

5 For simplicity, I focus on the case of grid cells with “$x$” (i.e., where the presence of $X$ is observed), but the same logic applies to counterfactual states (i.e., whether the grid cells would correctly capture the causal relationship between $X$ and $Y$ if “$x$” were observed).
that the dividing lines of grid data must be apolitical and atheoretical makes it even more challenging to use only one type of grid cell specification.

The third source of uncertainty in grid data is a measurement error. For example, the codebook of the PRIO Conflict Site (Hallberg 2011, 3) states:

A drawback with circular conflict zones is that they cover more territory than is actually affected by the conflict, including territories of neighboring countries. Due to the nature of armed conflict it may be impossible to gain information on the exact locations of armed encounters, occupied territories, and rebel bases.

A measurement error over the location of an event poses a dilemma about how small the size of grid cells should be. The smaller the grid cells, the more detailed micro-phenomena we can examine, but, at the same time, the greater the impact of a measurement error.

Figure 4 exemplifies why a smaller grid cell size might make inference more susceptible to a measurement error on the location of an event. “x” is a binary treatment variable taking a value of one, “(y)” is the true location of an outcome event but unobserved, and “y” is an observed, therefore measured binary outcome variable taking a value of one. “y” is recorded in a location different from “(y),” meaning it is measured with an error. If the size of grid cells is set smaller, as indicated by the dashed lines, the causal relationship between X and Y cannot be observed as a statistical correlation because of the measurement error. The true state would allow such a correlation to be observed under this grid cell specification, if “y” were measured without an error. Meanwhile, if a larger grid cell size indicated by the solid line (i.e., the aggregate of the four small cells) is used, the causal relationship between X and Y is observed as a statistical correlation even though “y” is measured with an error.6

3 A comprehensive robustness test by the aggregation of grid cells

To summarize the previous section, there are three sources of uncertainty in grid data: the size of grid cells, the locations where the dividing lines of grid data are drawn, and a greater effect of measurement errors due to finer grid cells. A comprehensive robustness test for grid data needs to capture all three sources of uncertainty. To do so, I propose (1) aggregating grid

---

6 Again, for simplicity, I focus on the case of grid cells with observed “x,” but the same logic applies to how to define grid cells for the correct specification of counterfactual states.
cells into several different sizes, and (2) varying the starting point of aggregation to change the locations of the dividing lines.

The size of grid cells can be changed, by aggregating original grid cells into a multiple of these grid cells. A theory might be unable to determine the single plausible size of grid cells, but it could determine a plausible range of grid cell sizes (Soifer 2019, 104). To reflect such a range, we can create the separate datasets of the plausible grid cell sizes through aggregation.

Grid cell aggregation also allows us to change the locations where dividing lines are drawn to place grid cells, even when we are using existing grid data that have predefined positions of grid cells. We can shift the starting point of aggregation. For example, in Fig. 5, if we aggregate the data by multiplying the sides of the original grid cell size by three and shift the aggregation starting point from the north-west to the south-east by zero, one, and two, we can create three different versions of the data. Note that not every grid cell has every side being the triple of the side of the original grid cell size, because the aggregated grid cell size cannot divide the entire map evenly, even if the size were theoretically correct (see the previous section). This is why it is necessary to change the starting point of aggregation for robustness tests.

By doing grid cell aggregation and shifting of the starting point of aggregation, measurement errors can also be accommodated. This is because, if the results are robust regardless of plausible aggregations, it suggests the measurement errors are not as significant as to threaten a substantive interpretation.

In short, if the estimates do not differ among the datasets of different and theoretically plausible grid cell specifications, it implies the estimates are robust. If they do differ, it encourages the researcher to further investigate why they differ. Is it because some of the grid cell sizes, and/or some of the ways to draw the dividing lines, are less appropriate than others? Is it because key variables are prone to measurement errors? Whether the results remain robust or not, the comprehensive robustness test provides more information than an analysis using the dataset of only one grid cell specification. The following section exemplifies the use of the test.

Fig. 5 Example of grid cell aggregation with shifting of the starting point. The single lines denote the original grid cell size. The double lines denote the aggregated grid cell size. x: a binary treatment variable \(X\) taking a value of one; y: a binary outcome variable \(Y\) taking a value of one. Locations not marked “x” or “y”: both binary variables taking a value of zero. Shaded cells: either \(X = 1, Y = 1\) or \(X = 0, Y = 0\)
4 Application

In this section, I apply the proposed robustness test to Theisen et al. (2011). Contrary to the conventional wisdom, they found “little evidence” that drought increases the likelihood of the onset of civil armed conflict (Theisen et al. 2011, 81). They employed statistical models using the grid data of the African continent from 1960 to 2004, based on the PRIO-GRID data (Tollefsen et al. 2012).

I aggregate the grid cells in the original data from the north-west edge of the map. In the aggregation process, each variable used in the original model must also be aggregated in a theoretically consistent way (for details of the variables used, see Theisen et al. 2011, 93–96). The cell-specific binary variables (the onset of conflict, drought, marginalized ethnic groups, and the capital city) are aggregated to take a value of one, if at least one of the grid cells to be aggregated has a value of one. The cell-specific continuous variables (a grid cell’s distance to a border and a grid cell’s population) are aggregated to take the mean of the observed values in the set of grid cells to be aggregated. The country-specific variables (infant mortality rate, an autocracy-democracy scale, and conflict history) are aggregated to take the mean of the observed values in the set of grid cells to be aggregated. In other words, if all these grid cells belong to one country, the mean is the same as the value of each of these cells. If the grid cells to be aggregated belong to different countries, the mean is the average of the values between/among these countries. If a variable takes a missing value for any of the grid cells to be aggregated to a larger one, the cells containing missing values are ignored when the aggregated value is computed; if all grid cells to be aggregated to a larger one have missing values, the aggregated value is also treated as missing.

The original grid cell size is approximately $55 \, \text{km} \times 55 \, \text{km} = 3025 \, \text{km}^2$, the one defined by the PRIO-GRID data (Tollefsen et al. 2012). In the aggregation process, it is important to consider the theoretically plausible maximum size. In particular, when the treatment and outcome variables are binary and their maximum value (which is 1) is used for aggregated grid cells, aggregating grid cells into an arbitrarily large size almost surely results in a positive statistical correlation between them. If the maximum grid cell size is not theoretically justified, the statistical correlation cannot be considered causal.

I multiply the sides of the original grid cell of Theisen et al. (2011) by two to six; thus, the plane of the aggregated grid cell becomes the square of the side multiplied by one of these values. In other words, the smallest aggregated cell size is $(55 \, \text{km} \times 2)^2 = 12,100 \, \text{km}^2$, while the largest is $(55 \, \text{km} \times 6)^2 = 108,900 \, \text{km}^2$. The longest straight line within the square is the diagonal, $a\sqrt{2}$, where $a$ is the length of one side; the diagonal of $108,900 \, \text{km}^2$ is approximately $467 \, \text{km}$.

I reason that the maximum grid cell size of $108,900 \, \text{km}^2$ is theoretically plausible, as follows. If a car moves 20 km per hour on average, it takes 23.35 hours to go all the way through the diagonal of 467 km. In reality, a car may not be able to go straight through the diagonal and at the constant rate of 20 km per hour. Nevertheless, even if we assume it takes three times longer, it is still 70.05 hours to cover from one corner to the other corner within the grid cell of $108,900 \, \text{km}^2$. In the data, the temporal unit of analysis is years. People might move over that distance within the time period of one year, if they were affected by drought and decided to move for water and survival. Such migration could increase the

More specifically, the test is applied to Model 2 in Theisen et al. (2011). Their replication dataset can be obtained from the Peace Research Institute Oslo website at https://www.prio.org/publications/5109 (accessed on January 7, 2022).
likelihood of conflict (Reuveny 2007). Rebel groups could also move as long a distance, if they were interested in recruiting drought-affected people whose economic difficulties might be great enough to make it less costly to participate in armed rebellion for economic gains or survival (see Theisen et al. 2011, 86). Finally, even though a few countries are smaller than 108,900 km² in terms of their land mass (e.g., Rwanda or Burundi), rebel groups often operate across borders as well.

When the grid cells are aggregated, I also vary the starting point of aggregation. I shift the grid cell on which the aggregation begins, from the north-west towards the south-east, by from one grid cell of the original size up to the aggregated grid cell size minus one. For example, if I multiply the sides of a grid cell size by three, I create three different versions of the data, where the shift of the aggregation starting point from the north-west to the south-east is zero grid cell of the original size (i.e. no shifting), one grid cell of the original size, and two grid cells of the original size.

Theisen et al. (2011) used a random subsample of 5% of the observations where there is no onset of civil armed conflict, most probably following the recommendation by King and Zeng (2001). King and Zeng (2001) point out that when the outcome variable is a rare event, it does not affect the estimation much, even when most of no-event observations are dropped. In the current case, using a subsample also helps reduce spatial autocorrelations. Thus, I follow this practice as well (the aggregation of grid cells is done before taking a subsample). I generate 30 random subsamples, making sure the results are not driven by a particular subsample; the reason for choosing 30 subsamples is simply to show a possible variation in the estimates and to keep the plots of the results legible at the same time. To estimate the effect of drought on the likelihood of civil armed conflict onset, I use logistic regression, as done in the original study. While Theisen et al. (2011) implemented robust standard errors clustered on countries, I cluster them on grid cells because the data used are a panel data format consisting of grid-cell years and some grid cells can belong to two countries.

Figure 6 presents the results. The y-axis is the point estimate of the average effect of drought on the likelihood of the onset of civil armed conflict on the log odds ratio scale. If the color of the symbol (∗) is darker, it means a smaller one-tailed p-value, for the point estimate that is a positive value greater than zero. The x-axis indicates grid cell sizes. As there are 30 subsamples per grid cell specification, there are 30 point estimates.

There is a large variation in terms of both point estimates and p-values, across different grid cell sizes and shifts. When I use the original grid cell size (cell size = 1), or a small aggregated size such as a cell size of two or three, almost all point estimates of the average effect of drought are negative and the p-values tend to be large. Thus, in these grid cell specifications, the finding of Theisen et al. (2011) holds: there is “little evidence” that drought increases the likelihood of the onset of civil armed conflict (81).

A difference appears if the grid cells are aggregated to larger sizes. The point estimates tend to indicate greater effects of drought on increasing the likelihood of the onset of civil

---

9 The statistical analysis was done on RStudio (RStudio Team 2020) running R version 4.1.2 (R Core Team 2021). The data visualization was done by the ggplot2 package (Wickham 2016).

10 Different grid cell sizes create different numbers of observations and change the baseline likelihood of the onset of civil armed conflict, although, in all grid cell sizes used here, the onsets of civil armed conflict remain rare events (less than 0.5% of the observations). Therefore, the comparison of effect sizes across models is more meaningful on the log odds ratio scale, the relative scale of an effect, than on the probability scale, the absolute scale of an effect.
armed conflict, and the $p$-values tend to become smaller. This result empirically points to uncertainty in the estimate of the causal effect, across the different grid cell sizes. Interestingly, this tendency does not apply to the cell size of six when the shift in aggregation is either four or five. This point empirically illustrates uncertainty in the estimate of the causal effect, across the different locations where the dividing lines of grid data are drawn.

For the Null Hypothesis Significance Testing, I re-color the $\times$ symbols in Fig. 6 based on statistical significance vs. non-significance at the threshold of $p < 5\%$ (see Fig. 7). There are many more models that produce statistical non-significance than those

---

**Fig. 6** Point estimates of the average effect of drought across 30 subsamples. Darker color of the symbol means a smaller $p$-value
that produce statistical significance. Thus, from this point of view, my test suggests that the original conclusion by Theisen et al. (2011) is fairly robust. Yet, it is controversial whether the Null Hypothesis Significance Testing is a plausible way to evaluate the uncertainty of an effect (see Amrhein et al. 2019; Gelman 2011; Suzuki 2022). If we take a $p$-value as a continuous measure of uncertainty (Lew 2012), the results here suggest that the original results of Theisen et al. (2011) are sensitive to the choice of grid cell sizes and to the starting point of grid cell aggregation. Whether researchers take

**Fig. 7** Re-coloring the symbols in Fig. 6 by statistical significance/non-significance
either viewpoint of statistical inference, the robustness test is useful to examine sensitivity in statistical estimation using grid data.

5 Conclusion

This note has first developed the methodological theory on how the three sources of uncertainty specific to grid data affect inference. While grid data have their own advantages, e.g., “allowing for units of observation that are identical in shape and completely exogenous to the feature of interest” (Tollefsen et al. 2012, 365), the uncertainty specific to grid data poses extra challenges to statistical analysis. To incorporate this uncertainty for robustness tests, I have proposed that we aggregate grid cells with different specifications (i.e., varying grid cell sizes and the starting point of aggregation) and analyze the datasets of different grid cell specifications.

I suggest researchers should not rely on the results from only one grid cell specification, unless they can theoretically justify the use of that particular grid cell specification. If there is a significant variation in the estimates across the datasets of different grid cell specifications, such as the one this methodological note has found, researchers should consider why different grid cell sizes produce different results, and the results from which grid cell size are theoretically more plausible than the others. If they cannot identify any theoretical reason to put greater credibility on specific grid cell specifications while facing a significant variation in the estimates, they can only conclude the results are inconclusive (Gross 2015; Kruschke 2018). In such a case, it might be worthwhile to look for empirical evidence from other methodological approaches such as qualitative and/or experimental research for further theorization on geographic causal mechanisms.11

Greater theorization on geographic causal mechanisms aligns well with the recent trend towards more focus on connecting a theory to a statistical model (Pearl and Mackenzie 2018; Keele et al. 2020; Lundberg et al. 2021). It is also possible that greater theorization will lead researchers to conclude that grid cells are not appropriate units of analysis. For example, in the case of conflict research, the artificial borders of grid cells might not always reflect theoretical borders relevant to conflict dynamics, such as the terrains that facilitate or hinder the movement of armed groups (Fearon and Laitin 2003). In such a case, political or theoretical geographic units of analysis might be preferred. Meanwhile, if there is a theoretical reason to consider grid cells as appropriate units of analysis, robustness tests such as that proposed in this methodological note should help evaluate the uncertainty specific to grid cell data.

6 Supplemental online material

The R code to replicate the analysis is available at https://akisatosuzuki.github.io/papers.html.

11 Fuzzy-set qualitative comparative analysis (Ragin 2000) is another promising method. For examples in political and conflict research, see Bretthauer (2015), Haesebrouck (2017). For other examples, see Kusa et al. (2021), Medina-Molina et al. (2022), Romero-Castro et al. (2022). I thank an anonymous reviewer for this point.
Acknowledgements. I would like to thank seminar participants at Dublin City University, the editors, and the anonymous reviewers, for their helpful comments, and Johan A. Dornschneider-Elkink for encouraging me to independently develop his initial idea on this topic. The accompanying R package, “rbstgrid,” has been co-developed with Johan A. Dornschneider-Elkink at the School of Politics and International Relations, University College Dublin, and is available at https://akisatosuzuki.github.io/programs.html. The views expressed are my own unless otherwise stated, and do not necessarily represent those of the institutes/organizations to which I am/have been related.

Funding. I would like to acknowledge the receipt of funding from the Irish Research Council (the Grant Number: GOIPD/2018/328) for the development of this work.

Declarations

Competing interests. The authors reports there are no competing interests to declare.

References

Amrhein, V., Greenland, S., McShane, B.: Retire statistical significance. Nature 567, 305–307 (2019)
Brethauer, J.M.: Conditions for peace and conflict: applying a fuzzy-set qualitative comparative analysis to cases of resource scarcity. J. Conflict Resolut. 59(4), 593–616 (2015)
Buhaug, H., Gleditsch, K.S., Holtermann, H., Ostby, G., Tollefsen, A.F.: It’s the local economy, stupid! Geographic wealth dispersion and conflict outbreak location. J. Conflict Resolut. 55(5), 814–840 (2011)
Fearon, J.D., Laitin, D.D.: Ethnicity, insurgency, and civil war. Am. Polit. Sci. Rev. 97(1), 75–90 (2003)
Gelman, A.: Causality and statistical learning. Am. J. Sociol. 117(3), 955–966 (2011)
Gross, J.H.: Testing what matters (if you must test at all): a context-driven approach to substantive and statistical significance. Am. J. Polit. Sci. 59(3), 775–788 (2015)
Haesebrouck, T.: NATO burden sharing in Libya: a fuzzy set qualitative comparative analysis. J. Conflict Resolut. 61(10), 2235–2261 (2017)
Hallberg, J.D.: PRIO conflict site 1989–2008 codebook: a geo-referenced dataset on armed conflicts. https://www.prio.org/download/datasetfile/57/ConflictSite%204-2010_v3%20Codebook.pdf (2011)
Hernán, M.A., Robins, J.M.: Causal Inference: What If. Chapman & Hall/CRC, Boca Raton, FL (2020)
Ito, G., Hinkkainen Elliott, K.: Battle diffusion matters: examining the impact of microdynamics of fighting on conflict termination. J. Conflict Resolut. 64(5), 871–902 (2020)
Keele, L., Stevenson, R.T., Elwert, F.: The causal interpretation of estimated associations in regression models. Polit. Sci. Res. Methods 8(1), 1–13 (2020)
Kikuta, K.: A new geography of civil war: a machine learning approach to measuring the zones of armed conflicts. Polit. Sci. Res. Methods 10(1), 97–115 (2022)
King, G., Zeng, L.: Logistic regression in rare events data. Polit. Anal. 9(2), 137–163 (2001)
Kruschke, J.K.: Rejecting or accepting parameter values in Bayesian estimation. Adv. Methods Pract. Psychol. Sci. 1(2), 270–280 (2018)
Kusa, R., Duda, J., Suder, M.: Explaining SME performance with fsQCA: the role of entrepreneurial orientation, entrepreneur motivation, and opportunity perception. J. Innov. Knowl. 6(4), 234–245 (2021)
Lee, D.W., Rogers, M.: Measuring geographic distribution for political research. Polit. Anal. 27(3), 263–280 (2019)
Lee, D.W., Rogers, M.Z., Soifer, H.D.: The modifiable areal unit problem in political science. Paper presented at the 116th APSA Annual Meeting, September 9th–13th (2020)
Lew, M.J.: Bad statistical practice in pharmacology (and other basic biomedical disciplines): you probably don’t know P. Br. J. Pharmacol. 166(5), 1559–1567 (2012)
Linke, A.M., Witmer, F.D.W., Holland, E.C., O’Loughlin, J.: Mountainous terrain and civil wars: geospatial analysis of conflict dynamics in the post-soviet caucasus. Ann. Am. Assoc. Geogr. 107(2), 520–535 (2017)
Lundberg, I., Johnson, R., Stewart, B.M.: What is your estimand? Defining the target quantity connects statistical evidence to theory. Am. Sociol. Rev. 86(3), 532–565 (2021)
Medina-Molina, C., Pérez-Macías, N., Gismera-Tierno, L.: The multi-level perspective and micromobility services. J. Innov. Knowl. 7(2), 1–11 (2022)
Neumayer, E., Plüumper, T.: Robustness Tests for Quantitative Research. Cambridge University Press, Cambridge (2017)

O’Loughlin, J., Wittmer, F.D.W.: The diffusion of violence in the North Caucasus of Russia, 1999–2010. Environ. Plann. A Econ. Space 44(10), 2379–2396 (2012)

Openshaw, S.: The Modifiable Areal Unit Problem. Geo Books, Norwich (1983)

Pearl, J.: On measurement bias in causal inference. In: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, 425–432. AUAI Press, Arlington, VA (2010)

Pearl, J., Mackenzie, D.: The Book of Why: The New Science of Cause and Effect. Allen Lane, London (2018)

Pickering, S.: Introducing SpatialGridBuilder: a new system for creating geo-coded datasets. Confl. Manag. Peace Sci. 33(4), 423–447 (2016)

R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org (2021)

Ragin, C.C.: Fuzzy-Set Social Science. The University of Chicago Press, Chicago, IL (2000)

Reuveny, R.: Climate change-induced migration and violent conflict. Polit. Geogr. 26(6), 656–673 (2007)

Romero-Castro, N., López-Cabarcos, M.A., Piñeiro-Chousa, J.: Uncovering complexity in the economic assessment of derogations from the European industrial emissions directive. J. Innov. Knowl. 7(1), 1–14 (2022)

RStudio Team: RStudio: integrated development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/ (2020)

Ruggeri, A., Dorussen, H., Gizelis, T.-I.: Winning the peace locally: UN peacekeeping and local conflict. Int. Organ. 71(1), 163–185 (2017)

Schutte, S., Donnay, K.: Matched wake analysis: finding causal relationships in spatiotemporal event data. Polit. Geogr. 41, 1–10 (2014)

Schutte, S., Weidmann, N.B.: Diffusion patterns of violence in civil wars. Polit. Geogr. 30(3), 143–152 (2011)

Shaver, A., Carter, D.B., Shawa, T.W.: Terrain ruggedness and land cover: improved data for most research designs. Confl. Manag. Peace Sci. 36(2), 191–218 (2019)

Soifer, H.D.: Units of analysis in subnational research. In: Giraudy, A., Moncada, E., Snyder, R. (eds.) Inside Countries: Subnational Research in Comparative Politics, pp. 92–112. Cambridge University Press, Cambridge (2019)

Sundberg, R., Melander, E.: Introducing the UCDP georeferenced event dataset. J. Peace Res. 50(4), 523–532 (2013)

Suzuki, A.: Presenting the probabilities of different effect sizes: towards a better understanding and communication of statistical uncertainty. arXiv: 2008.07478v3 [stat.AP] (2022)

Theisen, O.M., Holtermann, H., Buhaug, H.: Climate wars? Assessing the claim that drought breeds conflict. Int. Secur. 36(3), 79–106 (2011)

Tollefsen, A.F., Strand, H., Buhaug, H.: PRIO-GRID: a unified spatial data structure. J. Peace Res. 49(2), 363–374 (2012)

Von Uexkull, N., Croicu, M., Fjelde, H., Buhaug, H.: Civil conflict sensitivity to growing-season drought. Proc. Natl. Acad. Sci. 113(44), 12391–12396 (2016)

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, 2nd edn. Springer International Publishing, Cham (2016)

Wood, R.M., Sullivan, C.: Doing harm by doing good? The negative externalities of humanitarian aid provision during civil conflict. J. Polit. 77(3), 736–748 (2015)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.