

Abstract

Researchers in the conflict research community have become increasingly aware that we can no longer depend on state-aggregated data. Numerous factors at the sub-state level affect the nature of human interactions, so if we really want to understand conflict, we need to find more appropriate units of analysis. However, while many conflict researchers have realised this, actually taking the next step and performing data analysis on spatial data grids has remained a rather elusive goal for many because of the difficulty in learning the new techniques to perform such analyses. This paper introduces SpatialGridBuilder, a new, freely available, open-source system with the goal of empowering conflict researchers with no background in GIS methods to start their own spatial analyses. SpatialGridBuilder allows the researcher to: i) create entirely new spatial data sets, based on the needs of their own research; ii) import their own spatial data; iii) easily add a range of important variables to the data sets, including commonly used conflict variables, plus new variables which have not been presented before; iv) visualise graphical renderings of this data. Having done this, SpatialGridBuilder will then export the data set for the researcher to analyse using conventional statistical methods. This article introduces the new program, and demonstrates how it can be used to set up such a statistical analysis. It also shows how different results can be achieved by building grids of different resolutions, thereby encouraging researchers to choose grid resolutions appropriate to their research questions and data. The article also introduces a novel means of determining infrastructure complexity, using Google maps.

Keywords: GIS, spatial disaggregation

Introducing SpatialGridBuilder: A new system for creating geo-coded data sets

Steve Pickering.*

March 3, 2015

*Thanks go to Tobias Boehmelt, Ismene Gizelis, Kristin Skrede Gleditsch, Kaisa Hinkkainen, Shoko Kohama, Yin-Hsiu Liu, Kazuto Ohtsuki, Ulrich Pilster, Atsushi Tago and Pingping Xie for help and encouragement at various stages in this project.

1 Introduction

In recent years, conflict researchers have moved away from the state as a unit of analysis, looking instead for more appropriate units.¹ One recent expression of this disaggregation is the creation of the PRIO-GRID (Tollefsen, Strand and Buhaug 2012). Yet while many conflict researchers have managed to successfully use Geographic Information Systems (GIS) methods, there are many more who want to, but have not done so because of the steep learning curve in methods and software. A new generation of researchers looking at issues such as improvised explosive devices (IEDs), rebel bases, drug cultivation, mineral/ fossil fuel reserves and climate change are gathering data but have difficulty building it into a geo-spatial framework.

To try to solve this and to encourage conflict researchers to start using spatial data, this paper introduces SpatialGridBuilder, a freely available, open-source system for creating spatial data sets. The program can be downloaded from <http://sf.net/projects/spatialgridbuilder/>. The program uses a graphical user interface (see Figure 1), runs on MS Windows and is currently being ported to Linux and Mac OS. One of the inspirations for SpatialGridBuilder is Bennett and Stam's (2000) EUGene (Expected Utility Generation and Data Management) program. SpatialGridBuilder is quite different to EUGene in many ways: it is not motivated by expected utility theory, and it is not primarily focused on dyads. But in some ways, SpatialGridBuilder can be seen as a spatial version of EUGene. Its over-arching aim is to help the researcher to create a new data set, based on both built-in variables and, more importantly, researcher-provided variables. SpatialGridBuilder gives the researcher the tools to create a spatial data set precisely tailored to the needs of the research question, at the macro or micro level. Like EUGene, SpatialGridBuilder does not attempt to analyse your data for you. Instead, it helps the researcher to create a data set which can then be analysed using conventional statistical software, using all of the methods which researchers are already familiar with.

Figure 1 goes about here

1.1 Creating a new spatial data grid

A spatial data grid can be conceived of as being like a sheet of graph paper: a page full of empty boxes which can be given x and y coordinates. All spatial data grids have a resolution, which is commonly measured in degrees per grid cell. For the PRIO-GRID, this resolution is 0.5 degrees per grid cell, which for most analyses is sufficiently granular. Indeed, when a data set covers the entire world, 0.5 degrees

¹See, for example, O'Loughlin, 2002; Buhaug & Lujala 2005; Buhaug & Rød 2006; Cederman & Girardin 2007; Gleditsch 2007; Hegre & Raleigh 2007; Ormhaug 2007; Pugel 2007; Lujala 2009; Weidmann and Ward 2010; Aas Rustad, Buhaug, Falch and Gates 2011; Buhaug, Gleditsch, Holtermann, Østby and Tollefsen 2011; Wucherpfennig, Weidmann, Girardin, Cederman and Wimmer 2011; Dittrich Hallberg, 2012; Raleigh and Kniveton, 2012; Sundberg and Melander, 2013; Landis 2014.

per grid cell can become computationally slow to process, especially when many variables are involved, or when the data involved are time variant. However, in some cases, such as those involving small states or confined regions, 0.5 degrees per cell can be too coarse: at the equator, a half degree grid cell covers about 3063 km².

SpatialGridBuilder allows the researcher to decide the resolution and spatial confines of the grid. This is, in fact, the first decision the researcher must make when creating a new grid: what the resolution is, and which parts of the world are of interest. For instance, while a researcher looking at all of Africa may choose to use 0.5 degrees as the grid resolution, a researcher looking at East Timor may choose to use more granular grid cells, of perhaps 0.1 or 0.05 degrees per grid cell. SpatialGridBuilder allows the user to choose any resolution they like, and experiment with different resolutions.²

To try to illustrate how to go about creating a new data set and applying variables to it, the example of a researcher interested in China will be used.

1.2 Land mask

Figure 2 goes about here

When a researcher first creates a new grid, they are essentially creating a series of empty boxes: aside from an ID number, the latitude and longitude, plus the area of the grid cell (as this decreases based on the distance from the equator³), the grid contains no information at all. So, as has been mentioned, we will assume that our researcher is interested in China. To begin, it is necessary to determine China's 'bounding box': the coordinates within which China would fit (see Figure 2). By looking at a map or atlas of Asia, the researcher could determine that a box which starts at 55 degrees north of the equator, continuing to 15 degrees north of the equator, and runs from 70 degrees east to 140 degrees east of Greenwich would neatly contain China. The researcher then needs to choose a resolution. Because we are not dealing with the whole world, it is possible to select a higher resolution than that used by the PRIO GRID. However, China is a big country, so it would be unwise to use too high a resolution. Accordingly, the researcher chooses 0.1 degrees per grid cell. By entering these details into SpatialGridBuilder, it is possible to create an empty grid as is shown in Figure 3.

Figure 3 goes about here

At this stage, the grid is of little value. One of the first things we need, then, is a binary indicator of whether each cell is on land or not (in most instances, we will only be interested in land cells, although

²More grid cells is not always better: sometimes, this just increases the number of observations, which may not be appropriate. However, if you are looking at a small region and have data which have been coded to a precise level, it is usually better to build a higher resolution grid.

³For instance, a one degree grid cell at the equator covers an area of approximately 110km longitude by 110km latitude (12,100km²); whereas, at the poles, the latitude would remain 110km, but the longitude would be zero. Somewhere inbetween, such as a one degree grid cell centred on Oslo, would still have a latitude of 110km, but a longitude of 56km (6,160km²).

some researchers may be interested in naval conflict, or in exclusive economic zones (EEZs): as will be shown later, SpatialGridBuilder provides this option). The researcher can use SpatialGridBuilder to populate the grid with a land mask dummy variable, as is visualised in Figure 4.

Figure 4 goes about here

1.3 State masks

Even after introducing a land mask binary, the spatial grid may still contain cells that the researcher is not interested in. In the example above, while the bounding box has been drawn around China, the grid still contains cells from many other countries.⁴ To solve this problem, SpatialGridBuilder includes the option of adding state mask binaries. The researcher can select China, which will add a new dummy to the data grid identifying all of the grid cells in Chinese territory, marked up by the researcher's choice of identifier: Gleditsch-Ward (1999), Correlates of War (2011), or ISO 3166 (see Figure 5). The state maps are based on Weidmann, Kuse and Gleditsch's (2010) CShapes library.

Figure 5) goes about here

1.4 State identifiers

As well as providing state dummy variables with which to identify a single state, the program allows the researcher to add an additional variable which will identify all states, as is shown in Table ???. Again, these are based on Weidmann, Kuse and Gleditsch (2010) and can be identified using Gleditsch-Ward (1999), Correlates of War (2011), ISO or plain text state names.

Table ??? goes about here

Grid cells in which there is no state (i.e., the sea, or territories beyond the writ of Westphalia) are coded as 'NA'.

2 Importing new data

Now that the basic grid is ready, with dummy variables to determine whether each cell is land and if so, which state it is in, the researcher can now import their own data. SpatialGridBuilder offers four ways in which to do this: point data, radial point data, polygon data and raster data.

⁴Twenty in total: Bangladesh, Bhutan, India, Japan, Kazakhstan, Kyrgyzstan, Laos, Mongolia, Myanmar, Nepal, North Korea, Pakistan, Philippines, Russia, South Korea, Taiwan, Tajikistan, Thailand, Uzbekistan and Vietnam.

2.1 Using your own data 1: polygon data

Let's assume that our China researcher is looking at levels of pollution throughout the country. The researcher finds an ESRI shapefile showing high flouride pollution sources (USGS, n.d.). SpatialGridBuilder allows the researcher to import shapefiles and apply them to the grid (see Figure 6). It does this by determining whether the centre point of each grid cell falls within the polygons in the shapefile.

Figure 6 goes about here

2.2 Using your own data 2: point data

Let us now assume that our China researcher wants to see if there is some relationship between pollution and airports. The researcher takes the list of latitudes and longitudes of 9000 airports developed by Cazemier and van der Molen (n. d.) and imports them to SpatialGridBuilder in a simple .csv file. Once imported, these can be visually represented as in Figure 7.

Figure 7 goes about here

Again, SpatialGridBuilder will not tell the researcher whether there is a relationship between airports and pollution in China: this is where conventional statistical software should be used. But SpatialGridBuilder will help the researcher to create the data set which can then be imported into the statistical software.

2.3 Using your own data 3: radial point data

Finally, our China researcher wants to look at Asia more broadly and uses the PRIO Conflict Site 1989–2008 data (Dittrich Hallberg 2012). Conflict Site lists 791 conflicts in 75 countries, but our researcher just wants to look at conflicts in countries which have territory in Asia. This gives 426 conflict events in 24 countries.⁵

Figure 8 goes about here

Our researcher starts by creating a grid at 0.1 degrees per cell with a simple land mask, represented in Figure 8A. Then, identifiers for each of the 24 Asian conflict states are added, as is seen in Figure 8B. Each conflict event in Conflict Site is coded with a lat long coordinate pair, plus a radius, ranging from 50 to 1300 kilometres. Accordingly, our researcher imports these data and visualises them in Figure 8C. Finally, as our researcher is only interested in the Asian states coded by the PRIO Conflict Site data, the state identifiers are merged to form a PRIO-inclusion dummy. By then merging this variable with

⁵Afghanistan, Azerbaijan, Bangladesh, Cambodia, India, Indonesia, Iran, Iraq (and Kuwait), Israel, Laos, Lebanon, Myanmar, Nepal, Pakistan, Papua New Guinea, Philippines, Russia (and Soviet Union), Sri Lanka, Tajikistan, Thailand, Turkey, Uzbekistan and Yemen.

a dummy of the radial conflict zones, our researcher is able to create a mask which indicates territory which is both in a PRIO Conflict state and is within one of the conflict radii (Figure 8D).⁶

2.4 Using your own data 4: raster data

Sometimes, data are made available in a raster format: this is quite common for remote-sensing data, such as night lights, rainfall, topography and land-use. Accordingly, SpatialGridBuilder allows the import of raster data. The program will automatically determine how the data items in the raster file are separated (comma, semi-colon, tab or space). It will then ask the user which parts of the world are covered in the data. For instance, many raster-based global data sets will only cover the parts of the world in which people live; they will therefore exclude the polar regions in order to save memory and processing time. Once the user has entered the spatial bounds of the data set, SpatialGridBuilder will apply the raster data to the existing grid, based on the matching centre point of each grid cell (or, when the grid is of a different resolution to the raster data, the value from the nearest neighbour in the raster file).

3 Built-in variables

SpatialGridBuilder includes a number of variables commonly used by conflict researchers in their analyses. These are included to help researchers develop their data sets more quickly. They can be used as control variables, or as more substantive parts of hypotheses.

3.1 ACLED

To make it easier for the researcher to analyse conflict, ACLED (Armed Conflict Location Event Dataset) conflicts are built into the program (Raleigh, Linke, Hegre and Karlsen 2010). At time of writing, all ACLED conflicts for Africa have been included; other regions will be added in a later version. The researcher is able to select between the eight different conflict types⁷ or choose to include all ACLED conflict types. Of course, ACLED offers more ways to divide data than just these eight conflict types, and researchers may wish to have more fine-grained control over the conflicts. In this case, the researcher can modify ACLED in the conventional way (with a statistical package, spreadsheet, text editor, etc.) and then import the data into SpatialGridBuilder as point data (discussed above).

⁶This approach is useful when we want to exclude the bits of the world which are not useful to our analysis. Researchers should, however, consider the issue of spatial autocorrelation. This is discussed in the Appendix.

⁷Battle-Government regains territory; Battle-No change of territory; Battle-Rebels overtake territory; Headquarters or base established; Non-violent activity by a conflict actor; Non-violent transfer of territory; Riots/Protests; Violence against civilians

3.2 UC DP GED

Additionally, SpatialGridBuilder includes all conflicts from the UC DP GED (Uppsala Conflict Data Program Georeferenced Event Dataset: Sundberg, Lindgren and Pads kocimaite: 2010; Melander and Sundberg 2011). At time of writing, the only option is to include all conflicts. Future versions of SpatialGridBuilder will allow more control over this, but in the meantime, the researcher can still modify and import UC DP data using the method discussed above.

3.3 Distance

In the considerable research on the subject, distance has been shown to have an important effect on conflict. While many commentators have argued that globalisation has meant that distance is no longer an important factor for interaction opportunities, Gleditsch, Buhaug and Walter (2006) argue that the death of distance has probably been exaggerated. Indeed, data sets such as those developed by Gleditsch and Ward (2001) continue to form an important part of conflict research. Accordingly, the option to calculate a distance variable is included. This will determine the minimum distance (based on the haversine great circle formula) of each grid cell to any other variable the researcher has added to the grid. For instance, as our researcher is building a grid based on China, the distance of each grid cell to the nearest point on the border of China could be calculated (see Figure 9). This could also be useful for capital cities, conflict zones, or any other point or polygon data which the user wishes to use.

Figure 9 goes about here

3.4 Mountainous/ rugged terrain

Mountains have often been analysed in conflict research literature. Yet many conflict researchers have used data which suffer from a key problem: they are based on a binary which is of little relevance to conflict research (see the data sets presented in Collier and Hoeffler, 2000/ 2004, plus Fearon and Laitin 2003; all ultimately based on the research of Gerrard 1990/ 2000). While many conflict researchers start articles by saying they wish to analyse the relationship between *terrain* and conflict, they finish their articles by using the percentage of mountains as a proxy for this (often a state-aggregated proxy). SpatialGridBuilder solves this problem by introducing a linear measure of ruggedness.

Figure 10 goes about here

As can be seen in Figure 10, the program analyses each cell in the grid and looks at the elevation variance of that cell from its immediate neighbours, based on elevation data from the Shuttle Radar Topography Mission (Farr et al. 2007). This gives a variable which is of much more use to the conflict researcher than either the percentage of mountains, or simple elevation data. Indeed, looking at the

ruggedness rendering in Figure 11, we can see the ruggedness variable demonstrates a very clear border between China and Nepal.

Figure 11 goes about here

3.5 Infrastructure proxy

Figure 12 goes about here

Early attempts towards a measure of infrastructure can be seen in the work of Bueno de Mesquita (1981: 104) who looks at the distance over which a major military operation can be moved per day (Diehl (1985: 1209) describes Bueno de Mesquita's estimates as 'impressionistic, if not somewhat arbitrary'). Dupuy (1985) also adopts a similar approach.

More recently (and less impressionistically/ arbitrarily) both Buhaug and Rød (2006) and Raleigh and Hegre (2009) have used the ESRI's Digital Chart of the World to gather road data. This data set was released in the 1990s. The method employed by Buhaug and Rød is to take the logged total length of major roads in each grid cell, normalised by the country mean. They find that separatist conflict is more likely in regions with less than average road density, while governmental conflicts tend to occur in areas with a more developed road network.

Raleigh and Hegre employ a different method: they determine whether their grid cells include primary, secondary, or informal/ no roads. They find that '[c]onflict events are 47% less likely to happen in squares with no roads or only informal roads than in squares with primary roads.' It is worth quoting their explanation:

The results ran counter to our initial expectations – conflicts are assumed to occur in faraway and inaccessible regions. However, the finding may not be so counter-intuitive after all. First, battle events occur where rebel group and army units encounter each other. Such meeting places are normally reached by road. Second, rebel groups tend to target high-value places (villages, military installation, pipelines, mines, etc.), and roads also often connect such places. Third, there is also a reporting bias at play here – media report incidences primarily in accessible areas. (Raleigh and Hegre 2009: 234).

However, some concerns have been raised by users of the Digital Chart of the World as to the accuracy and completeness of its data (for detailed surveys of user feed-back on the DCW, see Langaas (1995) and Smith and Langaas (1995)). As such, this paper presents an alternative means of measuring the complexity, or level of development, of infrastructure based on the complexity of maps on the Google Maps server (see Figure 12). Google Maps include information on roads, buildings, etc., and the more information there is in each image, the more complicated the infrastructure is. This information is easy

to capture: as the maps themselves are stored in PNG format, a more complex map will inherently have a larger file size.

Figure 13 goes about here

A rendering of this infrastructure proxy can be seen in a grid of Asia, presented in Figure 13. This grid is built at a resolution of 0.1 degrees per grid cell, giving a size of 1300×660 grid cells. Accordingly, to determine the infrastructure complexity for this grid, SpatialGridBuilder sends 858,000 file size requests to the Google Maps servers. At first glance, the resulting grid looks rather like a night light image. SpatialGridBuilder also includes night light data (see section below) and while there are similarities between the images, there are also significant differences. Indeed, the correlation between the two is only 0.52.

It is hoped that researchers will use this approach as a valuable new method of determining the complexity of infrastructure. A first analysis of this new variable will be made at the end of this paper.⁸

3.6 Night lights

Another source which has often been used in literature as a proxy for not only infrastructure, but also population and economic development, is the brightness of night lights. SpatialGridBuilder uses NOAA data for the year 2000; future versions will include data from the years 1997–2010 for researchers interested in temporal variance.

3.7 Rainfall

As the recent special issue of the Journal of Peace Research on climate change and conflict (January 2012) made clear, there is an increasing expectation among many *outside* of the conflict research community that climate change will have an impact on conflict, but within the conflict research community, there is little evidence of a relationship. However, there is broad acknowledgement that more work needs to be done. Accordingly, SpatialGridBuilder includes rainfall data. While the spatial resolution of the rainfall data is quite low (2.5 degrees per grid cell), the data set makes up for this in its temporal coverage: it offers monthly rainfall data covering the years 1979–2011, based on Xie and Arkin (1997). By providing this data, the researcher is able to include climate-related data in models). More climate related data, e.g. temperature, plus higher resolution rainfall data will be included in future versions of SpatialGridBuilder.

⁸It should be noted that at time of writing, SpatialGridBuilder defaults to use Google map images at a zoom level of 14. Forthcoming versions of the program will allow this to be changed manually or automatically. Also, Bing and OpenStreetMap will be included for robustness checks.

3.8 Infant mortality

Following the suggestion by the State Failure Task Force group (Esty et al., 1998), Urdal (2005) argues that infant mortality is a better (inverse) proxy for development than more conventional economic measures, such as GDP or energy consumption per capita. Urdal goes on to find a strong relationship between high infant mortality rates and armed conflict. Accordingly, SpatialGridBuilder includes data on infant mortality (see Figure 14). However, as the purpose of SpatialGridBuilder is spatial disaggregation, it was necessary to use different data than the UN data used by Urdal. As such, the infant mortality measure is based on that compiled by Columbia University’s Center for International Earth Science Information Network (CIESIN).

Figure 14 goes about here

As well as giving an indication of the rates of infant mortality in China and surrounding countries, Figure 14 is useful in that it also shows some of the problems facing researchers using spatial methods in their analyses. First is the problem of missing data. The CIESIN data are coded in raster format, with the value ‘-9999’ indicating ‘NODATA’. At first glance at the data set, it would appear that the -9999 values indicate the oceans, which in most cases they do. However, it is only when we apply the data to a grid and apply a country mask, using a tool such as SpatialGridBuilder, that we notice that many of the missing data values are on land and in some cases comprise whole countries (looking again at Figure 14, we see that Taiwan is rendered in red, indicating missing data). As with non-spatial data, researchers must take measures to deal with missing values.

Looking at China, we can see great variation in the levels of infant mortality in the eastern part of the country, but as we move west, we can see that there are large regions with the same level of infant mortality. This is because in China, as with many countries, data are collected at the county level, and Chinese counties tend to grow larger as you move (north-) west. Even when spatial data are stored in a raster format, even with population data, that does not mean that the raster pixel gives a true indication of the value at that location: it is often inferred from the mean of a larger aggregate unit, such as a county. Researchers need to remember this when analysing spatial data, as the unit size and shape is usually not constant throughout a country, and are seldom comparable between countries. This does not mean that we cannot do spatial analysis at all; it just means we need to understand the limitations of our data before making any inferences.⁹

3.9 Using the built-in variables to construct new variables

Figure 15 goes about here

⁹Researchers should also consider the modifiable areal unit problem, or MAUP; see Openshaw 1983)

By combining SpatialGridBuilder with conventional statistical software, or indeed a simple spreadsheet, it is possible to create new variables. One such example is exclusive economic zones (EEZs). Part V of the UN Convention on the Law of the Sea defines EEZs as regions beyond the territorial sea of a state, in which each state has ‘sovereign rights for the purpose of exploring and exploiting’ over a region extending 200 nautical miles from the state’s coast (UNCLOS 1982). Figure 15 gives an illustration of China’s EEZ. While SpatialGridBuilder does not have an option to automatically calculate states’ EEZs, it is possible to do this by a) using SpatialGridBuilder to create a China dummy; b) using statistical software or a spreadsheet to create a ‘not China but still land’ dummy; c) using SpatialGridBuilder to calculate the distance of each cell to both China and the ‘not China’ cells; d) using statistical software or a spreadsheet to create a dummy for all cells where i) the cell is not on land; ii) the cell is 200 nautical miles (370.4 km) or less from China; iii) the distance from the cell to China is less than the distance to any other country.¹⁰ Clearly, EEZs are very politically contentious, and many states claim differing EEZs based on their claims to remote islands. By working with different understandings of state territory (e.g., with or without key islands), SpatialGridBuilder allows the researcher to calculate different EEZs for each state and thereby determine regions of the seas which may face future disputes. Indeed, the number of militarized interstate disputes involving small islands indicates that EEZs are a useful area for spatial conflict analysis.

4 Example

Now that the functions of SpatialGridBuilder have been introduced, it is possible to give an example of how it can be used. This example will look at conflict in Asia, based on the conflicts in ACLED Version 1,¹¹ during the period 1997–2010. One of the key advantages of SpatialGridBuilder is that it allows the researcher to construct grids of any resolution. ACLED only includes conflicts for six Asian countries (Afghanistan, Cambodia, Laos, Myanmar, Nepal and Pakistan), so as this only encompasses a relatively small part of the world, it is possible to build quite a high resolution grid. For this example, a grid with cells of 0.05 degrees will be built. Also, to give a comparison, and to highlight how different resolutions can lead to different results, a separate lower resolution grid will also be built of 0.5 degrees per grid cell (the same resolution as the PRIO Grid).

To begin, we first need to determine the ‘bounding box’ of our grid, using the method illustrated earlier in Figure 2. By consulting an atlas and looking at our six countries, we can see that northern and southern bounds of 38 and 6 degrees latitude, and western and eastern bounds of 60 and 108 degrees

¹⁰Future versions of SpatialGridBuilder will allow the user to do all of the steps within the program. Currently, it is possible to construct simple dummies in the program, but not based on multiple variables. SpatialGridBuilder comes with several tutorials, and details on how to calculate EEZs or similar variables are included in one of the tutorials.

¹¹More recent versions of ACLED exclude conflicts in Asia.

longitude will cover our six states nicely. We can then create two grids: a low (coarse) resolution grid (0.5 degrees per grid cell) and a high (fine) resolution grid (0.05 degrees per grid cell). Once the empty grids are created, we can populate them with a land mask, plus dummy identifiers for the six states. The first ten lines of the low and high resolution grids, plus visualisations of the two grids are presented in Figure 16.

Figure 16 goes about here

By importing these two data sets into a statistical package, we can start to do some analysis. For instance, we can make some preliminary observations as to how many grid cells we have in our two data sets, how many of those cells are on land, and how many of those cells are in each state. Additionally, as the ACLED data are restricted to these six states, we may want to create an ACLED dummy, so that our analysis will only look at territory within the six states. This can be done easily by merging the values of the six state dummies into a new ACLED dummy using straightforward commands in the statistical package.¹²

Table ?? goes about here

As can be seen in Table ??, the high resolution grid has approximately 100 times more grid cells than the low resolution grid.¹³ An important point to make here is that bigger is not always better: by increasing our resolution, we are increasing the number of observations, but this is not necessarily a good idea for spatial analysis. It is often the case that spatial data are only available at a resolution of approximately 0.5 degrees per grid cell, such as with the PRIO-GRID. In such a case, the researcher should make a grid of 0.5 degrees per grid cell. However, if the researcher has access to more granular data, then a grid at a higher resolution would be preferable.

Figure 17 goes about here

For the sake of this example, let us assume that we are interested in violence against civilians in urban areas in Asia. Accordingly, a measure of the level of infrastructure development would be useful for our analysis. While some of the world's largest cities may sprawl for many tens of kilometres, most urban areas will only spread for a few kilometres at most. If we use a low resolution grid, then, the chances are greater that we will miss these urban areas.¹⁴ Figure 17 illustrates this. The top row is based on Google Earth images of Parachinar, a small town (population approximately 70,000) in the Kurram Valley, close to the border with Afghanistan. ACLED records an incident of violence against civilians in Parachinar; the data set codes the location of the violence as having the coordinates 33.8975

¹²For instance, in R, the code would be `acled_mask <- AFG_mask + MMR_mask + NPL_mask + PAK_mask + KHM_mask + LAO_mask`

¹³Remember that ten times more degrees per grid cell applies longitudinally and latitudinally, or horizontally and vertically, so yields 100 times more grid cells. The slight differences from 100 are due to rounding.

¹⁴As was mentioned earlier, a grid built at 0.5 degrees per grid cell covers approximately 3063 km² at the equator (a cell of about 55.35 × 55.35km), which few cities would fully occupy. A grid built at 0.05 degrees per cell, however, would cover around 30.6 km² at the equator (a cell of 5.53km × 5.53km), which many urban areas would fully occupy.

degrees latitude, 70.1 degrees longitude (see Pakistan A). If we build a grid of 0.05 degrees per grid cell, the nearest location we can include is 33.9 degrees latitude, 70.1 degrees longitude (see Pakistan B), just 280 metres away. However, the nearest location in a grid built at 0.5 degrees per cell would be 34 degrees latitude, 70 degrees longitude (see Pakistan C), a forest area in the mountains, 14.67 kilometres away from Parachinar. Similarly, ACLED code an incident of conflict in Bangkok: the high resolution grid (Thailand B) located 0.47 kilometres away is also found in the centre of the city, while the low resolution grid cell is in farmland, 27.21 kilometres away.

Table ?? goes about here

Returning to SpatialGridBuilder, our researcher can now progressively add some other control variables to the models: ruggedness of terrain, tree cover, annualised rainfall, night light, population and infant mortality rate. Results are presented in Tables ?? and ?. The important point to note here is the difference in the infrastructure variable between the two grids. In the low resolution grid, the effect of the infrastructure variable maintains significance across most models, but completely loses significance in the final model H. However, in the high resolution grid, significance is maintained throughout the models. This is in line with expectations based on the example in Figure 11. The high resolution grid is more likely to capture a measure of infrastructure which is based in the same urban area as the one coded by ACLED, while the low resolution grid tends to find areas in more rural areas, which will have a lower value for the infrastructure variable.

The take-home message here is not that researchers should always build higher resolution grids; indeed, just building higher resolution grids for the sake of it merely increases the number of non-conflict events (thus skewing results), and leads to more time spent processing. Instead, researchers should spend time determining the appropriate resolution for their research before finalising their research design. This can be done by building test grids at multiple resolutions. If, like in this example, our researcher is interested in infrastructure, then he or she should start with a low resolution grid and run some spot checks to determine whether the grid is capturing urban areas; if it is not, gradually increase the resolution, and keep testing until sufficient urban areas are being captured. Alternatively, if all of the variables the researcher is interested in are at a relatively low resolution, then build the grid at that resolution. SpatialGridBuilder gives the researcher the freedom to make these decisions. Moreover, by allowing the researcher to build grids at multiple resolutions, SpatialGridBuilder not only provides additional opportunities for robustness checks, but also allows for a more inductive and flexible approach to predictive modelling.¹⁵

¹⁵The author is grateful to one of the anonymous reviewers for making this observation.

5 Conclusion

The past few years have seen a strong movement toward the use of GIS. This is a function of many things: the availability of computers that are able to work on very large data sets; the increasing degree to which the Internet has become an essential part of many researchers' lives; the willingness of many institutions, researchers and programmers to make their data and methods freely available; and finally, the evolution of GIS methods themselves as a result of these changes. Some excellent examples of geospatial software are being developed, but the steep learning curve has discouraged many researchers from using it. This paper has introduced a new means through which conflict researchers can start to introduce spatial data analysis into their work. It is hoped that by providing a program which is as straightforward to use as possible, but which does not in any way reproduce the statistical methods which researchers are already familiar with, this program will help researchers to start making new analyses which can help to further our understanding of conflict.

6 Bibliography

Aas Rustad, Siri Camilla, Halvard Buhaug, Falch and Scott Gates (2011). 'All Conflict is Local: Modeling Sub-National Variation in Civil Conflict Risk'. *Conflict Management and Peace Science*. 28(1): 15–40.

Anselin, Luc, 1988. *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

Anselin, Luc, & Bera, Anil K., 1998. 'Spatial dependence in linear regression models with an introduction to spatial econometrics.' *Statistics Textbooks and Monographs*, 155, 237–290.

Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon, 1996. 'Simple diagnostic tests for spatial dependence.' *Regional Science and Urban Economics* 26(1): 77–104.

Anselin, Luc, 2001. 'Spatial econometrics'. *A companion to theoretical econometrics*, 310330.

Bennett, D. Scott, and Allan Stam, 2000. "EUGene: A Conceptual Manual." *International Interactions* 26:179-204. Website: <http://eugenesoftware.org>.

Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gómez-Rubio, 2008. *Applied Spatial Data Analysis with R*. New York, Springer.

Buhaug, Halvard and Päivi Lujala, 2005. "Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War." *Political Geography*, 24: 399–418.

Buhaug, Halvard and Jan Ketil Rød, 2006. "Local Determinants of African Civil Wars, 1970–2001." *Political Geography* 25: 315–335.

Buhaug, Halvard, Kristian Skrede Gleditsch, Helge Holtermann, Gudrun and Andreas ForøTollefsen (2011). 'It's the local economy, stupid! Geographic wealth dispersion and conflict outbreak location.' *Journal of Conflict Resolution*. 55(5): 814–840.

Cazemier, Wiebe and Rowan Rodrik van der Molen, <http://www.halfgaar.net/localized-world-airport-codes>, licensed under the Creative Commons Attribution-Share Alike 3.0 License, downloaded 22 October 2013.

Center for International Earth Science Information Network (CIESIN), Columbia University; 2005 Global subnational infant mortality rates [dataset]. CIESIN, Palisades, NY, USA. Available at: http://www.ciesin.columbia.edu/povmap/ds_global.html

Cederman, Lars-Erik and Luc Girardin 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." *American Political Science Review* Vol. 101, No. 1: 173–185.

Collier, Paul, and Anke Hoeffler, 2000. "Greed and Grievance in Civil War," *World Bank Development Research Group Policy Research Working Paper* 2355.

Collier, Paul, and Anke Hoeffler, 2004. "Greed and Grievance in Civil War," *Oxford Economic Papers* 56: 563–595.

Correlates of War Project. 2011. "State System Membership List, v2011." Online, <http://correlatesofwar.org>.

Dittrich Hallberg, Johan 2012. PRIO Conflict Site 1989–2008: A Geo-Referenced Dataset on Armed Conflict. *Conflict Management and Peace Science* 29 219-232.

Esty, Daniel C., Jack A. Goldstone, Ted Robert Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamel Surko and Alan N. Unger, 1998. *State Failure Task Force Report: Phase II Findings*. McLean, VA: Science Applications International, for State Failure Task Force.

Farr, T. G., et al. (2007). "The Shuttle Radar Topography Mission," *Rev. Geophys.* 45, RG2004, doi:10.1029/2005RG000183.

Fearon, James D. and David D. Laitin, 2003. "Ethnicity, Insurgency, and Civil War," *The American Political Science Review*. Vol. 97, No. 1: 75–90.

Gerrard, A. J. W., 1990. *Mountain Environments: An Examination of the Physical Geography of Mountains*. London, Belhaven Press.

Gerrard, A. J. W., 2000. "What is a mountain?" mimeo, DECRG, World Bank, Washington, D. C., 2000.

Gleditsch, Kristian S. & Michael D. Ward, 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25: 393–413.

Gleditsch, Kristian S., & Michael D. Ward, 2001. "Measuring Space: A Minimum Distance Database." *Journal of Peace Research* 38: 749–68.

Gleditsch, Kristian S., 2007. "Transnational Dimensions of Civil War." *Journal of Peace Research*.

Vol. 44, No. 3: 293–309.

Gleditsch, Nils Petter, Halvard Buhaug & Barbara F. Walter, 2006. “The Death of Distance? The Globalization of Armed Conflict.” *Territoriality and Conflict In an Era of Globalization*. (187–216).

Gleditsch, Nils Petter, 2012. “Whither the weather? Climate change and conflict.” *Journal of Peace Research*. 49(1): 3–9.

Hegre, Håvard and Clionadh Raleigh, 2007. “Population Size, Concentration, and Civil War. A Geographically Disaggregated Analysis.” *World Bank Policy Research Working Paper* 4243 (Post-Conflict Transitions Working Paper No. 16).

Melander, Erik, and Ralph Sundberg, 2011, “Climate Change, Environmental Stress, and Violent Conflict - Introducing the UCDP Georeferenced Event Dataset”, Paper presented at the International Studies Association, March 16-19, Montreal, Canada.

Landis, Steven T. (2014). ‘Temperature seasonality and violent conflict: The inconsistencies of a warming planet.’ *Journal of Peace Research*. 51(5): 603–618.

Langaas, Sindre (1995). ‘Completeness of the Digital Chart of the World (DCW) database.’ Project report no. 2/1995, Norwegian Research Council, Project No. 31783/223, UNEP/ GRID-Arendal, Project No. 205026. <http://spatialnews.geocomm.com/whitepapers/re2.pdf>, accessed 14 March 2014.

Lujala, Päivi, 2009. “Deadly Combat over Natural Resources: Gems, Petroleum, Drugs, and the Severity of Armed Conflict.” *Journal of Conflict Resolution*. Vol. 53, No. 1: 50–71.

Moran, P. A. P. (1950). ‘Notes on Continuous Stochastic Phenomena.’ *Biometrika*. 37: 17–23.

NOAA, Image and data processing by NOAA’s National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency.

O’Loughlin, John, 2002. “Spatial Analysis in Political Geography,” in J. Agnew, K. Mitchell and G Ó Tuathail (eds.), *A Companion to Political Geography*. Oxford: Basil Blackwell.

Openshaw, S., 1983. “The Modifiable Areal Unit Problem.” *Catmog* 38, Study Group in Quantitative Methods (Institute of British Geographers). Norwich: Geo Books, 1983.

Ormhaug, Christin, 2007. “Disaggregating Conflict Data - Coding Challenges,” Paper presented for the International Studies Association Convention, Micro-Perspectives on Civil War I.

Pugel, James, 2007. “Spatial Aspects of Insurgencies: Examining Theory with GIS.” Paper presented at the annual meeting of the American Political Science Association, Chicago, IL.

Raleigh, Clionadh and Håvard Hegre (2009). ‘Population size, concentration, and civil war. A geographically disaggregated analysis.’ *Political Geography*. 28(4): 224–238.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research*. 47(5) 1–10.

Raleigh, Clionadh and Dominic Kniveton (2012). ‘Come rain or shine: An analysis of conflict and climate variability in East Africa’. *Journal of Peace Research*. 49(1): 51–64.

Smith, Christopher G., and Sindre Langaas (1995). ‘A survey of Digital Chart of the World (DCW) use and data quality’. Project report no. 3/1995, Norwegian Research Council, Project No. 31783/223, UNEP/ GRID-Arendal, Project No. 205026. <http://spatialnews.geocomm.com/whitepapers/re3.pdf>, accessed 14 March 2014.

Sundberg, Ralph, Mathilda Lindgren and Ausra Padskocimaite, 2010, “UCDP GED Codebook version 1.0-2011”, Department of Peace and Conflict Research, Uppsala University.

Sundberg, Ralph and Erik Melander, 2013. ‘Introducing the UCDP Georeferenced Event Dataset’. *Journal of Peace Research*. 50(4): 523–532.

Tollefsen, Andreas Forø; Håvard Strand & Halvard Buhaug (2012) PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*. 49(2): 363–374.

Urdal, Henrik 2005. People vs. Malthus: Population pressure, environmental degradation, and armed conflict revisited. *Journal of Peace Research*. 42(4): 417–434.

United Nations Convention on the Law of the Sea (UNCLOS), 1982. http://www.un.org/Depts/los/convention_agreements/texts/unclos/part5.htm, accessed 1 December 2014.

USGS, n.d. ‘High Fluoride Pollution Sources in the People’s Republic of China.’ *Coal Geology, Land Use, and Human Health in the People’s Republic of China*. Available on-line, <http://pubs.usgs.gov/of/2001/of01-318/>. Accessed 17 August 2014.

Weidmann, Nils B., Doreen Kuse, and Kristian Skrede Gleditsch. 2010. The Geography of the International System: The CShapes Dataset. *International Interactions*. 36 (1).

Weidmann, Nils B. and Michael D. Ward (2010). ‘Predicting conflict in space and time.’ *Journal of Conflict Resolution*. 54(6): 893–901.

Wucherpennig, Julian, Nils Weidmann, Luc Girardin, Lars-Erik Cederman and Andreas Wimmer (2011). ‘Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset.’ *Conflict Management and Peace Science*. 28(5): 423–437.

Xie, Pingping and Philip A. Arkin, 1997: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates and Numerical Model Outputs. *Bulletin of the American Meteorological Society*. 78: 2539–2558.

7 Appendix: radial data and spatial autocorrelation

A word of caution must be made when using radial data, such as PRIO Conflict Site. Radial data can be useful, but they introduce extreme spatial autocorrelation. Consider the three 8 x 8 grids presented

in Figure 18. In grid A, cells are clearly autocorrelated: the black cells are all next to each other, and the white cells are all next to each other. A standard measure of spatial autocorrelation is Moran's I, which in this case would have a value close to 1. Grid B, the chessboard pattern, is the other end of the spectrum: the correlation is perfectly negative as the black and white cells are completely dispersed. This yields a Moran's I close to -1. Finally, grid C has a pseudo-random distribution and accordingly its Moran's I is close to 0. When testing the grid based on the PRIO Conflict Site data, it was found that the Moran's I value was close to 1, indicating strong positive spatial autocorrelation. Accordingly, it is unwise to use PRIO Conflict Site grid cells as a dependent variable, as ordinary regression models do not control for spatial dependency. Instead, the PRIO Conflict Site cells could be used as an inclusion mask: instead of looking at all conflicts within a state, try looking at all conflicts within the conflict zone, but do not compare them with cells outside the zone. For more on spatial autocorrelation and Moran's I, see Moran, 1950; Bivand, Pebesma and Gómez-Rubio 2008. Luc Anselin's work is especially useful for identifying and dealing with spatial autocorrelation: see Anselin (1988); Anselin and Bera (1988); Anselin et al. (1996); Anselin (2001).

Figure 18 goes about here

8 Figures

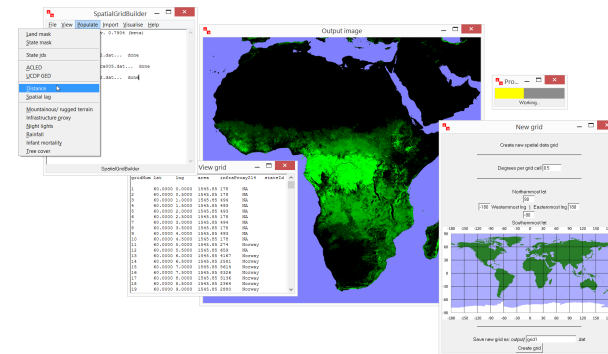


Figure 1: Typical SpatialGridBuilder user session, here showing a visualisation of tree coverage in Africa, a spatial grid of Europe and the creation of a new grid.

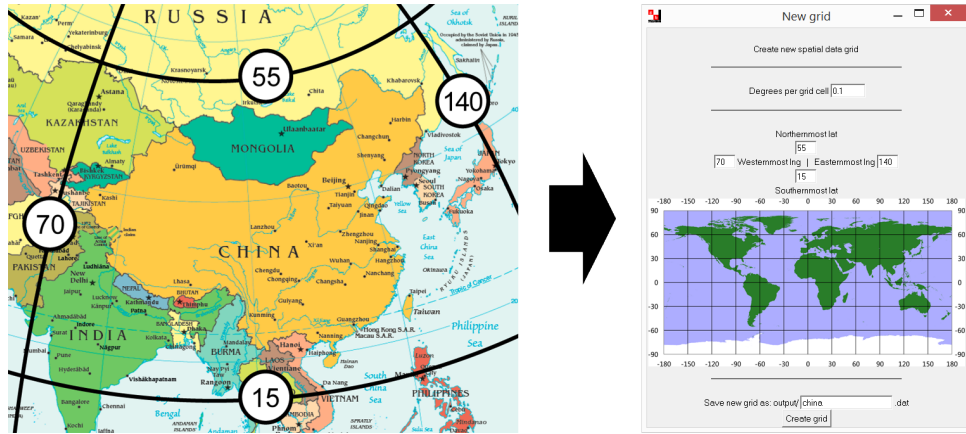


Figure 2: Getting started with SpatialGridBuilder: begin by determining the ‘bounding box’ of the area of interest (in this case, China). Map based on CIA map of Asia, 2013.

gridNum	lat	lng	area
1	55	70	70.93
2	55	70.1	70.93
3	55	70.2	70.93
4	55	70.3	70.93
5	55	70.4	70.93
6	55	70.5	70.93
7	55	70.6	70.93
...			
279999	15.1	139.8	119.4
280000	15.1	139.9	119.4

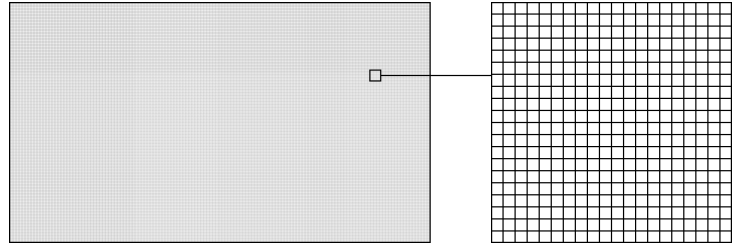


Figure 3: A newly created spatial data grid, shown as SpatialGridBuilder table output (left) and represented visually as an empty grid. At 0.1 degrees per cell, the grid is 700 cells wide by 400 high, making a total of 280,000 cells. The first cell in the grid will always be the north-west corner, moving one cell at a time first by column and then by row.

gridNum	lat	lng	area	land mask
1	55	70	70.93	1
2	55	70.1	70.93	1
3	55	70.2	70.93	1
4	55	70.3	70.93	1
5	55	70.4	70.93	1
6	55	70.5	70.93	1
7	55	70.6	70.93	1
...				
279999	15.1	139.8	119.4	0
280000	15.1	139.9	119.4	0

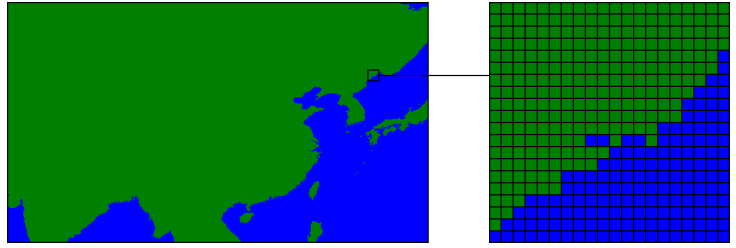


Figure 4: Land mask applied to the grid.

gridNum	lat	lng	area	land mask	CHN mask
1	55	70	70.93	1	0
2	55	70.1	70.93	1	0
3	55	70.2	70.93	1	0
4	55	70.3	70.93	1	0
5	55	70.4	70.93	1	0
6	55	70.5	70.93	1	0
7	55	70.6	70.93	1	0
...					
279999	15.1	139.8	119.4	0	0
280000	15.1	139.9	119.4	0	0

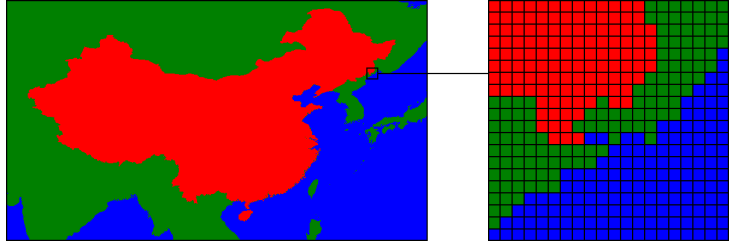


Figure 5: State mask applied to the grid (China). Data are imported by determining the country at the centre-point of each grid cell. Note that in border regions, it is possible for a cell may overlap two countries: this will be come more pronounced in lower resolution grids.

gridNum	lat	lng	area	land mask	CHN mask	flouride
1	55	70	70.93	1	0	0
2	55	70.1	70.93	1	0	0
3	55	70.2	70.93	1	0	0
4	55	70.3	70.93	1	0	0
5	55	70.4	70.93	1	0	0
6	55	70.5	70.93	1	0	0
7	55	70.6	70.93	1	0	0
...						
279999	15.1	139.8	119.4	0	0	0
280000	15.1	139.9	119.4	0	0	0

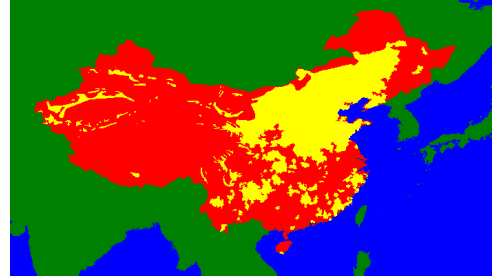


Figure 6: Adding a shapefile to the grid: now that a shapefile has been applied, we can visualise the grid as green for land, red for China, and yellow for flouride pollution.

gridNum	lat	lng	area	land mask	CHN mask	flouride	air ports
1	55	70	70.93	1	0	0	0
2	55	70.1	70.93	1	0	0	0
3	55	70.2	70.93	1	0	0	0
4	55	70.3	70.93	1	0	0	0
5	55	70.4	70.93	1	0	0	0
6	55	70.5	70.93	1	0	0	0
7	55	70.6	70.93	1	0	0	0
...							
279999	15.1	139.8	119.4	0	0	0	0
280000	15.1	139.9	119.4	0	0	0	0

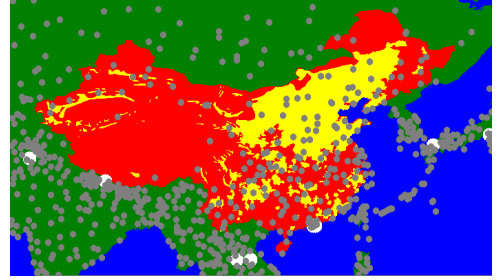


Figure 7: Applying point data to the grid: airports, based on Cazemier and van der Molen (n. d.). Note that the the size and brightness of some circles varies: SpatialGridBuilder gives the option of using weighted points when visualising data.

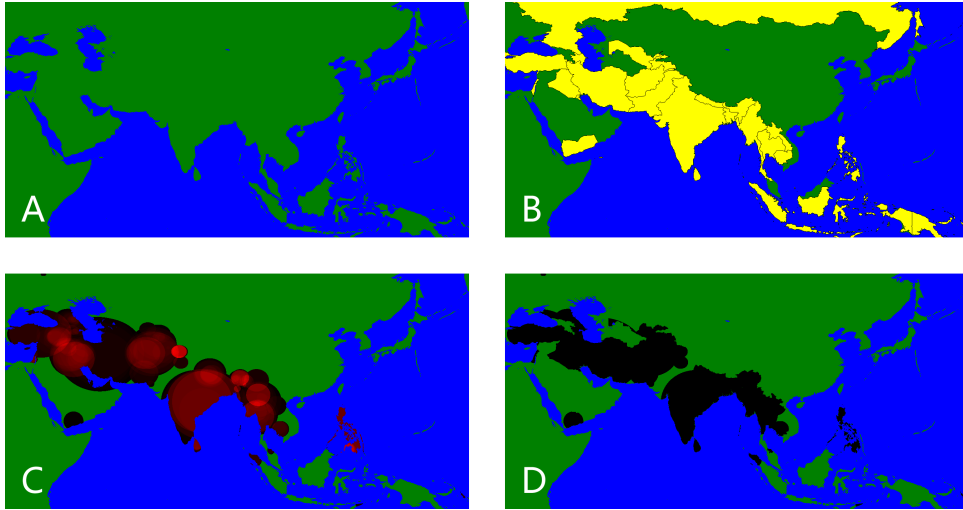


Figure 8: Asian states with conflicts in the PRIO Conflict Site 1989–2008 data set (Dittrich Hallberg 2012). ‘A’ constructs the grid at 0.1 degrees per grid cell and applies a land mask; ‘B’ creates masks for the 24 states covered by PRIO Conflict Site; ‘C’ plots the radial conflict zones (brighter colours equal more conflicts); ‘D’ merges the conflict states and the conflict zones to create a conflict-state-zone mask, thereby removing the radial zones which are outside of the PRIO Conflict Site states.



Figure 9: Distance of each grid cell to the nearest border point of China. The brighter the shade of grey, the greater the distance.

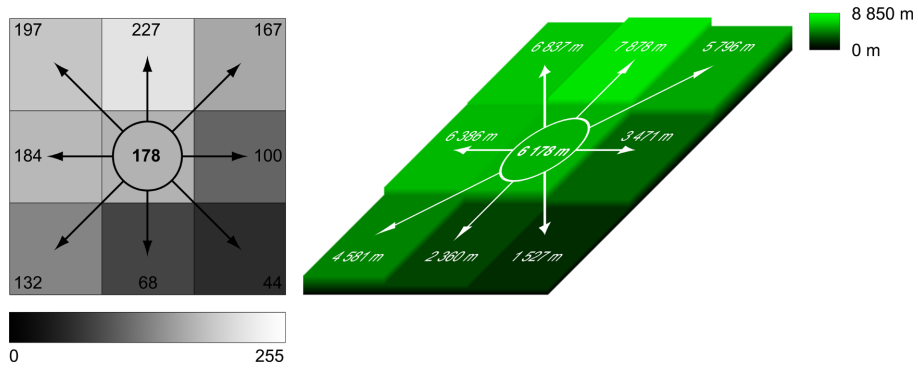


Figure 10: King's move method to determine ruggedness of terrain

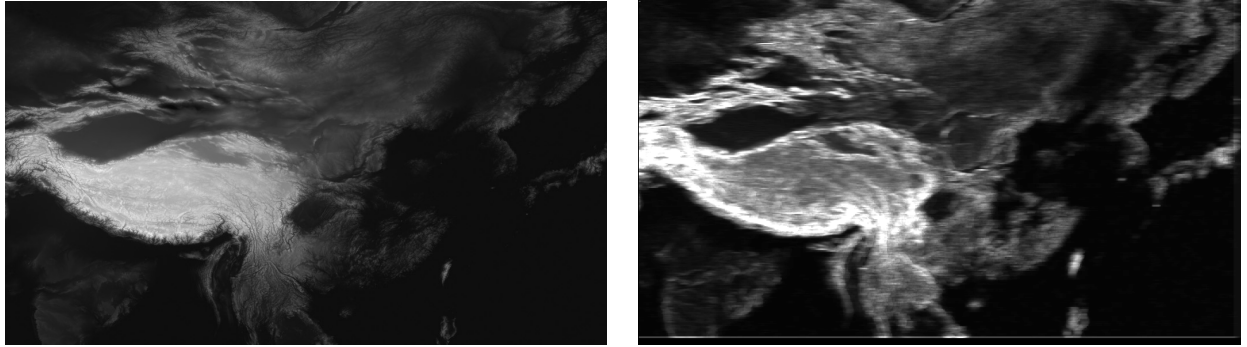


Figure 11: Simple elevation data (left), plus SpatialGridBuilder ruggedness rendering (right).



Figure 12: London, Tokyo and 'Area 51,' Nevada: Google Maps can be used as a proxy for the level of infrastructure development

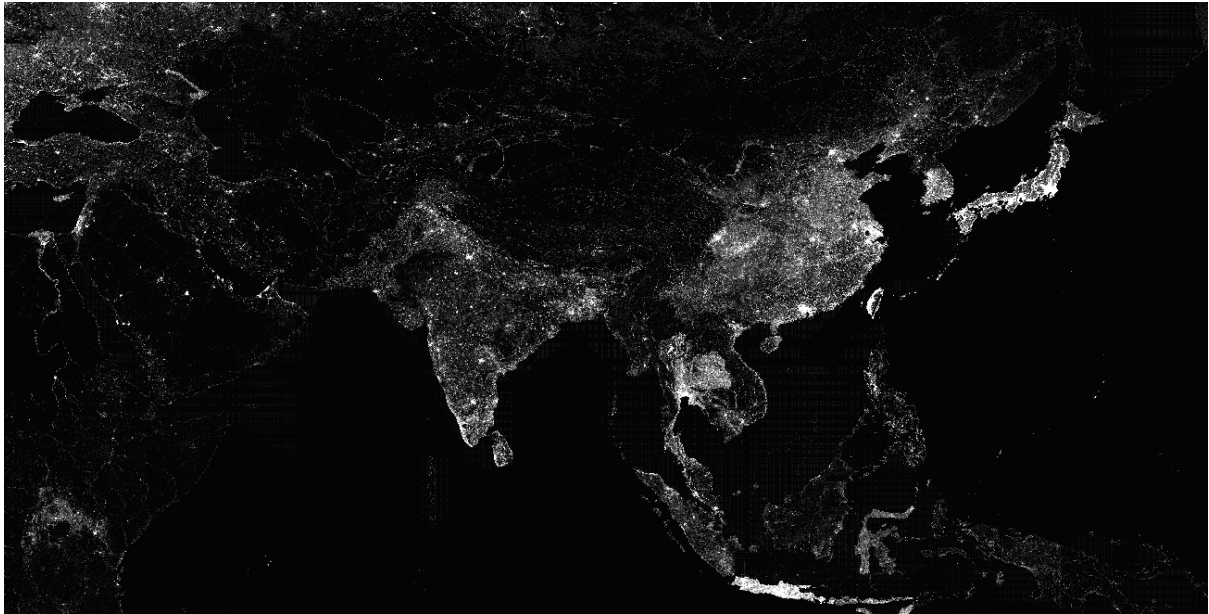


Figure 13: Infrastructure proxy variable, built by determining the file size of a separate Google map image for each individual grid cell (the construction of the grid in this image was based on an analysis of 858,000 Google map images).

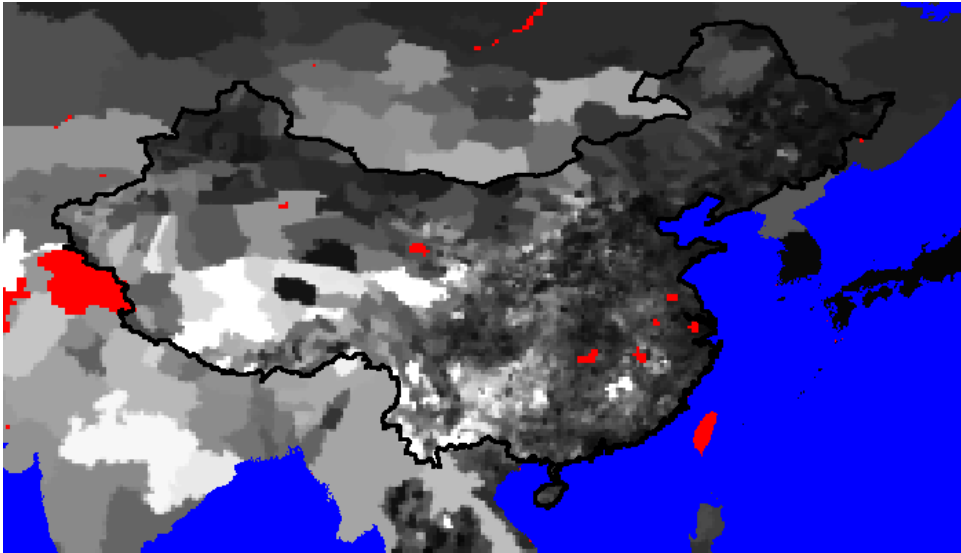


Figure 14: Infant mortality in China and surrounding countries. Brighter shade of grey indicates higher level of infant mortality. Red indicates inferred missing data.

grid Num	lat	lng	area	land_ mask	AFG_ mask	grid Num	lat	lng	area	land_ mask	AFG_ mask
1	38	60	2436.29	1	0	1	38	60	24.36	1	0
2	38	60.5	2436.29	1	0	2	38	60.05	24.36	1	0
3	38	61	2436.29	1	0	3	38	60.1	24.36	1	0
4	38	61.5	2436.29	1	0	4	38	60.15	24.36	1	0
5	38	62	2436.29	1	0	5	38	60.2	24.36	1	0
6	38	62.5	2436.29	1	0	6	38	60.25	24.36	1	0
7	38	63	2436.29	1	0	7	38	60.3	24.36	1	0
8	38	63.5	2436.29	1	0	8	38	60.35	24.36	1	0
9	38	64	2436.29	1	0	9	38	60.4	24.36	1	0
10	38	64.5	2436.29	1	0	10	38	60.45	24.36	1	0

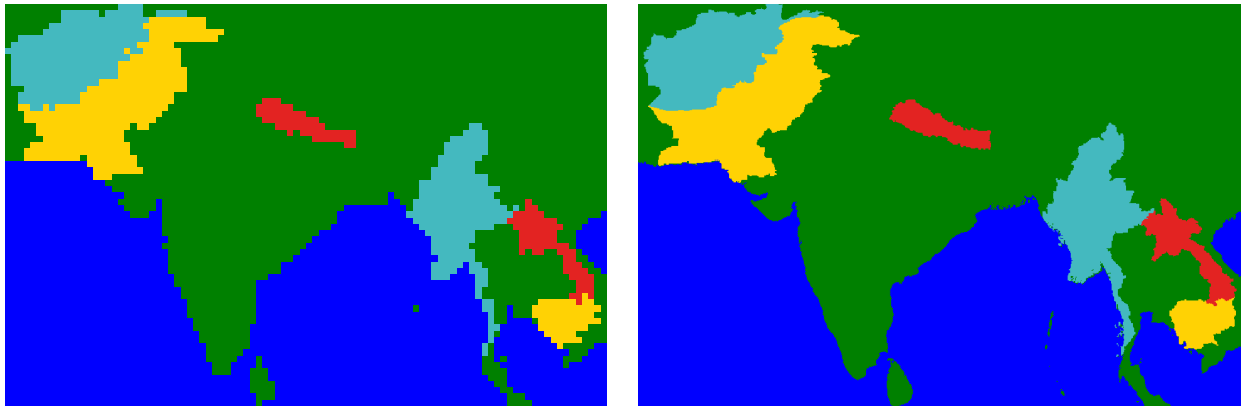


Figure 16: Low and high resolution grids of Asia identifying the six ACLED conflict states. First ten lines and just Afganistan represented in the tables for sake of space. For these grids, ISO 3166 codes have been used as state identifiers. Gleditsch-Ward and COW identifiers can also be used.

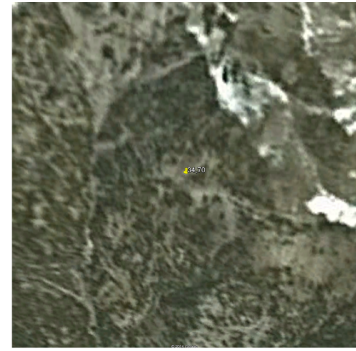
Pakistan



A: 33.8975° lat, 70.1° lng

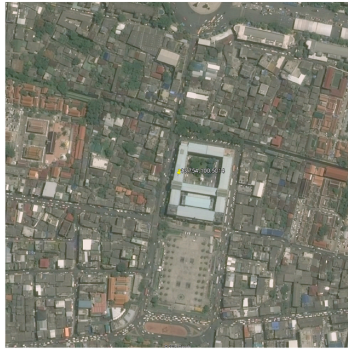


B: 33.9° lat, 70.1° lng

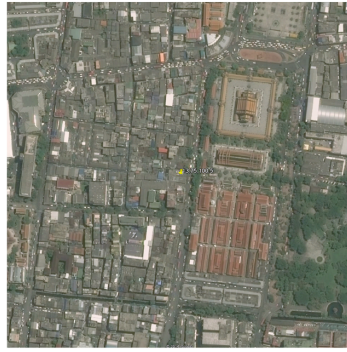


C: 34° lat, 70° lng

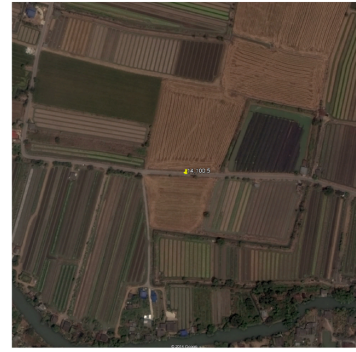
Thailand



A: 13.754° lat, 100.5014° lng



B: 13.75° lat, 100.5° lng



C: 14° lat, 100.5° lng

Figure 17: Comparison of two incidents of violence against civilians recorded in ACLED. The top row shows the location of an incident in Parachinar, Pakistan. ‘A’ indicates the actual coordinates recorded in ACLED; ‘B’ is the nearest coordinate pair in a 0.05 degree grid; ‘C’ is the nearest location in a 0.5 degree grid. The bottom row shows a separate occurrence of violence against civilians in Bangkok.

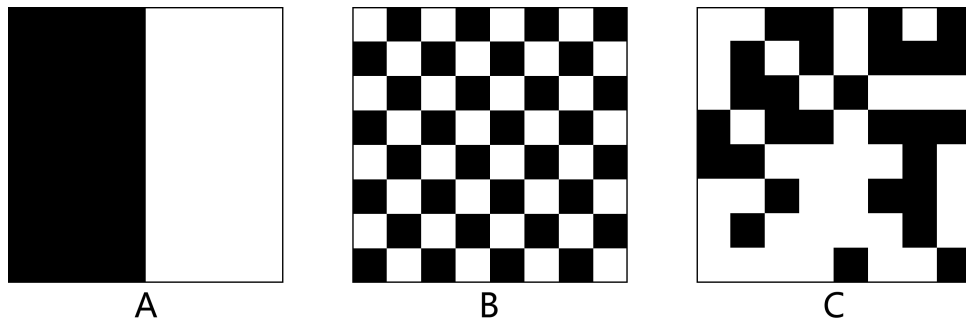


Figure 18: Explaining spatial autocorrelation. 'A' is positively autocorrelated, so would have a Moran's I value close to 1. 'B' is negatively correlated, so its Moran's I would be close to -1. 'C', the pseudo-random distribution, has a Moran's I close to 0.



Figure 15: Mathematically computed exclusive economic zone (EEZ) for China.