Effect of spatial data resolution on uncertainty

Mark Pogson*[1,2], Pete Smith[1]

[1] Institute of Biological & Environmental Sciences, University of Aberdeen, 23 St Machar Drive, Aberdeen AB24 3UU, UK

[2] Engineering, Sports and Sciences Academic Group, University of Bolton, Deane Road, Bolton BL3 5AB, UK

* Corresponding author:

email: m.pogson@bolton.ac.uk

Please cite as:

M Pogson, P Smith (2015). Effect of spatial data resolution on uncertainty. Environmental Modelling & Software 63:87–96, doi: 10.1016/j.envsoft.2014.09.021

Abstract

The effect that the resolution of spatial data has on uncertainty is important to many areas of research. In order to understand this better, the effect of changing resolution is considered for a range of data. An estimate is presented for how the average uncertainty of each grid value varies with grid size, which is shown to be in good agreement with observed uncertainties. The effect of bilinear interpolation is also investigated and is observed to provide no reduction in uncertainty relative to uninterpolated data. Finally, the effects of combining aggregated spatial data are found to obey standard properties of error propagation, which means that the presented estimate of uncertainty can be used to estimate resolution-related uncertainty in spatial model results, relative to the input data. The study quantitatively demonstrates the important role of the spatial autocorrelation of data in uncertainties associated with the resolution of spatial data.

Keywords: uncertainty; raster grid size; aggregation; resolution; interpolation; spatial autocorrelation

1. Introduction

Gridded spatial data are used as parameter inputs and outputs in all kinds of spatial models, including ecological, meteorological and hydrological (Fischer and Wang, 2011; Pogson et al., 2012; Yang et al., 2008). The spatial detail, or resolution, of data affects how well they represent reality, as well as how accurately they can be combined with other spatial data within models to make predictions. The use of gridded data, or rasters, is central to a variety of other disciplines such as biological and medical imaging (Sensen and Hallgrímsson, 2009), hence the effect of data resolution on uncertainty is important for a wide range of research. It is of particular importance to environmental modelling due to the crucial role of spatial data (Beale et al., 2010), the inherent limitations of environmental spatial data (Schmidt et al., 2006), the computational requirements of many models (Wood, 2006), and the large spatial scales often considered (Yue et al., 2011). The present study therefore considers the spatial resolution of data from the perspective of environmental modelling, but the generic formulation is applicable to any field where rescaling and combining of rasters of different resolutions is performed.

Model uncertainty, also referred to as error, can be affected by data resolution not just in terms of the uncertainty of the original data, but also by combining multiple rasters, interpolating data to higher resolutions (Stampfl et al., 2007), and lowering resolution for reasons of computation or compatibility with other rasters. Environmental models commonly take a number of spatial data inputs; for example, a soil organic matter model might require a range of meteorological, land use and soil data (Smith et al., 1997). The model then performs calculations using these data, which amount to different combinations of the data; furthermore, if the data are of different resolutions, there may also be some form of aggregation or interpolation. The way in which uncertainty in the input data is propagated by these operations is crucial to understand the resultant uncertainty in model outputs. The effects of aggregating spatial data are particularly important for spatial optimisation models, which commonly require very low resolution data due to their high computational demands (Wang et al., 2012).

Previous work has investigated the resolution which is appropriate to represent different types of data (Hengl, 2006), as well as the effect that data resolution has on results from specific models (Booji, 2005; Chaubey et al., 2005; Pisoni et al., 2010), the accuracy of particular datasets and derived values (Vaze et al., 2010), error propagation in the production of rasters from observed data (Huevelink, 1993; Lark, 2000; Knotters et al., 2010), the behaviour of metrics calculated from rasters (Stein et al., 2001) and the level of detail that can reasonably be modelled (Pogson, 2011). The effect of changing

the resolution of gridded data has also been considered in a number of ways, such as by fractal dimension (Bian, 1993). However, the way in which uncertainty changes with the resolution of gridded data has not been predicted for the general case.

This paper considers a range of data operations which are used in spatial models, namely aggregation (lowering resolution), interpolation (increasing resolution) and combination (using a number of datasets to perform calculations). The study first examines the average change in uncertainty of individual grid values caused by aggregating spatial data to lower resolutions, with the aim of predicting the mean uncertainty of values relative to the original, for any grid size. That is, if an aggregated value *v'* is used to represent a parameter value at point *x*, the aim is to predict the expected uncertainty of *v'* relative to that of the original value *v* at the same point. Consideration of interpolating rasters to higher resolutions and combining aggregated or interpolated rasters is performed in order to see how the uncertainty estimate for aggregating individual rasters can be used to estimate resolution-related uncertainty in model outputs relative to input uncertainty.

Spatial data are considered in this paper as any regular grid of values. While it is uncommon for spatial datasets to be obtained directly from observation, the consideration of uncertainty in the present study is simply relative to the uncertainty of the input data; how the data were obtained or generated does not matter for the present purposes. An estimate is presented for how uncertainty varies with resolution for different types of data distribution, thus enabling prediction of uncertainty for any aggregated grid size. The estimate is tested with a number of artificially generated rasters, as well as data from a widely-used environmental dataset. The effects of interpolating and combining rasters are then investigated by using a number of representative examples. The findings of the study enable quantitative prediction of uncertainties introduced by rescaling gridded spatial data.

2. Methods

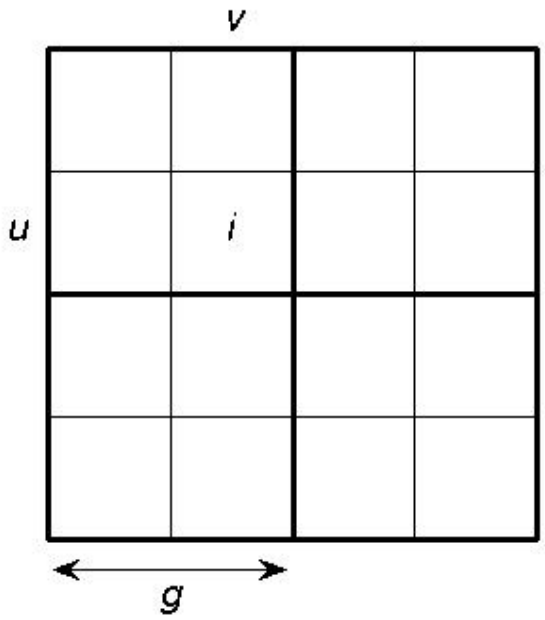
2.1 Raster and aggregation definition

We first define an n_1-by-n_2 raster A containing $N = n_1 n_2$ cells; a value n is also defined as the larger of n_1 and n_2. For convenience, each cell in A is identified by a single index i, which is a unique value combining the row index u and column index v. The mean of A is thus:

$$\overline{A} = \frac{1}{N}\sum_{i=1}^{N} A_i \qquad (1)$$

The resolution of A is reduced to an n'_1-by-n'_2 raster A', with n' the larger of n'_1 and n'_2, where $n' < n$. This is achieved by aggregating square groups of cells in A (square in terms of number, rather than distance), as commonly performed by a geographical information system (GIS); therefore each cell in A' is an aggregate of g-by-g cells in A (as shown in Fig. 1), and the grid size (the side length of each cell) in A' is g times larger than in A, so $n' = n/g$, where $g > 1$. The present study only presents results for aggregation by the mean of cells (i.e. each cell in A' is the mean of g-by-g cells in A), although use of the mode is also considered in the Discussion. No specific assumption is made of the shape of the grid cells themselves, but the shape must be such that it is not changed by the described aggregation; this is the case for any square or rectangular cells, assuming the grid is regular and aligned. Depending on the raster dimensions and rescaling factor g, some aggregated grid cells at the edge of the raster may include a non-square aggregation of cells, but for a sufficiently large raster the effects of this are not deemed important. Integer values of g are most straightforward to implement but non-integer values are also possible by appropriately weighted allocation of grid cells that span aggregation boundaries.

Fig. 1. Grid for original n-by-n raster A (faint lines, $n = 4$), and grid for aggregated n'-by-n' raster A' (heavy lines, $n' = 2$, $g = 2$). For convenience, each cell in A is identified by a single index i, a unique value which combines the row index u and column index v.



In the case of a raster which contains nulls (i.e. values which are either non-numerical or are arbitrary numbers used to signify something other than their numerical value), null values must be excluded from calculations as they cannot be aggregated with other cells (at least, not by aggregation according to the mean). In this case, N is the total number of non-null values in the raster,which means $N < n_1 n_2$ if null values are present.

2.2 Uncertainty definition

Uncertainty E for any aggregated grid size is defined in the present study as the mean absolute error between the original and aggregated raster (Witten and Frank, 2005):

$$E(g) = \frac{1}{N}\sum_{i=1}^{N}\left|A_i - A'_{i'}\right| \qquad (2)$$

where index *i'* denotes the cell in raster *A'* which corresponds to cell *i* in raster *A* (hence there are g^2 values of *i* for each *i'*). Note that although the described aggregation method is unbiased, $E > 0$ for any aggregated raster (assuming it is not uniform) due to the use of absolute differences, which do not cancel out.

The mean absolute error represents the expected difference between a cell in *A* and the corresponding cell in *A'*; it therefore gives the average magnitude of uncertainty caused by using *A'* instead of *A*. Mean absolute error is used instead of other metrics, such as mean squared error, as it provides a very tangible measure of error (i.e. the average magnitude of difference between aggregated and original values). However, the choice of metric generally has a relatively straightforward effect on results, as described in the Discussion.

2.3 Spatial autocorrelation

The effect of spatial resolution on uncertainty is clearly dependent on the spatial distribution of values within the raster. For example, if neighbouring cell values are similar to each other, the uncertainty resulting from aggregating them into a single cell would be smaller than if they were highly dissimilar.

The spatial autocorrelation of a raster is a measure of how closely related the values in the raster are, with a weighting according to the distance between cells. Moran's *I* is a widely used metric for spatial autocorrelation (Anselin, 1995; Moran, 1950), defined for raster *A* as:

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(A_i - \overline{A})(A_j - \overline{A})}{\sum_{i=1}^{N}(A_i - \overline{A})^2} \qquad (3)$$

where w_{ij} is a spatial weight between cells *i* and *j*. For the purpose of this study, the weight is defined as inversely proportional to distance (as discussed below):

$$w_{ij} = \frac{1}{1 + r_{ij}} \qquad (4)$$

where r_{ij} is the distance between cells i and j, and the value of 1 in the denominator is included to avoid division by 0 when $r_{ij} = 0$. Since all calculations are performed on a regular grid, the distance between cells is defined in terms of Taxicab geometry (Krausse, 1987):

$$r_{ij} = \left|i_u - j_u\right| + \left|i_v - j_v\right| \qquad (5)$$

where i_u and j_u are the row indices of cells i and j respectively; similarly, i_v and j_v are the column indices. Because r_{ij} is dependent only on cell indices, the method is valid for any shaped grid cell (so long as aggregation does not change cell shape), since aggregation is performed according to the number of cells rather than physical dimensions.

The choice of weight function is a compromise made to accommodate all possible length scales. Any two cells are either aggregated together (and hence should have weight function 1 for the purposes of this study), or they are not (and hence should have weight function 0). Since this would require a separate autocorrelation value to be obtained for each grid size of interest, as well as knowledge of the alignment of original and aggregated grids, it would be far less useful for making predictions. Therefore inverse proportion to distance is used instead, as it represents the decreasing probability of cells being aggregated together if they are further apart (which varies inversely with grid size), but unlike the more common inverse-square weight function, does not over-penalise cells which are distant but could still be aggregated together given a sufficiently large grid size. It is clear from this that any prediction of uncertainty will be limited to some extent, but this is for the benefit of general applicability.

By definition, values for I range from -1 (completely dispersed) to +1 (completely correlated); a value of 0 indicates no correlation. For example, a chessboard-like pattern would tend to $I = -1$ (depending on the weight function, as discussed below), a uniform area would tend to $I = 1$, and independently assigned random values would tend to $I = 0$. However, the value of I depends on the method of spatial weighting. The definition used here makes values of $I < 0$ very unlikely, since although values may be locally dissimilar, the weight function gives unusually large weighting to more distant cells, and hence differences between cells are on average smaller because distant cells are liable to be similar in value for dispersed patterns. For example, the nearest neighbours (in terms of Taxicab geometry) of a white square on a chessboard are all black, but at greater distances the distribution of

white and black squares is almost equal. This means that highly dispersed data may not be represented well by the autocorrelation method used here, but this is intentional since the weight function is chosen to represent all possible length scales. Furthermore, such patterns are uncommon in nature (Beale et al., 2010). The performance of the method on spatially dispersed data will be considered as part of the Results in order to test its potential limitations.

2.4 Estimated uncertainty

We seek to predict uncertainty E as a function of grid size g relative to the original raster, so that the uncertainty defined in Eq. 2 can be estimated for any grid size without obtaining the aggregated raster. The maximum possible uncertainty E_G is obtained by aggregating A to a single cell; i.e. using a single value to represent the entire area. Using the mean to aggregate cells, E_G is given by Eq. 2, with $A'_{i'} = \bar{A}$ for all i:

$$E_G = \frac{1}{N}\sum_{i=1}^{N}\left|A_i - \overline{A}\right| \qquad (6)$$

For any choice of grid size, uncertainty relative to the original raster will clearly lie between a minimum of 0 (when $g = 1$) and a maximum of E_G (when $g \geq n$). How $E(g)$ behaves between these values is dependent on the distribution of data in the raster.

For an uncorrelated raster ($I = 0$), the difference between $E(g)$ and E_G will change in proportion to the inverse square of grid size due to the number of aggregated cells present (since grid size is a linear measure, and aggregation is based on square groups of cells). As a general estimate, the relationship $E(g) = E_G(1-g^{-2})$ satisfies this behaviour. In particular, $E(g)=0$ when $g=1$, and $E(g)$ increases with grid size as described above, approaching a maximum of E_G for large g.

For a correlated raster ($I > 0$), $E(g)$ will approach E_G more slowly, since local fluctuations between values are smaller. Likewise, for a dispersed raster ($I < 0$), $E(g)$ will tend to E_G more quickly, since the spatial distribution of data is similar at all scales, thus $E(g) \approx E_G$ for $g > 1$. However, due to the nature of the weight function used in this study, occurrences of $I < 0$ are unlikely, as explained above.

To reflect the effect of spatial autocorrelation on uncertainty, we introduce a factor of g^I/n^I for $E(g)$, which tends to 1 as either I tends to 0 or g tends to n, hence satisfying the behaviour described above.

The approximation only holds for $I \geq 0$, but this is not considered restrictive due to the spatial weight used, and will be investigated as part of the Results. Thus for any raster *A*, we propose that the mean absolute uncertainty introduced by aggregating the raster to a larger grid size may be estimated as:

$$E(g) = E_G(1 - g^{-2})\frac{g^I}{n^I} \qquad (7)$$

This result addresses the major aim of the study, which is to predict uncertainty in aggregated rasters for any grid size. Since Eq. 7 is for the aggregation of cells, the condition exists that $g \geq 1$; this is explored further in the Results. Both *g* and *n* are dimensionless, hence the units of uncertainty *E(g)* are the same as the units of the maximum uncertainty E_G, which for the definition in Eq. 6 are the same as the units of the original values in raster *A*. Uncertainty due to aggregation *E(g)* can be added to the uncertainty of the original raster in order to obtain total uncertainty. As indicated above, any prediction of uncertainty for different grid sizes will contain flaws, in part due to characterising autocorrelation as a single value for all length scales, but also due to being unable to generically represent all possible types of data distribution, as well as implementation-specific issues relating to the alignment of grids. It is worth noting that the only part of Eq. 7 which is specific to the choice of uncertainty metric is the value of E_G, as considered in the Discussion.

In summary, Eq. 7 is formulated to satisfy the basic properties of how uncertainty is likely to change according to grid size, and is intended to work as a general estimate for the effects of aggregating spatial data. As described above, its formulation is guided in part by conjecture. Limitations in its applicability are most likely to apply to spatially dispersed rasters and rasters where the alignment of the aggregated grid is important due to the highly regular nature of the data; it should also be noted that using a single value of Moran's *I* for all length scales means Eq. 7 will not be perfectly accurate at any single grid size, as described above. The effectiveness of Eq. 7 is evaluated for a number of rasters in the Results, including cases which are expected to test its limitations.

In order to apply Eq. 7 to estimate the effect of aggregating any raster *A* to a larger grid size, only five pieces of information are required, all based on the original raster *A* itself (i.e. the aggregated raster *A'* does not need to be obtained to estimate its uncertainty):

1. The grid size rescaling factor *g*, where $g \geq 1$.
2. The number *n* of grid cells on the longer of the two sides of raster *A*.
3. The number *N* of non-null cells in raster *A*.

4. Moran's *I* for raster *A*, as described by Eq. 3-5 (note the inverse-distance definition of spatial weight in Eq. 4). If the raster is particularly large, this may be computationally expensive to calculate, and therefore it may be necessary to estimate *I* from smaller samples of the raster (as discussed further in Section 2.8).

5. The maximum possible aggregated uncertainty E_G of raster *A*. In the present study we use the mean absolute difference between each raster value and the mean raster value, as described by Eq. 6; for other metrics, please see the Discussion. E_G is typically less computationally expensive to calculate than *I*, but may still be estimated from smaller samples if necessary.

2.5 Interpolating rasters

Aggregated raster *A'* may be interpolated to the same resolution as *A* by bilinear interpolation (Buss, 2003); the uncertainty of the interpolated raster *A''* is then explicitly calculated from Eq. 2 by using it in place of *A'*; in the case of interpolating to the original raster resolution, $i' = i$. This is performed to evaluate the effectiveness of interpolation as a means to increase resolution. An example of interpolation is provided in Section 2.8, and the effect of the choice of interpolation method is considered in the Discussion.

2.6 Combining rasters

In order to investigate uncertainties when rasters are combined (such as by addition or multiplication), raster *A* may be created by combining *M* separate rasters A_m. Raster *A'* is then created by combining the *M* corresponding aggregated rasters A'_m in the same way. For example, if *A* is obtained by summing rasters A_1 and A_2 (i.e. corresponding cells in the rasters are added together), then *A'* is obtained by summing A'_1 and A'_2, which are the aggregated rasters of A_1 and A_2 respectively. The uncertainty of the combined raster is calculated from Eq. 2 as normal.

Uncertainties of combined rasters are expected to follow standard properties of error propagation. Because operations are performed on individual cells in the rasters, and each cell has an unknown uncertainty around the mean uncertainty for the raster, uncertainties are treated as independent and random. The covariance of separate rasters is assumed to be zero for the rasters under consideration in this study. Hence for sums and differences (Taylor, 1997):

$$E(g) = \sqrt{\sum_{m=1}^{M} \left(E_m(g)\right)^2} \qquad (8)$$

and for products and quotients:

$$\frac{E(g)}{\overline{A}'} = \sqrt{\sum_{m=1}^{M} \left(\frac{E_m(g)}{\overline{A}'_m} \right)^2} \qquad (9)$$

where $E_m(g)$ is the mean uncertainty of raster A'_m, $\bar{A}'_m$ is the mean of raster A'_m, and $E_m(g) << \bar{A}'_m$ in the case of Eq. 9. Similar formulae exist for other mathematical functions and for data with non-zero covariance.

2.7 Raster values

In order to investigate the effect of resolution on uncertainty for a range of spatial data distributions, the following rasters are considered, as shown in Fig. 2:

(a) Highly dispersed values in a chessboard-like pattern.

(b) Highly correlated values, defined as two distinct uniform areas.

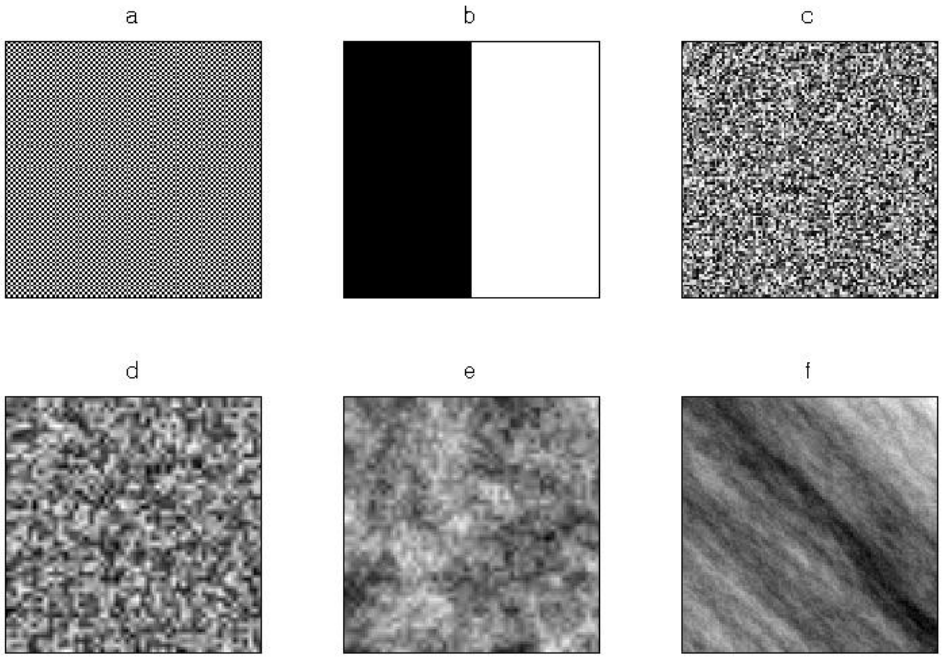
(c) Highly uncorrelated, independently assigned random values.

(d) An implementation of Perlin noise (Perlin, 1985), with large short-scale fluctuations.

(e) As (d), but with smaller short-scale fluctuations and larger long-scale fluctuations.

(f) Values defined by diagonal random fluctuation from neighbouring cells.

Fig. 2. Artificially generated rasters under consideration: (a) highly dispersed, (b) highly correlated, (c) highly uncorrelated, (d) Perlin noise with large short-scale fluctuations, (e) as (d) but with smaller short-scale fluctuations, (f) fluctuation from diagonally neighbouring cells. Values range from 0 (black) to 1 (white).



Rasters (a)-(c) are intended to provide a range of different spatial autocorrelations in order to test Eq. 7 across its parameter range, and identify potential limitations, while rasters (d)-(f) are perhaps of more interest, since they represent locally correlated but noisy values, similar to most data which are obtained from nature. All rasters are square, with $n = 100$ and $N = 10^4$. Please see the Appendix for further details of the rasters.

In order to provide an environmental example, a raster from the Harmonized World Soil Database (HWSD) version 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) is also used to test the validity of Eq. 7. Values are extracted from HWSD for soil carbon content in the top 30cm layer of the dominant soil type in each 30-by-30 arc second grid cell in the UK, as a percentage of soil mass. Like almost all environmental spatial datasets, HWSD data are obtained from a combination of practical observation and data processing, and therefore include an element of artificiality, but this is not considered important for the purposes of the study. The HWSD raster includes null values for grid cells without soil (as shown in Fig. 5a). The entire raster for the UK is 1401-by-1249 grid cells, with $N \approx 4.8\times10^5$ non-null grid cells out of a total of over 1.7×10^6 grid cells, and $n = 1401$.

2.8 Assessment methods

Each raster is aggregated to a number of larger grid sizes in order to assess the effectiveness of Eq. 7. Estimated uncertainties are compared against observed uncertainties by use of the r^2 correlation coefficient (Edwards, 1976):

$$r^2 = \frac{\left(\sum_{i=1}^{L}(o_i - \bar{o})(e_i - \bar{e})\right)^2}{\left(\sum_{i=1}^{L} o_i - \bar{o}\right)\left(\sum_{i=1}^{L} e_i - \bar{e}\right)} \qquad (10)$$

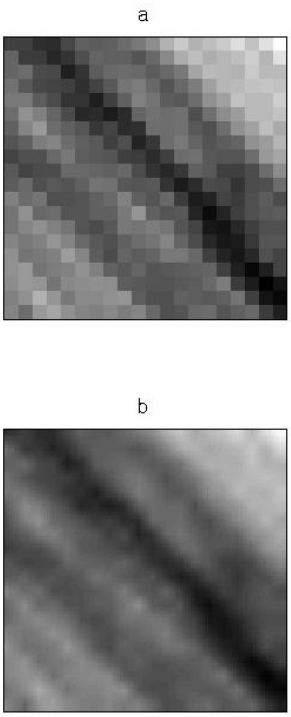
where L is the number of grid sizes considered, o_i and e_i are the observed and estimated uncertainty respectively for the i^{th} grid size, and $\bar{o}$ and $\bar{e}$ are the mean of observed and estimated uncertainties respectively. By definition, values for r^2 range from -1 (perfectly uncorrelated) to +1 (perfectly correlated). For the artificially generated rasters (a)-(f), all possible integer values of g from 1 to n are considered, hence $L = 100$. For the larger HWSD raster, 30 evenly spaced integer values of g from 1 to n are considered, hence $L = 30$.

All calculations, including raster creation and cell aggregation, are programmed in Fortran. HWSD values are extracted using ArcGIS; the resultant raster is exported as a text file which is read in by the Fortran program. Results from the Fortran program are plotted using Matlab.

For reasons of computation, the autocorrelation I for the HWSD raster is estimated from ten randomly located 10-by-10 samples of data, rather than using the whole raster. Only samples which include at least one non-null value are considered (i.e. any sample which contains only nulls is discounted and another one is taken until ten non-null samples are obtained). The mean of sampled I values is used to provide the global I value; this method is expected to give a reasonable estimate of the actual value of I for the raster. Clustered samples, rather than the same number of randomly dispersed points, are used to estimate I due to the importance of local values to the calculation of I. Due to the dimensions of each sample, only small distances relative to the total raster size are represented by the autocorrelation value; the effectiveness of this is noted in the Results and Discussion.

In order to illustrate the aggregation and interpolation methods used in this study, an example of each is shown in Fig. 3.

Fig. 3. Aggregation and interpolation of raster (f). (a) raster aggregated to $g = 5$; (b) aggregated $g = 5$ raster interpolated to original resolution by bilinear interpolation.



3. Results

3.1 Single rasters

The effect that grid size has on uncertainty (as defined in Eq. 2) for rasters (a)-(f) is shown in Fig. 4, along with the estimated uncertainty obtained from Eq. 7. For simplicity, a single instance of each raster is used, which is roughly equivalent to several instances of smaller rasters (it is not exactly equivalent due to the reduction of edge effects and the presence of greater length scales). In order to emphasise small changes in resolution, which are generally of most interest, a logarithmic scale is used for the g-axis. Uncertainty values are divided by the mean of the original raster $\bar{A}$ for generality; uncertainties are divided by $\bar{A}$ rather than E_G so that the maximum uncertainty of each raster relative to its mean can be seen. Shown also in Fig. 4 are results for bilinear interpolation of aggregated rasters back to the original grid size. Values for I are shown in Table 1, along with values for the correlation coefficient r^2 between observed and estimated uncertainty for each raster.

Fig. 4. Uncertainty as a function of aggregated grid size g for rasters (a)-(f), as shown in Fig. 2. Uncertainties are divided by the mean of each original raster for generality. Dashed line ('Obs.'):

observed uncertainty; dotted line ('Int.'): observed uncertainty following interpolation; solid line ('Est.'): estimated uncertainty according to Eq. 7. Values for I are shown in Table 1, along with values for the correlation coefficient r^2 between observed and estimated uncertainty for each raster.

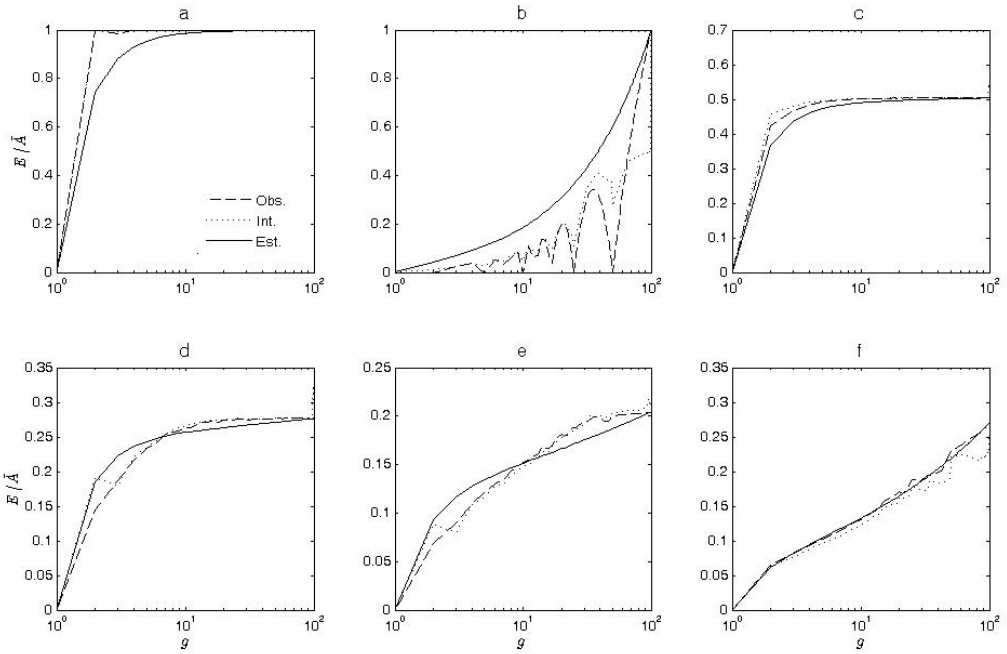


Table 1. Moran's I spatial autocorrelation value for each raster (a)-(f), and correlation coefficient r^2 between observed and estimated uncertainty.

Raster	(a)	(b)	(c)	(d)	(e)	(f)
Moran's I (as defined in Eq. 3-5)	0.002	0.726	0.008	0.029	0.126	0.304
Correlation coefficient r^2 between observed and estimated uncertainty	0.925	0.776	0.982	0.952	0.933	0.993

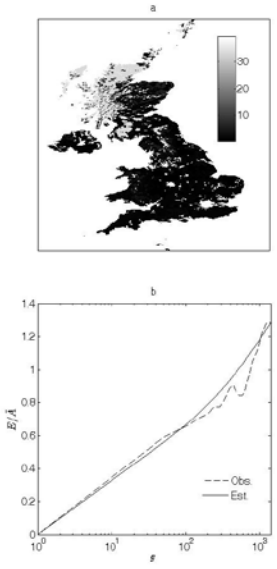
It is evident from both Fig. 4 and Table 1 that estimated uncertainty is close to observed uncertainty for all the rasters considered, and is particularly close in the case of raster (f), which is highly correlated. Very good agreement is also found in the case of raster (c), which is highly uncorrelated; this is due to the near-uniform spatial autocorrelation across the raster, without significant local patterns. Even in the case of raster (a), which is highly dispersed, estimated uncertainty is reasonably close to observed uncertainty, despite the formulation of Eq. 7 having limited applicability for spatially dispersed rasters. As expected, the uncertainty for raster (a) approaches E_G too slowly because the value of I is close to that for uncorrelated data, due to the definition of the weight function in Eq. 4. A value of $I < 0$ would of course make Eq. 7 invalid, hence the limitation is inherent to the

approximation used; furthermore, such patterns are uncommon in nature, and the issue only exists at small scales. It is noteworthy that observed uncertainties for interpolated rasters are very similar to those for uninterpolated rasters in all cases and at all resolutions, even for rasters (d) and (e) which are generated in part by bilinear interpolation.

Fluctuations in observed uncertainties in Fig. 4 are partly due to the particular spatial distributions of data at different scales, but also due to edge effects and varying commensurability between the original and sampled raster grid sizes, which changes the alignment of grids. It is for these reasons that r^2 for raster (b) is relatively low, as the highly ordered raster is particularly sensitive to these effects: uncertainty is highly dependent on whether aggregated grid cells cross the divide between uniform areas. Because such uniform patterns are unusual in nature, this issue is not considered to be important, and it is unavoidable without explicit consideration of the alignment of aggregated grids, which would be at the expense of generality.

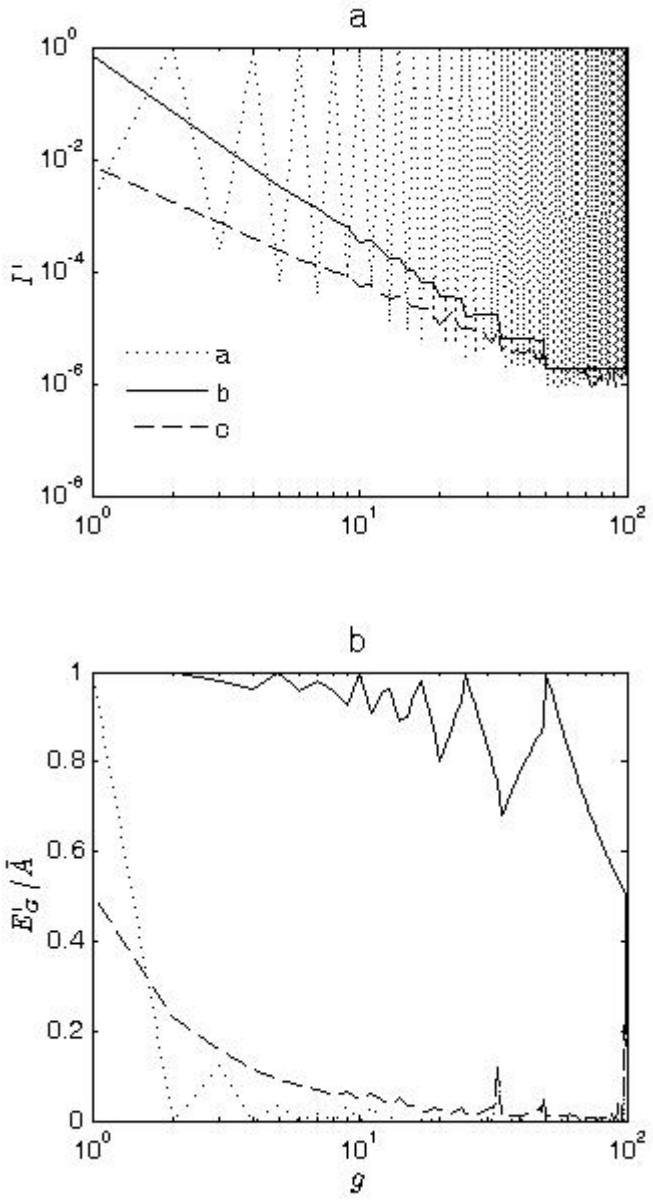
Similarly to Fig. 4 for artificially generated rasters, Fig. 5 shows the effect of grid size on uncertainty for HWSD soil carbon data in the UK. Moran's *I* for the raster is estimated as 0.25 from 10 randomly located samples, with a standard deviation of 0.27. Eq. 7 is again found to be in good agreement with observed uncertainties, with $r^2 = 0.943$; this demonstrates the ability of Eq. 7 to estimate uncertainties from spatial aggregation of environmental datasets. It also suggests that estimating Moran's *I* from small samples of the original raster is an adequate method for the purposes of Eq. 7, and it provides good results for grid sizes which are far larger than the distances used in the estimate of *I*. The distribution of HWSD data and the method to estimate *I* are considered further in the Discussion.

Fig. 5. Uncertainty as a function of grid size *g* for HWSD soil carbon values in the UK. (a) Map of soil carbon as a percentage of soil mass in top 30cm layer (n.b. white denotes null values); data from HWSD version 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). (b) Uncertainty as a function of aggregated grid size. Dashed line ('Obs.'): observed uncertainty; solid line ('Est.'): estimated uncertainty according to Eq. 7. Moran's *I* is estimated as 0.25, and the correlation coefficient between observed and estimated uncertainty is $r^2 = 0.943$.



Results in Fig. 4 and Fig. 5 show uncertainties relative to the original raster *A*. However, this provides no information on uncertainties associated with the grid size of *A* itself; i.e. what is the uncertainty of *A* relative to a grid size tending to zero. It is therefore of interest to consider how spatial autocorrelation and maximum uncertainty vary with grid size, and hence how predictions of uncertainty are dependent on the grid size of the original raster. The symbols *I'* and E'_G are introduced to represent values for sampled raster *A'* which are equivalent to *I* and E_G for raster *A*, as described in Eq. 3-6. These are plotted in Fig. 6 for rasters (a)-(c), which are used to provide a wide range of spatial autocorrelation properties.

Fig. 6. Graphs of spatial autocorrelation and maximum uncertainty as a function of grid size *g* for rasters (a)-(c). (a) Spatial autocorrelation *I'* (n.b. logarithmic scale on both axes). (b) Maximum uncertainty E'_G as a fraction of the raster mean.



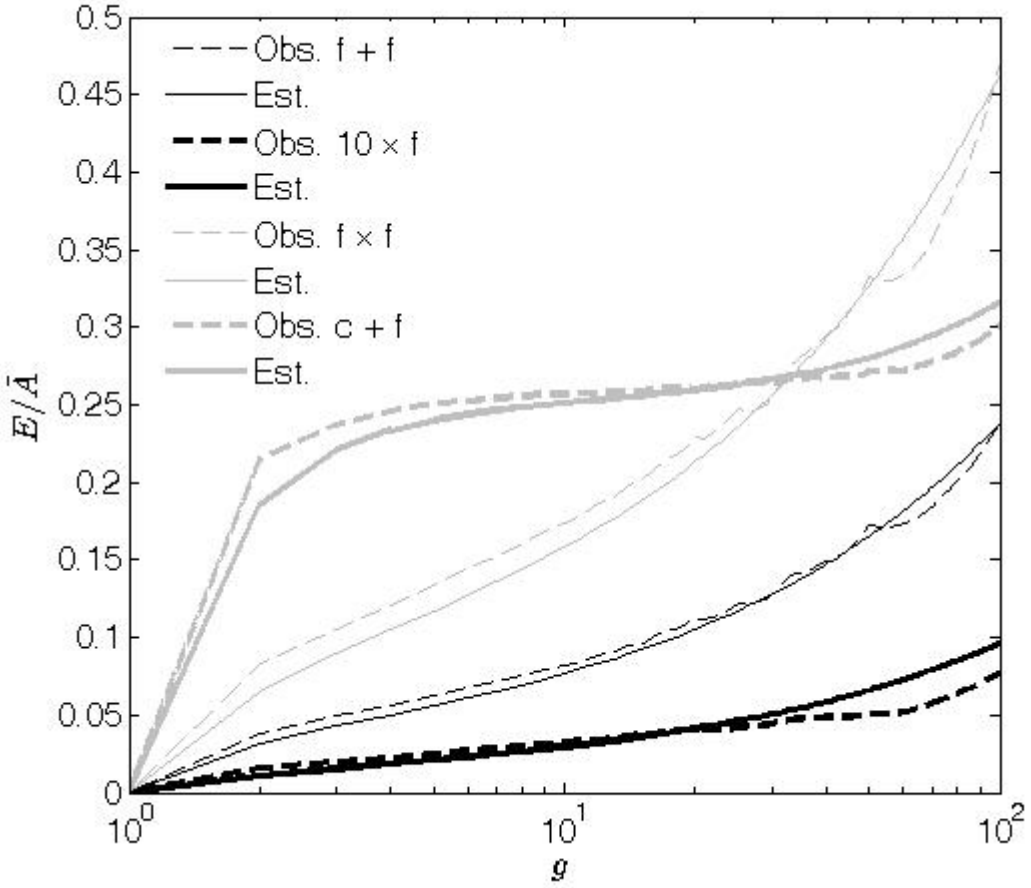
It is evident from Fig. 6 that values for I' and E'_G vary significantly with grid size. As grid size increases, I' tends to 0 (except when $g = n$, hence $n' = 1$ and therefore $I = 1$ by definition); correlation between values therefore decreases as resolution decreases. Fluctuations in I' are very large for raster (a) because odd values for g tend to preserve a chessboard pattern, while even values tend to a uniform pattern; this suggests that spatial autocorrelation for highly dispersed data is particularly sensitive to changes in grid size. Increasing grid size also causes E'_G to tend to zero. By comparison of Fig. 6b with Fig. 4a-c, E'_G appears to vary approximately as $E_G - E(g)$, where E_G is the maximum uncertainty of the original raster. This seems reasonable given the inverse dependence on the number

of grid cells present; for example, if there is only one grid cell present in *A'*, there is no difference between cells in *A'* (hence $E_G = 0$), but the average absolute difference between *A'* and each cell in *A* is at a maximum, namely E_G. The consequences of these results are considered further in the Discussion.

3.2 Combined rasters

The effects of combining multiple rasters are shown in Fig. 7. Rasters (c) and (f) are used as their uncertainties are predicted well by Eq. 7, and their autocorrelation values approximate the expected range of real-life data. Different versions of raster (f) are created by using different random seeds, hence these rasters are distinct but possess very similar values of I and E_G. Uncertainties for individual rasters are estimated from Eq. 7, and combined using Eq. 8-9, depending on the method of combination. Observed uncertainties are calculated directly from the combined rasters. Fig. 7 shows that estimates obtained from Eq.7 for individual rasters can be combined for multiple rasters by using standard formulae for error propagation, which provides close agreement with observed uncertainties.

Fig. 7. Combination of rasters as a function of grid size g. Uncertainty is shown as a fraction of the mean of the combined raster. Dashed line ('Obs.'): observed uncertainties of combined rasters; solid line ('Est.'): estimated uncertainties from individual rasters combined according to standard equations for propagation of error. The combination method and rasters are listed in the graph (n.b. 10×f means 10 different versions of raster (f) are added together, whereas f×f means 2 different versions of raster (f) are multiplied together).



4. Discussion

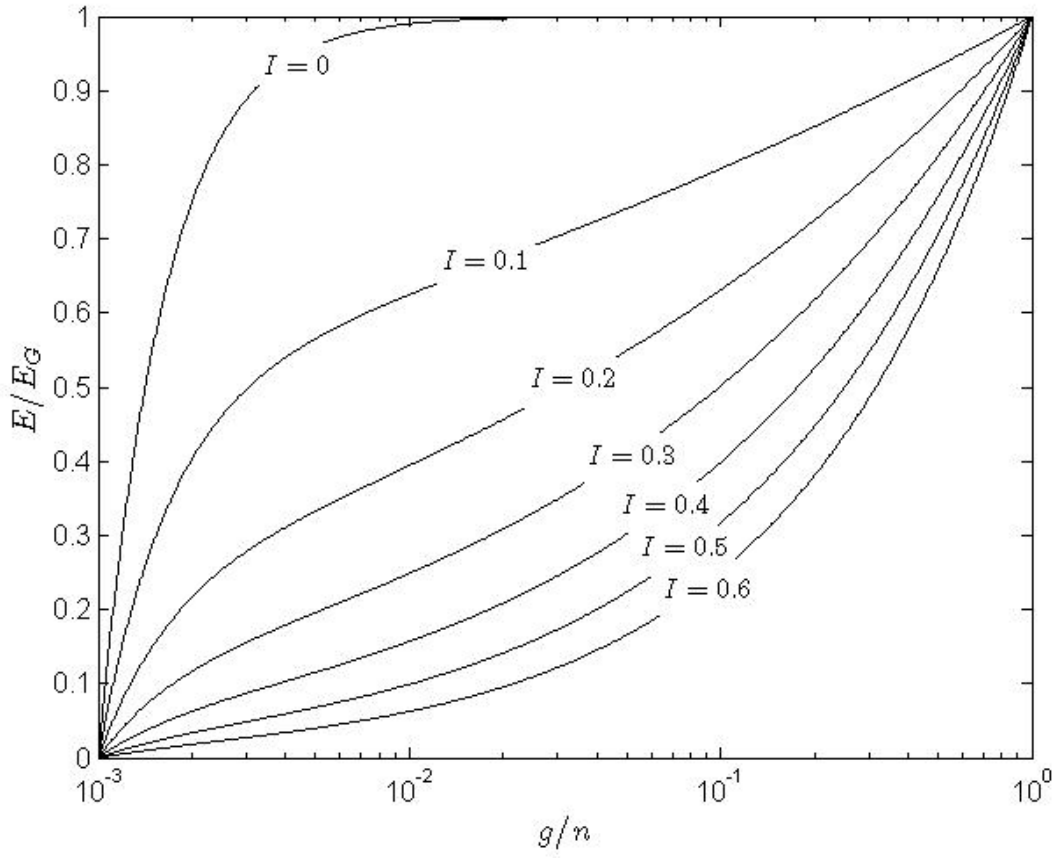
4.1 Single rasters

Results in Fig. 4 and Table 1 suggest a good match between observed uncertainties and estimates obtained from Eq. 7, with r^2 typically over 0.9 for both artificially generated rasters and environmental data based on observation. This means that for any given raster, the uncertainty caused by aggregating grid cells can be reasonably estimated for any grid size. Small discrepancies are present at different length scales due to the difficulty of defining a spatial autocorrelation weight function which is able to represent all possible grid sizes, as stated in the Methods, but the effect of this is generally small.

In order to illustrate the predictions of Eq. 7 for any raster, its results are plotted in Fig. 8 for a number of spatial autocorrelation values. To provide general applicability, uncertainties are plotted as a fraction of maximum aggregation uncertainty, and the grid size is plotted as a fraction of maximum

grid size. Fig. 8 clearly illustrates the strong effect that spatial autocorrelation has on uncertainties resulting from aggregating grid cells, and the changing effect of aggregation on uncertainty at different length scales. It is worth noting that not only does uncertainty (as a fraction of the maximum) increase faster with grid size for rasters with low spatial autocorrelation, but the maximum uncertainty itself is likely to be larger too, hence the effect of spatial autocorrelation on aggregation uncertainty is often even larger than shown.

Fig 8. Uncertainty *E* as a function of grid size *g* for any raster, as predicted by Eq. 7. Uncertainty is plotted as a fraction of maximum aggregation uncertainty E_G, and grid size as a fraction of maximum possible grid size *n*. Results are shown for a number of raster spatial autocorrelation values *I*, as labelled.



4.2 Interpolation

It is evident from Fig. 4 that bilinear interpolation has almost no effect on uncertainty relative to an uninterpolated raster, as the observed uncertainties for both are almost the same for all grid sizes (the

uncertainty for interpolated rasters would be expected to be smaller were the interpolation method effective at increasing raster resolution). This suggests that there is no benefit in using interpolation to increase the resolution of a raster, a technique which is sometimes used for compatibility with other data used in a model (Mishra et al., 2013). For example, while the raster shown in Fig. 3b appears to be of a higher resolution than the raster shown in Fig. 3a, both represent the original raster shown in Fig. 2f equally well in terms of mean absolute error. Although the present study has considered only one of a number of possible interpolation methods (Li and Heap, 2008), it is unlikely that any other method would behave significantly better (Carletti et al., 2000). The only exception to this is if further information exists for the expected spatial properties of the data, in which case this information might be able to help increase the resolution of a raster, but using information only from the raster itself is unlikely to have a beneficial effect. The definition of uncertainty used in the present study does not account for the statistical properties of the distribution of raster values, only the difference with original data at each cell location, but it is possible that interpolation might improve this. It should be noted that the main purpose of interpolation is to extrapolate data samples or estimate missing data values rather than increase resolution, although there is some overlap between the two concepts. It is clear from this that the resolution of datasets which contain interpolated values may be misleading, depending on the degree of interpolation.

4.3 Environmental data example

From Fig. 5a it is evident that there are several distinct regions in the HWSD raster in which values are similar. This explains the large standard deviation of Moran's *I* obtained from samples of the raster: local values of *I* will be large if the sample is within a single region, but small if it crosses regions. Although the size of each sample (10-by-10 cells) is very small compared with the total raster size, the global estimate of *I* appears good given the agreement of Eq. 7 with observed uncertainties, and the agreement holds at length scales far greater than the lengths considered in any sampled calculation of *I*. Because spatial weights to calculate *I* are greater for closer cells, it is clear that local values have a dominant effect on the calculation. Therefore by having a sufficiently large number of samples to capture the range of local spatial distributions of data in the raster, using only a small number of grid cells in each sample appears sufficient to provide a reasonable estimate of *I*, which is useful as the size of each sample (rather than the number of samples) is the computationally expensive part.

4.4 Choice of metric

For simplicity, results have only been presented for the mean absolute error. Results for other metrics are likely to scale very similarly since most would have the same dependence on the number of cells being aggregated. For example, using the root mean square uncertainty gives similar results (not shown for simplicity); observed uncertainties are effectively rescaled according to the maximum uncertainty E_G, which is defined in terms of the uncertainty metric being used (in the case of root mean square error, the value of E_G is simply the standard deviation of values in A). Similarly, only results for aggregation by the mean have been presented, although results for aggregation by the mode are comparable; calculation of E_G simply needs to be changed according to the aggregation method (here the mode of A would simply be used in place of $\bar{A}$ in Eq. 6). Uncertainties for aggregation by the mode tend to be slightly larger than for the mean since the mode is by definition less representative of all the cells being aggregated, particularly for highly uncorrelated data (results not shown).

4.5 Original raster resolution

Results shown in Fig. 6 suggest that the estimate of uncertainty by Eq. 7 is highly dependent on the original raster resolution, because values for spatial autocorrelation I' and maximum uncertainty E'_G vary greatly with the resolution of the raster. Estimates of uncertainty due to changes in resolution are therefore relative only to the original raster, and estimates cannot be made for grid sizes smaller than the original raster, because Eq. 7 requires values for I and E_G at the highest resolution, and these cannot be estimated from larger grid sizes. The consequence of this is that uncertainties associated with the resolution of the original raster itself (relative to the underlying parameter it represents) cannot be estimated by this method. This is unsurprising as the available information is limited by the original raster, much as interpolation is not in itself effective at increasing the resolution of a raster. However, it is possible to say qualitatively that rasters with a low spatial autocorrelation are likely to contain greater uncertainty due to resolution than rasters of the same grid size with a high spatial autocorrelation. Furthermore, an uncertainty value should ideally be provided for the original raster, based on both the resolution and methodology used to create it, in which case estimates from Eq. 7 may be added to this for uncertainties introduced by aggregation to larger grid sizes.

4.6 Combined rasters

The effect of combining aggregated rasters is investigated in Fig. 7. Uncertainties for individual rasters are estimated from Eq. 7 and combined using Eq. 8-9. These estimates are found to be in very close agreement with observed uncertainties. Estimates of uncertainty for any combination of

aggregated rasters can therefore be made by first using Eq. 7 for each individual raster, and then combined by using the appropriate equation for error propagation. This allows the aggregation-related uncertainty in model results to be estimated relative to the uncertainty in input data.

4.7 Application

The implications of Eq. 7 are of course specific to the data in question. For example, consider a raster of temperatures on a 1km grid, with E_G = 2℃, n = 1500 and I = 0.2. If A were aggregated to a 5km grid (i.e. g = 5), then according to Eq. 7 this would introduce mean absolute error E = 0.61℃. If the original data were less correlated, say I = 0.1, but with all other values equal, then E = 1.09℃ for the same grid aggregation, which is almost double the uncertainty. For the I = 0.2 data to have the same uncertainty would require an aggregated grid size of greater than 70km. This surprising result demonstrates the importance of the spatial autocorrelation of data when considering uncertainties due to changing grid size: highly uncorrelated data on a high resolution grid are liable to contain greater uncertainty due to the grid size than highly correlated data on a low resolution grid.

Although it would be possible to explicitly calculate uncertainty in aggregated rasters on a case-by-case basis, perhaps using a sampling technique to reduce computation, the method described by Eq. 7 offers a number of benefits. It is simple to apply and it only requires calculation of Moran's I; while this can be computationally intensive, it is a common calculation and it need only be performed once for a given raster. However, the method was formulated for reasons other than simplicity and computational efficiency. The method provides an insight into how uncertainty varies with grid size for any raster, and the effectiveness of Eq. 7, as demonstrated in the Results, shows that spatial autocorrelation alone is sufficient to estimate uncertainties due to spatial aggregation. Eq. 7 also provides information on the rate of change in uncertainty with changing grid size, as evident in Fig. 8. The non-linear response of uncertainty to grid size means there are ranges in which grid size can be increased with little effect on uncertainty, and others where small changes have a large effect on uncertainty, as demonstrated in the above hypothetical example with temperatures. This is valuable information in order to balance computing requirements with precision in results.

The limits of the applicability of Eq. 7 are noted in its formulation as being for rasters with spatially dispersed data, where the definition of spatial autocorrelation used in the method is less valid, and highly correlated rasters, where the alignment of the aggregated grid to the original grid is especially important; it is also noted that due to the nature of using a single spatial autocorrelation value, the

method is not perfectly accurate at any single grid size. These limitations are explored in the Results section. The extreme cases of rasters (a) and (b) show that even when the assumptions of Eq. 7 are stretched to their limits, the method still performs well overall. The accuracy of Eq. 7 tends to vary across different length scales, which can be due to changing commensurability between original and aggregated grid sizes, or local differences in spatial autocorrelation. These effects are particularly evident in the extreme examples of rasters (a) and (b), which are highly ordered; however, for natural or natural-looking data, these limitations tend to be minor, as shown by rasters (d)-(f) and the HWSD raster.

4.8 Related uncertainties

A single measure of uncertainty for aggregated rasters has been provided, but the uncertainty of individual grid cells will of course vary. Prediction of the range of uncertainties for a particular aggregated raster might be possible based on local fluctuations of Moran's *I* in the original raster, but this is beyond the scope of the present study.

It should be emphasised that resolution is only one of a number of sources of uncertainty, and a higher resolution dataset is not necessarily more accurate than a lower resolution one for the same parameter.

A benefit of higher resolution data which has not been considered in this study is that representation of spatial boundaries may be improved. However, if resolution is increased for this or any other reason, such as for convenience to combine with other data, interpolation is probably both unnecessary and misleading.

5. Conclusion

The effect that spatial resolution has on uncertainty has been investigated for a range of data distributions. An estimate of how uncertainty changes with increasing grid size has been presented and found to be in good agreement with observed uncertainties for a range of spatial data, with r^2 correlation coefficient values typically over 0.9. The estimate is straightforward to apply to any raster without any cell aggregation being performed, and it can be used for a range purposes, such as to determine the minimum grid size required for a spatial optimisation model to work within a specified uncertainty limit. It should be noted that estimates of uncertainty are relative to the original raster,

and not the underlying parameter itself; quantitative estimates of underlying uncertainties are not possible, but the study illustrates that rasters with low spatial autocorrelation are likely to contain greater uncertainty due to their resolution than rasters of the same grid size with high spatial autocorrelation.

By using a number of examples to provide a range of possible data distributions, the use of bilinear interpolation to increase the spatial resolution of data has been observed to have no significant effect on mean absolute error compared to uninterpolated data. Therefore the uncertainty of an interpolated raster can be reasonably assumed to have the same uncertainty as the original raster. The uncertainties resulting from combining spatially aggregated data obey standard properties of error propagation; this means that the presented estimate of uncertainty can be used to estimate uncertainty in spatial model results, relative to the uncertainty of the input data. The study quantitatively demonstrates that the spatial autocorrelation of data plays an important role in uncertainties associated with the resolution of spatial data, and therefore that grid size alone cannot be used to infer resolution-related uncertainties.

Acknowledgements

PS is a Royal Society-Wolfson Research Merit Award holder. This work is based on the Ecosystem Land Use Modelling & Soil Carbon GHG Flux Trial (ELUM) project, which was commissioned and funded by the Energy Technologies Institute (ETI). The authors wish to thank Dr Natasha Savage, University of Liverpool, for her helpful conversations.

References

Anselin, L., 1995. Local Indicators of Spatial Association – LISA. Geogr. Anal. 27, 93-115

Beale, C.M., Lennon, J.J, Yearsley, J.M., Brewer, M.J., Elston, D.A., 2010. Regression analysis of spatial data. Ecol. Lett. 13, 246-264

Bian, L., 1997. Multiscale nature of spatial data in scaling up environmental models, in: Quattrochi, D.A., Goodchild, M.F. (Eds.), Scale in Remote Sensing and GIS, CRC Press, pp. 13-26

Booji, M.J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. J. Hydrol. 303, 176-198

Buss, S.R., 2003. 3-D computer graphics: a mathematical introduction with OpenGL, Cambridge University Press, Cambridge, UK, pp99-108

Carletti, R., Picci, M., Romano, D., 2000. Kriging and bilinear methods for estimating spatial pattern of atmospheric pollutants. Environ. Monit. Assess. 63, 341-359

Chaubey, I., Cotter, A.S., Costello, T.A., Soerens, T.S., 2005. Effect of DEM data resolution on SWAT output uncertainty. Hydrol. Process. 19, 621-628

Edwards, A.L., 1976. An Introduction to Linear Regression and Correlation. W. H. Freeman, San Francisco, USA, pp. 33-46

FAO/IIASA/ISRIC/ISSCAS/JRC, 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria

Fischer, M.M., Wang, J., 2011. Spatial data analysis: models, methods and techniques. Springer, Heidelberg, Germany, pp1-5

Hengl, T., 2006. Finding the right pixel size. Comput. Geosci. 32, 1283-1298

Heuvelink, G.B.M., 1993. Error propagation in quantitative spatial modelling, applications in geographical information systems. Netherlands Geographical Studies, Knag/ Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht.

Knotters, M., G.B.M. Heuvelink, T. Hoogland & D.J.J. Walvoort, 2010. A disposition of interpolation techniques. Wageningen, Statutory Research Tasks Unit for Nature and the Environment, WOt-werkdocument 190.

Krause, E.F., 1987. Taxicab geometry: an adventure in non-euclidean geometry. Dover Publications, New York, USA, pp2-5

Lark, R.M., 2000. Regression analysis with spatially autocorrelated error: simulation studies and application to mapping of soil organic matter, Int. J. Geogr. Inf. Sci. 14, 247-264

Li, J., Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists. Geoscience Australia, Record 2008/23, Canberra, Australia, pp. 4-10

Mishra, U., Torn, M.S., Fingerman, K., 2013. Miscanthus biomass productivity within US croplands and its potential impact on soil organic carbon. GCB Bioenergy 5, 391-399

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17-23

Perlin, K., 1985. An image synthesizer. Comp. Graph. (Proceedings of ACM SIGGRAPH 85) 19, 287-296

Pisoni, E., Carnevale, C., Volta, M., 2010. Sensitivity to spatial resolution of modelling systems designing air quality control policies. Environ. Modell. Softw. 25, 66-73

Pogson, M., 2011. Modelling Miscanthus yields with low resolution input data. Ecol. Model. 222, 3849-3853

Pogson, M., Hastings, A., Smith, P., 2012. Sensitivity of crop model predictions to entire meteorological and soil input datasets highlights vulnerability to drought. Environ. Modell. Softw. 29, 37-43

Schimdt, C., Rounsevell M.D.A., La Jeunesse I., 2006. The limitations of spatial land use data in environmental analysis. Environ. Sci. Policy 9, 174-188

Sensen, C.W., Hallgrímsson, B., 2009. Advanced imaging in biology and medicine: technology, software environments, applications. Springer, Heidelberg, Germany, pp131-170

Smith, P., Smith, J.U., Powlson, D.S., McGill, W.B., Arah, J.R.M., Chertov, O.G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D.S., Jensen, L.S., Kelly, R.H., Klein-Gunnewiek, H., Konarov, A.S., Li, C., Molina, J.A.E., Mueller, T., Parton, W.J., Thornley, J.H.M., Whitmore, A.P., 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81, 153-225

Stampfl, P.F., Clifton-Brown, J.C., Jones, M.B., 2007. European-wide GIS-based modelling system for quantifying the feedstock from Miscanthus and the potential contribution to renewable energy targets. Glob. Change Biol. 13, 2283-2295

Stein, A., Riley, J., Halberg, N., 2001. Issues of scale for environmental indicators. Agr. Ecosyst. Environ. 87, 215-232

Taylor, J.R., 1997. An introduction to error analysis: the study of uncertainties in physical measurements. University Science Books, Sausalito, USA, pp73-79

Vaze, J., Teng, J., Spencer, G., 2010. Impact of DEM accuracy and resolution on topographic indices. Environ. Modell. Softw. 25, 1086-1098

Wang, S., Hastings, A., Smith, P., 2012. An optimization model for energy crop supply. GCB Bioenergy 4, 88-95

Witten, I.H., Frank, E., 2005. Data mining: practical machine learning tools and techniques, second edition. Morgan Kauffman Publishers, San Francisco, USA, pp176-179

Wood, S.N., 2006. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics 62, 1025-1036

Yang, Y., Atkinson, P., Ettema, D., 2008. Individual space–time activity-based modelling of infectious disease transmission within a city. J. R. Soc. Interface 6 759-772

Yue, T.-X., Jorgensen, S.E., Larocque, G.R., 2011. Progress in global ecological modelling. Ecol. Model. 222, 2172-2177

Appendix

Each raster (a)-(f) shown in Fig. 2 is an n-by-n grid of values $A_{u,v}$ where row and column indices are denoted by u and v respectively, and $n = 100$. The method to create each raster is described below. Random values $r(a, b)$ follow a uniform distribution between a and b.

Raster (a): $A_{u,v} = 1$ if u and v are both odd, or u and v are both even; else $A_{u,v} = 0$.

Raster (b): $A_{u,v} = 0$ if $v \leq 50$; else $A_{u,v} = 1$.

Raster (c): $A_{u,v} = r(0, 1)$ for all u, v.

Rasters (d) and (e): These matrices are based on a simple implementation of Perlin noise (Perlin, 1985). First define the number of 'octaves' T (where different octaves have different distances between random values in the raster), and then create an n-by-n matrix B for each octave. The final raster A is obtained from the sum of the B matrices. For each B matrix, random values are assigned to a number of points defined from a list d, with the number of points dependent on the octave. All other values in B are calculated from bilinear interpolation between these points.

For octave t, where $t = 1, 2, \ldots, T$:

$m = 2(1 + T - t)$

$p = 1+\text{int}(n^2 / m)$, where int() rounds down to the nearest integer

$d(1) = 1$

For $i = 2, 3, \ldots, p - 1$:

$d(i) = i * m$

$d(p) = n$

For $i = 1, 2, \ldots, p$:

For $j = 1, 2, \ldots, p$:

$B_{d(i),d(j)} = 2 * k * r(-0.5, 0.5)$

For all u, v not included in d:

$B_{u,v} = 2 * t *$ bilinear(d) , where bilinear(d) is bilinear interpolation between the 4 nearest points defined from the list d

Following summation of B matrices to obtain A, normalise values to the range [0, 1]:

$x = \min(A)$, where $\min(A)$ is the minimum value of $A_{u,v}$ in A

$y = \max(A) - \min(A)$, where $\max(A)$ is the maximum value of $A_{u,v}$ in A

$A_{u,v} = (A_{u,v} - x) / y$

For raster (d), $T = 3$, and for raster (e), $T = 10$. Due to the described implementation, adding octaves tends to dampen high 'frequency' (short length scale) noise, as lower frequencies dominate.

Raster (f): Values are defined sequentially from the 1st corner of the raster, in the following order.

If $(u, v) = (1, 1)$, assign a random value:

$A_{1,1} = r(-0.5, 0.5)$

Else if $u = 1$, assign a random deviation from the preceding column value in row 1:

$A_{1,v} = A_{1,v-1} + 0.1r(-0.5, 0.5)$

Else if $v = 1$, assign a random deviation from the preceding row value in column 1:

$A_{u,1} = A_{u-1,1} + 0.1r(-0.5, 0.5)$

Else assign a random deviation from the mean of preceding row and column values:

$A_{u,v} = (A_{u-1,v-1} + A_{u,v-1} + A_{u-1,v})/3 + 0.1r(-0.5, 0.5)$

Finally, normalise values to the range [0, 1], as described above for rasters (d) and (e).