# Quantifying and using spatial errors in the SRTM 30m DEM

Jonathan Rougier

School of Mathematics, University of Bristol

October 14, 2016, file compiled from `SRTM2.Rnw`

**Abstract**

Spatial error in the SRTM DEM plays a crucial role in constraining uncertainty about elevations, especially in remote or poor regions, where little other information is available. Spatial error covariance functions are estimated from a figure in the paper 'A global assessment of the SRTM performance' (Rodriguez et al., 2006, this Journal). The range of the near-field error is found to be approximately 2km. The use of these error covariance functions is demonstrated in a reanalysis of an upscaling calculation for a hydraulic model of a floodplain.

Keywords: SRTM DEM, spatial error, covariance function, flooding

## 1 Introduction

The Shuttle Radar Tomography Mission (SRTM) provides a Digital Elevation Map (DEM) at 30m resolution over a large part of the earth's surface, within latitudes $\pm 60°$. Most of the earth's inhabitants fall within its footprint, and it is a vital resource for many applications involving natural hazard risk assessment, particularly in remote or poor regions, where the SRTM DEM is the primary source of information about elevations. The original SRTM dataset is currently being reprocessed to produce an improved DEM called NASADEM, to be released in 2017. See `http://www2.jpl.nasa.gov/srtm/` for more details.

Risk assessment involves making choices under uncertainty. Where possible, all sources of uncertainty that can be quantified should be included. One such source of uncertainty for natural hazards concerns elevations, which will often provide boundary conditions for assessing the footprint of a hazard event; for example a flood, lahar, landslide, avalanche, earthquake, or wildfire (see, e.g., Rougier et al., 2013). As explained in section 2, the SRTM DEM and its spatial error structure play a crucial role in constraining uncertainty about elevations.

In a large and carefully executed study, E. Rodriguez and colleagues at the JPL (Rodriguez et al., 2005, 2006) collected and analysed data on the error of the SRTM DEM, primarily through a ground campaign covering thousands of miles, to collect 'ground truth' GPS data. Although containing a wealth of information, the published results did not explicitly quantify an object of interest to statisticians and hazard scientists, namely the near-field spatial covariance function of the SRTM DEM errors. (Typically only the near-field covariances are required, since most hazard events play out on scales of a few kilometres.) Section 3 fits spatial covariance functions to the error data contained in Figure 6 of Rodriguez et al. (2006), which also appears as Figure 3.18 of Rodriguez et al. (2005). Section 4 demonstrates the importance of these spatial covariance functions in upscaling the SRTM to larger blocks, as is often done in flood modelling.

One hopes that the ground-truth data collected by Rodriguez et al. will be used again, to assess the forthcoming NASADEM's errors. With access to the raw data, scientists at JPL will be able to provide a more accurate assessment of the NASADEM error covariance functions than given here, although possibly expressed in the same functional form (eq. 5 and Table 1). In the meantime, however, the covariance functions produced here could serve for the NASADEM as well, because, as shown below, the structure of the spatial error is large enough that it would be better to approximate it with the covariance functions from the old product, than to ignore it.

## 2 Statistical interpretation of the SRTM DEM

This short section makes a simple but conceptually important point. Consider a region, such as a floodplain, which is tiled by SRTM 30m pixels. For concreteness, and consistency with section 4, suppose that this region comprises a 9km $\times$ 9km block, i.e. a $300 \times 300$ grid of SRTM 30m pixels.

There are actually *two* statistical objects defined on this region. First, there is a prior probability distribution for the true elevation at each pixel. This might be represented by an expectation vector $m$ and a variance matrix $T$, where $m_i$ is the the expected elevation of the $i$th pixel, and $T_{ij}$ is the covariance between elevation at the $i$th pixel and at the $j$th pixel, where $i, j = 1, \ldots, 300^2$.

Second, there is the conditional distribution of the SRTM DEM given the true elevations accepting that the shuttle flights, the primary data, the JPL scientists, their algorithms and their computers together represent an inexact instrument for measuring elevation. The expected error of the SRTM DEM is taken to be zero at every pixel. But it is also necessary to specify a $(300^2 \times 300^2)$ variance matrix $\Sigma$, where $\Sigma_{ii}$ is the variance of the error for at $i$, and $\Sigma_{ij}$ is the covariance between the error at pixel $i$ and the error at pixel $j$.

Formally, the probability distribution for the true elevations is updated by conditioning on the SRTM DEM. Mathematics confirms what many people find intuitive, and perhaps accept without question. Namely, that if the prior variance matrix $T$ is *much* larger than the error variance matrix $\Sigma$, then the updated expectation of the true elevations is the SRTM DEM, and the updated variance of the true elevations is the SRTM error variance. The mathematics is given in Rougier and Zammit-Mangion (2016, notably Theorem 3), where this result is termed the 'plug-in' update. In a nutshell,

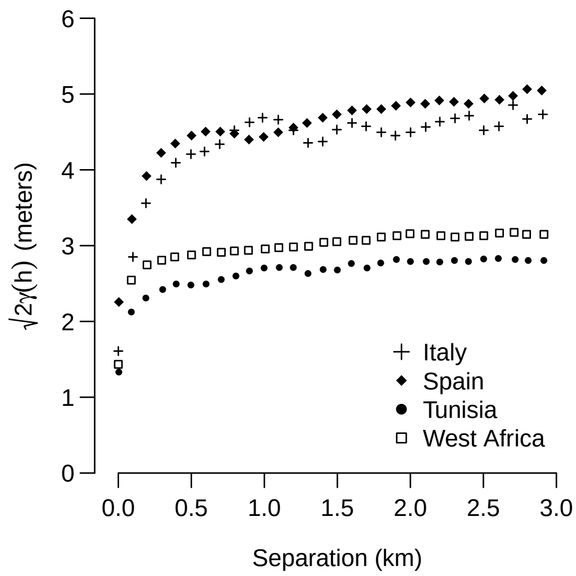


Figure 1: Digitized version of Rodriguez et al. (2006), Figure 6, re-created in the statistical computing environment R (R Core Team, 2016). The vertical axis shows the square root of the 'structure function', i.e. the square root of the empirical variogram.

> If the SRTM DEM is $\widehat{m}$ and the SRTM error variance is $\Sigma$, then, in the absence of further information, it is reasonable to represent the true elevations as having expectation $\widehat{m}$ and variance $\Sigma$.

This is not to rule out a better update for the elevations, should more information be available: it is merely the default approach.

Modern practice in spatial statistics tends not to use this default approach. Instead, an informative prior distribution is used to encode smoothness assumptions about the elevations through the structure of the prior variance matrix $T$; see, e.g., the textbooks of Banerjee et al. (2004), Rue and Held (2005), or Cressie and Wikle (2011). Zammit-Mangion et al. (2014, 2015) give a flavour of this approach, which is highly technical. In the absence of a computational statistician, the default approach seems to be a natural compromise. But it requires a function to specify the covariance between the errors at any two pixels in the region of interest.

## 3 Spatial modelling of the SRTM DEM errors

Let $\delta h(x)$ be the SRTM DEM error in metres for the 30m pixel containing location $x$. Assume that for $\|x - x'\| \leq 3$km (i.e. a distance of 100 pixels) the error process $\delta h$ can be modelled as an intrinsic stationary and isotropic process (see, e.g., Cressie, 1991, ch. 3), so that

$$\mathbb{E}\left\{\delta h(x) - \delta h(x + \Delta)\right\} = 0 \tag{1a}$$

$$\mathbb{V}\left\{\delta h(x) - \delta h(x + \Delta)\right\} = 2\gamma(d) \tag{1b}$$

where $\Delta$ is an offset, and $d := \|\Delta\|$. The function $\gamma$ is termed the 'semivariogram', and $2\gamma$ is the 'variogram'. The empirical variogram is referred to by Rodriguez et al. (2006) as the "structure function" (p. 254). Rodriguez et al. plot the empirical structure function for four regions in their Figure 6, recreated here in Figure 1.

Expanding out the lefthand side of (1b) shows that

$$\mathbb{V}\left\{\delta h(x) - \delta h(x + \Delta)\right\} = 2\left\{\,\mathrm{C}(0) - \mathrm{C}(d)\right\} \tag{2}$$

or

$$\gamma(d) = \mathrm{C}(0) - \mathrm{C}(d), \tag{3}$$

where C is the covariance function of $\delta h$, termed the 'error covariance function' below. The usual practice in Statistics is to use 'C' to denote the covariance function, although Rodriguez et al. (2005, eq. 1.2) use it for the correlation function.

Inspecting shape of the curves in Figure 1 suggests that the error covariance function has the general form of 'Nugget + Exponential'. The Exponential correlation function with range $r$ is implemented as

$$\mathrm{corr}(d; r) = \exp(-3d/r), \tag{4}$$

because $\exp(-3) = 0.050\,$(3dp); i.e., the correlation drops to 0.05 at a distance $d = r$. However, on trying to fit this model, it is clear than one Exponential term will not suffice. Instead, two Exponential terms are required, one short-range one, which is taken to have a range of 300m, and one longer-range one, which is taken to have a range of 3km. Thus the chosen

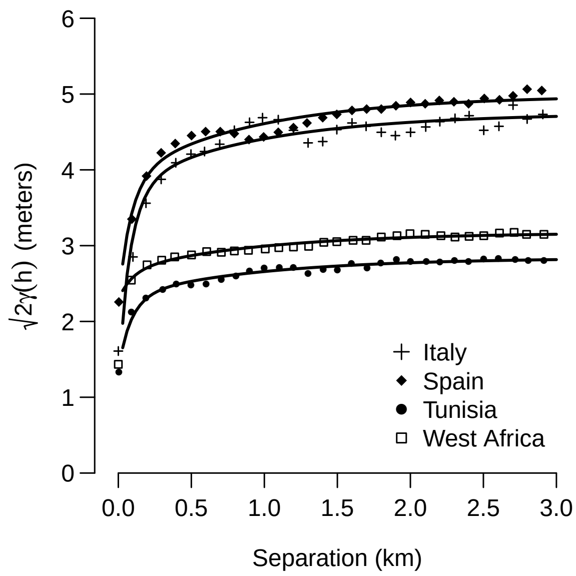


Figure 2: Like Figure 1 but with fitted covariance functions. See (5) and Table 1.

error covariance function is

$$C(d) = \sigma_0^2 \, \mathbb{1}(d = 0) + \sigma_1^2 \exp(-3d/300) + \sigma_2^2 \exp(-3d/3000). \qquad (5)$$

where $\mathbb{1}$ is the indicator function. This form, with its nugget plus two Exponential components, does not reflect just the spatial characteristics of the SRTM instrument, but also artefacts introduced by JPL scientists in post-processing the raw data, for example to reduce the mean bias.

The error covariance function in (5) is fitted to each of the curves in Figure 1, by minimising the sum of squared deviations. There is no theoretical reason for this choice of loss function, and it is retained only because the fits look plausible. The result is shown in Figure 2, with the values of the $\sigma$'s given in Table 1. The correlation functions are shown in Figure 3, showing that the range of the error is about 2km, for all four regions. The error covariance function for other regions can be chosen from the most similar of these four regions, or interpolated between them. The fitted covariance function can be tentatively extended up to the distance at which the very long-wave-length baseline roll errors start to de-correlate.

The penultimate row of Table 1 shows the estimated standard devia-

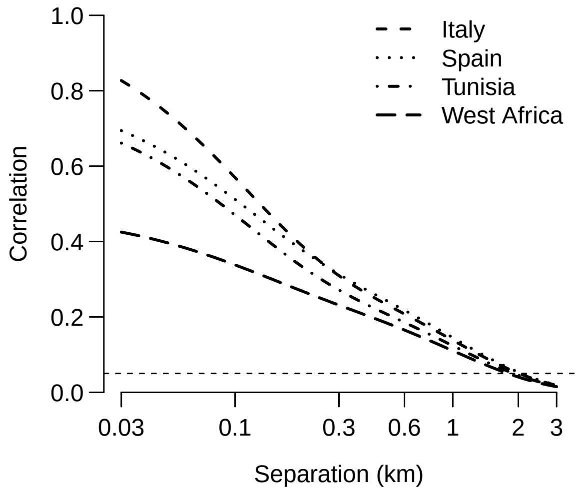


Figure 3: Fitted correlation functions, logarithmic scale. The horizontal dashed line shows that the range is about 2km in each region.

tion of the single-pixel uncertainty, and the final row shows the single pixel 90% error, which is, by convention, 1.64 times the standard deviation of the single-pixel uncertainty (Rodriguez et al., 2005, sec. 1.6.1). According to these estimates, the near-field error of the SRTM (i.e. ignoring the contribution of baseline roll errors) has a 90% error of less than 6m for the four regions Italy, Spain, Tunisia, and West Africa. This is consistent with the conclusion of of Rodriguez et al. that the SRTM DEM substantially outperforms its target of a 90% error of 16m, because the baseline roll error is only of the order of 2m (Rodriguez et al., 2006, p. 254).

## 4 Illustration: Upscaling

The SRTM DEM is widely used in hydraulic modelling of floodplains. Hydraulic models are often run at resolutions lower than 30m, and the standard procedure is to aggregate the SRTM 30m pixels into larger blocks by taking the arithmetic mean.

Neal et al. (2012, sec. 4.1.1) provide an instructive example. In this study, the requirement is to aggregate the SRTM pixels into large enough blocks that the difference in elevation between two contiguous blocks has a

Table 1: Estimated $\sigma$ values for the covariance function given in (5), in metres.

| | Italy | Spain | Tunisia | West Africa |
|---|---|---|---|---|
| $\sigma_0$ | 0.00 | 1.52 | 0.92 | 1.62 |
| $\sigma_1$ | 2.65 | 2.29 | 1.35 | 0.95 |
| $\sigma_2$ | 2.06 | 2.22 | 1.16 | 1.23 |
| Total | 3.36 | 3.53 | 2.01 | 2.24 |
| 90% err. | 5.51 | 5.78 | 3.29 | 3.68 |

95% probability of being less than 1m. Neal et al. use the approach outlined in section 2, treating the SRTM DEM error variance as uncertainty about the true elevation. But they assume that the range of the SRTM DEM spatial error is less than 90m, whereas we find it to be 2km. Thus they compute the standard deviation of the elevation of a 100-pixel block as one tenth of the standard deviation of the elevation of a one pixel block. This leads them to choose 900m blocks. This choice can be reexamined using the error covariance functions from section 3, to find the probability that two contiguous 900m blocks have a difference in elevations which is less than 1m.

First, we require a formula for the covariance of the elevations between two rectangular blocks of pixels organised into a regular grid. Let $X$ be the elevations for a $(r \times c)$ rectangular block of pixels,

$$X := \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1c} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rc} \end{pmatrix}. \tag{6}$$

Let $X(i, j)$ be the elevations for a rectangular block at grid location $(i, j)$, indexed from $(0, 0)$ in the bottom lefthand corner. Let $Y_{ij} := \text{mean}\{X(i, j)\}$, the arithmetic mean of the elevations in block $(i, j)$. Then, under stationary and isotropy,

$$\mathbb{C}(Y_{ij}, Y_{kl}) = \text{mean}\left\{\Sigma(|i-k|, |j-l|)\right\} \tag{7}$$

where '$\mathbb{C}$' is the covariance operator, and $\Sigma(i, j)$ is the $(rc \times rc)$ covariance matrix between the elevations of the pixels in block $(0, 0)$ and the elevations

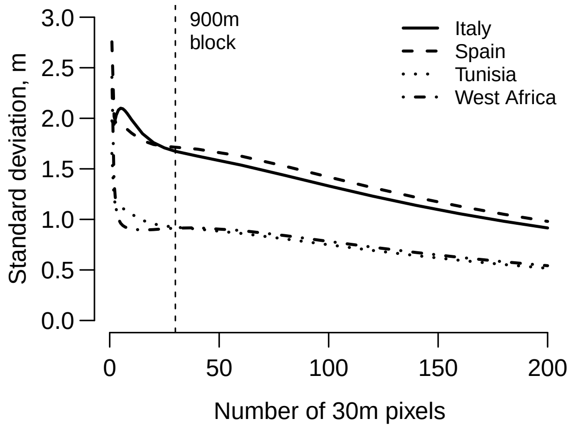


Figure 4: The standard deviation of the difference in elevation between two contiguous blocks, as a function of block width (square blocks).

of the pixels in block $(i, j)$. This result is derived in the Appendix. Adopting the approach outlined in section 2, the elements of $\Sigma(i, j)$ are computed directly from one of the error covariance functions fitted in section 3. Based on (7), the variance of the difference between elevations in two contiguous blocks is

$$\begin{aligned}\mathbb{V}(Y_{00} - Y_{10}) &= \mathbb{V}(Y_{00}) - 2\,\mathbb{C}(Y_{00}, Y_{10}) + \mathbb{V}(Y_{10}) \\ &= 2\,\text{mean}\,\big\{\Sigma(0,0) - \Sigma(1,0)\big\}. \qquad (8)\end{aligned}$$

```
## Screen chunk 'sdelev' out after running once
```

The result is shown in Figure 4, for the four regions. *A priori*, we expect these curves to have complicated shapes, because of the offsetting effects in (8). A larger block implies a smaller $\Sigma(0, 0)$, lowering the standard deviation, but also a smaller $\Sigma(1, 0)$, raising it. For blocks of 900m (a $30 \times 30$ block of 30m pixels), the difference in elevations between contiguous blocks has a standard deviation of 0.91m, based on the West Africa curve. This implies that the probability of the difference being within 1m is 0.73, using a Multivariate Normal distribution, rather than 0.95. But the Figure shows that it is hard to get this probability higher without going to huge blocks, which would result in much less physical boundary conditions. This is because the range of the SRTM DEM error is 2km, roughly 70 pixels. To drive the probability upwards towards 0.95, the block width would need to be much larger than 2km, so that most of the pixels within a block were uncorrelated with each other.

The question remains, therefore, of how to deal with substantial uncertainty about the true elevations, given that this cannot be suppressed by blocks of a physically reasonable size. One approach is to propagate elevation uncertainty through the modelling, by running the model not just once, on the upscaled SRTM DEM values, but on a set of candidate elevations all of which are consistent with the SRTM DEM values and the error covariance function. For example, suppose the floodplain is organised into 100 blocks of 900m each,

$$\begin{array}{cccc} Y_{09} & Y_{19} & \dots & Y_{99} \\ Y_{08} & Y_{18} & \dots & Y_{98} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{00} & Y_{10} & \dots & Y_{90} \end{array} \tag{9}$$

In general the blocks do not have to form a square or a rectangle; this is just for ease of visualisation in Figures 5 and 6. The variance matrix of $(Y_{00}, \dots, Y_{99})$ can be used to generate perturbations to add to the upscaled SRTM DEM, in order to provide a set of possible elevations.

Nine such perturbations are shown in Figure 5, generated using the West Africa values from Table 1 and a Multivariate Normal distribution. The line segments separate blocks whose perturbations differ by more than 1m; i.e. 1m of flood will not cross a line segment from a darker to a lighter block, if these perturbations are applied to a flat DEM.

By way of contrast, Figure 6 shows the effect of treating the blocks as uncorrelated. It is easy to see by eye that the connectivity of the two cases is completely different. This is because the range of the SRTM DEM error is longer than one 900m block, so that contiguous blocks are positively correlated, and hence the difference in their elevations is smaller than it would be were they uncorrelated. Hydraulic models of floodplains are extremely sensitive to connectivity, and therefore the presence of substantial spatial error in the SRTM DEM is an important consideration for flood risk assessment.

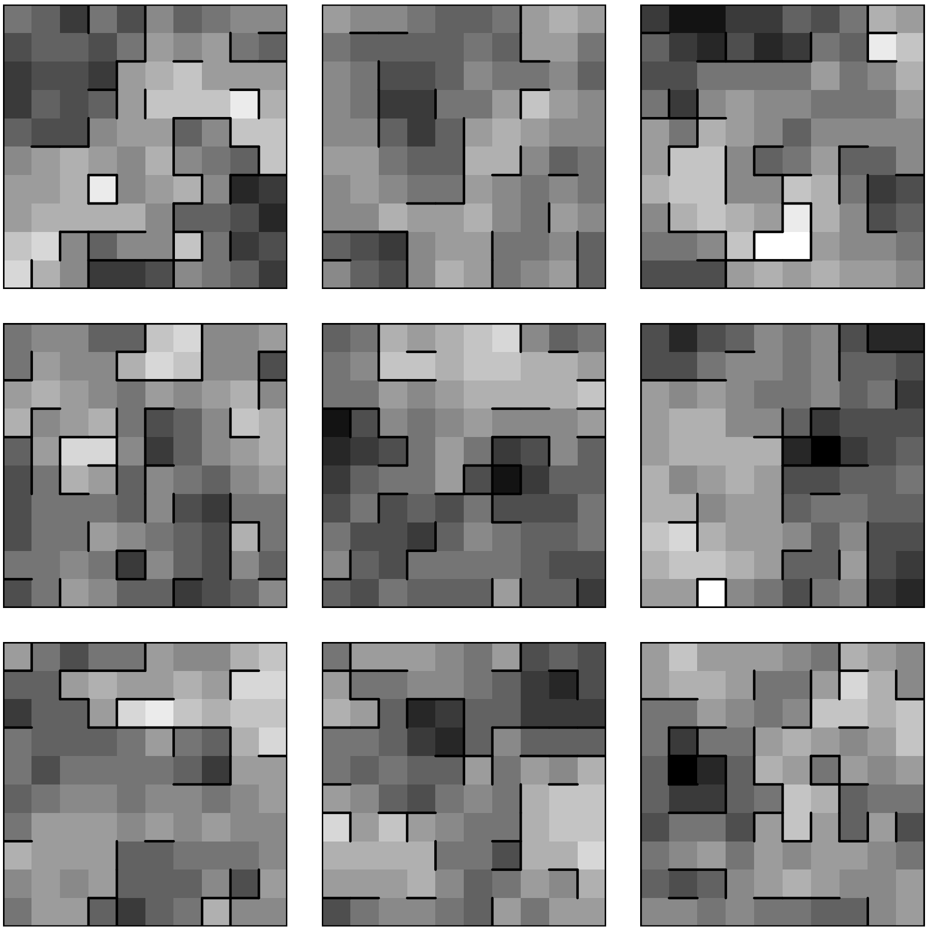

Figure 5: Nine perturbations to the SRTM DEM on upscaled blocks of 900m $\times$ 900m. Each panel represents a region of 9km $\times$ 9km. The greyscale goes from $-3$m (black) to 3m (white) in steps of 0.5m. Line segments separate blocks whose perturbations differ by more than 1m.

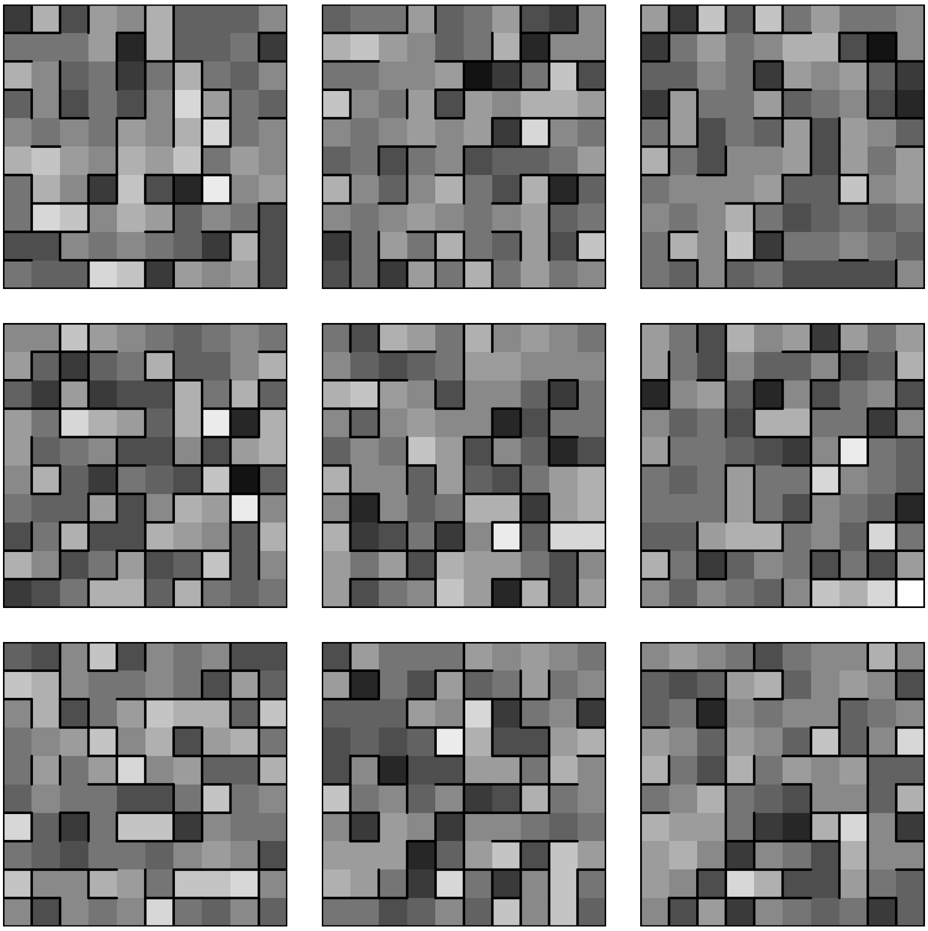

Figure 6: The same as Figure 5, except with all of the between-block covariances set to zero. The connectivity is completely different.

## A Appendix: More details on the upscaled covariances

This Appendix provides more details about (7). Let 'vec' be the operator that stacks the columns of a matrix into a vector, and define $\vec{X} := \operatorname{vec} X$. By stationarity,

$$\mathbb{C}\{\vec{X}(a,b), \vec{X}(a+i, b+j)\} \tag{10}$$

depends only on $(i,j)$; denote this covariance matrix as $\Sigma(i,j)$.

Under stationarity and isotropy, all values for $\Sigma(i,j)$ can be computed from $\Sigma(i,j)$ restricted to the case $(i,j) \geq \mathbf{0}$. For example, let $R_c$ be the permutation matrix which reverses columns in $\vec{X}$. To derive this matrix, start with $X$, an $r \times c$ rectangle of pixels. Reverse the columns in X by post-multiplying by the $c \times c$ matrix $D$, which has 1s down the anti-diagonal (i.e., $D$ is the identity matrix with the columns reversed). Then $R_c$ solves $R_c\vec{X} = \operatorname{vec}(XD)$. Because distances are preserved under reflection, and because the covariance function is stationary and isotropic,

$$\Sigma(-i,j) = \mathbb{C}\{\vec{X}(0,0), \vec{X}(-i,j)\} = \mathbb{C}\{R_c\vec{X}(0,0), R_c\vec{X}(i,j)\} = R_c\Sigma(i,j)R_c^T. \tag{11}$$

The form of $R_c$ will not be required below, beyond the fact that it is a permutation matrix.

Let $Y_{ij}$ be the upscaled elevation in block $(i,j)$, so that $Y_{ij} = \mathbf{1}^T\vec{X}(i,j)/(rc)$, where $\mathbf{1}$ is an $(rc)$-vector of ones. Then

$$\begin{aligned}
\mathbb{C}\{Y_{ij}, Y_{kl}\} &= \mathbb{C}\{Y_{00}, Y_{k-i,l-j}\} \\
&= \mathbb{C}\{\mathbf{1}^T\vec{X}(0,0)/(rc), \mathbf{1}^T\vec{X}(k-i,l-j)/(rc)\} \\
&= \frac{\mathbf{1}^T\Sigma(k-i,l-j)\mathbf{1}}{(rc)^2}.
\end{aligned} \tag{12}$$

Now suppose that $k - i < 0$ and $l - j \geq 0$. Being a sum, the numerator is invariant to permutations of the elements of $\Sigma(k-i, l-j)$. Or, to put it differently, $\mathbf{1}^T R_c = \mathbf{1}^T$. So after substituting from (11), the $R_c$ matrices in the numerator disappear, and all that happens is $k-i$ becomes $|k-i|$. Applying the same reasoning to the case where $l - j < 0$, or twice if both $k - i < 0$ and $l - j < 0$, gives

$$\mathbb{C}\{Y_{ij}, Y_{kl}\} = \frac{\mathbf{1}^T\Sigma(|k-i|, |l-j|)\mathbf{1}}{(rc)^2}, \tag{13}$$

which holds for all $(i, j, k, l)$. The expression on the righthand side is simply the arithmetic mean of the elements of $\Sigma(|k - i|, |l - j|)$, as stated in (7).

## References


Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Boca Raton, Florida: Chapman & Hall/CRC.

Cressie, N. (1991). *Statistics for Spatial Data.* New York: John Wiley & Sons.

Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data.* John Wiley & Sons, Inc., Hoboken NJ, USA.

Neal, J., Schumann, G., and Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, 48. W11506, doi:10.1029/2012WR012514.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rodriguez, E., Morris, C., and Belz, J. (2006). A global assessment of the SRTM performance. *Photogrammetric Engineering & Remote Sensing*, 72(3):249–260.

Rodriguez, E., Morris, C., Belz, J., Chaplin, E., Martin, J., Daffer, W., and Hensley, S. (2005). An assessment of the SRTM topographic products. Technical Report JPL D-31639, Jet Propulsion Laboratory, Pasadena, CA.

Rougier, J., Sparks, R., and Hill, L., editors (2013). *Risk and Uncertainty Assessment for Natural Hazards.* Cambridge University Press, Cambridge, UK.

Rougier, J. and Zammit-Mangion, A. (2016). Visualization for large-scale Gaussian updates. Forthcoming in the *Scandinavian Journal of Statistics.*

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability.* Chapman & Hall/CRC, Boca Raton FL, USA.

Zammit-Mangion, A., Rougier, J., Bamber, J., and Schön, N. (2014). Resolving the Antarctic contribution to sea-level rise: A hierarchical modelling framework. *Environmetrics*, 25:245–264.

Zammit-Mangion, A., Rougier, J., Schön, N., Lindgren, F., and Bamber, J. (2015). Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics*, 26:159–177.