Bias in the variance of gridded datasets leads to misleading conclusions about changes in climate variability

Santiago Beguería · Sergio M. Vicente-Serrano · Miquel Tomás-Burguera · Marco Maneta

November 19, 2015

Abstract Many studies addressing climate change and climate variability over large regions rely on gridded data. Grids are preferred to station-based datasets because they help avoiding bias arising from the irregular spatial distribution of the observations. However, while spatial interpolation techniques used for constructing gridded data are good at preserving the mean of the data, they do not offer an adequate representation of their variance. In fact, the grid's variance depends largely on the spatial density of observations used for constructing it. Most global and regional climate datasets are characterized by large temporal changes in the number of observations available for interpolation, with a strong reduction in the last thirty years. These changes in the sample size result in changes in the variance of gridded data that are merely an effect of the interpolation process, and ignoring this fact may lead to erroneous conclusions about changes in climate variability and extremes. We discuss this problem and we demonstrate its importance with a widely-used global dataset of temperature and precipitation. We propose to move from interpolation techniques towards statistical simulation approaches that provide a better representation of climate variability when constructing climatic grids.

Keywords gridded data · climatic grids · climate change · climate variability · spatial variance

1 Underestimation of spatial variance in gridded datasets

Most studies addressing climate variability and change at regional, hemispheric and global scales rely on gridded data. In the context of this article, gridded data z' refers to synthetic data resulting from interpolating the values of a finite set of spatially scattered individual measurements z ($z \equiv z(s), s \in A \subset \mathbb{R}^d$) on a regular mesh of points or grid ($z' \equiv z'(s), s \in B \subset \mathbb{R}^d$), as represented in Figure 1. Spatial interpolation techniques used to produce gridded datasets range from the nearest neighbor method (Hulme 1992); simple arithmetic averaging over rectangular

S. Beguería · M. Tomás-Burguera
 Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas
 (EEAD-CSIC), Zaragoza, Spain

S. M. Vicente-Serrano
 Instituto Pirenaico de Ecología-Consejo Superior de Investigaciones Científicas
 (IPE-CSIC), Zaragoza, Spain

M. Maneta
 Geosciences Department, University of Montana, Missoula MT, USA

cells (Jones et al 2012); weighted averaging such as the inverse distance method (New et al 2000; Mitchell and Jones 2005; Hansen et al 2006, 2010), the angular distance weighting method (Shepard 1968; Caesar et al 2006; Alexander et al 2006; Donat et al 2013), or other ad-hoc schemes based on climatological knowledge (Bussieres and Hogg 1989; Huffman et al 1997; Chen et al 2002; Rudolf and Schneider 2005; Huffman et al 2009; Richard 2013); linear interpolation (Harris et al 2014); regression (Daly et al 2008); thin-plate splines (New et al 1999); or kriging (Haylock et al 2008; Rigor et al 2000; Hofstra et al 2008; Cowtan and Way 2014).

The sets z and z' are discrete samples and estimates respectively of the (unknown) continuous spatial field of interest Z ($Z \equiv Z(s), s \in \mathbb{R}^d$). Gridded data have a number of advantages over irregular observational datasets that explain their wide use in climate studies. Being regular spatial structures, gridded data are more adequate for representing spatially continuous variables, and they result in more satisfying maps than those depicting (irregular) point clouds. But there are other, more substantial, reasons. When computing areal statistics such as the mean regional or global temperature (Jones et al 2012; Hansen et al 2001; Richard et al 2012) grids constitute optimal sampling structures in terms of spatial representation, avoiding biases that arise from the irregular distribution of the observations. For instance, the number of meteorological observatories tends to be related to the population density and the development level of a region. Hence, integrating directly the raw point cloud (e.g., by computing the mean of the measurements) would result in statistics biased toward the climatic conditions of highly populated and developed regions. Gridded variables are also best suited for driving agricultural, eco-hydrological or other spatially-explicit simulation models, as well as for a number of statistical techniques such as principal component analysis, and are a natural choice for the validation of global and regional climate models (Haylock et al 2008). These and possibly other reasons explain the wide and increasing use of gridded datasets of climatic variables.

Gridded data also have drawbacks that need to be understood to avoid erroneous interpretations of derived statistics. While interpolation is good at modeling the mean of the data (the expected value at each grid node), it does not offer an adequate representation of its variance (deviations from the mean values). When each grid node is assigned its expected value, the result is a field with reduced spatial variability (a smoother surface) than the (unobserved) actual field (Grimes and Pardo-Igúzquiza 2010). This is clearly the case of the interpolated field z' in Figure 1(b), which is a smoothed version of the actual field Z in 1(a). The probability density curves in Figure 1(c) show that, while the mean of Z is accurately predicted by the mean of z' , its variance is highly underestimated.

The variance of gridded variables depends largely on the size of the sample z used for interpolating the grid. (It also depends on the degree of spatial correlation or coherence of the climatic variable, but we will assume that this is constant in the remaining discussion.) This is shown in Figure 2, where a large number of realizations of z'_i were produced based on random samples z_i ($1 \leq i \leq 1000$), for three different sample sizes n . Area statistics such as the mean and standard deviation were then computed from the gridded fields z'_i , and probability curves describing the posterior distribution of these statistics over the ensemble of simulations were created. Results of this experiment show that the mean of the gridded data $\mu(z')$ is an unbiased estimator of the actual field's mean value $\mu(Z)$, the sample size only affecting the uncertainty of this estimation (Figure 2 (a)). However, the standard deviation of the gridded data $\sigma^2(z')$ is a biased estimator of the actual field's variance $\sigma^2(Z)$ (Figure 2 (b)). There is a systematic underestimation of the true field's variance, and this bias increases as the sample gets smaller. Bias in the estimation of the field's variance is propagated to other statistics that rely on it such as the fraction of the field above or below a threshold value (Figures 2 (c) and 2(d)).

Understanding the dependence of the grid variance on the number of observations is especially important when analyzing change through a time-series of grids. This is so because the number

of observations used in the interpolation very often change from one grid to the next, leading to change in the relative amount of spatial smoothing between grids and therefore affecting the sequence (time-series) of grid variances. The smoothing effect is not only relevant for computing area statistics such as in the above example, since it also affects time variances of local time-series that can be constructed at a single grid node. For example, in Figure 3 the observed time series of precipitation anomalies at one meteorological station was interpolated from nearby stations in a densely sampled region. The number of neighboring stations used in the interpolation was randomly decreased from 50 to 5 stations. Comparison of Figures 3 (b) and 3 (c) illustrates the significant alteration introduced by changing the sample size. While the original series did not show significant trends in the median, 1σ and -1σ values, the reconstructed series showed a clear reduction in the variability of precipitation, with decreasing / increasing trends in the 1σ / -1σ values. These examples motivate our claim that some conclusions based on climate statistics and trends obtained from gridded datasets using conventional interpolation methods may be biased. Here we test such hypothesis using a widely-used global climate gridded data.

2 Analysis of a global gridded dataset

It is well established that the frequency and magnitude of extreme events such as heat waves and droughts in a context of climate change depend more on changes in the variance than in the mean of climatological variables (Katz and Brown 1992; Schär et al 2004). Therefore, the question of whether climate variability is increasing with climate change is a crucial one. Recently a number of studies reported little or no increase in climate variability at the global scale, both for temperature (Huntingford et al 2013) and for precipitation (Sun et al 2012; Greve et al 2014), and no relevant change in the occurrence of extreme events such as droughts has also been highlighted (Sheffield et al 2012). Since these and other similar studies relied on gridded data, a relevant question is whether or not their analysis may be affected by variations in the number of observations during their study periods, and most notably by the data reduction found in all datasets since the 1980's.

To address this issue we conducted the analysis of a well-known and widely used global dataset: the CRU TS 3.21, consisting of monthly gridded values of several climatic variables over land at a half-degree latitude and longitude spatial resolution, for the period between 1901 and 2012 (Harris et al 2014). We focused on temperature and precipitation, although our conclusions can be extrapolated to other variables or datasets. Analysis of the size of the observational datasets used for constructing the grids reveals large temporal changes, with the number of stations peaking between the 1970's and the 1980's followed by a fast decline in the last two decades (Figure 4). The geographical distribution of the observations shows large differences within the study period (Figure 5), which in addition to the variation in the number of ground stations results in a variable proportion of cells with in-cell data (i.e., with ground station data within 0.5 degrees distance) and interpolated cells (cells with ground station data within the correlation decay distance for that variable) (Figure 6). The CRU TS dataset also uses 'dummy' stations in data-scarce regions, resulting in 'dummy grid cells' with values relaxed to their climatological normals. This dummy grid cells, which have a zero variance, may exert a large influence on the global variance of the grid, so we masked these cells out before further analysis.

In order to assess the influence of sample size variation on the global variability of gridded precipitation and temperature we transformed the monthly time series of both variables at each grid node into standardized variates with mean equal to zero and standard deviation equal to one (series of anomalies), with respect to a reference period between January 1961 and December 1990. We masked out desert areas where computation of climatic anomalies is problematic (mostly

in the precipitation dataset), and we computed the global, latitude-weighted, spatial standard deviation for each month in the gridded time series of anomalies. The area standard deviation can be seen as a measure of how spatially variable was the climate in a given month, at the global scale. That is, a high standard deviation ($\sigma(zt) > 1$) implies a large number of places experiencing unusually warm (wet) or cold (dry) conditions as compared to the local long-term climatologies, while a low standard deviation ($\sigma(zt) < 1$) implies values that were closer to the long-term averages in more grid nodes than usual.

The resulting time series are characterized by a steady increase of $\sigma(zt)$ until the 1960's to 1980's, followed by a decline (stronger in the case of precipitation) until the present (Figure 7a). These results coincide with those from a recent study reporting no increase in global temperature's variability for the period 1958–2006, but a decrease if the period since the 1980's was considered (Huntingford et al 2013). Similarly, a decreasing trend in global precipitation's variability was reported for the period 1940–2009 (Sun et al 2012).

It is of interest to determine whether these changes in the global variability of temperature and precipitation were due to changes in the left tail of the distribution (lower temperatures and less precipitation, respectively), in the right tail (higher temperatures and more precipitation), or both. This can be assessed by considering the time evolution of the fraction of the land surface with values below or above a given threshold (Figure 7b and 7c). An overall increase was found for the fraction of wet ($P > 1\sigma$) and dry ($P < -1\sigma$) land before the 1980's. In the last three decades, however, the fraction of wet land remained almost stationary while the dry land showed a decreasing pattern. This result coincides with recent studies reporting little change in global drought since 1950 (Sheffield et al 2012). With respect to temperature, a (non-linear) decay of the fraction of land under cold ($T < -1\sigma$) conditions was found, while the fraction of land under under warm ($T > 1\sigma$) conditions increased, most notably since the 1970's.

Although tempting, it would be misleading to use these results to achieve climate insight. The similarities between the time series shown in Figure 7 (a) and that of the observational sample size in Figure 4 are noticeable and they should raise concern about the previous results. There is not an exact match between both curves, since the global climate curves did not fall to their lowest values found at the beginning of the study period as it is seen in the case of the number of observations. This discrepancy may be a reflection of a real shift in the global climatology, although it could also be explained, at least in part, by changes in the geographic distribution of the meteorological observatories, which could introduce additional bias. In fact, while some very densely sampled regions such as the US or Europe recorded a drastic reduction in the number of observations in the last decades, other regions such as Asia experienced an increase in data density.

Is it possible to avoid statistical artifacts that mislead our interpretation of the global climate variability? One solution could be to restrict the analysis to only those grid cells that had at least one observation within grid-size distance, thus avoiding interpolation of data over large distances. This approach may not be valid for regional analysis since large areas are left with no data, but it has been used for analyzing changes in global and hemispheric mean temperature (Jones et al 2012).

By applying this approach to the CRU TS 3.21 dataset we obtained results that are substantially different from the previous calculations based on the complete grids. The standard deviation of global temperature and precipitation depicted a subtle but steady increase in both cases, with almost linear trends during the whole study period (Figure 8a). The change in temperature's variability was 0.0050σ per decade, while for precipitation it reached 0.0085σ , both significant at the $\alpha = 0.05$ confidence level. The increase in precipitation's variability was mostly due to an increasing fraction of land under wetter-than-normal conditions (0.0028σ per decade, p -value < 0.01), while the fraction of land under dry conditions remained stable (Figure 8b). The

fraction of land under warm conditions increased, with an accelerated pattern since the 1980's and a period of stagnation since 2000, while the cold area shrunk (Figure 8c).

Although users of gridded datasets are typically unfamiliar with (or ignore) the problem of artificial changes in variance that arise from time-varying station density this is a known problem to dataset developers, who have addressed it in several studies. For instance the CRUTEM temperature dataset, by the same authors of the CRU TS data used here, is available in its original form and as a variance-corrected version (Jones et al 2001; Brohan et al 2006; Osborn and Jones 2014). The CRUTEM dataset performs no spatial interpolation, and the grid cell values are computed as the arithmetic mean of the observations within the grid cell.

Although it can be argued that this approach is preferable to using the whole grids including spatially interpolated grid cells, it still suffers from the aforementioned problem of spatial representation. Some regions of the world are less represented than others and their relative weight in the global statistic change over time as new meteorological observatories appear and others are decommissioned. Therefore, the computed values are also biased estimations of the actual climate's variability, and they may even lead to errors on the estimation of global temperature trends (Cowtan and Way 2014).

3 From interpolation to simulation

There is clearly a need for new, better methods for constructing climatic grids. Conditional stochastic simulation has potential for producing gridded datasets that are less sensitive to the number and spatial location of the stations (Lantujoul 2002). Unlike interpolated grids, conditional simulations are semi-random representations of the expected field of a sampled variable. They include a fixed effect that depends on the values measured at neighboring observatories, a non-spatial random effect that represents the local variability (variation at micro-scales smaller than the typical sampling distances) or measurement error, and a spatial random effect. The fixed part dominates at grid nodes close to observations, while the random part becomes important as the distance between nodes and observations increases. Far from observations this random effect simulates the local variability of the actual (unobserved) field.

This can be seen in the example depicted in Figure 9a and 9b, where two random simulations of global land temperature anomalies for year 2012 are shown. Monthly temperature anomalies were calculated for the observatories in the CRU TS 3.21 dataset as standard deviations with respect to the baseline period between January 1961 and December 1990. Data series with more than 25% of missing values during the reference period were removed, representing 2.6% of the original series. Conditional Gaussian simulation using the re-parameterized and marginalized posterior sampling (RAMPS) algorithm (Smith et al 2008) was then used for obtaining one-degree grids based on the 2012 anomalies. The two realizations of the field are almost identical in observatory-dense regions such as the US or Europe, while less dense regions show larger differences between the simulated fields. The local variability is preserved in these simulations, avoiding the artificial smoothing that is typical of interpolated fields. This is achieved by constraining the simulated fields to a given spatial correlation model (usually based on a semi-variogram, i.e. a function describing the spatial dependence of a spatial random field), which is fitted to the observations. There are several approaches to conditional simulation such as the turning bands method (Mantoglou and Wilson 1982); sequential (Deutsch and Journl 1992; Pyrcz and Deutsch 2014) and truncated (Mathieu et al 1993) simulation; circulant embedding (Dietrich and Newsam 1993; Chan and Wood 1997); simulated annealing (Srinivasan and Caers 2000); or, more recently, bayesian geostatistical modeling (Yan et al 2007; Smith et al 2008; Banerjee et al 2008).

Conditionally simulated random fields have some convenient properties from the point of view of statistical inference. On the one hand, they honor the data since the simulated fields have values close to observations at grid nodes nearby observations. On the other hand, the spatial random effect is constructed in such a way that it constitutes an approximation of the expected local variability of the data. As such, simulated fields are good at representing the expected spatial variability of the data, while they are not as good at predicting the expected value of the data at nodes far from observations. In this respect, their properties are almost opposite to those of interpolated grids.

However, since a large number of simulations can be easily drawn it is possible to post-analyze an ensemble of simulated fields to derive posterior statistics. For example, the field obtained by averaging the ensemble of simulated fields at each grid node provides unbiased estimations of the values at the grid nodes, being equivalent to the grid obtained, for instance, by ordinary kriging (Figure 9c). Being an averaged field, it shows the typical smoothed surface characteristic of interpolated fields. The field obtained by computing the standard deviation of the ensemble of simulations represents the magnitude of the spatial random effect at each grid node, and it is equivalent to the so-called kriging variance field (Figure 9d).

Local (at the grid node level) integration does not offer any substantial advantages over the most sophisticated deterministic interpolation methods, such as kriging. However, area statistics such as the global mean, the global standard deviation, or the fraction of land above or below a given threshold value can be calculated for each simulation and then integrated over the ensemble of simulations. The calculations of these statistics occur for each individual simulation before the ensemble is smoothed during the integration. Unlike global statistics obtained from interpolated (smoothed) fields, probability density functions obtained from the ensemble of conditional simulations provide an unbiased estimation of those statistics. This is illustrated in Figure 10, where posterior density curves of the mean and standard deviation estimates of global land temperature in 2012 are shown. These curves were constructed from an ensemble of 100 conditional simulations such as those depicted on Figures 9a and 9b. According to these results, both the global temperature's mean and standard deviation were significantly higher than the climatological averages over the period 1960–1990. The increased variance was related to a significantly larger fraction of land with temperature higher than 1σ (54%) and a low fraction of land with values lower than -1σ (3.5%), as seen in Figures 9c and 9d.

4 Conclusion

The construction of regular spatio-temporal datasets (grids) based on irregularly located and time varying sets of climatological observations poses a number of technical problems that need be correctly understood. While spatial interpolation techniques normally used for constructing gridded datasets of climatological variables are good at preserving the mean of the data they result in smoothed surfaces compared to the reality, i.e. they underestimate the field's variance. The variance bias depends on the spatial density of the observations, so it becomes a relevant problem when analyzing time series of climatic grids, since the number and location of climatic observations they are based on very often exhibit large time variation. As a consequence, artificial variability is introduced on any analysis based on the variance of the data.

We have illustrated this problem with synthetic data, and we demonstrated its relevance in one of the most widely-used global gridded datasets, the CRU TS 3.21. We have shown that biases in the dataset arising from variations in the number of observations lead to a misleading interpretation of changes in the variability of global temperature and precipitation variability. As the demand for interpolated grids to study spatial aspects of climate change grows, the need

for better gridding techniques that do not result in a biased estimation of the field's variance also increases. We propose that statistical simulation approaches may be better than simple spatial interpolation methods for providing an adequate representation of the spatial and temporal variability of climatic data. Further work is needed in order to adapt current simulation techniques to the particularities of climate data, but the development of approximate efficient algorithms opens new options that should be explored in depth by climate scientists.

Acknowledgements This work has been supported by the research projects CGL2011-24185 and CGL2014-52135-C3-1-R, financed by the Spanish Ministerio de Economía y Competitividad (MINECO) and EU-FEDER. M. Maneta acknowledges support by the Montana Institute on Ecosystems, NSF EPSCoR Track-1 EPS-1101342, NSF GSS-1461576 and EPS-IIA-1443108 (INSTEP 3). The work of M. Tomás-Burguera was supported by a doctoral grant by MINECO. We declare that we have no competing interests.

References

- Alexander L, Zhang X, Peterson T, Caesar J, Gleason B, Klein Tank A, Haylock M, Collins D, Trewin B, Rahimzadeh F, et al (2006) Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres* (1984–2012) 111(D5)
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(4):825–848
- Brohan P, Kennedy JJ, Harris I, Tett SF, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research: Atmospheres* (1984–2012) 111(D12)
- Bussieres N, Hogg W (1989) The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. *Atmosphere-Ocean* 27(3):521–541
- Caesar J, Alexander L, Vose R (2006) Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres* (1984–2012) 111(D5)
- Chan G, Wood AT (1997) Algorithm as 312: An algorithm for simulating stationary gaussian random fields. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46(1):171–181
- Chen M, Xie P, Janowiak JE, Arkin PA (2002) Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology* 3(3):249–266
- Cowan K, Way RG (2014) Coverage bias in the hadcrut4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*
- Daly C, Halbleib M, Smith JJ, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. *International journal of climatology* 28(15):2031–2064
- Deutsch CV, Journel AG (1992) *Geostatistical software library and user's guide*. New York
- Dietrich C, Newsam G (1993) A fast and exact method for multidimensional gaussian stochastic simulations. *Water Resources Research* 29(8):2861–2869
- Donat M, Alexander L, Yang H, Durre I, Vose R, Dunn R, Willett K, Aguilar E, Brunet M, Caesar J, et al (2013) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The hadex2 dataset. *Journal of Geophysical Research: Atmospheres* 118(5):2098–2118
- Greve P, Orlowsky B, Mueller B, Sheffield J, Reichstein M, Seneviratne SI (2014) Global assessment of trends in wetting and drying over land. *Nature geoscience* 7(10):716–721
- Grimes DI, Pardo-Igúzquiza E (2010) Geostatistical analysis of rainfall. *Geographical Analysis* 42(2):136–160
- Hansen J, Ruedy R, Sato M, Imhoff M, Lawrence W, Easterling D, Peterson T, Karl T (2001) A closer look at united states and global surface temperature change. *Journal of Geophysical Research: Atmospheres* (1984–2012) 106(D20):23,947–23,963
- Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M (2006) Global temperature change. *Proceedings of the National Academy of Sciences* 103(39):14,288–14,293
- Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. *Reviews of Geophysics* 48(4)
- Harris I, Jones P, Osborn T, Lister D (2014) Updated high-resolution grids of monthly climatic observations—the cru ts3. 10 dataset. *International Journal of Climatology* 34(3):623–642
- Haylock M, Hofstra N, Klein Tank A, Klok E, Jones P, New M (2008) A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres* (1984–2012) 113(D20)

- Hofstra N, Haylock M, New M, Jones P, Frei C (2008) Comparison of six methods for the interpolation of daily, european climate data. *Journal of Geophysical Research: Atmospheres* (1984–2012) 113(D21)
- Huffman GJ, Adler RF, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneider U (1997) The global precipitation climatology project (gpcp) combined precipitation dataset. *Bulletin of the American Meteorological Society* 78(1):5–20
- Huffman GJ, Adler RF, Bolvin DT, Gu G (2009) Improving the global precipitation record: Gpcp version 2.1. *Geophysical Research Letters* 36(17)
- Hulme M (1992) A 1951–80 global land precipitation climatology for the evaluation of general circulation models. *Climate Dynamics* 7(2):57–72
- Huntingford C, Jones PD, Livina VN, Lenton TM, Cox PM (2013) No increase in global temperature variability despite changing regional patterns. *Nature*
- Jones P, Osborn T, Briffa K, Folland C, Horton E, Alexander L, Parker D, Rayner N (2001) Adjusting for sampling density in grid box land and ocean. *Journal of Geophysical Research* 106(D4):3371–3380
- Jones P, Lister D, Osborn T, Harpham C, Salmon M, Morice C (2012) Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research: Atmospheres* (1984–2012) 117(D5)
- Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. *Climatic change* 21(3):289–302
- Lantuéjoul C (2002) *Geostatistical simulation: models and algorithms*. 1139, Springer
- Mantoglou A, Wilson JL (1982) The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Research* 18(5):1379–1394
- Mathieu Y, Verdier F, Houel P, Delmas J, Beucher H (1993) A subsurface and outcrop study. In: *Subsurface Reservoir Characterization from Outcrop Observations: Proceedings of the 7th IFP Exploration and Production Research Conference Held in Scarborough, April 12-17, 1992*, Editions TECHNIP, 51, p 145
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference on Applied Climatology*, American Meteorological Society Boston, MA, vol 17, pp 179–183
- Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International journal of climatology* 25(6):693–712
- New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. part i: Development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate* 12(3):829–856
- New M, Hulme M, Jones P (2000) Representing twentieth-century space-time climate variability. part ii: Development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate* 13(13):2217–2238
- Osborn T, Jones P (2014) The crutem4 land-surface air temperature data set: construction, previous versions and dissemination via google earth. *Earth System Science Data* 6(1):61–68
- Pyrz MJ, Deutsch CV (2014) *Geostatistical reservoir modeling*. Oxford university press
- Richard A, et al (2012) A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*
- Richard M (2013) Berkeley earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview*
- Rigor IG, Colony RL, Martin S (2000) Variations in surface air temperature observations in the arctic, 1979-97. *Journal of Climate* 13(5):896–914
- Rudolf B, Schneider U (2005) Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. In: *Proc. Second Workshop of the Int. Precipitation Working Group*, pp 231–247
- Schär C, Vidale PL, Lüthi D, Frei C, Häberli C, Liniger MA, Appenzeller C (2004) The role of increasing temperature variability in european summer heatwaves. *Nature* 427(6972):332–336
- Sheffield J, Wood EF, Roderick ML (2012) Little change in global drought over the past 60 years. *Nature* 491(7424):435–438
- Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*, ACM, pp 517–524
- Smith BJ, Yan J, Cowles MK (2008) Unified geostatistical modeling for data fusion and spatial heteroskedasticity with r package ramps. *Journal of Statistical Software* 25(10):1–21
- Srinivasan S, Caers J (2000) Conditioning reservoir models to dynamic data: A forward modeling perspective. In: *SPE annual technical conference*
- Sun F, Roderick ML, Farquhar GD (2012) Changes in the variability of global land precipitation. *Geophysical Research Letters* 39(19)
- Yan J, Cowles MK, Wang S, Armstrong MP (2007) Parallelizing mcmc for bayesian spatiotemporal geostatistical models. *Statistics and Computing* 17(4):323–335

Fig. 1: A (second order stationary) random spatial field Z representing the (unknown) spatial distribution of a climatic variable (a); a gridded spatial field z' obtained by ordinary kriging from a finite random sample z (circles) (b); probability density curves of Z and z' (c). Despite its discrete nature (point estimates), the gridded field z' is represented as a continuous surface (single-coloured cells) for convenience.

Fig. 2: Probability density curves of: (a) the mean $\mu(Z')$ and (b) standard deviation $\sigma(Z')$ of 1000 gridded spatial fields Z' , from random samples of varying size n ; (c) fraction of gridded spatial fields Z' with values above $1\sigma(Z')$ and (d) below $-1\sigma(Z')$. The red vertical lines indicate the true values of each statistic in spatial field Z .

Fig. 3: Station network surrounding station 5036 in the CRU TS 3.21 raw dataset (near Springfield, Illinois, in red) (a); observed monthly time series of precipitation anomalies (standardized precipitation index, SPI (McKee et al 1993)) at station 5036 (b); monthly times series of SPI at station 5036, reconstructed by ordinary kriging from a progressively decreasing number of neighboring stations (c). Black lines on (b) and (c) show linear quantile regression trends corresponding to 1, 0 and -1 SPI.

(a)(b)

Fig. 4: Monthly time series of the number of ground stations in CRU TS 3.21 dataset for (a) precipitation and (b) temperature, based on the station count files (*.st0).

Fig. 5: Geographic distribution of precipitation stations in the CRU TS 3.21 precipitation dataset: 1901 (a); 1921 (b); 1941 (c); 1961 (d); 1981 (e); 2012 (f). Colors in (b)–(f) represent the number of stations per 1x1 degree grid node: one (grey), two (yellow), three (orange), four or more (red).

(a)(b)

Fig. 6: Monthly time series of the number of grid cells with in-cell, interpolated and dummy data in the (a) precipitation and in the (b) temperature grids. Based on the analysis of the station count files (*.st0) and on the 1961-1990 normal climatologies of the CRU TS dataset.

(a)

(b)

Fig. 7: Monthly time series of: (a) global precipitation and temperature standard deviation (in standardized units); (b) fraction of the land surface under wet ($P > 1\sigma$) and dry ($P < -1\sigma$) conditions; (c) fraction of the land surface under warm ($T > 1\sigma$) and cold ($T < -1\sigma$) conditions. Monthly values (dots) and loess-smoothed curves are shown for each variable. Based on the CRU TS 3.21 dataset, with dummy-cells masked out.

(a)

(b)

Fig. 8: Same as Figures 7a, 7b and 7c, but considering only the grid cells with station data within one grid-size distance (0.5 degrees).

Fig. 9: Land temperature anomalies in 2012: (a) and (b), two gaussian simulations conditioned to the observations in the CRU TS 3.21 raw dataset; (c) average of 100 simulations, equivalent to kriging mean; (d) standard deviation of 100 simulations, equivalent to kriging variance.

Fig. 10: Posterior density curves of areal statistics of global land temperature anomalies in 2012, based on 100 conditional simulations: (a) field's mean; (b) field's standard deviation; fraction of land with values (c) below and (d) above 1σ . The curves were cut on the extremes to include 95% of the density.

Underestimation of spatial variance in gridded datasets leads to biased conclusions about changes in climate variability

Online supplementary material

Santiago Beguería · Sergio M. Vicente-Serrano · Miquel
Tomás-Burguera · Marco Maneta

June 24, 2015

1 Effect of the sample size on the interpolated field's variance

Although spatial interpolators are (in general) good estimators of the field's expected value, they are poor estimators of the field's variance. In this example we demonstrate the effect of the sample size in both the mean and the variance of the field.

Let's consider a (second order stationary) random spatial field $Z = Z(s), s \in \mathbb{R}^d$ such as the one depicted in Figure 1. Z is completely characterized by its mean m , variance σ and correlation function ρ . Z and its statistical properties are unknown, but we have a finite sample of Z at points $z = z(s), s \in V$ inside a spatial domain $V \subset \mathbb{R}^d$ such as the one depicted in Figure 1. The variable Z follows (approximately) a Gaussian distribution, i.e. Gaussian with zero mean and unit variance. The field Z was generated using (unconditional) Gaussian simulation, with a spatial correlation structure defined by a spherical semi-variogram [1]. Let's consider the field Z to be the 'true' and unknown variable that we want to estimate based on a sample $z_i = (z_1, z_2, \dots, z_n)$ consisting of n measurements taken at random locations z_j within the field's spatial domain. We can generate a high number $m = 1000$ of such samples, and approximate the field Z'_i by interpolating each sample using, for instance, ordinary kriging (OK). Each Z'_i is an optimum estimation of the real (but unknown) field Z , based on the information drawn from each sample Z_i . Finally, we can estimate the density curves of global statistics estimated from the Z'_i fields. We repeat this operation for different sample sizes, n .

The result of these calculations is shown on Figure 2. Two things become apparent: The first one is that, irrespective of the sample size, OK is a good (unbiased) estimator of the mean of Z . The second one is that the variance of Z is underestimated by OK for all sample sizes. The underestimation increases as the number of points used in the interpolation decreases. It is of interest to note that the estimated variance does not increase linearly with the number of points in the sample. This can be seen from the fact that to achieve a similar degree of increase

S. Beguería · M. Tomás-Burguera
Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas
(EEAD-CSIC), Zaragoza, Spain

S. M. Vicente-Serrano
Instituto Pirenaico de Ecología-Consejo Superior de Investigaciones Científicas
(IPE-CSIC), Zaragoza, Spain

Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas
(EEAD-CSIC), Zaragoza, Spain · M. Maneta
Geosciences Department, University of Montana, Missoula MT, USA

in the variance the sample size first goes from 100 to 250 (a factor of 2.5) but then it must increase from 250 to 1000 samples (i.e. a factor of 4). The reason for this is that the samples are not independent due to spatial correlation present in Z , and thus any new observations carry progressively less information as the sample gets larger. Biases in the estimation of the field's variance is propagated to other statistics that rely on the spatial variability, such as the percent of land above or below a certain threshold value.

2 Effect of dummy grid cells on the variance bias of the CRU TS 3.21 dataset

The CRU TS dataset uses dummy stations with zero anomalies (i.e., they are set to the local climatological mean) in regions and months where no ground observations exist within the correlation decay distance for the variable of interest. Use of dummy data has of course a big impact on the variance of the grid on data scarce regions and on the whole dataset, as stated by the authors of the dataset [2]. It is not possible to completely exclude the effect of these dummy stations, since their influence spreads to nearby grid cells during the interpolation process. However, it is possible to identify those 'dummy grid cells' by comparing their values with their corresponding monthly climatologies for the period 1961-1990 (also provided as part of the CRU TS dataset). To avoid excessive variance bias related to the dummy cells we masked them out to conduct our analysis (as seen on Figure 7 of the main article).

It is interesting to compare these results with those using the whole dataset, i.e. including the dummy grid cells. The results shown on Figure 3 are based on the complete CRU TS 3.21 dataset, and can be compared to those on Figure 7 of the main article. Comparison of the two figures reveals little changes, indicating that use of dummy cells in the CRU TS dataset does not have a large effect on the overall variance of the dataset. Although this may seem counter-intuitive, it is probably an effect of the relatively low number of dummy cells, as compared to the more abundant interpolated cells.

3 Divergent evolution of the distribution tails for temperature and precipitation

It is of interest to determine whether changes in the global variability of temperature and precipitation are due to changes in the left tail of the distribution (lower temperatures and less precipitation, respectively), in the right tail (higher temperatures and more precipitation), or in both. This can be assessed, for example, by calculating the fraction of the grid with values below or above a certain threshold.

The monthly time series of the fraction of land surface under wet and dry conditions, as defined by values of the *SPI* below -1 and above 1 , respectively, reveals a diverging time pattern (Figure 4). When considering the complete grid, there is a very marked increase in the fraction of wet land, especially since the 1980's, while the fraction of land under dry conditions seemed to shrink after a period of increase. This last result is similar to recent studies reporting little change in global drought since 1950 [3]. When using only the grid nodes with in-cell station data, however, the acceleration of the wet land trend is replaced by a steady linear increase (0.0028σ per decade, $p\text{-value} < 0.01$), while the fraction land on dry conditions remains stable. Hence, the observed increase in the global precipitation variability is due mostly to an increase in the wet areas.

With respect to temperature, a fluctuating decay is found for the fraction of land under cold conditions, especially when only grid nodes with in-cell data are considered (Figure 5). At the same time, an increasing trend is found in the fraction of land under warm conditions. This

opposing trend of cold and warm areas is especially notorious in the period between 1980 and 2000. In the last decade both curves show a steady behavior in the last decade.

These results, however, should be interpreted with care, since the grid nodes with in-cell station data are not evenly distributed across space leading to a problem of spatial representation (see explanation in the main article).

4 Conditional simulation of global land temperature anomalies in 2012

Unlike standard gridding methods such as regression, Delaunay triangulation, inverse distance weighting, kriging, etcetera that produce a single output, stochastic or simulation methods generate a large number of plausible realizations of the spatial field that is being approximated. These plausible realizations capture the local structure (variability) of the field that is lost in standard gridding methods. For instance, in Figure 7 we show two conditional simulations of the temperature anomalies over land for year 2012. These simulations are the sum of a fixed effect that depends on the values measured at neighboring stations as determined by a spatial correlation model, a non-spatial random effect (measurement error plus variation at micro-scales smaller than the typical sampling distances), and a spatial random effect. The spatial correlation model is usually determined by a semi-variogram, a function describing the spatial dependence of a spatial random field, that may include a nugget effect representing random variation at zero distances (measurement error plus local variation, Figure). Once a particular semi-variogram model is chosen (there are many different models with varying properties), it is fitted to the observations. Then, They were obtained using the re-parameterized and marginalized posterior sampling (RAMPS) algorithm [4,5].

If measurement errors are low, the fixed effect dominates at grid nodes around observations, resulting in regions of low variability between realizations of the random field. As the distance between nodes and observations increases, so does the relative effect of the spatial random term. Far from observations this random effect simulates the local variability of the actual (unobserved) field. Clearly, each realization of the random field will be different, especially at locations far from observations, but the multiple realizations can be used to determine a better (unsmoothed) approximation of the variability of the actual field.

Conditionally simulated random fields have some nice properties from the point of view of statistical inference. On the one hand, they honor the data since the gridded field will have values close to observations in nodes nearby stations. On the other hand, the spatial random effect is constructed in such a way that constitutes the best approximation of the expected spatial variability of the data. This is usually achieved by fitting an adequate variogram model (i.e. a model of the variance of the field as a function of the spatial distance between points) to the data. As such, simulated fields are good at representing the expected spatial variability of the data, while they are not as good at predicting the expected value of the data at nodes which are far from the observations. In this respect, simulated grids have properties that are almost opposite to those of deterministic grids.

Since a large number of equally-probable simulations can be easily drawn, it is possible to post-analyze the ensemble of simulated fields to derive statistics, such as expected values and uncertainties for each grid node (Figure 8). The map of the mean obtained by averaging all the simulated fields provides an unbiased estimation of the expected values at the grid nodes, and it is equivalent to the map of expected values obtained, for instance, by ordinary kriging. Such map shows the typical smoothed surface characteristic of interpolated fields. The map of the standard deviation of the simulated fields is related to the local uncertainty, i.e. it represents the magnitude of the spatial random effect at each grid node. As expected, the local standard deviation is

relatively low in areas well covered by observations and increases in data-poor regions. The map of the standard deviation of the simulations is equivalent to the so-called kriging variance field.

Global statistics can be calculated from the simulated fields, too (Figure 9). Posterior probability density functions of statistics such as the global mean, the global standard deviation, or the fraction of land above or below a given threshold value can be obtained from the ensemble of realizations. A key aspect is that unlike global statistics obtained from interpolated fields, which carry a biased estimation of the field's variability that is a function of the number of observations, probability density functions obtained from the ensemble of conditional simulations provide an unbiased estimation (provided the model characterizing the spatial variability is well determined). The only effect of the number of observations is on the uncertainty related to the global statistics (i.e., the variance of their posterior distributions), but not on the bias.

References

1. Cressie N.A.C. *Statistics for Spatial Data*. Wiley, John & Sons, Inc. (1993).
2. Harris I., Jones P., Osborn T., Lister D. Updated high-resolution grids of monthly climatic observations – The CRU TS 3.10 dataset. *International Journal of Climatology* **34**(3), 623–642 (2014).
3. Sheffield J., Wood E.F., Roderick M.L. Little change in global drought over the past 60 years. *Nature* **491**, 435–438 (2012).
4. Yan J., Cowles M.K., Wang S., Armstrong M. Parallelizing MCMC for Bayesian Spatiotemporal Geostatistical Models, *Statistics and Computing* **17**, 323–335 (2007).
5. Smith B. J., Yan J., Cowles M. K. Unified Geostatistical Modeling for Data Fusion and Spatial Heteroskedasticity with R Package ramps, *Journal of Statistical Software* **25**, 1–21 (2008).

Fig. 1: A random Gaussian spatial field Z .

Fig. 2: Probability density curves of global statistics of Z as estimated from random samples of varying size n : (a) mean; (b) standard deviation; fraction of land with values (c) below and (d) above 1σ . Each curve is constructed from $m = 1000$ fields interpolated by OK from the same number of random samples Z_i . Red vertical lines indicate the true values of each statistic, as computed from the field Z .

Fig. 3: Monthly time series of: (a) global precipitation and temperature standard deviation (in standardized units); (b) fraction of the land surface under wet ($P > 1\sigma$) and dry ($P < -1\sigma$) conditions; (c) fraction of the land surface under warm ($T > 1\sigma$) and cold ($T < -1\sigma$) conditions. Monthly values (dots) and loess-smoothed curves are shown for each variable. As opposed to Figures 6 and 7 of the main article, these figures are based on all the grid cells in the CRU TS 3.21 dataset.

(a)

(b)

Fig. 4: Fraction of the land surface experiencing wet ($SPI > 1$) and dry ($SPI < -1$) conditions, based on (a) all the grid cells except the dummy cells; and (b) only on grid nodes with in-cell station data. Monthly values (dots) and loess-smoothed curves.

(a)

(b)

Fig. 5: Monthly time series between 1901 and 2012 of the fraction of the land surface experiencing cold and warm conditions, based on (a) all the grid cells except the dummy cells; and (b) only on grid nodes with in-cell station data. Monthly values (dots) and loess-smoothed curves.

Fig. 6: Empirical semi-variogram (dots) and fitted spherical semi-variogram model (dashed line) for the 2012 global temperature anomalies. Semi-variogram parameters: 1806 km (range), 0.683 (partial sill), 0.526 (nugget).

Fig. 7: Two random Gaussian simulations conditioned to the observed temperature anomalies over land in year 2012. Based on CRU TS 3.21 raw dataset, were the anomalies were calculated as standard deviations with respect to the baseline period 1961—1990.

Fig. 8: Mean (a) and standard deviation (b) fields of land temperature anomalies in year 2012, based on $m = 100$ conditional Gaussian simulations of the field.

Fig. 9: Posterior density curves of global statistics of land temperature anomalies in 2012 based on $m = 100$ conditional Gaussian simulations: (a) field's mean; (b) field's standard deviation; fraction of land with values (c) below and (d) above 1σ .