

Comments on: Spatiotemporal Models for Skewed Processes

Marc G. Genton¹ and Amanda S. Hering²

October 28, 2016

We would first like to thank the authors for this paper that highlights the important problem of building models for non-Gaussian space-time processes. We will hereafter refer to the paper as SGV, and we also would like to acknowledge and thank them for providing us with the temporally detrended temperatures, plotted in their Figure 1, along with the coordinates of the twenty-one locations and the posterior means of the parameters for the MA1 model. We find much of interest to discuss in this paper, and as we progress through points of interest, we pose some questions to the authors that we hope they will be able to address.

1 Other Approaches

SGV present an extension of the spatial skewed model proposed by Zhang and El-Shaarawi (2010). Indeed, there has not been much work to build skewness into models for the *space-time* setting. Some additional tools that could be used in the spatiotemporal context to model skewed processes are transformations, copulas, and quantile regression. First, let $Y(\mathbf{s}, t) = \psi\{Z(\mathbf{s}, t)\}$, where $Z(\mathbf{s}, t)$ is a Gaussian spatio-temporal random field, and ψ is a one-to-one transformation. Here, $\psi(\cdot)$ could be the exponential function, a power function, or the Tukey *g*-and-*h* transformation; see Xu and Genton (2017) and references therein for the purely spatial case.

Secondly, a factor copula model, $Y(\mathbf{s}) = Z(\mathbf{s}) + V_0$, where $Z(\mathbf{s})$ is a replicated Gaussian spatial random field and V_0 is a common factor that does not depend on the location \mathbf{s} has been proposed by Krupskii et al. (2016). If $V_0 = |Z_0|$, $Z_0 \sim N(0, 1)$, then a skew-Gaussian random field is obtained. Some identifiability issues with this model when applied to purely spatial data (i.e., with no replicates) were discussed by Genton and Zhang (2012). They proposed some simple remedies. With an appropriate choice of the distribution of the common factor, V_0 , the common factor copula spatial model has both tail dependence and asymmetric dependence between the two tails. An extension to the spatio-temporal context is under way (Krupskii and Genton, 2016, in preparation).

¹Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia.
E-mail: marc.genton@kaust.edu.sa

²Department of Statistical Sciences, Baylor University, Waco, TX 76798, USA.

Finally, in the non-Gaussian setting, one of the primary interests is the behavior in the tails of the distribution. To model such behavior, quantile regression methods are often employed, and recently Sun et al. (2016) introduced a new quantile function estimator for spatial and temporal data with a special Lasso penalty designed to accommodate this type of dependence. In that work, only one realization of the process is needed to estimate the quantile function.

2 Other Applications

Temperature distributions can display skewness (Anderson et al. 2016), but other variables that are more highly skewed may benefit more from models such as SGV’s. In fact, the residuals from Figure 1 in SGV, in which the effects of altitude and the annual temperature cycle have been removed, are not strongly skewed. The boxplots in SGV’s Figure 1 do display a few outliers, and some medians are not in the center of each box. However, using these residuals, we fit the skew-normal (SN) and also the skew- t distributions and test for the presence of skewness using the built-in functions in the R package `sn` (Azzalini 2016). Figure 1 presents the estimated skewness of the SN distribution on the $[-1, 1]$ scale, each with an associated 95% confidence interval, and Table 1 summarizes the agreement between this figure and SGV’s Figure 6. No adjustment is made for multiple testing. The direction of the skewness and its significance agrees with that plotted in SGV’s Figure 6 in some of the locations (8 out of 21), but in many instances, the SN estimated skewness is not significant when SGV estimate it to be positive or negative (12 locations), and in one case, the SN estimates the opposite direction of the skewness. Figure 2 shows the histograms of the residuals when the SGV model and the SN estimates agree on the direction and significance of the skewness (left) and the case when they disagree (right). In the right-hand panel, the SGV model indicates that Urussanga is negatively skewed, but in our opinion, the histogram indicates that the skewness is in the opposite direction, and the SN distribution supports this opinion. Note that when a skew- t distribution is fit to the residuals at each location, the degrees of freedom parameter is very large, with two exceptions when it is between 15 and 20. Thus, there is not much evidence that these residuals present heavy tails. This approach of initially fitting the skew-normal distribution to detrended residuals could be a simple test to determine if fitting a model that accounts for skewness is warranted.

Other applications, such as wind, could benefit greatly from fitting a skewed process. For example, in the Irish Wind dataset of Haslett and Raftery (1989), daily averages of wind speed at eleven meteorological stations in Ireland are recorded. In Gneiting et al. (2007), a square root transformation is applied after which a common seasonal trend is removed along with location-specific means. They found that the square root transformation stabilized the variance across locations and made the marginal distributions approximately normal. In addition, Hering et al. (2015) used a Gaussian copula to handle the strong right-skew in wind speed distributions, and the strength of the skewness did vary spatially (results not shown). One of the advantages of not transforming the data and modeling the skewness directly, as SGV do, is that an estimate of the skewness is obtained, and this is informative for understanding the process holistically.

Q1: This discussion leads us to question how much skewness in the data can the SGV modeling approach handle? Even though the temperatures are not extremely skewed, the MA1 and MA2 models still show improvement in the predictions for the withheld last year of observations. We wonder how well such a model would do in the presence of extreme skewness and if there are any limitations. The motivation for this line of questioning stems from the comment in SGV that the skewness is not very strong in the simulated data (Section 3.2).

Secondly, the computational requirements of the SGV model appear to be high given that only 20 datasets are simulated in the simulation study. Furthermore, the dataset that SGV use only has 21 locations and 120 monthly measurements, totalling 2,520 observations, which is not a large spatial dataset, and 90% of the observations are used to train the model with only 10% being withheld for prediction comparisons. **Q2: Thus, what are the computational limitations of this method? How much data is recommended for training, and how much larger can the dataset be before the limits of computation are reached?**

3 Spatial Covariance

Using the residuals from SGV Figure 1, we explore their purely empirical spatial dependence (under the assumption that temporal dependence has been removed as illustrated in SGV’s Figure A.3). The top left-hand panel of Figure 3 shows the empirical spatial semivariogram with each curve based on the residuals in a different month, colored from past (red) to present (purple). There are 115 months and therefore 115 curves. Many of the semivariograms do not level off, an indicator that some spatial trend may be present. The top right-hand panel of Figure 3 summarizes these curves with a functional boxplot (Sun and Genton 2011), assuming that the temporal dependence of the residuals has been removed, making the functions independent. This assumption is verified by SGV’s Figure A.3 that plots the autocorrelation function of each location and does not demonstrate any significant lags. Otherwise, an adjustment would be needed in the computation for identifying the functional outliers (Sun and Genton 2012). Here, the functional boxplot identifies four curves as unusual with respect to the others. The values of the residuals are plotted in space for each of these four months in Figure 4. In May 2002, the two closest locations are very dissimilar, causing the large spike for the shortest distance in Figure 3. In the other three cases, a strong spatial gradient is present with northern locations being hotter than southern locations. This is quite commonly observed for other points in time, and it is not clear that the SGV model will account for this, given that the effect of altitude has already been removed.

In fact, Figure 5 shows the relationship between latitude and the residuals from SGV’s Figure 1. Fitting a simple linear regression model to these values results in 42% of the variability being explained. Once this trend is removed from the residuals, the semivariograms are computed again and shown in the bottom row of Figure 3. The effect is that now many of the empirical semivariograms level off, indicating that the non-constant spatial trend has been removed. **Q3: As a point of interest, how do the model fits for MA1 and MA2 change when latitude**

is included as a covariate in the trend?

Now, we investigate further the spatial features of the SGV residuals along with the implied spatial covariance of the SGV model. Figure 6 shows the empirical spatial correlation between every pair of locations. Namely, the points plotted are

$$\hat{C}(h) = \frac{1}{n_h} \sum_{t=1}^{n_h} \{R(\mathbf{s}, t) - \bar{R}\} \{R(\mathbf{s}', t) - \bar{R}\},$$

where \bar{R} is the overall mean of the residuals across space and time; n_h is the number of pairs of points for each distance h with $\|\mathbf{s} - \mathbf{s}'\| = h$; and $R(\mathbf{s}, t)$ is the value of the residual at location \mathbf{s} and time t . For each pair of locations, n_h is 115. $\hat{C}(h)$ is then scaled by the square root of the sample variance at each location to obtain the empirical correlation. There are $(n^2 - n)/2 = (21^2 - 21)/2 = 210$ unique distances and therefore 210 plotted points. Overlaid on this is the fitted exponential model in red, which appears to drop near the origin too quickly and also does not approach the x -axis fast enough. It has an effective range of 1,118 km. The Matérn model in blue has a smoothness parameter that is very large, so it does not drop as rapidly as the exponential model does near the origin, and it also reaches the x -axis more quickly with an effective range of 610 km. The fit of the Matérn model nearly coincides with the fit of a Gaussian correlation model. The additional flexibility that the Matérn model's smoothness parameter gives it clearly improves the fit.

Alternatively, when the Matérn model's smoothness parameter is fixed to one, this equates to the Whittle model, and it is also overlaid on Figure 6 in green. With an effective range of 863 km, it lies in between the exponential and Matérn models. We make this comparison because North et al. (2011) motivate and advise using a Whittle model due to the physical dynamics of temperature. In fact, in that paper, the authors compare the spatial “decorrelation lengths” at various temporal frequencies for two regions with relatively flat terrain, one on land and one in the ocean. For one month averages, fitting the Whittle model, the distance at which the observations become uncorrelated are 1700 and 1500 km for land and ocean, respectively. For the Brazilian data, the geography is more complex, so we suspect that the effective range will be much smaller than 1700 km, and all three of the models fitted here support this conclusion.

Finally, using the fitted posterior means of the parameters for MA1 reported by SGV, we can obtain the implied spatial correlation of their model. The covariance given by SGV is

$$\begin{aligned} \text{cov}\{K_t(\mathbf{s}), K_t(\mathbf{s}')\} &= \frac{2}{\pi} \sigma(\mathbf{s})\sigma(\mathbf{s}') \left[\sqrt{1 - \rho_\eta^2(h, \phi_\eta)} + \rho_\eta(h, \phi_\eta) \arcsin\{\rho_\eta(h, \phi_\eta)\} - 1 \right] \\ &\quad + V_t \rho(h, \phi) + \tau I(h), \end{aligned} \tag{1}$$

where $\sigma(\mathbf{s})$ is a spatially varying scale controlling the skewness at each location; V_t is a temporally varying variance; τ is a nugget effect; and $I(h)$ is an indicator function of $h = \|\mathbf{s} - \mathbf{s}'\|$. Here ρ and ρ_η are two correlation functions that SGV chose to be of exponential form with range parameters ϕ and ϕ_η , respectively. If $\sigma(\mathbf{s}) = 0$, then we retrieve the usual Gaussian model

with exponential spatial correlation satisfying $\rho'(0, \phi) = -1/\phi$ at the origin. When substituting the posterior means of the parameters for MA1 into Equation (1), we can obtain the implied spatial correlation of this model. We do not overlay this fitted covariance on the empirical spatial covariance of the residuals in our Figure 6 because the fitted trend of MA1 can differ from the trend fit in Section 1.1 of SGV. However, it would be nice if the authors can provide a plot of empirical correlations of the residuals of their detrended MA1 model in order to compare it with Figure 6. **Q4: Is the spatial dependence of the detrended MA1 model still strong, as we see from the residuals from SGV's Figure 1?**

The fitted SGV correlations are plotted for each pair of locations in Figure 7 (left). The strength of the implied spatial dependence is relatively weak, which contradicts the relatively strong spatial dependence shown in Figure 6. The practical range for this fitted model is 93 km. When overlaying the spatial correlation, $\rho_\eta(\cdot)$ with $\hat{\phi}_\eta = 33.31946$ plotted in blue, we see that the transformation of $\rho_\eta(\cdot)$ in the square brackets of Equation (1) dampens the spatial dependence of the corresponding exponential model since all of the gray dots lie below the blue line. The right-hand panel of Figure 7 shows all of the pairwise SGV fitted correlations with those less than 0.05 rounded down to zero. Only seven pairs of locations have fitted correlation values greater than 0.05, so are these values good estimates of the empirical spatial correlations, as suggested be plotted in Q4 above?

Finally, in Figure 8, we plot the posterior mean of $\delta(\mathbf{s})$ of the MA1 model at each location. Here, $\delta(\mathbf{s}) = \alpha(\mathbf{s})/\sqrt{1 + \alpha(\mathbf{s})}$ and $\alpha(\mathbf{s}) = \sigma(\mathbf{s})/\sqrt{V + \tau}$. We find it interesting that there also appears to be a spatial trend in the skewness and wonder if the authors can comment on this. **Q5: In particular, how does the interpolation of $\sigma(\mathbf{s})$ to a new location occur, as these are needed to make predictions at new locations?** In the case study, one year of observations is withheld, and forecasts are made at the existing 21 locations. **Q6: How well does spatial prediction of the model perform?** In other words, if a location is withheld and not used to fit the model, how well does the fitted model predict the temperatures at this withheld location? If this could be repeated across multiple locations (or all 21), then a good sense of the spatial prediction ability of the model could be established.

4 Conclusion

We thank the authors again for motivating a stimulating discussion of skewed models for space-time data and congratulate them on their proposed model. Our goal in this discussion has primarily been to probe more deeply into the interpretation of the skewness and spatial dependence properties that their model implies. We believe that their responses to the questions that we have posed will ultimately provide readers with a better understanding of their model so that it may gain traction and application in the sciences.

References

- Anderson, A. N., Browning, J. M., Comeaux, J., Hering, A. S., and Nychka, D. (2016), “A simulation study to compare statistical quality control methods for error detection in historical radiosonde temperatures,” *International Journal of Climatology*, **36**: 28–42.
- Azzalini, A. (2016). “The R package ‘sn’: The Skew-Normal and Skew-t distributions (version 1.4-0),” <http://azzalini.stat.unipd.it/SN>.
- Genton, M. G., and Zhang, H. (2012), “Identifiability problems in some non-Gaussian spatial random fields,” *Chilean Journal of Statistics*, **3**: 171–179.
- Gneiting, T., Genton, M. G., and Guttorp, P. (2007), “Geostatistical space-time models, stationarity, separability and full symmetry,” in Finkenstaedt, B., Held, L. and Isham, V. (eds.), *Statistics of Spatio-Temporal Systems*, Chapman & Hall/CRC Press, Monographs in Statistics and Applied Probability, 151–175.
- Haslett, J., and Raftery, A. E. (1989), “Space-time modelling with long-memory dependence: assessing Ireland’s wind-power resource (with discussion),” *Applied Statistics*, **38**, 1–50.
- Hering, A. S., Kazor, K., and Kleiber, W. (2015), “A Markov-switching vector autoregressive stochastic wind generator for varying spatial and temporal scales,” *Resources*, **4**: 70–92.
- Krupskii, P., Huser, R., and Genton, M. G. (2016), “Factor copula models for replicated spatial data,” <https://arxiv.org/pdf/1511.03000.pdf>.
- Krupskii, P., and Genton, M. G. (2016), “Factor copula models for data with spatio-temporal dependence,” manuscript in preparation.
- North, G. R., Wang, J., and Genton, M. G. (2011), “Correlation models for temperature fields,” *Journal of Climate*, **24**: 5850–5862.
- Sun, Y., and Genton, M. G. (2011), “Functional boxplots,” *Journal of Computational and Graphical Statistics*, **20**: 316–334.
- Sun, Y., and Genton, M. G. (2012), “Adjusted functional boxplots for spatio-temporal data visualization and outlier detection,” *Environmetrics*, **23**: 54–64.
- Sun, Y., Wang, H. and Fuentes, M. (2016), “Fused adaptive Lasso for spatial and temporal quantile function estimation,” *Technometrics*, **58**: 127–137.
- Xu, G., and Genton, M. G. (2017), “Tukey g -and- h random fields,” *Journal of the American Statistical Association*, to appear.
- Zhang, H., and El-Shaarawi, A. (2010). “On spatial skew-Gaussian processes and applications,” *Environmetrics*, **21**: 33–47.

Table 1: Instances in which the significance and sign of the SGV estimated skewness agrees with the estimate of skewness of the skew-normal parameter, plotted in Figure 1.

		Skew-Normal		
		Positive	None	Negative
SGV	Positive	4	8	0
	None	0	3	0
	Negative	1	4	1

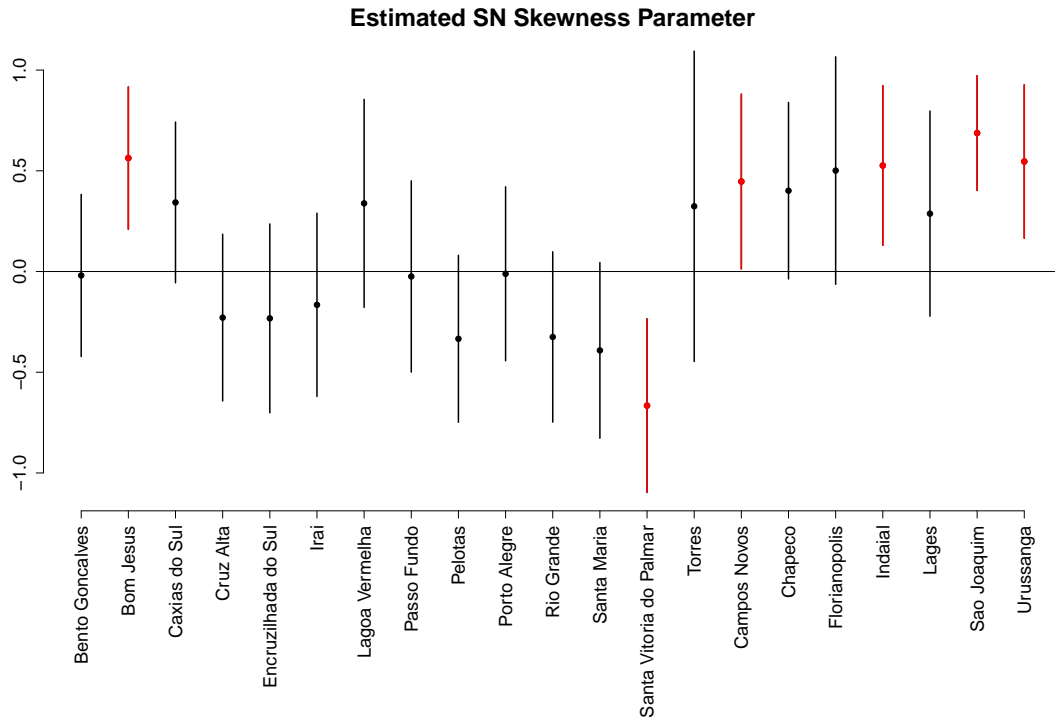


Figure 1: Estimated skewness parameter α from a skew-normal distribution along with 95% confidence intervals. Those intervals that do not cross the line of symmetry are colored in red.

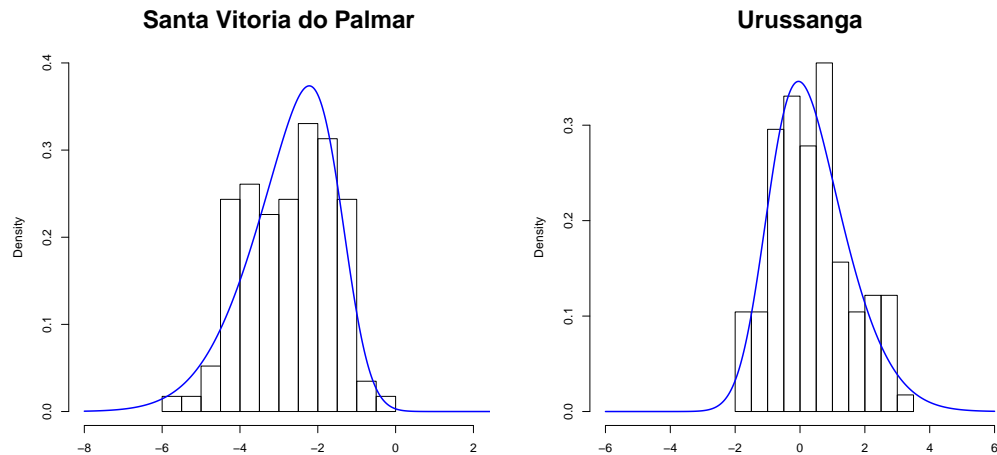


Figure 2: Distribution of the residuals for locations Santa Vitoria do Palmar (left) and Urussanga (right). On the left is an example wherein both the SGV model and the SN distribution estimate negative skewness. On the right, the SGV model indicates negative skewness while the SN distribution estimates positive skewness. The estimated SN distributions are overlaid in blue.

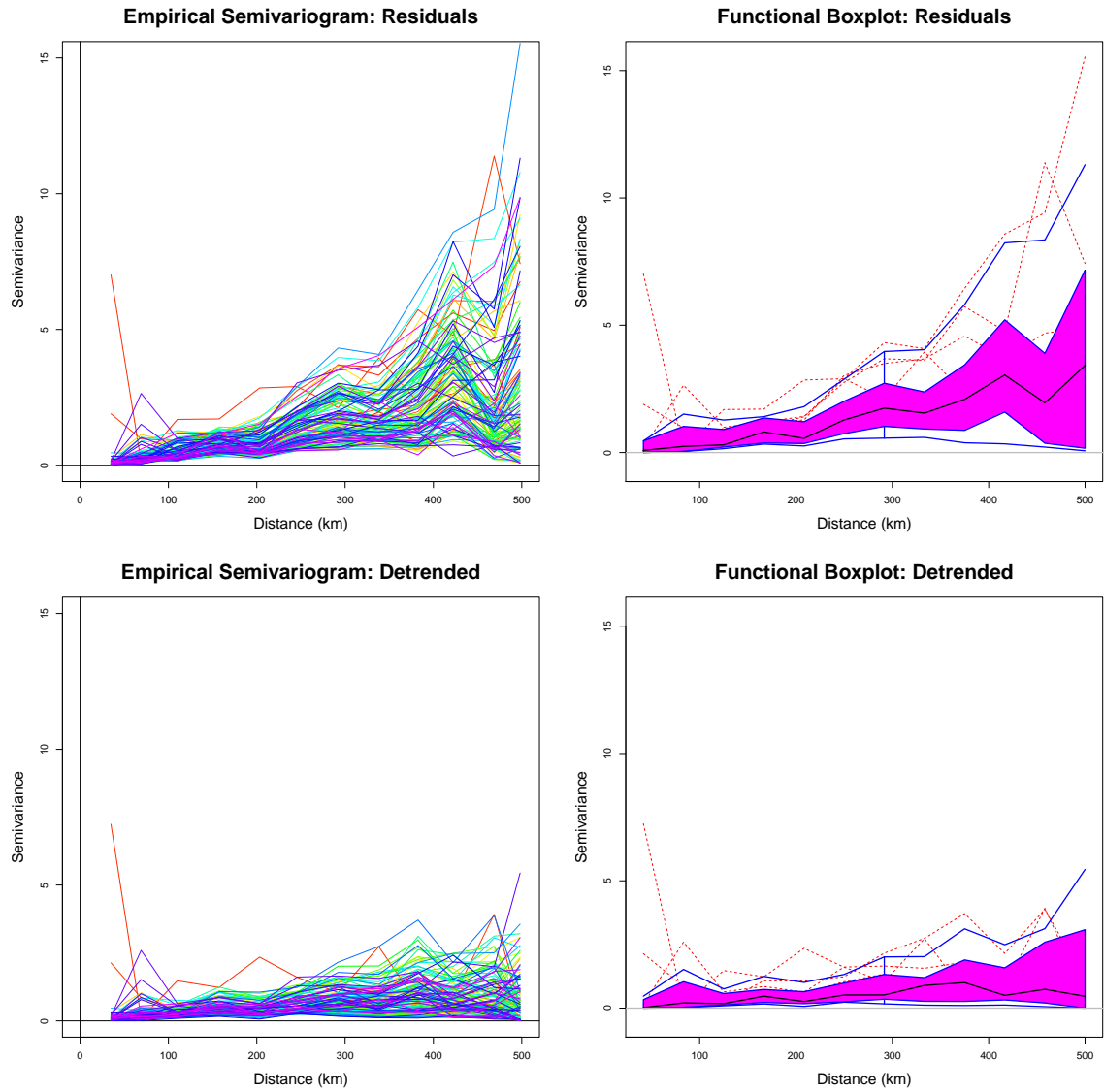


Figure 3: Empirical spatial semivariograms for each point in time, colored from red to purple (left) and the functional boxplot summary of these curves (right). Top row is based on the residuals from SGV's Figure 1, and the bottom is based on the spatially detrended residuals.

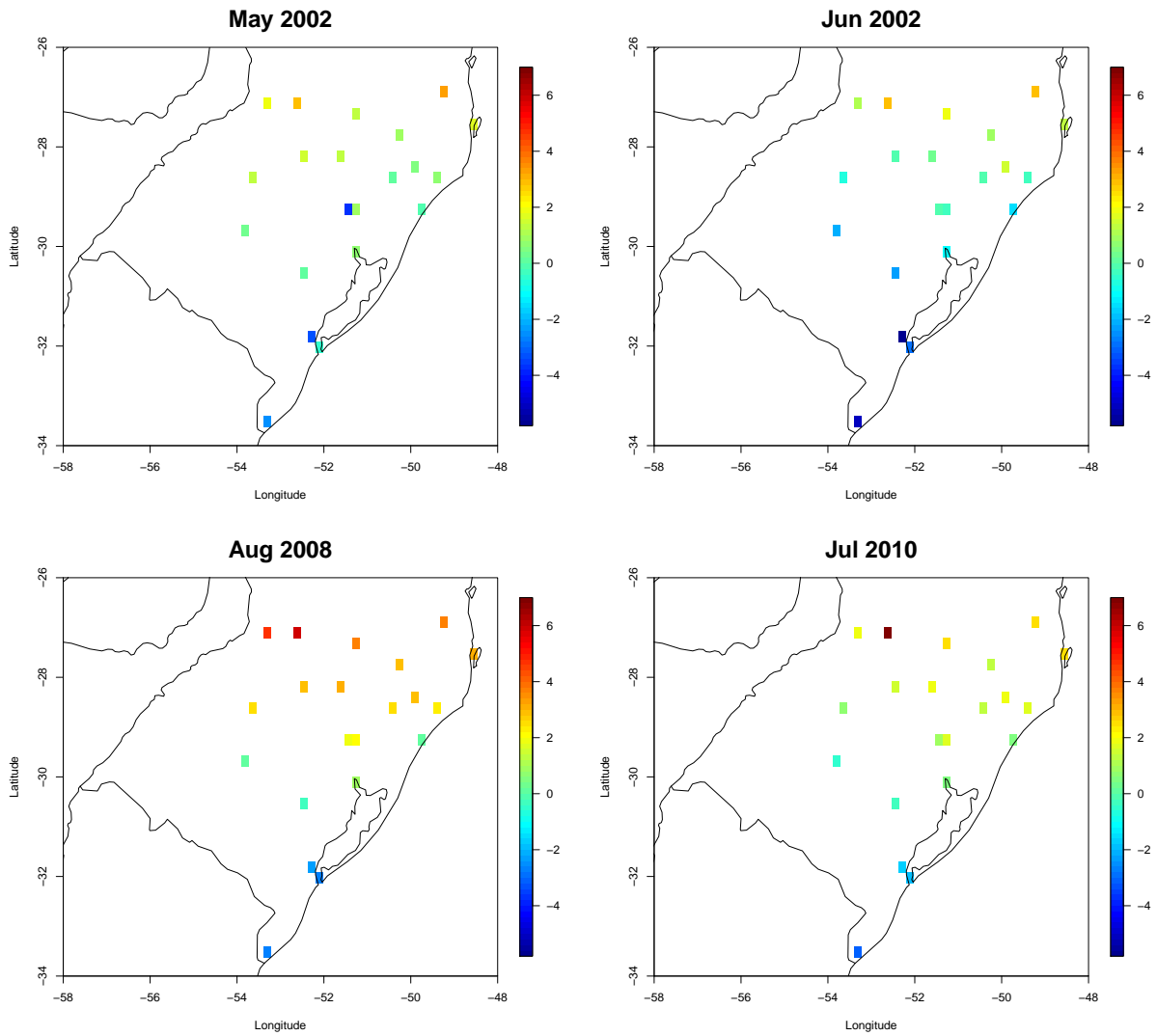


Figure 4: Spatial plot of residuals whose empirical semivariograms are identified as outliers by the functional boxplot.

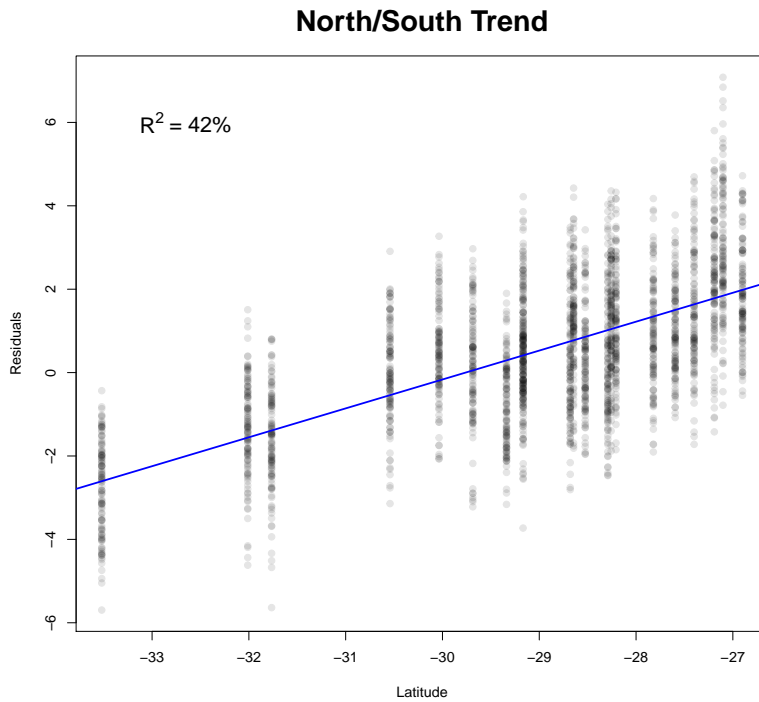


Figure 5: Relationship between the residuals in SGV's Figure 1 and latitude.

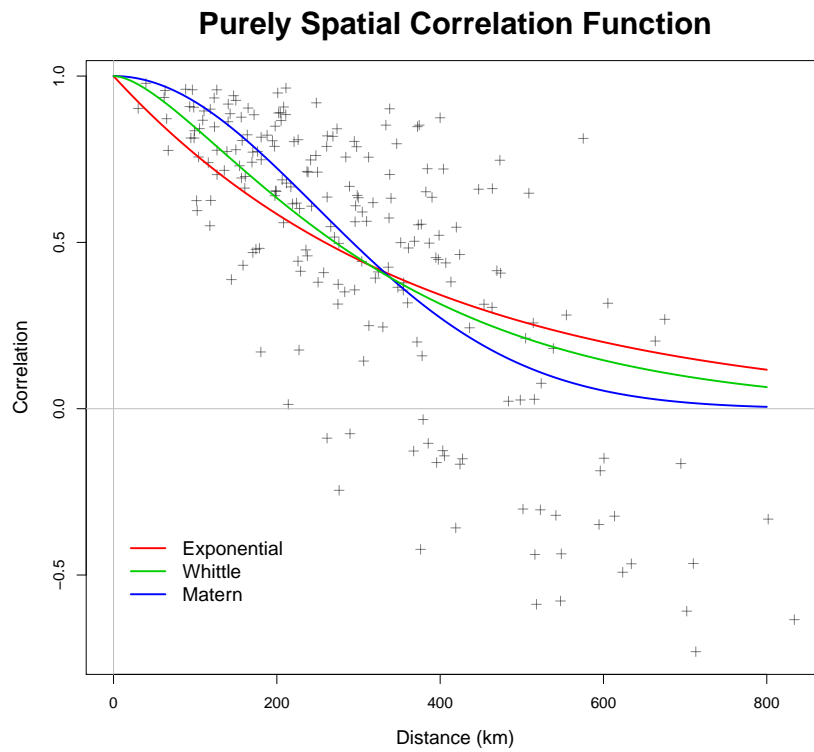


Figure 6: Empirical spatial correlations and the fitted exponential, Whittle, and Matérn correlation functions with effective ranges of 1,118, 863, and 610 km, respectively.

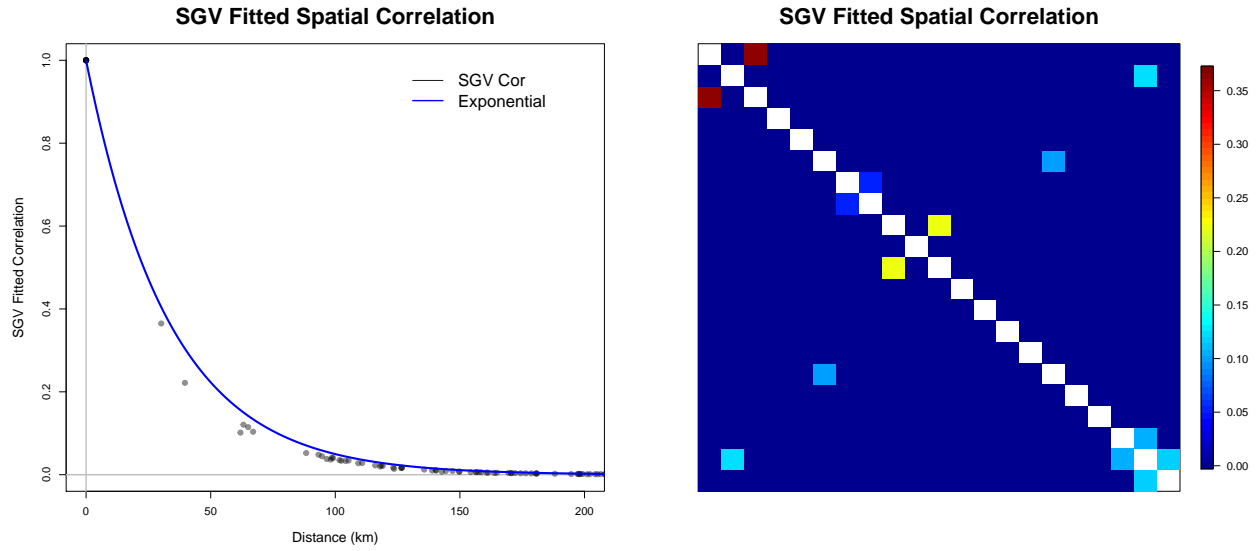


Figure 7: On the left, fitted values of the SGV correlation for each of the 21 locations are plotted (gray dots are at observed distances between pairs of points) along with an exponential function (blue) with the estimated ϕ_η parameter from the SGV model. On the right are all of the estimated SGV pairwise correlations with correlations less than 0.05 rounded to zero.

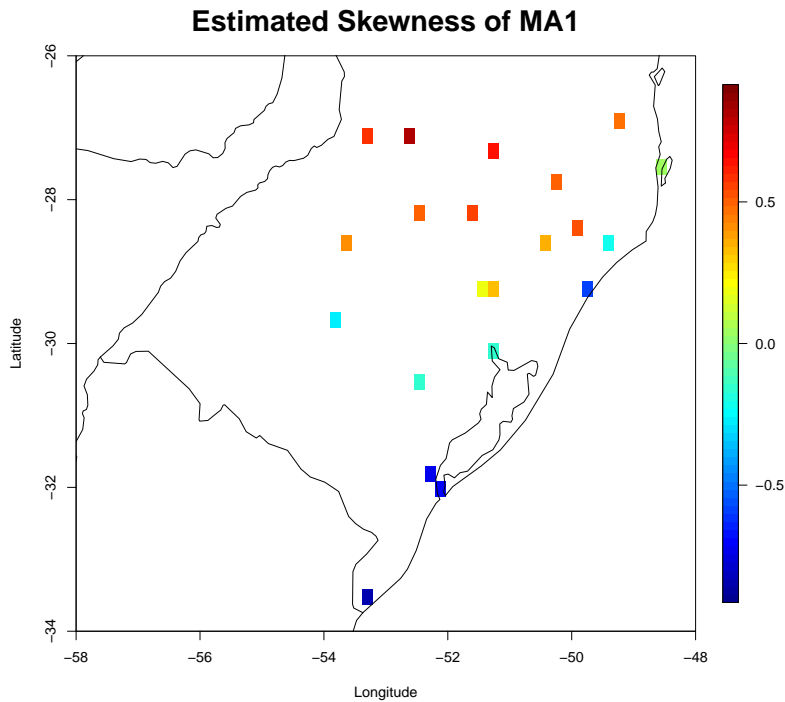


Figure 8: The posterior mean of $\delta(\mathbf{s})$ of the MA1 model, plotted for each location. Note, the average across the four different prior specifications is plotted here, but the plot of the posterior mean for each individual prior distribution follows the same pattern shown here.