Alternative Spatio-temporal models for daily rainfall prediction in Navarre

Militino A. F. [1], Ugarte M. D.[2], Goicoa, T. [3] and Genton M. [4]

[1] *Department of Statistics and Operations Research; militino@unavarra.es*
[2] *Department of Statistics and Operations Research; lola@unavarra.es*
[3] *Department of Statistics and Operations Research; tomas.goicoa@unavarra.es*
[4] *Computer, Electrical and Mathematical Sciences and Engineering (CEMSE)*
King Abdullah University Science and Technology
Kingdom of Saudi Arabia

Abstract. *To predict accumulated daily rainfall in a particular location where no rain gauges are available is important for agriculture, meteorology, traffic networks, environment and many other areas. However, statistical modelling of rain is not trivial because a high variability is presented within and between days. In this work we analyze the performance of alternative spatio-temporal models. The sampled data consist of daily observations taken in 87 manual rainfall gauges during the 1990-2010 period in Navarre, Spain. The accuracy and precision of the interpolated data is checked with data of 28 automated rainfall non-sampled gauges of the same region but placed in different locations than the manual rainfall gauges. Interpolations will be mapped on a squared grid of 1km^2 grid over the whole study region and to assess the prediction performance of the models a continuous ranked probability score (CRPS) has also been calculated.*

Keywords. *Kriging; State-space models; Statistical models; Thin-plate splines*

1 Spatio-Temporal Models

For introducing the stochastic models to be used in the spatio-temporal interpolation, we assume that $\mathbf{z}_{st} = (z(s_1,t_1), z(s_1,t_2), \ldots, z(s_1,t_n), \ldots, z(s_n,t_1), \ldots, z(s_n,t_T))$ is the spatio-temporal process that has been observed at n geographical locations, $s_1, \ldots, s_n$, (87 in this case), at time t_j, from $j = 1, \ldots, T$, varying from the 1st of January, 1990 until the 31st December, 2010. We analyze three models: the spatio-temporal kriging, the thin-plate spline, and the state-space. A brief description of these models is presented.

The linear trend of kriging is given by a linear combination of the planar coordinates and the orthometric elevation, $\mathbf{h}_s$, all of which are time-invariant. To include the time dimension we introduce a new covariate called $\mathbf{a}_{st}$, computed as the average precipitation of each 5-day period within each month between 1990 and 2010. The value of $\mathbf{a}_{st}$ is the same for all the days in the same period. We therefore have 72 different average rainfall values, although alternative periods are possible (Militino et al., 2003). The total daily rainfall on a fixed day t, and at location s is then modelled as

$$\mathbf{z}_{st} = \mu_{st} + \varepsilon_{st} = \beta_{0,t}\mathbf{1} + \beta_{1,t}\mathbf{x}_s + \beta_{2,t}\mathbf{y}_s + \beta_{3,t}\mathbf{h}_s + \beta_{4,t}\mathbf{a}_{st} + \varepsilon_{st}, \quad (1)$$

where μ_{st} is the linear trend, $\mathbf{1}$ is a vector of ones, $\mathbf{x}_s$ and $\mathbf{y}_s$ are the spatial coordinates in $\mathbb{R}^2$, and $\varepsilon_{st} \sim N_n(\mathbf{0}, \Sigma(d))$. The spatial covariance structure is accounted for in the model error. $\Sigma(d)$ can be estimated between known alternatives as Matérn, exponential or spherical covariance matrices (Militino and Ugarte, 2001; Apanasovich et al., 2012). We used Model (1) to make predictions for every day of the year 2010 with the 87 automated rainfall sampled stations. To validate the model we used data from the 87 manual rainfall gauges and the non sampled set of 28 automated rainfall gauges.

The thin-plate spline model used here is an additive model similar to the spatio-temporal kriging model (1), where we do not necessarily assume a linear relationship of the total rainfall with the planar coordinates, but a smoother relationship fitted by splines and expressed by $f(\mathbf{x}_s, \mathbf{y}_s)$. The thin-plate spline model is given by

$$\mathbf{z}_{st} = f(\mathbf{x}_s, \mathbf{y}_s) + \beta_{1t}\mathbf{h}_s + \beta_{2t}\mathbf{a}_{st} + \varepsilon_{st}, \quad (2)$$

where $\varepsilon_{st} \sim N_n(\mathbf{0}, \Sigma(d))$, and $\mathbf{h}_s$ and $\mathbf{a}_{st}$ are the same covariates defined in spatio-temporal kriging model. The spline is obtained as a weighted average of the observed data because the optimal estimate of $f(\mathbf{x}_s, \mathbf{y}_s)$ turns out to be linear in the observations. The solution to this minimization problem is identical to the universal kriging predictor of $f(\mathbf{x}_s, \mathbf{y}_s)$ under a certain intrinsic random function model that yields the second-order thin-plate smoothing spline as the optimal predictor (Stein, 1991).

The state-space model is a spatio-temporal linear model that simultaneously accounts for spatial and temporal dependence. It is given by a transition equation and a state equation:

$$\begin{aligned} \mathbf{z}_{st} &= \beta_{0,t}\mathbf{1} + \beta_{1,t}\mathbf{x}_s + \beta_{2,t}\mathbf{y}_s + \beta_{3,t}\mathbf{h}_s + v_t + \varepsilon_{st} \\ v_t &= v_{t-1} + \eta_t, \end{aligned} \quad (3)$$

where $\mathbf{z}_{st}$ is the spatio-temporal process of rainfall, the error process $\varepsilon_{st} \sim N_n(\mathbf{0}, \Sigma(d))$ and $\Sigma(d)$ is a spatial-covariance matrix similar to the other models (Cameletti et al., Fàzzo and Finazzi, 2011). The unobservable latent temporal process, v_t, shows the temporal dynamics of data through a Markovian random walk. Finally, $\eta_t \sim N(0, \sigma_\eta)$, quantifying the uncertainty of the state estimate given the n observations. The transition equation incorporates the spatial dependence and the state equation takes into account the temporal dependence. Therefore, this state-space model can be interpreted as a spatio-temporal kriging model with a separable spatio-temporal covariance function.

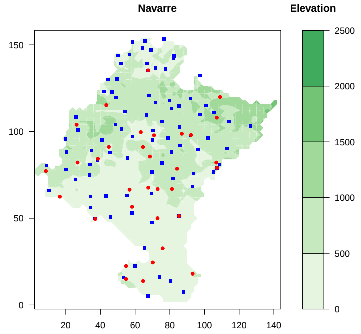

Figure 1: Map of manual gauge rainfall in blue square dots and automatic gauge rainfall stations in red circular dots located in Navarre according to the elevation.

2 Conclusions

The performance of these models is done using real data from the sampled manual rainfall stations and an additional set of automated rainfall gauge stations, located at different sites. We compared the predicted rainfall calculated with the three models in the 365 days of the year 2010. The continuous ranked probability score (CRPS) (Gneiting and Raftery, 2007, or Gneiting et al., 2007) is also used. Both criteria rank the state-space model as the best model when predictions are made from non-sampled stations.

Acknowledgements We thank to Francisco Cerveto and the Department of "Desarrollo Rural y Medio Ambiente" of the Government of Navarre. This research was supported by the Spanish Ministry of Science and Education (project MTM2011-22664 jointly sponsored with Feder grants). Research reported in this publication was also supported by King Abdullah University of Science and Technology.

References

[1] Cameletti M., Ignaccolo R., Bande S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. Environmetrics, 22, 985-996.

[2] Fassò, A., Finazzi, F.: Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. Environmetrics 22, 735-748. (2011)

[3] Gneiting T, Raftery AE. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**(477):359-378.

[4] Gneiting T, Genton M, Guttorp, P. 2007. Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistical Methods for Spatio-Temporal Systems*:151-175.

[5] Kokic, P., Crimp S. and Howden M. 2011. Forecasting climate variables using a mixed-effect state-space model. *Environmetrics* **22**:409-419.

[6] Militino A. F., Palacios, M.B. and Ugarte, M.D. 2001. Robust Predictions of Rainfall in Navarra, Spain. *Geoenv III - Geostatistics for Environmental Applications*, 79-90. Kluwer Academic Publishers.

[7] Militino A. F., Palacios, M.B. and Ugarte, M.D. 2003. Robust Trend Parameters in a Multivariate Spatial Linear Model. *Test*, **12**(2):445-457.

[8] Moral, F. J. 2010. Comparison of different geostatistical approaches to map climate variables: application to precipitation. *Interational Journal of Climatology* **30**:620-631.

[9] Newlands, N. K., Davidson, A., Howardc A., and Hilld H. 2011. Validation and inter-comparison of three methodologies for interpolating daily precipitation and temperature across Canada. *Environmetrics* **22**:205-223.

[10] Tait, A. Henderson, R., Turner, R., Zheng, X. 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *International Journal of Climatology* **26**:2097-2115.

[11] Stein M. 1991. A Kernel Approximation toi the Kriging Predictor of a Spatial Process. *Ann. Inst. Statist. Math.* **43**(1):61-75.