METMA V International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

Spatio-temporal modelling of daily air temperature in Catalonia

M. Saez[1,2*], M.A. Barceló[1,2], A. Tobias[3], D. Varga[1,4] and R. Ocaña-Riola[5]

[1]*Research Group on Statistics, Applied Economics and Health (GRECS), University of Girona, Spain; marc.saez@udg.edu*
[2]*CIBER of Epidemiology and Public Health (CIBERESP)*
[3]*Institute of Environmental Assessment and Water Research (IDAEA), Spanish Council for Scientific Research (CSIC), Barcelona, Spain*
[4]*Geographic Information Technologies and Environmental Research Group, University of Girona, Spain*
[5]*Andalusian School of Public Health (EASP), Granada, Spain*
[*]*Corresponding author*

Abstract. *We propose a model for describing the mean function as well as the spatio-temporal covariance structure of 11 years of daily average temperature data from 190 stations throughout Catalonia, with daily data covering the period 1998-2008.*

Keywords. *Average temperature; Spatio-temporal covariance; 2-stage Bayesian approach; Spatial variability.*

1 Introduction

We propose a model for describing the mean function as well as the spatio-temporal covariance structure of 11 years of daily average temperature data from 190 stations throughout Catalonia, with daily data covering the period 1998-2008.

We include as explanatory variables of the mean function, a longterm trend, annual harmonics, latitude, altitude and winds. We controlled temporal autocorrelation by means of ARMA models. For the spatial covariance structure we used the Matérn family of covariance functions and a nugget term. For final model estimation we used a 2-stage Bayesian approach. In the firststage, we assume the stations are spatially independent. In the second-stage we model the spatio-temporal covariance, using the interim posterior from the residuals of the model in the first-stage as prior distributions of replications of a spatial process. We allow spatial parameters to also vary with time, trying different space-time interactions.

Using the estimates we interpolate the temperature to a grid that covers all the Catalan territory.

During the period, we estimated a slightly increase in the daily average temperature, corresponding to a 15-years increase of 0.262 ºC (95% credibility interval 0.046, 0.478). Spring months were those with a maximum increase on temperature all over the period. There was an increase in the spatial variability of daily air temperature.

With the model we could not only estimate spatio-temporal variations of daily average temperature but also other like the number of heatwaves and other periods of high temperature and the early onset of spring.

2 Methods

We specified the daily average temperature spatio-temporal process as follows:

$$Y_{it} = X_{it}'\beta_{it} + \frac{\Phi_i(B)}{\Theta_i(B)}\varepsilon_{it} \tag{1}$$

$$cov(\varepsilon_{it}, \varepsilon_{it}') = M\left(\left|i - i'\right|, r_t^2\sigma_t^2, \rho_t, \nu_t\right) + \left(1 - r_t^2\right)\sigma_t^2 I\left(i = i'\right) \tag{2}$$

where Y_{it} denotes the average temperature on day t (measured from 1 January 1994 to 31 December 2008) at station i; and X_{it} contains explanatory variables.

In particular, we include as explanatory variables of the mean function (1), a long-term trend (*trend=1, ...,5479*, where *1* corresponded to 1 January 1994 and *5479* to 31 December 2008), annual harmonics $cos(2\pi ntrend/365)$ and $sin(2\pi ntrend/365)$, with $n = 1,2,3,4,5,6$, corresponding to periods 12, 6, 4, 3, 2.4 and 2 months, respectively; latitude and altitude of the station; and daily measurements of wind, average wind speed and predominant direction, on each station.

All parameters associated with the explanatory variables, β, were allowed to vary between stations and some of them (those associated to the long-term trend, the annual harmonics and the measurements of wind) were also allowed to vary with year (this is the reason of the sub indexes in the parameters in (1)).

Meteorological conditions that may persist from one day to another could influence air temperature, leading to temporal correlation. We controlled this autocorrelation by means of ARMA models. Note that we allowed that the parameters of the ARMA models could vary between stations (see the second summand in the right part of (1)).

For the spatial covariance structure we used the Matern family of covariance functions and a nugget term, for fixed t over i (see (2)). M was the Matern function; $\sigma_t^2 I$ denotes the sill (the total variance of the innovation process) at time t; $r_t^2\sigma_t^2$ is the variance of the spatially correlation portion of the process; $\left(1 - r_t^2\right)\sigma_t^2$. corresponds to the *nugget* (the variability unique to a given station); ρ_t is the range of the process (the size of the region where the process is significantly correlated); and ν_t is the smoothness of the process (specifically, we will try $\nu_t = 1,2,3$).

For final model estimation we used a 2-stage Bayesian approach. In the firststage, we assume the stations are spatially independent and we model the mean function of daily average temperature controlling for possible temporal autocorrelation. In the second-stage we will model the spatio-temporal covariance, using the interim posterior from the residuals of the model in the first-stage as prior distributions of replications of a spatial process. We will allow spatial parameters to also vary with time, trying different space-time interactions. This procedure would allow borrowing strength from other temporary and spatially neighbour areas.

Using the estimates we will interpolate the temperature to a grid that covers all the Catalan territory. With the model we could estimate also the number of heatwaves and other periods of high temperature and the early onset of spring.

3 Results

We could give here some preliminary results. The median number of stations with daily (temperature) data was 155 (minimum 4, maximum 166, first quartile 153, third quartile 159).

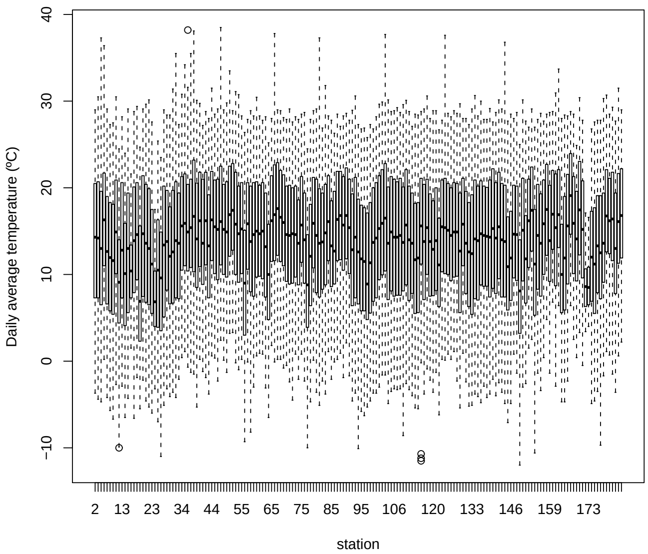

Figure 1: Daily average temperature in 190 stations all over Catalonia.

In the first part of the model, (1), all parameters were estimated statistically significant. We estimated an ARMA(2,1), i.e. $\frac{\left(1-\phi_{1i},B-\phi_{2i}B^2\right)}{1-\theta_i B}$, with parameters (on average) $\phi_1 = 0.4932045, \phi_2 = 0.1571704$ and $\theta_1 = 0.4314618$.

During the period, we estimated a slightly increase in the daily average temperature (1.66% annualized on average), corresponding to a 15-years increase of 0.262 ºC (95% credibility interval 0.046, 0.478). There was a 15-year (estimated) increase only 48.65% of all stations (90 of the 185 stations), whereas in the rest of the stations the (estimated) decrease was not statistically significant.

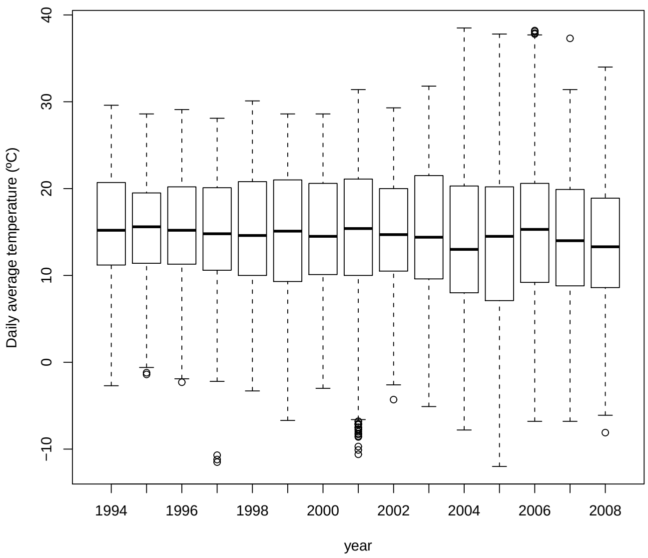

Figure 2: Annual variation of daily average temperature, 1994-2008.

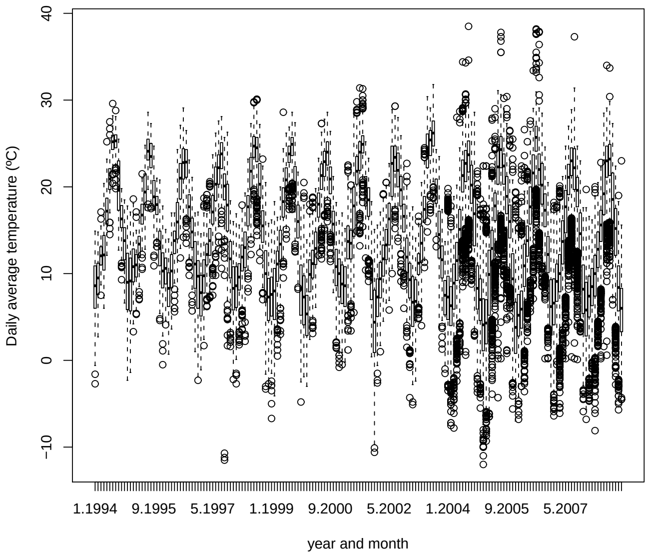

Figure 3: Monthly variation of daily average temperature, 1994-2004.

With respect to the 1994, Spring months were those with a maximum increase on temperature all over the period (June 15-years increase of 0.142 ^{o}C, April 0.030 oC and May 0.027^{o}C). December and February (-0.466 ^{o}C) followed by January (-0.374 ^{o}C) and November (-0.368) were the months where the decrease was the maximum.

There was an increase in the spatial variability of daily air temperature (estimated standard deviation equal to 5.733 ^{o}C in 1994 and 7.032 ^{o}C in 2008), denoting the existence of space-time interactions. In fact, it could be observed a geographical pattern in the variations of daily average temperature. It seems that the variation of the temperature depended on the latitude, with a maximum in the latitude corresponding to Badalona (the town adjacent to Barcelona in the North).

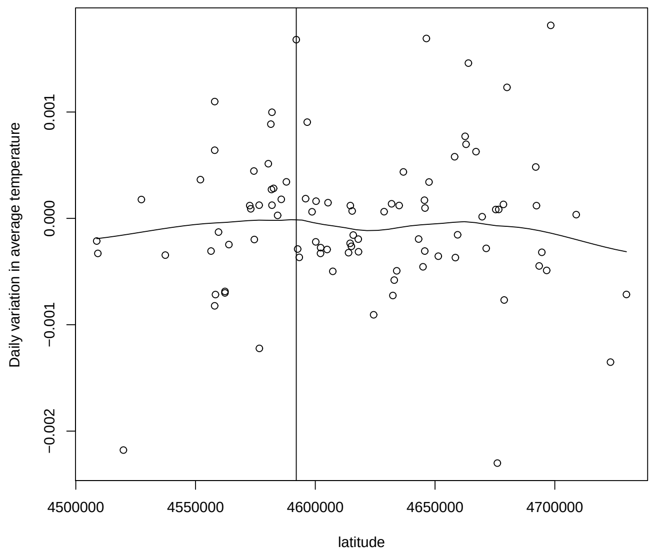

Figure 4: Estimated variation in daily average temperature and latitude of the station.

Acknowledgments. We would thank the *Àrea de Climatologia, Servei Meteorològic de Catalunya* for providing meteorological data.

References

[1] Alpuim T, El-Shaarawi A (2009). Modeling monthly temperature data in Lisbon and Prague. *Environmetrics* **20(7)**, 835-852.

[2] Huerta G, Sanso B, Stroud JR (2004). A spatiotemporal model for Mexico city ozone levels. *Journal of the Royal Statistical Society, Series C* **53(2)**, 231-248.

[3] Im HK, Rathouz PJ, Frederick JE (2009). Space-time modelling of 20 years of daily air temperature in the Chicago metropolitan region. *Environmetrics* **20(5)**, 494-511.

[4] Martino S, Rue H (2010). Case studies in Bayesian computation using INLA. *Complex Data Modeling and Computationally Intensive Statistical Methods* (in press). [http://www.r-inla.org/papers]

[5] Rue H, Martino S, Chopin N (2008). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion) *Journal of the Royal Statistical Society, Series B* **71**, 319392.

[6] Stein ML (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.