A Periodic Mixed Linear State Space Model to Monthly Long-term Temperature Data

M. Costa[1] & M. Monteiro

ESTGA - Águeda School of Technology and Management, University of Aveiro, Portugal

CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro, Portugal

Abstract

In recent decades, the world has been confronted with the consequences of global warming; however, this phenomenon is not reflected equally in every part of the globe. Thus, the warming phenomenon must be monitored in a more regional or local scale. This paper analyzes monthly long-term time series of air temperatures in three Portuguese cities: Lisbon, Oporto and Coimbra. We propose a periodic state space framework, associated with a suitable version of the Kalman filter; which allows for the estimation of monthly warming rates taking into account the seasonal behavior and serial correlation. Results about the monthly mean of the daily mid-range temperature time series show that there are different monthly warming rates. The greatest annual mean rise was found in Oporto with 2.17°C whereas, in Lisbon and Coimbra, it was respectively, 0.62°C and 0.55°C, per century.

[1]Corresponding author: M. Costa, ESTGA & CIDMA, Universidade de Aveiro, Apartado 473, 3754–909 Águeda, Portugal, Email: marco@ua.pt

Keywords: Air temperature, Climate change, Kalman filter, Portuguese cities, Seasonality, Time series analysis

1 Introduction

The rise in global temperature has been of increasing concern for several authorities. Accordingly to the Intergovernmental Panel on Climate Change, the world's greenhouse gas emissions are continuing increasing and on the current scenario, global temperature rise will far exceed the limit goal of 2°C that countries have agreed with in order to avoid the most dangerous impacts on climate change. In the United Nations (UN) Framework, countries adopted the Paris Agreement, on 12 December 2015 in France, at the UN Climate Change Conference, where parties committed to take ambitious actions to keep global temperature rise below 2°C by the end of the Century (UN, 2016).

Hence, the monitoring and the analysis of the temperature rise, at the global, regional or local levels, have been a challenge for the scientific community. Global and European average time series for monthly temperature have been gridded using different interpolation techniques; and the results unveiled that there has been especially relevant climate change in the Iberian Peninsula. Indeed, a high warming has been observed in the past 50 years over the Iberian Peninsula and, over the past 30 years it occurred mainly in summer, EEA (2018). Hence, in the European context the analysis of local time series has a special interest in order to monitor temperature rise. Furthermore, the analysis of these time series avoids interpolation uncertainties associated with global or regional time series.

Temperature data can be daily, monthly or annual depending on both the scale, nature, and the theme under consideration. Several studies based on different types of periodicity of data have also been considered in the literature. For instance, the following works were based on daily temperature data: Poppick et al. (2016) studied changes in the distribution of daily temperatures in an ensemble of general circulation model (GCM) runs predicting changes in both means and variability; Trevin (2013) developed a new homogenized

daily maximum and minimum temperature data set for Australia; among others (Fischer, 2015; Kleiber et al., 2013; Ngo and Horton, 2016; Trigo and Palutikof, 1999; Wang et al., 2013).

A considerable volume of research based on monthly temperature data has also been published. Considering monthly means of daily minimum temperature and daily maximum temperature data from Albert Park, Auckland from 1910 to 1986, Withers and Nadarajah (2015) analyzed the efficiency of modeling observations that are grouped into weekly means, monthly means and annual means. Monthly average temperature series in two widely separated European cities, Lisbon (1856–1999) and Prague (1841–2000), were examined in Alpuim and El-Shaarawi (2009). Other works investigated monthly temperatures (Abbasnia and Toros, 2016; Bengtsson and Cavanaugh, 2008; Liu et al., 2016; Reich, 2012; William et al., 2012). Annual temperatures were investigated in Xu et al. (2015), and Moreno et al. (2013) analyzed the change points in the annual mean temperature in central England between 1659 and 2011 using a consistent online Bayesian procedure for detecting change points.

The most applied statistical models in the analysis of climate time series, in particular of temperatures data sets, are the linear models and time series models such as Autoregressive Integrated Moving Average models (ARIMA) (Alpuim and El-Shaarawi, 2009; Bližňák et al., 2015; Freitas et al., 2015). As an extension of the linear models, state space models have been largely applied to the modeling of environmental data, since they incorporate a versatile stochastic structure that allow for the integration of temporal dependence. The statistical robustness and predictive ability of state space models make them the most promising avenue towards a new type of modern statistical modeling (Patterson et al., 2008). State space models are commonly used to analyze data sets with measurement errors as environmental data (temperature, precipitation, etc.) (Tandeo et al., 2011). Recent works have considered a state space approach to model temperature time series data sets. An additive, structural state-space model was considered in Bengtsson and Cavanaugh (2008) in order to represent the monthly temperature mean as a sum of an overall constant mean, a seasonal component, a monthly temperature anomaly, and a white noise term. An extension of the linear and Gaussian state space was proposed in Tandeo et al. (2011) to analyze time series with irregular

time sampling in the modeling of sea surface temperature data from a particular satellite. The monthly minimum temperatures in the State of Rio de Janeiro were analyzed by Castro Morales et al. (2013) using a spatial-temporal model, whose temporal trend is modeled through state space models.

This paper is organized as follows. Section 2 introduces methodological aspects, namely, the model description, the periodic Kalman filter version, the parameters estimation and inference aspects. In Section 3 an application to three Portuguese cities (Lisbon, Oporto and Coimbra) is presented. Data analyzed was studied by Morosova and Valente (2012a) in order to detect and correct non-climatic homogeneity breaks, long-term temperature data series measured in these three Portuguese cities. These three datasets are available for studies of climate variability in Morosova and Valente (2012b). The work finishes with a discussion of the results in Section 4.

2 The Periodic Mixed Linear State Space Model

In this section a state space model is proposed, which incorporates a flexible structure in order to accommodate the different characteristics of the monthly temperature data. Nevertheless, the proposed model is presented in a general form which allows for its adaption to other datasets in different contexts of the environmental area or, for instance, of the economic area.

2.1 The model

The *Periodic Mixed Linear State Space* (PMLSS) model considers an observable variable Y that is collected with a regular sampling procedure, i.e., the model assumes equidistant time in observations. Usually, the variable Y is observed during N years and each year has S seasons. Thus, we denote $Y_t \equiv Y_{s,n}$ with $t = 1, 2, ..., T$, $n = 1, 2, ..., N$ and $s = 1, 2, ..., S$, where n is the year associated with time t and s is the respective season. With this notation, when t corresponds to the first season of year n than the previous time, which is the season S of year $n-1$, can be denoted, to simplicity, as season 0 of year n, i.e., $Y_{0,n} \equiv Y_{S,n-1}$.

The model is formulated as follows:

$$Y_{s,n} = \begin{bmatrix} 1 & S(n-1)+s \end{bmatrix} \begin{bmatrix} a_{s,n} \\ X_{s,n} \end{bmatrix} + D_{s,n}\beta + e_{s,n} \tag{1}$$

$$\begin{bmatrix} a_{s,n} \\ X_{s,n} - \mu_s \end{bmatrix} = \begin{bmatrix} \phi_a & 0 \\ 0 & \phi_s \end{bmatrix} \begin{bmatrix} a_{s-1,n} \\ X_{s-1,n} - \mu_{s-1} \end{bmatrix} + \begin{bmatrix} \omega_{s,n} \\ \varepsilon_{s,n} \end{bmatrix}. \tag{2}$$

In the *observation equation*, Eq. 1, the observable variable $Y_{s,n}$ represents the temperature observation in the s^{th} season of the n^{th} year, that is, the $[S(n-1)+s]^{\text{th}}$ observation of the time series. The line matrix $\mathbf{H}_{s,n} = \begin{bmatrix} 1 & S(n-1)+s \end{bmatrix}$ is a design matrix of known values, where $S(n-1)+s$ represents the time. The random vector $\mathbf{X}_{s,n} = \begin{bmatrix} a_{s,n} & X_{s,n} \end{bmatrix}'$ has a Periodic Vector Autoregressive of order 1 structure, PVAR(1), and it includes a non-periodic autoregressive process, $a_{s,n} = a_t$, to incorporate the serial correlation; and the Periodic Autoregressive Process of order 1, PAR(1), $X_{s,n}$ that represents the stochastic slopes. The seasonal pattern is incorporated by fixed seasonal coefficients through vector $\beta = [\beta_1\ \beta_2\ ...\ \beta_S]'$, (e.g., $S = 12$ in the case of monthly data). However, these seasonal fixed effects associated to the stochastic slopes in the trend component causes the seasonal pattern to change over time. The $1 \times S$ line matrix $D_{s,n}$ is a design matrix with zeros and ones, as an indicator function in order to associate the respective seasonal coefficient β_s, with $s = 1, 2, ..., S$, to the variable $Y_{s,n}$. The observation error $e_{s,n}$ is a Gaussian white noise process with variance $\text{Var}(e_{s,n}) = \sigma_e^2$.

In the *state equation*, Eq. 2, the *state* vector $\mathbf{X}_{s,n}$ follows a PVAR(1) process with mean $\mu_{\mathbf{X}_{s,n}} = \begin{bmatrix} 0 & \mu_s \end{bmatrix}'$, where μ_s is the mean of the slope of season s; Φ_s is the autoregressive matrix $\Phi_s = \text{diag}\{\phi_a, \phi_s\}$, where ϕ_a is the autoregressive coefficient of the AR(1) process, $\{a_{s,n}\}$, and ϕ_s is the autoregressive coefficient associated with the slope of the season s. The vector of errors $\zeta_{s,n} = \begin{bmatrix} \omega_{s,n} & \varepsilon_{s,n} \end{bmatrix}'$ has a multivariate

Gaussian distribution with a covariance matrix $\Sigma_{\zeta_{s,n}} = \text{diag}\{\sigma^2_\omega, \sigma^2_{\varepsilon,s}\}$, such that,

$$\text{Cov}(\varepsilon_{s,n}, \varepsilon_{s-i,n}) = \begin{cases} \sigma^2_{\varepsilon,s}, & i = 0 \\ 0, & i \neq 0 \text{ for } i = 1, 2, ..., S \end{cases}$$

and processes $\{\omega_{s,n}\}$ and $\{\varepsilon_{s,m}\}$ are uncorrelated, that is $\text{E}(\omega_{s,n}\varepsilon_{r,m}) = 0$, for all s, r, n and m. When in the AR(1) process $\{a_{s,n}\}$ $|\phi_a| < 1$, which represents the serial correlation, the process is stationary with zero mean and variance $\text{Var}(a_{s,n}) = \sigma^2_\omega(1 - \phi^2_a)^{-1}$.

The PAR(1) process, $\{X_{s,n}\}$, which represents the periodic stochastic slopes, is cyclostationary when $\left|\prod_{k=1}^{S} \phi_k\right| < 1$, (Gardner et al., 2006; Monteiro et al., 2010; Obeysekera and Salas, 1986). In this case, $\text{E}(X_{s,n}) = \mu_s$ and

$$\text{Var}(X_{s,n}) = \sigma^2_s = \left\{ \sigma^2_{\varepsilon,s} + \sum_{i=1}^{S-1} \left(\prod_{j=1}^{i} \phi^2_{s-j} \sigma^2_{\varepsilon,s-i} \right) \right\} \left(1 - \prod_{k=1}^{S} \phi^2_k \right)^{-1}, \tag{3}$$

with the convention $\phi_{-i} = \phi_{S-i}$ and $\sigma^2_{\varepsilon,-i} = \sigma^2_{\varepsilon,S-i}$, for $i = 0, 1, \cdots, S-1$.

The model (1) – (2) assumes the periodic mixed effect state space representation

$$Y_{s,n} = \mathbf{H}_{s,n}\mathbf{X}_{s,n} + D_{s,n}\beta + e_{s,n} \tag{4}$$

$$\mathbf{X}_{s,n} = \mu_{\mathbf{X}_{s,n}} + \Phi_s(\mathbf{X}_{s-1,n} - \mu_{\mathbf{X}_{s-1,n}}) + \zeta_{s,n}. \tag{5}$$

The model (4)–(5) allows for the incorporation of some characteristics which make the model versatile. Furthermore, this state space formulation associated to the Kalman filter deals well with missing values, since the maximum likelihood estimates can be obtained through the EM-algorithm. In the temperature modeling context, the mixed effects approach associated with the intra-annual seasonality is a simple way of modeling the seasonality that naturally exists in this variable. For instance, in Kokic et al. (2011) the

potential of using a linear mixed-effect state-space model for statistical downscaling of climate variables compared to the frequently used approach of linear regression is shown; Ursu and Pereau (2016) considered Seasonal Autoregressive Moving Average (SARMA) models which represent a class of stationary models with large lag autocorrelations that are invariant with respect to the season. However, it is reasonable to admit that the long-term temperature time series have rises with periodic mean and covariance functions with respect to time. Therefore, the use of periodic autoregressive modeling of the slope of the temperature time series linear trend is an appropriate option. Nevertheless, when there are good estimates of periodicities, one may first deseasonalize the data and then model the resulting temperatures anomalies.

2.2 The Kalman filter adaptation to the PMLSS model

A state space model has, in its structure, a latent process, the state, which is not observable and needs to be predicted. The most usual procedure for this prediction is the Kalman filter algorithm. This algorithm computes, at each time, the optimal estimator of the state vector based on the available information until t and its success lies on the fact that it is an online estimation procedure. The optimal properties can be guaranteed only when all model's parameters Θ are known and the normality of errors is valid (Harvey, 1996; Shumway, 2017). However, if the normality is dropped, then the Kalman filter predictors are the best linear unbiased estimators (BLUE). When parameters of the state space model are estimated, the uncertainty associated with the Kalman filter estimators are underestimated and some procedures can be implemented (Costa and Monteiro, 2016; Rodríguez and Ruiz, 2012).

Briefly, the Kalman filter is an iterative algorithm that produces, at each time, an estimator of the state vector $\mathbf{X}_{s,n}$, which is given by the orthogonal projection of the state vector onto the observed variables up to that time. Considering the matrix representation of the PMLSS model (4) – (5), let $\widehat{\mathbf{X}}_{s|s-1,n}$ denote the estimator of $\mathbf{X}_{s,n}$ based on the observations $Y_{1,1}, Y_{2,1}, ..., Y_{s-1,n}$ and let $\mathbf{P}_{s|s-1,n}$ be its covariance matrix, i.e. $\mathrm{E}[(\widehat{\mathbf{X}}_{s|s-1,n} - \mathbf{X}_{s,n})(\widehat{\mathbf{X}}_{s|s-1,n} - \mathbf{X}_{s,n})']$, the Mean Square Error (MSE) matrix. Since the orthogonal

projection is a linear estimator, the forecast of the observable vector $Y_{s,n}$ is given by

$$\widehat{Y}_{s|s-1,n} = \mathbf{H}_{s,n}\widehat{\mathbf{X}}_{s|s-1,n} + D_{s,n}\beta \tag{6}$$

with MSE given by $\omega_{s,n} = \mathbf{H}_{s,n}\mathbf{P}_{s|s-1,n}\mathbf{H}'_{s,n} + \sigma_e^2$.

When $Y_{s,n}$ is available, the prediction error or *innovation*, $\eta_{s,n} = Y_{s,n} - \widehat{Y}_{s|s-1,n}$, is used to update the estimate of $\mathbf{X}_{s,n}$ (*filtering*) through the equation

$$\widehat{\mathbf{X}}_{s|s,n} = \widehat{\mathbf{X}}_{s|s-1,n} + \mathbf{K}_{s,n}\eta_{s,n}, \tag{7}$$

where $\mathbf{K}_{s,n}$ is called the Kalman gain matrix and is given by $\mathbf{K}_{s,n} = \mathbf{P}_{s|s-1,n}\mathbf{H}'_{s,n}\omega_{s,n}^{-1}$. Furthermore, the MSE of the updated estimator $\widehat{\mathbf{X}}_{s|s,n}$, represented by $\mathbf{P}_{s|s,n}$, verifies the relationship $\mathbf{P}_{s|s,n} = \mathbf{P}_{s|s-1,n} - \mathbf{K}_{s,n}\mathbf{H}_{s,n}\mathbf{P}_{s|s-1,n}$. The forecast for the state vector $\mathbf{X}_{s+1|s,n}$ is given by the equation

$$\widehat{\mathbf{X}}_{s+1|s,n} = \mu_{s+1} + \Phi_{s+1}(\widehat{\mathbf{X}}_{s|s,n} - \mu_s) \tag{8}$$

and its MSE matrix is $\mathbf{P}_{s+1|s,n} = \Phi_{s+1}\mathbf{P}_{s|s,n}\Phi'_{s+1} + \Sigma_{\zeta_{s,n}}$.

The Kalman filter algorithm is initialized with $\mathbf{X}_{1|0,1}$ and $\mathbf{P}_{1|0,1}$. When the state process is stationary, the Kalman filter algorithm can be initialized considering that initial state vector $\widehat{\mathbf{X}}_{1|0,1} = \text{diag}\{0, \mu_1\}$ and the covariance matrix $\mathbf{P}_{1|0} = \text{diag}\{\sigma_\omega^2, \sigma_1^2\}$ according to Eq. 3. In the non-stationarity case, the initialization of the Kalman filter can be incorporated in the estimation procedure or can be specified in terms of a diffuse or non-informative prior (Harvey, 1996).

2.3 Gaussian maximum likelihood estimation of parameters

Under the assumptions that the initial state, the state noise process $\{\varepsilon_{s,n}\}$ and the observation noise process $\{e_{s,n}\}$, are normal and mutually independent and considering, without loss of generality, N complete years

in the realization $\mathcal{Y} = (Y_{1,1}, Y_{2,1}, ..., Y_{S,N})'$, the logarithm of the conditional Gaussian likelihood function is computed as follows

$$\log(\mathcal{L}(\Theta; \mathcal{Y})) = -\frac{NS}{2}\log(2\pi) - \frac{1}{2}\sum_{n=1}^{N}\sum_{s=1}^{S}\log(\omega_{s,n}) - \frac{1}{2}\sum_{n=1}^{N}\sum_{s=1}^{S}\frac{\eta_{s,n}^2}{\omega_{s,n}}, \tag{9}$$

where $\Theta = (\beta_1, ..., \beta_S, \sigma_e^2, \mu_1, ..., \mu_S, \phi_a, \phi_1, ..., \phi_S, \sigma_\omega^2, \sigma_{1,\varepsilon}^2, ..., \sigma_{S,\varepsilon}^2)'$ is the τ-vector, $\tau = 4S+3$, of the unknown parameters to be estimated.

The maximum likelihood (ML) estimates are obtained upon maximizing the log-likelihood function, that is, $\Theta_{\text{ML}} = \text{argmax}_\Theta \log(\mathcal{L}(\Theta; \mathcal{Y}))$. As the log-likelihood function $\log(\mathcal{L})$ is nonlinear, it is possible to obtain the ML estimates using numerical algorithms. Further details on parameters estimation can be obtained in the Appendix.

3 Application to homogenized monthly time series of air temperature

In this section, we apply the periodic mixed linear state space modeling to the homogenized monthly time series of air temperature in three Portuguese cities produced by Morosova and Valente (2012a) and available in Morosova and Valente (2012b). Based on these time series, we constructed monthly mean temperature series based on maximum and minimum of the homogenized time series.

3.1 Data description

This section provides a description of the data that motivates this study. Morosova and Valente (2012a) developed and provided a data set with long-term time series of monthly data of temperatures in the three Portuguese cities. This data set is the result of a procedure that detected and corrected non-climatic homogeneity breaks in an original data set. As mentioned by these authors Morosova and Valente (2012a), long instrumental climatological records assume a paramount role in the studies of atmospheric conditions

variation, providing vital information about climate variability, trends and cycles. However, long-term series often contain inhomogeneities originated by changes in instruments, station locations and surrounding environment, observation routines and methods of preliminary data treatment. Thereof, the data set available after the detection and correction of these inhomogeneities is useful to develop accurate climate studies. The procedure adopted in Morosova and Valente (2012a) is based, in a first step, on statistical methods for detection of non-climatic breaks and their corrections dT were computed considering time interval around the break and smoothing of 12 monthly correction values dT by 3-month adjacent averaging to achieve a reasonable variation of dT throughout the year; so, the correction of each time series only depends on itself.

The work developed in Morosova and Valente (2012a) produced a set of homogenized time series of monthly averages of daily minimum (T_{min}) and daily maximum (T_{max}) temperatures in three Portuguese cities – Lisbon, Oporto and Coimbra– based on the original data sets, which were tested in order to detect and correct the homogeneity breaks. Our work focuses on the monthly mean of the daily mid-range temperature time series, computed as $T_{aver} = (T_{min} + T_{max})/2$. This climatology variable was used in other previous works on temperature evolution in these Portuguese cities, for instance in Alpuim and El-Shaarawi (2009); Espírito Santo et al. (2014); Ramos et al. (2011); or in other contexts (Bengtsson and Cavanaugh, 2008; Perry and Hollis, 2005).

Relatively to the Oporto data series, the original data set was measured by the *Instituto Geofísico (Observatório Meteorológico da Serra do Pilar) da Universidade do Oporto (IGUP)*, from 1888 to 2001, comprising 114 years. The original data series of Lisbon were measured at the *Instituto Geofísico do Infante D. Luís (IGIDL)*, from 1856 to 2008, corresponding to a data set length of 153 years. The original data set of Coimbra was measured at the *Instituto Geofísico da Universidade de Coimbra (IGUC)*, from 1865 to 2005, with a length of 141 years.

Notice that Lisbon, Oporto and Coimbra have different climate characteristics (see geodesic coordinates in Table 1). While Lisbon has a humid temperate climate with dry and hot summer, Oporto and Coimbra have a temperate climate with dry and mild summers. Nonetheless, Oporto is located along the Atlantic

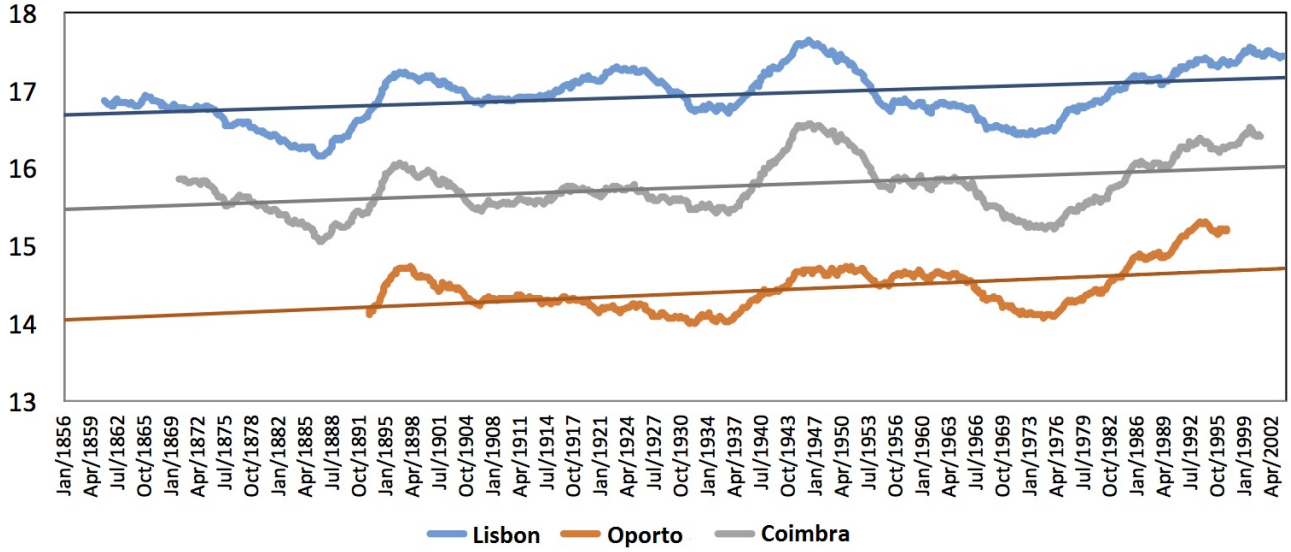

Figure 1: Ten-year moving average temperature series, with a linear trend adjustment.

Ocean in the northern coast, while Coimbra is located approximately 40 km from the Atlantic Ocean and close to the cordillera Montejunto – Serra da Estrela System, highest mountain of mainland Portugal with an altitude of 1993 m.

—— table 1 ——

Since the time series are quite long, Figure 1 shows the 10-year moving average of the time series in order to facilitate a visual inspection of an over-all behavior. Figure 1 shows that the rise in temperature varies on time. Figure 2 represents, for each city, the least squares parameters estimates of the 12 linear regression models to the respective annual time series associated with each month of the year. These estimates indicate that the temperature rise is not the same through every month of the year. In fact, in the previous work Alpuim and El-Shaarawi (2009), based on monthly temperatures data of Lisbon (1856–1999), and Prague, results showed that the rise in temperature is not equal in each month of the year. Thus, this means that the rise in temperature has two types of variability: from monthto month over the year and over time.

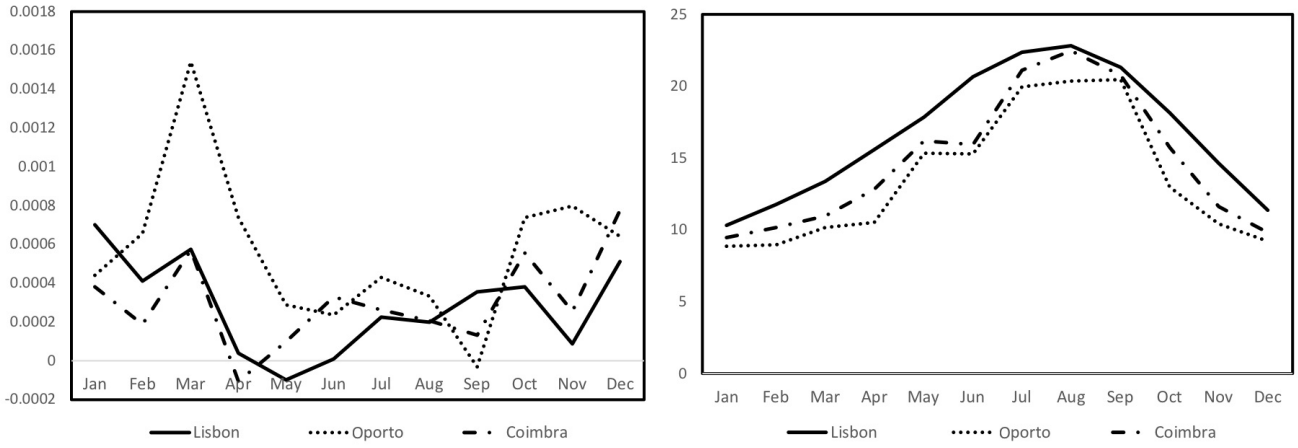

Figure 2: Least squares estimates of the linear model for each month of the year, for each city; (left) estimates of the slopes and (right) estimates of levels.

3.2 Models adjustments and validation

A first approach to model the time series under analysis was to consider, based on the exploratory analysis exposed in section 3.1, a PMLSS model (1) – (2) without the autoregressive state $\{a_{s,n}\}$ component, since it was expected that the serial correlation would be accommodated through the state vector of stochastic slopes with a PAR(1) structure with $S = 12$ seasons – the months of the year. The log-likelihood maximization procedure was initialized using the least squares estimates of the linear regression models $Y_t^{(s)} = b_s + t \cdot m_s + \xi_t^{(s)}$, where $t = s + 12k$ for $s = 1, 2, ..., 12$ and k is an integer, for the monthly subseries. Hence, the initial values for the fixed effects were the estimated levels, $\widehat{\beta}_s^{(0)} = \widehat{b}_s$, and for the mean of the periodic stochastic slopes were the estimated slopes in the regression models, $\widehat{\mu}_s^{(0)} = \widehat{m}_s$ (see Figure 2).

In the initial model validation phase, the analysis of the observed innovations series $\{\widehat{\eta}_{s,n}\}$ showed the existence of a remaining temporal correlation. In fact, both sample autocorrelation (ACF) and partial autocorrelation (PACF) functions of innovations $\{\widehat{\eta}_{s,n}\}$ of this model in three cities (Lisbon, Oporto and Coimbra) have indicated that there is a temporal correlation. Therefore, innovations series do not present the characteristics of a white noise as it was assumed; sample ACF and PACF of the innovations series show

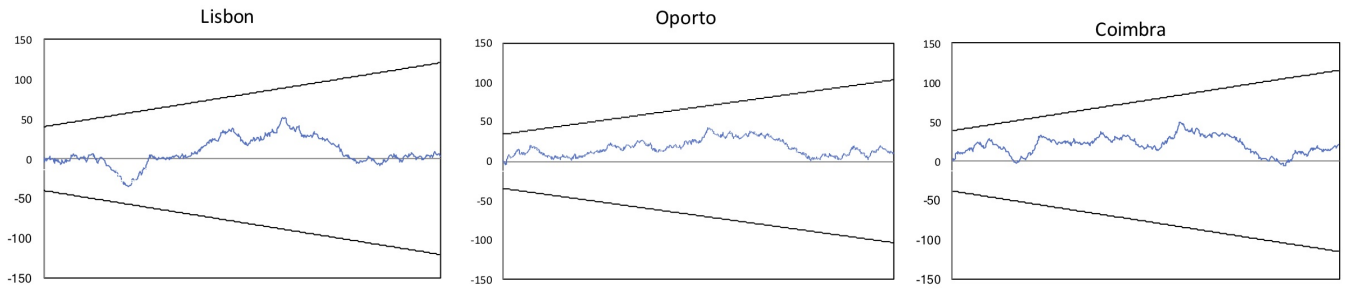

Figure 3: Plots of the cumulative sum (CUSUM).

that an autoregressive process should be incorporated into the model under analysis.

Thus, the estimation procedure described in Section 2.3 was implemented for three models (1) – (2) with the autoregressive process, using the Generalized Reduced Gradient (GRG2) method in the maximization step. In the optimization procedure, several disturbances were added to the initial values in order to assess the stability of the procedure. In each time series, the maximum likelihood procedure converged, with several significant digits, to the same solution. Before the discussion and interpretation of the results, a set of procedures were performed in order to validate the models and to evaluate their assumptions. In a global analysis, all models adjusted very well to data, since all assumptions are verified and they also present high values of the respective coefficients of determination. Innovations $\eta_{s,n}$ were assumed to have a conditional Gaussian distribution $\eta_{s,n} = Y_{s,n} - \widehat{Y}_{s|s-1,n} \sim N(0, \omega_{s,n})$. The normality of residuals series were tested considering both the Kolmogorov–Smirnov (K-S) test and the Jarque–Bera (JB) test. In all cases, normality was not rejected considering the usual 5% to the significance. In fact, in the K-S test ,all p-values were greater than 0.20 and in the JB test p-values were equal to 0.74, 0.21 and 0.10, respectively for Lisbon, Oporto and Coimbra. Moreover, histograms and QQ–plots with 95% confidence envelopes of standardized innovations allow to conclude that their empirical distributions are consistent with the Gaussian curve.Furthermore, the three innovations series did not present serial correlation, as assumed, since empirical ACF and PACF indicated that innovations are compatible with a white noise process.

Moreover, plots of cumulative sum (CUSUM) with a significance level of 5% were obtained to three residuals series. A considerable amount of information can be obtained simply by inspecting these plots; the CUSUM procedure is particularly valuable for detecting structural change (Harvey, 1996). Figure 3 shows the plots of CUSUM and the graphical analysis indicates that there are no structural changes in residuals series.

In order to assess models adjustment, coefficients of determination associated with the one-step ahead forecasts were computed, i.e., $r^2 = \text{corr}(Y_{s,n}; \widehat{Y}_{s|s-1,n})^2$. All models have coefficients of determination greater than 90%. Indeed, in Lisbon, the adjusted model has a coefficient of determination for the one-step-ahead forecast of $r^2 = 0.937$, while, in Oporto, this coefficient is $r^2 = 0.919$ and in Coimbra $r^2 = 0.922$ (see Table 2).

—- TABLE 2 —-

Furthermore, for each model the coverage probability for the empirical one-step-ahead 95% confidence intervals were computed and are close to the confidence level considered. The coverage probability values were 95.15%, 95.76% and 94.98%, in Lisbon, Oporto and Coimbra, respectively.

3.3 Warming rates estimates based on the PMLSS models adjustment

Figure 4 represents the estimated curves of the seasonal fixed effects according to the Gaussian maximum likelihood estimates of β_i, with $i = 1, 2, ..., 12$. These estimates show that Lisbon and Coimbra have a similar pattern over the year, nonetheless Lisbon has the highest values of monthly means. Results associated with Oporto indicate that the annual temperature curve in this city has a different pattern in comparison with the other two cities and also that Oporto has comparatively the lowest monthly values of temperature.

The monthly means μ_s, with $s = 1, 2, ..., 12$, of the stochastic slopes $X_{s,n}$, have an environmental special interest, since they quantify the monthly rise of the long-term temperature time series. Note that these parameters are the mean parameters of the PAR(1) model. For each month s, based on the ML estimate of $\widehat{\mu}_s$, we compute $100 \times 12 \times \widehat{\mu}_s$, which are represented in Figure 5 (see also Table 3), in order to compare

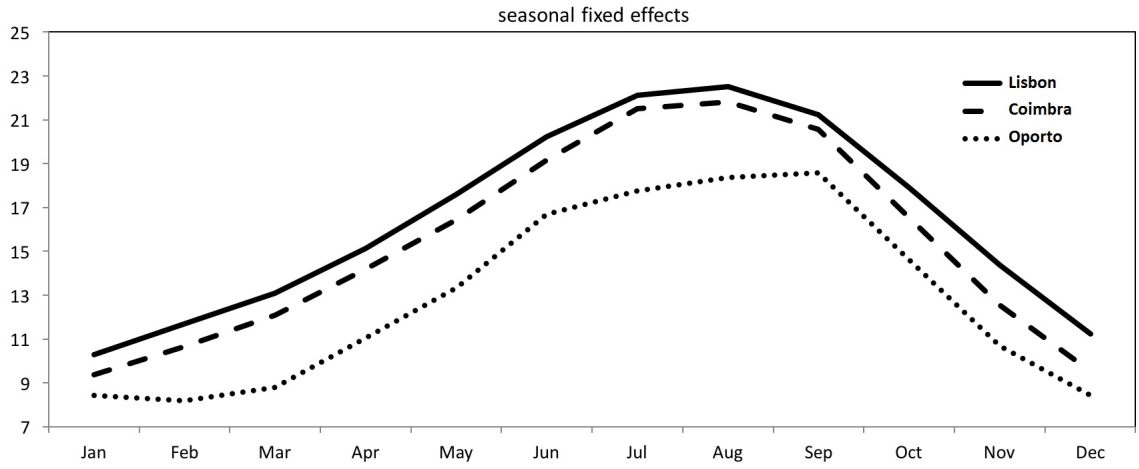

Figure 4: Maximum likelihood estimates of the seasonal fixed effects of temperature for Lisbon, Oporto and Coimbra.

the mean rise of temperature in the different cities using a reference time unit. These values represent 12 estimates of the mean rise of temperature in 100 years, that is, a mean rise per century. Hence, their average is an estimate of the mean increase of annual temperature in a century, $1/12 \sum_{s=1}^{S}(100 \times 12 \times \widehat{\mu}_s) = 100 \times 12 \times \widehat{\mu}$. In fact, if we compute, for each time series (city), the overall monthly mean rise of temperature $\mu = 1/12 \sum_{s=1}^{12} \mu_s$, then the mean μ_s in the PMLSS model can be rewritten as $\mu_s = \mu + \mu_s^*$, where μ is the overall monthly mean rise of temperature and μ_s^* is the effect of the month s, with $\sum_{s=1}^{12} \mu_s^* = 0$.

—-TABLE 3—-

The adjusted model for Lisbon estimates the mean increase, per century, in the annual temperature at 0.621°C, whereas in the corresponding models for Coimbra and Oporto the estimate is 0.545°C and 2.166°C, respectively. The annual rise in temperature in Lisbon and Coimbra is similar and less than 1°C, while in Oporto we estimated a rise exceeding 1°C, per century. Relatively to Lisbon, we can compare the annual rise of temperature with two other works. In the first work, Alpuim and El-Shaarawi (2009) used linear models with correlated errors to model Lisbon monthly temperatures time series between 1856

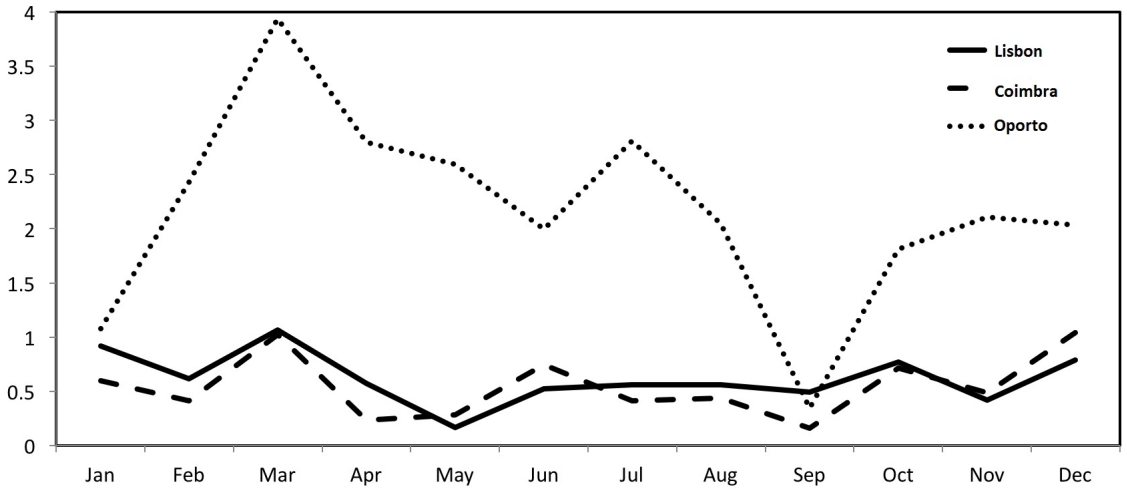

Figure 5: Estimates of monthly warming rates, per century, in Lisbon, Oporto and Coimbra.

and 1999 collected at the Climatological Archive of the Portuguese Meteorological Institute and they have obtained the value of 1.024°C as an estimate to the annual rise per century. That is, the adjusted PMLSS model to the homogenized data produced by Morosova and Valente (2012a) obtained a lower estimate than Alpuim and El-Shaarawi (2009), although both indicate a mean rise of temperature in Lisbon in the last decades. The linear model that was applied only considered a AR(1) structure in residuals series, once slopes are deterministic, whereas in the PLMSS model two types of serial correlation were considered: a VAR(1) structure in slopes and a month-to-month AR(1) structure. More recently, Costa and Monteiro (2015) obtained an estimate of the mean rise of 0.427°C, per century, in Lisbon, based on the same dataset of this work considering a dynamic linear model without a periodic structure for the stochastic slopes, which may explain the difference between both estimates.

Regarding the Oporto and Coimbra monthly mean temperature data, there are no other studies which allow us to make a direct comparison. In fact, the Lisbon series is usually chosen to make comparative European or global climate studies, because it is the best documented series and is generally considered representative of the Portuguese average climate, although one should keep in mind that there may be some

regional variations (Miranda et al., 2002). In a more thorough analysis, we can also observe the estimates of monthly mean rises in temperature for each month of the year since this value changes according to the month. In Lisbon, from April until September, the temperature rise is less than the annual rise average, whereas between December and March the temperature rise have values on average or greater than the annual average. So, mean temperatures are increasing every month, but the warming rate in winter months is greater. The largest warming rate per century is found in March with an estimated value of 1.066°C and the lowest warming rate is estimated in May with the value of 0.167°C. The monthly mean rises of temperature in Coimbra are close to the corresponding values in Lisbon time series and with a similar shape (empirical linear correlation of $r = 0.700$). In Coimbra, the largest warming rate per century is in December with 1.044°C, very close to the estimate rise in March with 1.024°C. Nevertheless, the lowest warming rate per century is found in September with 0.158°C . In Oporto, eventhough annual warming is more significant than in Lisbon or Coimbra, results show that there is a clear annual cycle in the warming rate. In fact, excluding June, which has a rate similar to the annual average, the remaining months (between February and July) have warming rates greater than the annual rate average, while from August until January warming rates are lower. Therefore, in Oporto there are two periods of the year with different warming rates.

3.4 Global analysis of the adjusted models

The proposed PMLSS model has three types of variability sources. One is related to the errors of the periodic state associated with the stochastic slopes, $\varepsilon_{s,n}$, with variance $\sigma^2_{\varepsilon,s}$; another is associated with the error $\omega_{s,n}$ of the autoregressive process $\{a_{s,n}\}$; and the last source of variability is the observation error $e_{s,n}$ with variance σ^2_e.

—— TABLE 4 ———

Table 4 presents Gaussian maximum likelihood estimates of all variances present in the PMLSS models. Lisbon and Coimbra have observations errors $e_{s,n}$ with similar estimates of their variances while in the time series of temperatures of Oporto this estimate is greater. Coimbra presents, each month, the highest

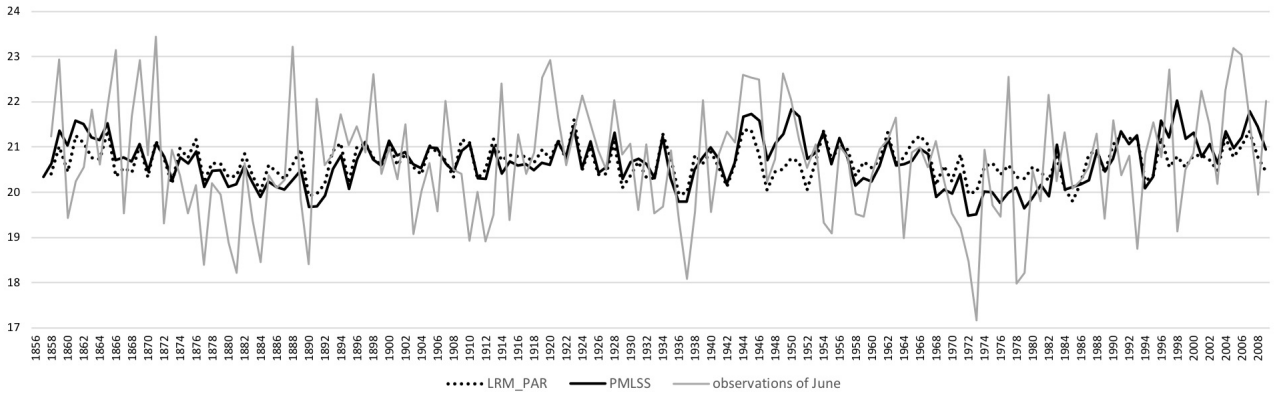

Figure 6: Trend estimates of June in Lisbon obtained by PMLSS model and the multiple linear regression model with PAR errors.

estimated variance of the periodic state errors. As expected, the autoregressive processes $a_{s,n}$ has errors with small variances, since this component incorporates the month-to-month serial correlation. Estimates of the autoregressive parameters involved in the PMLSS model are presented in Table 5. On the one hand, all three autoregressive processes $a_{s,n}$ are estimated as stationary, since estimates of ϕ in three cities verify comfortably the stationarity condition $|\widehat{\phi}| < 1$. On the other hand, the PAR(1) process, that represents the periodic stochastic slopes, is stationary as well, since we have $\prod_{s=1}^{12} \widehat{\phi}_s < 1$ in Lisbon, Coimbra and Oporto.

—— TABLE 5 ——

A competing multiple linear regression model with PAR(1) errors, MLR_PAR, was adjusted to perform a comparison of trends estimates. This model is an improvement of the linear regression model considered in the initialization of the log-likelihood maximization procedure, that is, $Y_{s,n} = b_s + t \cdot m_s + \xi_{(s,n)}$ where $\xi_{(s,n)}$ follows a PAR(1) process. Analyzing the PMLSS model trend estimates justaposed against MLR_PAR trends estimates, we concluded that the these estimates are more similar when the month's subseries trend is approximately linear (fixed). Unlike this, when in a month trend is not linear, the PMLSS model captures these changes in a more dynamic way than the MLR_PAR model. For instance, Figure 6 shows trend estimates of June in Lisbon obtained by PMLSS model and the MLR_PAR. The PMLSS model has a PAR

structure in the slopes process, which allows for local linear trends, while the MLR_PAR model incorporates a periodic structure in errors.

Since PMLSS models adjusted very well to three time series, their structural components can be used to monitor the rise in temperature both in a large scale, for instance in a secular period as was done before, or in a on-line procedure, that is, in a month-to-month analysis.

In an online scheme, the Kalman filter provides one-step-ahead forecasts for monthly temperature, $\widehat{Y}_{s|s-1,n}$ with the respective empirical confidence intervals at $(1-\alpha)\times 100\%$ level

$$Y_{s,n} = \widehat{Y}_{s|s-1,n} \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\omega}_{s,n}}$$

where $z_{1-\alpha/2}$ represents the normal quantile of probability $1-\alpha/2$. The unobservable structural components can be predicted through the Kalman filter equation and the associated confidence intervals. Hence, confidence intervals for the stochastic slopes, $\{X_{s,n}\}$, can be computed through

$$X_{s,n} = \widehat{X}_{s|s,n} \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{p}^{(2,2)}_{s|s,n}}$$

while the confidence interval for the AR(1) process, $\{a_{s,n}\}$, which incorporates the month-to-month serial correlation is

$$a_{s,n} = \widehat{a}_{s|s,n} \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{p}^{(1,1)}_{s|s,n}}$$

where $\widehat{p}^{(i,i)}_{s,n}$ represents the ith element, with $i=1,2$, of the diagonal of the MSE matrix $\widehat{\mathbf{P}}_{s|s,n}$.

Figure 7 represents predictions with the respective empirical 95% confidence intervals of the monthly mean temperature in Lisbon and the filtered predictions of the unobservable components obtained with the adjusted PMLSS model in last decade of available data. These results allow for the monitoring of monthly rise of temperature in mean temperature measurement, as well as in the monthly ratio of warming filtering errors of observation.

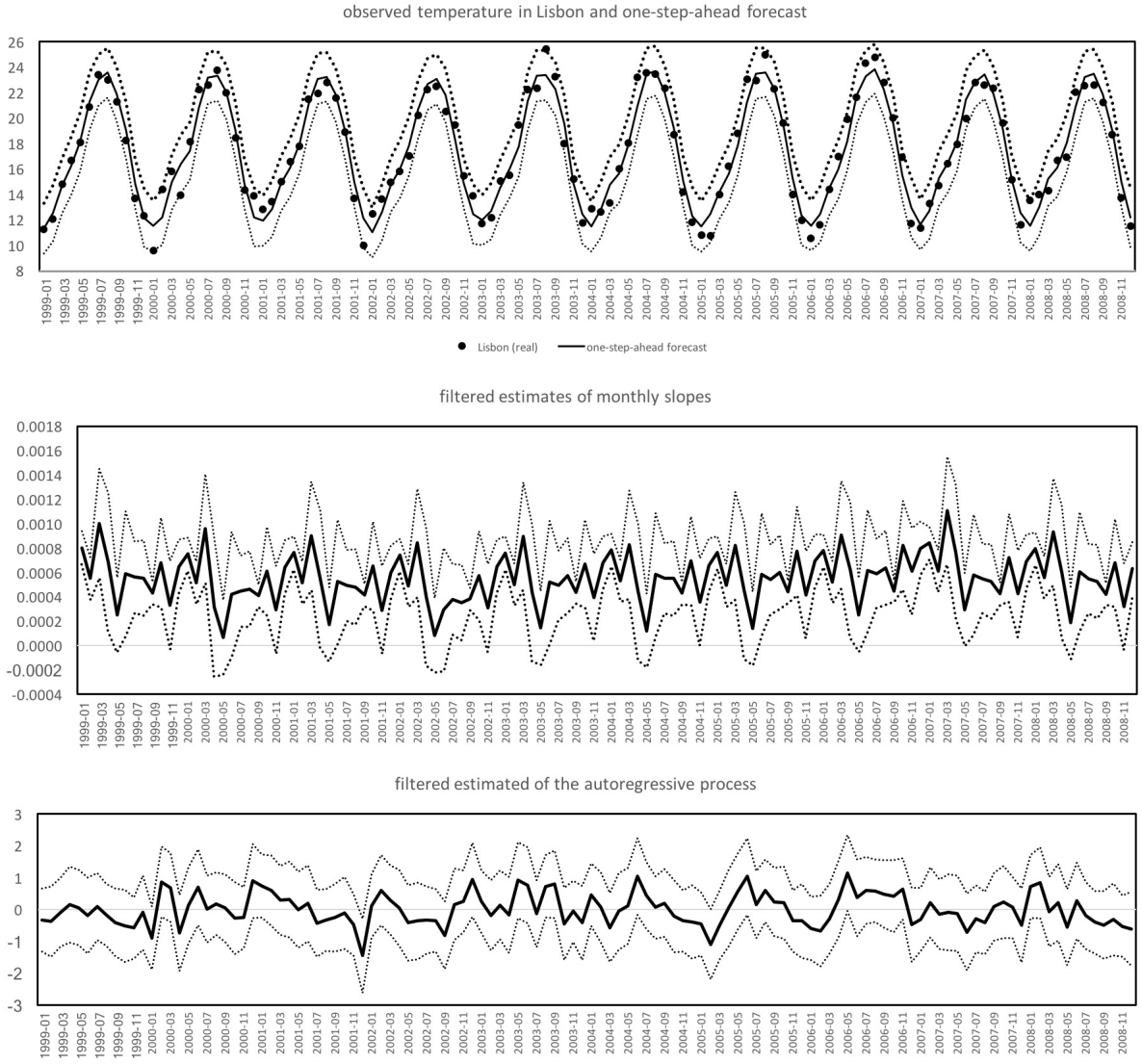

Figure 7: (top) Observed temperature (points) in Lisbon, $Y_{s,n}$, with the one-step-ahead forecast, $\widehat{Y}_{s|s-1,n}$; (middle) filtered estimates of the PAR(1) process, $\widehat{X}_{s|s,n}$; (below) filtered estimates of autoregressive process, $\widehat{a}_{s|s,n}$, with empirical 95% confidence intervals, related to the last decade (January 1999 to December 2008).

4 Conclusions and discussion

In this work, we have proposed a periodic state space model to analyze long-term temperature time series. The proposed framework incorporates fixed and stochastic effects, namely, the seasonality behavior with fixed levels effects and dynamic slopes over the months of the year. Previous works indicated that temperature rise differs between months, as well as an autoregressive process in order to accommodate the serial correlation. The state space representation allows for the achievement of accurate one-step-ahead forecasts for the monthly mean temperature and for the unobservable structural components, as well as filtered estimates of states. These estimates are computed by a periodic version of the Kalman filter algorithm.

The proposed model was adjusted to monthly mean of the daily mid-range temperature long-term time series of three Portuguese cities: Lisbon, Oporto and Coimbra. Main results showed that Lisbon and Coimbra had a similar annual rise of temperature: 0.6212°C and 0.5454°C, per century, respectively. Additionally, it was estimated that, in Oporto, there is a significant mean rise in temperature of 2.1655°C per century. In a more refined analysis, it was found that there is a similarity between mean monthly rises in Lisbon and Coimbra while Oporto has a unique rise pattern. In fact, Oporto has greater warming rates in spring and in the beginning of summer, and lower warming rates at the end of summer until the middle of winter. In Lisbon and Coimbra the biggest warming rates were found in winter and in the beginning of spring and the lowest rise rates were found from the middle of spring until the beginning of autumn.

The PMLSS model accommodates the most common properties of environmental. The seasonality is incorporated in two ways: seasonal levels as fixed effects and in the periodic process of monthly slopes; the month-to-month serial correlation is accommodated by an autoregressive process. The proposed model can be improved in order to include spatial-variation when a set of time series of monthly mean temperature are available. If the number of locations is significant and locations are geographically scattered, a spatial structure can be incorporated in the model via the error in the observation equation, through the specification of a spatially structured variance-covariance matrix (Castro Morales et al., 2013). Alternatively, the spatial

structure can be incorporated in an additive way according to an isotropic Gaussian process as proposed in Cunha et al. (2017).

5 Acknowledgements

The authors would like to thank the anonymous referees for many helpful critics and suggestions that contributed to improve this paper. Authors were partially supported by Portuguese funds through the CIDMA - Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology ("FCT– Fundação para a Ciência e a Tecnologia"), within project UID/MAT/04106/2013.

6 Appendix

In model (6)–(7) the dimension of the vector of parameters Θ can be quite large, depending on the number of seasons S of the year that is being considered in the model. Taking into account that temperature data in this work has a monthly periodic structure, that is $S = 12$, we have $\tau = 51$ parameters to be estimated, which makes the optimization problem computationally unstable. Hence, in this case a partitioned algorithm with leapfrog (non-simultaneous) iterations is adopted to replace the original estimation problem into a series of problems of lower dimension (Smyth, 1996). Without loss of generality, lets assume that Θ is partitioned into subvectors Θ_1 and Θ_2, allowing for the updating equation to be written as

$$\begin{aligned} \Theta_1^{(k+1)} &= \underset{\Theta_1}{\operatorname{argmax}} \log(\mathcal{L}_1(\Theta_1^{(k)}, \Theta_2^{(k)}; \mathcal{Y})) \\ \Theta_2^{(k+1)} &= \underset{\Theta_2}{\operatorname{argmax}} \log(\mathcal{L}_2(\Theta_1^{(k+1)}, \Theta_2^{(k)}; \mathcal{Y})). \end{aligned}$$

In the model under analysis, the natural partition of Θ is considering Θ_1 the fixed effects β_s; Θ_2 the slopes averages μ_s, Θ_3 the autoregressive parameters ϕ_s,$s = 1, \cdots, S$, and ϕ_a and Θ_4 with $S + 2$ variances

parameters.

So, the ML estimates are computed through the iterative procedure:

Step 1 Obtain least squares estimates of intercept, slope and error variance of a linear trend model for each month, $Y_t^{(s)} = b_s + t \cdot m_s + \xi_t^{(s)}$, and $\xi_t^{(s)}$ is a random error with $\text{var}(\xi_t^{(s)}) = \sigma^2_{s,\xi}$, where $t = s + Sk$ for $s = 1, 2, ..., S$, $k = 0, ..., N-1$.

Step 2 Consider initial values to subvector of fixed effects and slopes averages such that $\Theta_1^{(0)} = \{\widehat{b}_1, ..., \widehat{b}_S\}$ and $\Theta_2^{(0)} = \{\widehat{m}_1, ..., \widehat{m}_S\}$.

Step 3 Initialize the subvector of the autoregressive parameters taking initial values satisfying the cyclo-stationary of the process $\{X_{s,n}\}$ and the AR(1) process, for instance $\phi_k^{(0)} = 0.5^{1/S}$ and $\phi_a^{(0)} = 0.5$, that is, $\Theta_3^{(0)} = \{0.5, 0.5^{1/S}, ..., 0.5^{1/S}\}$.

Step 4 Take initial values to the subvector of variances $\Theta_4^{(0)} = \{\sigma_e^{2(0)}, \sigma_\omega^{2(0)}, \sigma_{1,\epsilon}^{2(0)}, ..., \sigma_{S,\epsilon}^{2(0)}\}$ considering the estimates $\widehat{\sigma}^2_{s,\xi}$ as the total variability in each month, for instance, $\widehat{\sigma}^{2(0)}_{s,\epsilon} = \widehat{\sigma}^{2(0)}_e = \widehat{\sigma}^2_{s,\xi}/2$, and a small value to monthly variability of slopes, $(\sigma_\omega^{2(0)} = 10^{-10})$.

At this step, vector Θ is fully initialized with values $\Theta^{(0)} = \{\Theta_1^{(0)}, \Theta_2^{(0)}, \Theta_3^{(0)}, \Theta_4^{(0)}\}$;

Step 5 Initialize the periodic Kalman filter according to stationary properties of the state process; without loss of generality, considering that the first observation is related to the season $s = 1$, take $\widehat{\mathbf{X}}_{1|0,1} = [\ 0 \quad \widehat{\mu}_1^{(0)}\ ]'$ and $\mathbf{P}_{1|0,1} = \text{diag}\{\widehat{\sigma}_\omega^{2(0)}(1 - \widehat{\phi}_a^{2(0)})^{-1}, \widehat{\sigma}_1^2\}$, where $\widehat{\sigma}_1^{2(0)}$ is computed by Eq. 3.

Step 6 Compute the periodic Kalman filter estimators by equations Eq. (6), Eq. (7) and Eq. (8); and the log-likelihood $\log(\mathcal{L}(\Theta^{(0)}; \mathcal{Y}))$ by Eq. (9).

Step 7 Use a computational routine in order to perform, in a leapfrog algorithm, the direct numerical

maximization of the log-likelihood function to obtain $\Theta_j^{(k+1)}$, with $j = 1, 2, 3, 4$, by:

$$\begin{aligned}
\Theta_1^{(k+1)} &= \underset{\Theta_1}{\operatorname{argmax}} \log(\mathcal{L}_1(\Theta_1^{(k)}, \Theta_2^{(k)}, \Theta_3^{(k)}, \Theta_4^{(k)}; \mathcal{Y})) \\
\Theta_2^{(k+1)} &= \underset{\Theta_2}{\operatorname{argmax}} \log(\mathcal{L}_2(\Theta_1^{(k+1)}, \Theta_2^{(k)}, \Theta_3^{(k)}, \Theta_4^{(k)}; \mathcal{Y})) \\
\Theta_3^{(k+1)} &= \underset{\Theta_3}{\operatorname{argmax}} \log(\mathcal{L}_3(\Theta_1^{(k+1)}, \Theta_2^{(k+1)}, \Theta_3^{(k)}, \Theta_4^{(k)}; \mathcal{Y})) \\
\Theta_4^{(k+1)} &= \underset{\Theta_3}{\operatorname{argmax}} \log(\mathcal{L}_4(\Theta_1^{(k+1)}, \Theta_2^{(k+1)}, \Theta_3^{(k+1)}, \Theta_4^{(k)}; \mathcal{Y}))
\end{aligned}$$

In this step the parameters estimates are available in iteration $k + 1$, $\Theta^{(k+1)}$.

Step 8 Repeat steps 5 to 7 up to the convergence.

References

Abbasnia, M. and Toros, H. (2016). Future changes in maximum temperature using the statistical downscaling model (SDSM) at selected stations of Iran. *Model. Earth Syst. Environ.*, 68. doi: 10.1007/s40808-016-0112-z.

Alpuim, T. and El-Shaarawi, A. (2009). Modeling monthly temperature data in Lisbon and Prague. *Environmetrics*, 20:835–852. doi: 10.1002/env.964.

Bengtsson, T. and Cavanaugh, J. (2008). State-space discrimination and clustering of atmospheric time series data based on Kullback information measures. *Environmetrics*, 19:103–121. doi: 10.1002/env.859.

Bližňák, V., Valente, M., and Bethke, J. (2015). Homogenization of time series from portugal and its former colonies for the period from the late 19th to the early 21st century. *International Journal of Climatology*, 35:2400–2418.

Castro Morales, F., Gamerman, D., and Paez, M.S. (2013). State space models with spatial deformation. *M.S. Environ Ecol Stat*, 20:191–214.

Costa, M. and Monteiro, M. (2015). Dynamic linear modeling of homogenized monthly temperature in Lisbon. In Daniels, J. A., editor, *Advances in Environmental Research*, pages 203–215. Nova Science Publishers.

Costa, M. and Monteiro, M. (2016). Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters. *J Stat Plan Infer*, 176:22–32. doi: 10.1016/j.jspi.2016.04.002.

Cunha, M., Gamerman, D., Fuentes, M., and Paez, M. (2017). A non-stationary spatial model for temperature interpolation applied to the state of Rio de Janeiro. *J Roy Stat Soc C-App*, 66:919–939.

EEA (2018). European Environment Agency (2018). Global and European temperatures. Available as a website at https://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature-8/assessment.

Espírito Santo, F., Lima, I., Ramos, A., and Trigo, R. (2014). Trends in seasonal surface air temperature in mainland Portugal, since 1941. *International Journal of Climatology*, 34:1814–1837.

Fischer, M. (2015). Predictable components in australian daily temperature data. *Journal of Climate*, 28:5969–5984.

Freitas, C., Dedekind, M., and Brill, B. (2015). A reanalysis of long-term surface air temperature trends in New Zealand. *Environmental Modeling & Assessment*, 20:399–410. doi: 10.1007/s12665-015-5139-3.

Gardner, W., Napolitano, A., and Paura, L. (2006). Cyclostationary: half a century of research. *Signal Process*, 86:639–697.

Harvey, A. (1996). *Forecasting structural time series models and the Kalman filter*. Cambridge University Press.

441 Kleiber, W., Katz, R., and Rajagopalan, B. (2013). Daily minimum and maximum temperature simulation
442 over complex terrain. *Annals of Applied Statistics*, 7:588–612.

443 Kokic, P., Crimp, S., and Howden, M. (2011). Forecasting climate variables using a mixed-effect state-space
444 model. *Environmetrics*, 22:409–419. doi: 10.1002/env.1074.

445 Liu, G., Shao, Q., Lund, R., and Woody, J. (2016). Testing for seasonal means in time series data. *Envi-*
446 *ronmetrics*, 27:198–211. doi: 10.1002/env.2383.

447 Miranda, P., Coelho, F.E.S., Tomé, A.R., Valente, M.A., Carvalho, A., Pires, C., Pires, H.O., Pires, V.C.,
448 and Ramalho, C. (2002). 20th century portuguese climate and climate scenarios. In FD, S., K, F., and R,
449 M., editors, *Climate Change in Portugal: Scenarios, Impacts and Adaptation Measures (SIAM Project)*,
450 pages 23–83. Gradiva.

451 Monteiro, M., Pereira, I., and Scotto, M. (2010). Integer-valued autoregressive processes with periodic
452 structure. *J Stat Plan Infer*, 140:1529–1541.

453 Moreno, E., Girn, J., and Garc´ ia Ferrer, A. (2013). A consistent on-line Bayesian procedure for detecting
454 change points. *Environmetrics*, 24:342–356. doi: 10.1002/env.2207.

455 Morosova, A. and Valente, M. (2012a). Homogenization of Portuguese long-term temperature data series:
456 Lisbon, Coimbra and Oporto. *Earth System Science Dat*, 4:187–213.

457 Morosova, A. and Valente, M. (2012b). Homogenization of Portuguese long-term temperature data series:
458 Lisbon, Coimbra and Oporto. PANGAEA. doi:10.1594/PANGAEA.785377.

459 Ngo, N. and Horton, R. (2016). Climate change and fetal health: The impacts of exposure to extreme
460 temperatures in New York City. *Environmental Research*, 144:158–164.

461 Obeysekera, J. and Salas, J. (1986). Modeling of aggregated hydrologic time series. *Journal of Hydrology*,
462 86:197–219.

463 Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models
464 of individual animal movement. *Trends in Ecology & Evolution*, 33:87–94.

465 Perry, M. and Hollis, D. (2005). The development of a new set of long-term climate averages for the UK.
466 *International Journal of Climatology*, 25:1023–1039.

467 Poppick, A., McInerney, D.J., Moyer, E.J., and Stein, M.L. (2016). Temperatures in transient climates:
468 Improved methods for simulations with evolving temporal covariances. *Annals of Applied Statistics*,
469 10:477–505.

470 Ramos, A., Trigo, R.M., and Santos, F.E. (2011). Evolution of extreme temperatures over Portugal: and
471 future scenarios. *Climate Research*, 48:177–192.

472 Reich, B. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental
473 processes. *J Roy Stat Soc C-App*, 61:535–553.

474 Rodríguez, A. and Ruiz, E. (2012). Bootstrap prediction mean squared errors of unobserved states based
475 on the Kalman filter with estimated parameters. *Computational Statistics and Data Analysis*, 56:62–74.

476 Shumway, R.H.and Stoffer, D. (2017). *Time Series Analysis and Its Applications With R Examples*. Springer,
477 New York, 4th edition edition.

478 Smyth, G. (1996). Partitioned algorithms for maximum likelihood and other non-linear estimation. *Statistics*
479 *and Computing*, 6:201–216.

480 Tandeo, P., Ailliot, P., and Autret, E. (2011). Linear gaussian state-space model with irregular sampling:
481 application to sea surface temperature. *Stochastic Environmental Research and Risk Assessment*, 25:793–
482 804.

483 Trevin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of*
484 *Climatology*, 33:1510–1529.

485 Trigo, R. and Palutikof, J. (1999). Simulation of daily temperature for climate change scenarios over
486 Portugal: a neural network model approach. *Climate Research*, 13:45–59.

487 UN (2016). United Nations. http://www.un.org/sustainabledevelopment/climate-change. accessed in
488 December 7, 2017.

489 Ursu, E. and Pereau, J. (2016). Application of periodic autoregressive process to the modeling of the Garonne
490 river flows. *Stochastic Environmental Research and Risk Assessment*, 30:1785–1795. doi: 10.1007/s00477-
491 015-1193-3.

492 Wang, W., Shao, Q., Yang, T., Peng, S., Yu, Z., Taylor, J., Xing, W., Zhao, C., and Sun, F. (2013).
493 Changes in daily temperature and precipitation extremes in the Yellow River basin, China. *Stochastic*
494 *Environmental Research and Risk Assessment*, 27:401–421.

495 William, C., Menne, M.J., and Thorne, P.W. (2012). Benchmarking the performance of pairwise homog-
496 enization of surface temperatures in the United States. *Journal of Geophysical Research*, 117. doi:
497 10.1029/2011JD016761.

498 Withers, C. and Nadarajah, S. (2015). Estimating trend from seasonal data: is daily, monthly or annual
499 data best? *Environmetrics*, 26:488–501. doi:10.1002/env.2356.

500 Xu, Y., Gao, X., Shi, Y., and Botao, Z. (2015). Detection and attribution analysis of annual mean temper-
501 ature changes in China. *Climate Research*, 63:61–71.

Table 1: Geodesic coordinates of three stations in Lisbon, Oporto and Coimbra

Station	Latitude	Longitude	Altitude
Lisbon	$38^o\ 43'$ N	$9^o\ 09'$ W	77 m
Porto	$41^o\ 08'$ N	$8^o\ 36'$ W	93 m
Coimbra	$40^o\ 12'$ N	$8^o\ 25'$ W	141 m

Table 2: Coefficients of determination and the log-likelihood of PMLSS models of Lisbon, Coimbra and Oporto.

	Lisbon	Coimbra	Porto
r^2	0.937	0.922	0.919
$-\log(\mathcal{L})$	2744.35	2795.64	2152.73

Table 3: Estimates of the mean rise of temperature Celsius degree per century, in Lisbon, Coimbra and Oporto.

	Lisbon	Coimbra	Porto
Jan	0.918	0.596	1.076
Feb	0.615	0.411	2.427
Mar	1.066	1.024	3.939
Apr	0.574	0.232	2.801
May	0.167	0.283	2.594
Jun	0.521	0.747	1.998
Jul	0.559	0.412	2.810
Aug	0.562	0.437	2.047
Sep	0.492	0.158	0.343
Oct	0.771	0.715	1.811
Nov	0.420	0.487	2.109
Dec	0.789	1.044	2.031
overall mean	0.621	0.545	2.166

Table 4: Estimates of the variances in the PMLSS models of Lisbon, Coimbra and Oporto.

	month	Lisbon	Coimbra	Porto
$\widehat{\sigma}^2_{\varepsilon,s}$	Jan	0.404	0.488	0.330
	Feb	0.662	1.073	0.909
	Mar	0.457	1.061	0.749
	Apr	0.554	1.082	0.661
	May	0.751	1.174	0.633
	Jun	0.592	1.017	0.394
	Jul	0.438	0.782	0.475
	Aug	0.345	0.491	0.295
	Sep	0.582	1.077	0.658
	Oct	0.603	1.398	1.029
	Nov	0.305	0.836	0.720
	Dec	0.795	0.920	0.871
$\widehat{\sigma}^2_{\omega}$		9.013×10^{-10}	9.992×10^{-12}	6.694×10^{-10}
$\widehat{\sigma}^2_{e}$		0.548	0.562	0.643

Table 5: Estimates of autoregressive parameters in the PMLSS models of Lisbon, Coimbra and Oporto.

	month	Lisbon	Coimbra	Porto
$\widehat{\phi}_s$	Jan	0.549	0.914	0.428
	Feb	1.231	2.162	3.029
	Mar	2.703	2.376	1.265
	Apr	1.369	0.961	0.907
	May	0.537	0.402	1.165
	Jun	1.763	1.729	0.755
	Jul	0.566	0.232	1.326
	Aug	1.078	1.485	0.712
	Sep	0.221	0.035	0.227
	Oct	4.176	7.668	2.398
	Nov	1.011	4.142	1.231
	Dec	0.596	0.730	1.087
$\prod_{s=1}^{12} \widehat{\phi}_s$		0.804	0.869	0.900
$\widehat{\phi}_a$		0.381	0.307	0.351