XIV Congreso Galego de Estatística e Investigación de Operacións
Vigo, 24, 25 e 26 de outubro de 2019

CHANGE POINT ANALYSIS IN A STATE SPACE FRAMEWORK TO MONTHLY TEMPERATURE DATA IN EUROPEAN CITIES

Magda Monteiro[1,2], Marco costa[1,2]

[1] ESTGA - Águeda School of Technology and Management, University of Aveiro
[2] CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro

Abstract

In this work, we present time series of monthly average temperatures in several European locations which were statistically analyzed using a state space approach, where it is considered a model with a deterministic seasonal component and a stochastic trend. The analysis of smoother prediction of the stochastic trend and its comparison in a temporal viewpoint can reveal patterns about warming in Europe. The temperature rise rates in Europe seem to have increased in the last decades when compared with longer periods, hence a change point detection method is applied to the trend component in order to identify these possible changes in the monthly temperature rise rates. The adopted methodology pointed out, for most series a change point in the late eighties.

Keywords : climate change, monthly data, temperature data, time series analysis, state space modeling, Change point detection

1. INTRODUCTION

Global warming and the associated climate changes are topics in the world's agenda, which are receiving much attention from the scientific community, due to their major social, economic and health impacts on human life.

These changes are already being felt all over the world, with the increase of natural catastrophes such as droughts, floods, tropical storms, among others. Nevertheless, global warming effects vary around the world, therefore this phenomenon must be monitored at a smaller scale. In particular, given our location, we are interested in understanding these phenomena at Europe scale. In this sense, the analysis of local temperature time series is an important topic that can contribute to a better understanding of climate dynamics at smaller scales, allowing for the identification of their change patterns. Thus, this analysis can be very useful for the application of efficient environment monitoring processes.

This work is a preliminary analysis of long monthly average temperature time series in fifteen European cities in order to find change point associated with rise rates changes of temperature. The Data is available at the Climate Data Online and comprises the period between January 1900 and December 2017, making a total of 1416 observations. The location of the cities and data descriptive statistics can be found in Table 1. The major data of this work has already been analyzed in Costa and Monteiro(2017), where state space models associated to the Kalman filter algorithm and also a cluster procedure to smoother trend levels were applied in order to investigate patterns on the temperature rise in Europe. A conclusion therein is that temperature rise rates in Europe seem to have increased in the last decades being important to find the change points and use the information from there to get accurate temperature rise rates. Therefore, in the present work it is applied a change point detection procedure of maximum type in order to assess if there is a change point in the temperature trend pattern for each city under analysis.

There are several works on change point detection in climate or environmental settings. For instance, Reeves et al.(2007) makes a review of this topic for models with independent and identically distributed (i.i.d.) errors and Jarušková and Antoch(2018) applied several change points procedures to Klementinum temperature. In line with this last work, it will be considered basic statistical tests with the applications of maximum type statistics to the stochastic trend of the temperature time series in Europe.

city (country)	coord.	min 1900 to 2017	max 1900 to 2017	average 1900 to 1917	average 2000 to 2017	missing values (%)
Berlin (Germany)	52.47N, 13.30E	-10.7	23.0	8.7	9.9	2.3
Bucharest (Romania)	44.40N, 26.10E	-11.5	27.5	10.1	11.1	2.6
Copenhagen (Denmark)	55.68N, 12.55E	-6.7	21.6	8.6	9.6	0.4
Dublin (Ireland)	53.43N, 6.25W	0.0	17.1	9.2	9.6	2.2
Kiev (Ukraine)	50.40N, 30.57E	-15.3	26.0	8.1	9.4	0.9
Lisbon (Portugal)	38.72N, 9.15W	8.0	25.1	16.9	17.4	3.4
Minsk (Belarus)	53.93N, 27.63E	-16.8	22.6	6.1	7.4	1.8
Nantes (France)	47.17N, 1.60W	-1.2	23.3	12.0	12.5	0.4
Prague (Czech Rep.)	50.10N, 14.25E	-13.1	22.0	7.9	9.1	8.2
S. Petersburg (Russia)	59.97N, 30.30E	-18.3	24.4	5.3	6.5	0.0
Talin (Estonia)	59.42N, 24.80E	-15.5	21.5	5.3	6.5	5.4
Vienna (Austria)	48.25N, 16.37E	-9.5	24.4	10.1	11.3	0.2
Vilnius (Lithuania)	54.63N, 25.10E	-17.1	21.8	5.8	7.3	2.6
Zagreb (Croatia)	45.82N, 15.98E	-7.0	25.9	11.8	13.0	1.4
Zurich (Switzerland)	47.38N, 8.57E	-8.7	22.4	8.7	9.8	0.0

Table 1: Characterization of time series data set.

2. MODEL DESCRIPTION

In order to model, in each city, the monthly temperature time series, it is considered a model with a deterministic seasonal component and a stochastic trend, defined by

$$Y_t = \tau_t + \sum_{i=1}^{12} \beta_i S_{t,i} + e_{1,t} \tag{1}$$

$$\tau_t = \mu + \phi(\tau_{t-1} - \mu) + e_{2,t} \tag{2}$$

where $\{\tau_t\}$ is the trend component that follows a stationary autoregressive process with mean μ and the autoregressive parameter $|\phi| < 1$; the error $e_{i,t}$, with $i = 1, 2$, follows a white noise process ($E(e_{i,t}) = 0$, $var(e^2_{i,t}) = \sigma^2_i$ and $E(e_{i,t}e_{i,r}) = 0$, for $t \neq r$), and are uncorrelated errors, that is $E(e_{1,t}e_{2,r}) = 0$, $\forall t, r$. The error $e_{1,t}$ is called the observation error and it can be seen as a measure error, whereas the error $e_{1,t}$ is called the state error and translates the randomness of the trend component.

Note that state τ_t is an unobservable variable and its predictions must be obtained. For that, the state process $\{\tau_t\}$ in the model (1)-(2) can be predicted by Kalman filter and Kalman smoother estimators.

This model was also used in Costa and Monteiro(2017) to which a clustering procedure was applied to understand the patterns associated to the climate in Europe. In terms of parameter estimation, similar to what was done in this previous work, a classical decomposition approach was used combining the least squares estimation of the seasonal parameters β_i, with $i = 1, ..., 12$, with a distribution-free estimators developed to state space models in order to estimate the remaining parameters.

2. CHANGE POINT DETECTION

In order to discover an abrupt change in the behaviour of the temperature rise rates, a basic statistical test applying maximum type statistics to detect a change in location is performed. If there are no structural changes in the temperature trend, it is expected that τ_t has no changes over time in its mean. Therefore, it is applied a statistical test to the smoother predictions, $\tau_{t|n}$, of τ_t having into into account that $\tau_1, \tau_2, ..., \tau_n$ are an AR(1) process and that smoother predictions are the predictors of the non-observable component that present lower mean square error when compared with the other Kalman filter predictions.

Briefly, if we are in the presence of an AR(1) process, $\{X_t\}$, the hypotheses of the test of *maximum type* are

H_0: X_t are an $AR(1)$ sequence normally distributed variables according to the same $N(\mu, \sigma^2)$

vs

H_1: there is a time point $k \in \{1, ..., n-1\}$ such that $X_1, ..., X_k$ are distributed according to $N(\mu_1, \sigma^2)$ and $X_{k+1}, ..., X_n$ are distributed according to $N(\mu_2, \sigma^2)$.

Assuming σ^2 is unknown, then the test statistic $T(n)$ is the maximum of the absolute values of two sample t-test statistics

$$T(n) = \max_{1 \le k < n} |T_k| = \max_{1 \le k < n} \sqrt{\frac{(n-k)k}{n}} \; \frac{\left|\overline{X}_k - \overline{X}_k^*\right|}{s_k}$$

where

$$\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \; \overline{X}_k^* = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \text{ and } s_k = \sqrt{\frac{1}{n-2}\left[\sum_{i=1}^{k}\left(X_i - \overline{X}_k\right)^2 + \sum_{i=k+1}^{n}\left(X_i - \overline{X}_k^*\right)^2\right]}.$$

The null hypothesis can be rejected if the statistic $T(n)$ is greater than the critical value. The exact distribution of $T(n)$, for independent variables, is very complex and Worsley(1979) was able to calculate the true critical values only for the number of observations n less than 10. Alternatively, approximate critical values can be computed by other methods as the Bonferroni inequality, simulation or the asymptotic distribution.

Antoch et al.(1997) shows that when random variables X_1, X_2, ..., X_n are not independent but form an ARMA sequence then the asymptotic critical values of the test statistics considering independence have to be multiplied by $\sqrt{2\pi f(0)/\gamma}$, where $\gamma = varX_t$ and $f(\cdot)$ denotes the spectral density function of the corresponding ARMA process. Especially for an AR(1) sequence, the critical values should be multiplied by $\left[(1+\phi)(1-\phi)^{-1}\right]^{1/2}$ where ϕ is the first autoregressive coefficient, (Jarušková,1997).

The present work uses critical values obtained by the asymptotic distribution which expression can be found in Jarušková (1997) and multiplyed by the AR(1) factor with the estimates of ϕ.

For each city under study, these basic change point procedure detected a change and Table 2 summarizes these changes and also present the mean of smoother predictors before and after changes. The majority of temperature time series presents a change point in the late eighties. Lisbon presents a change point much earlier, in the previous decade, and Bucharest presents a change point about twenty years later.

Note that western cities present the lowest mean values of $\tau_{t|n}$ and because of that the shorter distances between averages before and after the detected change point with differences below 0.5 ^{o}C. In the opposite site are the major eastern cities with differences above 1^{o}C.

This work is at an early stage and there are some issues that need further research. For instance we are applying the change point test not directly to the stochastic trend component, that is unobservable, but to the smoother predictions. The way the Kalman smoother predictions are built induces a strong dependence between consecutive values and their use to evaluate the test assumptions associated with the trend component might be misleading, so it is necessary to study this subject in depth. Another issue to be studied is the possibility of other change points, since it was only analyzed the at most one change point case. Finally, the confrontation between the obtained results and other different change points methodologies will be the target of future work.

city (country)	change point month/year	mean **bef. c.–p.**	mean **after c.–p.**
Berlin (Germany)	9/1987	0.491	1.426
Bucharest (Romania)	9/2006	0.517	1.732
Copenhagen (Denmark)	9/1987	0.416	1.376
Dublin (Ireland)	10/1988	0.281	0.721
Kiev (Ukraine)	12/1989	0.506	1.479
Lisbon (Portugal)	7/1978	0.357	0.843
Minsk (Belarus)	11/1987	0.659	1.814
Nantes (France)	7/1987	0.359	0.827
Prague (Czech Rep.)	8/1987	0.563	1.457
S. Petersburg (Russia)	12/1988	0.572	1.643
Talin (Estonia)	12/1988	0.507	1.596
Vienna (Austria)	8/1987	0.520	1.297
Vilnius (Lithuania)	11/1987	0.715	2.325
Zagreb (Croatia)	12/1992	0.389	1.274
Zurich (Switzerland)	8/1987	0.578	1.329

Table 2: Summary of change point detection.

ACKNOWLEDGEMENTS

Authors were partially supported through Portuguese funds by the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology (“Fundação para a Ciência e a Tecnologia” [FCT]), within project UID/MAT/04106/2019.

REFERENCES

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

Antoch, J., Hušková, M. and Z. Práškova (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference.* **60**, 291–310.

Costa, M. and Monteiro, M. (2016). Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters. *Journal of Statistical Planning and Inference* **176**: 22 – 32.

Costa, M. and Monteiro, M. (2017). Statistical Modeling of an Air Temperature Time Series of European Cities, In Advances in Environmental Research. vol. 59, ed. Justin A. Daniels, 213 – 236. ISBN: 978-1-53612-701-0. New York: Nova Science Publishers.

Costa M. and Monteiro M. (2018). A periodic mixed linear state-space model to monthly long-term temperature data. *Environmetrics* e2550 https://doi.org/10.1002/env.2550

Jarušková, D. (1997). Some problems with application of change point detection methods to environmental data. Environmetrics, 8:469–483.

Jarušková, D. and Antoch, J. (2018). Changepoint analysis of Klementinum temperature series. Environmetrics. 2019;e2570. https://doi.org/10.1002/env.2570.

Reeves, J. Chen, J., Wang, X.L., Lund, R.B. and Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology.* 46, 900–915.

Worsley, K. J. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association.* **74**: 365–367.