Statistical detection of smooth climate change without prior information

Aurélien RIBES[1], Jean-Marc AZAÏS[2], Serge PLANTON[1]

[1]: *CNRM/GAME (Météo France, CNRS), 42 Av. Gaspard Coriolis, 31057 Toulouse, France;*
[2]: *University of Toulouse; IMT; LSP; UPS, F 31062, TOULOUSE Cedex 9, France.*
aurelien.ribes@cnrm.meteo.fr

Abstract:

We investigate inference in the model (1), in order to estimate the first smooth principal component, and to test its significance. This model and this test are used to detect climate change over the Mediterranean basin.

I. Statistical framework:

We consider the functional model:

$$\psi_{s,t} = m_s + g_s\mu(x_t) + \varepsilon_{s,t}, \qquad s = 1,\dots,S,\ t = 1,\dots,T, \tag{1}$$

where $x_t \in [0,1]$ are known real numbers that can be random or deterministic (eg equally spaced), and where ε is a centered Gaussian noise, the covariance of which satisfies

$$\mathrm{Cov}(\varepsilon_{s,t}, \varepsilon_{s',t'}) = C_{s,s'}\,\delta_{t,t'}, \tag{2}$$

where C is an unknown $S \times S$ real matrix, and where $\delta_{t,t'} = \mathbb{1}_{t=t'}$. g is an unknown S-dimensional real vector, and $\mu(.)$ is an unknown real function. The problem discussed here is to test the hypothesis

$$\mathrm{H}_0 : \text{“}g\,\mu(.) = 0\text{”} \quad \text{vs} \quad \mathrm{H}_1 : \text{“}g\,\mu(.) \neq 0\text{”}. \tag{3}$$

II. Testing strategy:

Our goal is to adapt the likelihood ratio test (LRT) to the model (1) (a LRT cannot be directly implemented). An adaptation can be proposed by using the penalized -2 log-likelihood function

$$pl(m, g, \mu(.), C) = l(m, g, \mu(.), C) + \rho\,\mathrm{pen}(\mu(.)), \tag{4}$$

where ρ is the smoothing parameter. The use of penalization is classical for estimating functions, in order to require some regularity. We here will focus on penalties of the form

$$\mathrm{pen}(\mu(.)) = \int_0^1 \left(\mu^{(q)}(t)\right)^2 dt. \tag{5}$$

Note that $q = 2$ is often chosen in order to penalize roughness. Finally, we will construct a test based on the variable

$$v = \min_{H_1} pl(m, g, \mu(.), C) - \min_{H_0} pl(m, C). \tag{6}$$

III. Computing v:

• Smoothing splines

Under H_1, the computation of v requires to evaluate the maximum penalized likelihood estimates:

$$(\widehat{m}, \widehat{g}, \widehat{\mu}(.), \widehat{C}) = \underset{\substack{(m,g,\mu(.),C)\\ g^T C^{-1} g = 1}}{\mathrm{Argmin}}\ pl(m, g, \mu(.), C). \tag{7}$$

The constrain $g^T C^{-1} g = 1$ is added in order to make the problem well-posed. **?** showed that smoothing splines minimize quantities of this form. For minimizing (7), splines of order $q+1$ are needed, that can be defined as piecewise polynomials of order $q+1$ with knots $(x_t)_{(t=1,\dots,T)}$. Assuming that $\mu(.)$ is a spline, one only requires to estimate a vector (its coordinates in the spline basis).

• First order conditions

The first order conditions of (7) can be derived by performing the spectral analysis of some matrix. Denoting $\widehat{\mu} = (\widehat{\mu}(x_1), \dots, \widehat{\mu}(x_T))$, $\Pi = (I_T - \frac{1}{T}\mathbb{1}_T\mathbb{1}_T^T)$, and $\mathrm{ev}_1(M)$ the first eigenvector of the matrix M, these conditions can be written

$$\widehat{m} = \frac{1}{T}\Psi\mathbb{1}_T, \tag{8}$$

$$\widehat{g} = \lambda.\mathrm{ev}_1\left(\Psi\Pi\Gamma_\rho\Pi\Psi^T\widehat{C}^{-1}\right), \ \text{ with } \lambda \text{ given by } \widehat{g}^T\widehat{C}^{-1}\widehat{g} = 1, \tag{9}$$

$$\widehat{\mu} = \Gamma_\rho\Pi\Psi^T\widehat{C}^{-1}\widehat{g}, \tag{10}$$

$$\widehat{C} = \frac{1}{T}\left(\Psi\Pi - \widehat{g}\widehat{\mu}^T\right)\left(\Psi\Pi - \widehat{g}\widehat{\mu}^T\right)^T, \tag{11}$$

where Γ_ρ is the matrix of a scalar product in the spline space. Although explicit solutions are not known, this system of equation allows to numerically compute $(\widehat{m}, \widehat{g}, \widehat{\mu}(.), \widehat{C})$, using sequences of estimates that converge to the true values.

IV. Null-distribution of v

The null-distribution of v isn't known, even in the case $S = 1$, when the variable v has an explicit form (see **?**, **?** and **?**, **?** for similar issues). Let $\mathcal{D}_{m,C}$ be this null distribution. It can be shown that

$$\mathcal{D}_{m,C} = \mathcal{D}, \quad \forall\, m, C,\ \det C \neq 0, \tag{12}$$

that is to say that the null-distribution of v doesn't depend on the true values of the parameters m and C. This result allows to compute $\mathcal{D}$ easily via bootstrap.

V. Application of the method:

The method is applied to a temperature dataset covering the Mediterranean basin over the 1900-2006 period (HadCRUT3v). The method allows to perform a detection study without using prior information from coupled general circulation models (CGCMs).

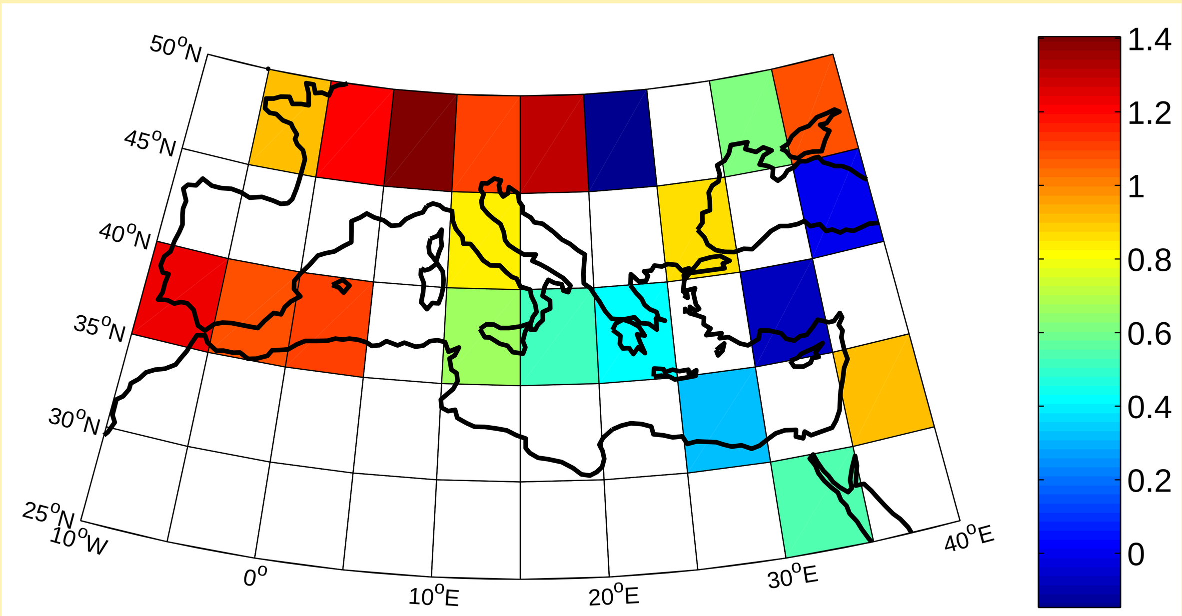

FIG 2: Estimate of the spatial distribution of change g.
The estimated vector $\widehat{g}$ is represented after renormalization (the estimated change in mean temperature between 1900 and 2006 is represented at each point). White grid-points are not taken into account due to a lack of data over the 1900-2006 period. This spatial pattern is not sensitive to the choice of ρ.

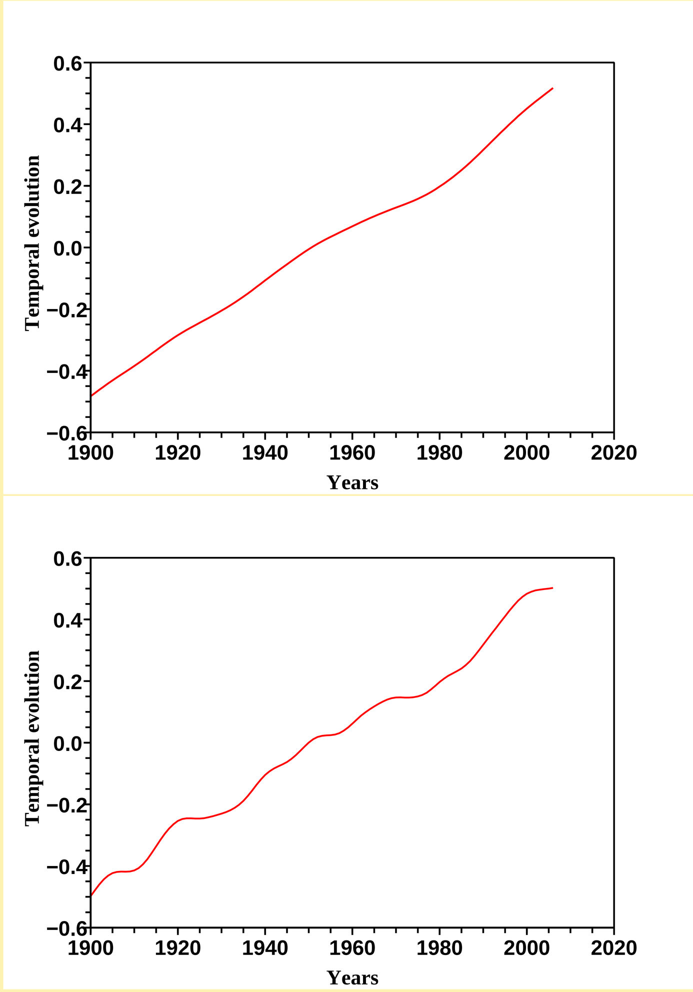

FIG 1: Estimates of the temporal evolution $\mu(.)$.
The temporal evolution function $\mu(.)$ is estimated twice, using two different values for the smoothing parameter ρ. The estimate deduced using the strongest value (top figure), is quasi-linear over time. A second estimate is computed using a smaller ρ, in order to determine whether this result is robust when relaxing the constrain (bottom figure). It can be seen that even with a number of degrees of freedom relatively high, the estimated temporal evolution is quasi-linear. Both signals are strongly significant (p-value $< 10^{-3}$).

Two main results can be highlighted:

- The change is very significant over the studied region.
- The estimation of the temporal evolution from the regional mean and from the whole dataset yield different results. The curves presented here are closer to some CGCMs outputs.

Conclusion:

We showed a way to make statistical inference in order to determine whether the first smooth principal component is significant. The inference can be based on functional penalized likelihood. Estimation and test are both useful for climate change study, in order to describe the change, and evaluate its significance. The application of the method over the Mediterranean area leads to new results.

References:

Crainiceanu C, Ruppert D, Claeskens G, Wand M (2005) Exact likelihood ratio tests for penalized splines. Biomatrika 92(1):91–103
Liu A, Wang Y (2004) Hypothesis testing in smoothing spline models. Journal of statistical computation and simulation 74(8):581–597
Wahba G (1990) Spline models for observational data. Society for Industrial and Applied Mathematics (SIAM)

INSTITUT de MATHEMATIQUES de TOULOUSE

METEO FRANCE
Toujours un temps d'avance