Prediction mimicking Gaussian model

Azzouz Dermoune
Joint work with Khalifa Es-sbaie and our PhD students Mohammed Es-esbaie and Jabrane Moustaaid (Marrakech)

Université de Sciences et Technologies de Lille.
Laboratoire Paul Painlevé-UMR-CNRS 8524, France.

NASCA 2018, Kalamata

Problem

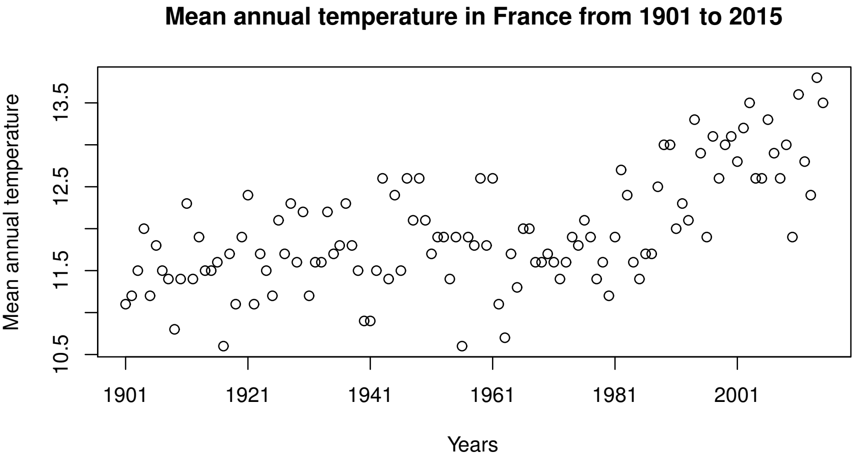
Having the time series

$$\mathbf{s}^{(n)} = (s_1, \ldots, s_n)^\top \in \mathbb{R}^n \quad \text{column vector,}$$

how to predict the next observation s_{n+1} ?

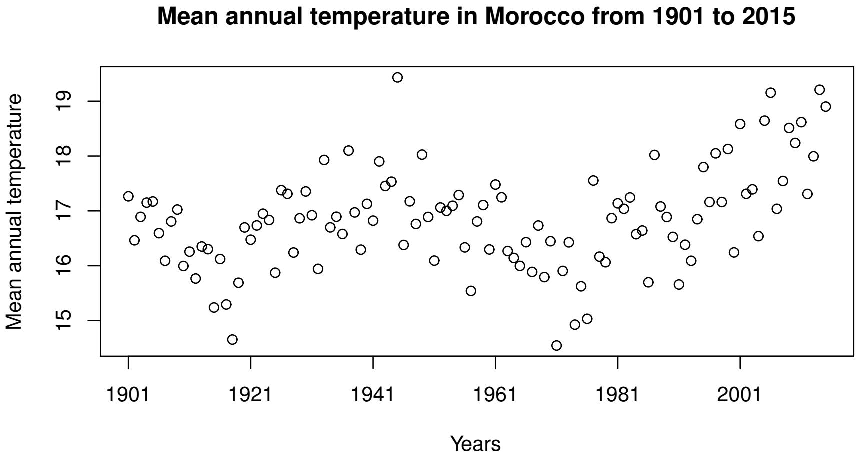
Example

Figure : Mean annual temperatures in France from 1901 to 2015.



Example

Figure : Mean annual temperatures in Morocco from 1901 to 2015.



Predictors

$$\hat{s}_{n+1} = g(n+1 \mid s_1, \ldots, s_n).$$

The maps $g(\cdot \mid s_1, \ldots, s_n) : [1, n+1] \to \mathbb{R}$ are given by a wide variety of methods, e.g.

Interpolation

Take a good function $g(\cdot \mid s_1, \ldots, s_n) : [1, n+1] \to \mathbb{R}$ such that $g(i \mid s_1, \ldots, s_n) = s_i$, with $i = 1, \ldots, n$.

Penalization and smoothing

$$g(\cdot \mid s_1, \dots, s_n) = \arg\min_{h \in \mathcal{F}} \{ \sum_{i=1}^{n} |h(i) - s_i|^2, \\ C(h) \le \delta \} \\ = \arg\min_{h \in \mathcal{F}} \{ \sum_{i=1}^{n} |h(i) - s_i|^2 + \lambda C(h) \}.$$

Probabilistic method

$f(s_1, \ldots, s_{n+1}, \theta) \geq 0$ is a PDF on $\mathbb{R}^{n+1}$ if

$$\int_{\mathbb{R}^{n+1}} f(s_1, \ldots, s_{n+1}, \theta) ds_1 \ldots ds_{n+1} = 1.$$

Probabilistic method

Each PDF $f(s_1, \ldots, s_{n+1}, \theta)$ on $\mathbb{R}^{n+1}$ produces two predictors :

$$s(n+1, CEXP) =: \frac{\int_{\mathbb{R}} s_{n+1} f(s_1, \ldots, s_{n+1}, \theta) ds_{n+1}}{\int_{\mathbb{R}} f(s_1, \ldots, s_{n+1}, \theta) ds_{n+1}}$$

$=$ conditional expectation.

$$s(n+1, MAP) =: \arg\max\{f(s_1, \ldots, s_{n+1}, \theta) : \quad s_{n+1} \in \mathbb{R}\}$$

$=$ maximum a posteriori.

Theoretical performance

Using the PDF f, we generate a sequence $(s_1^1, \ldots, s_{n+1}^1)$, $\ldots$, $(s_1^N, \ldots, s_{n+1}^N)$ with N very large such that

$$\frac{1}{N}\sum_{i=1}^{N} |s_{n+1}^i - s^i(n+1, CEXP)|^2$$

$$\approx \min\{\frac{1}{N}\sum_{i=1}^{N} |s_{n+1}^i - g(s_1^i, \ldots, s_n^i)|^2 : \quad g\}$$

$$\approx \int_{\mathbb{R}^{n+1}} (s_{n+1} - s(n+1, CEXP))^2 f(s_1, \ldots, s_{n+1}, \theta) ds_1 \ldots ds_{n+1},$$

Theoretical mean squared prediction error (TMSPE).

Gaussian model

If

$$f(s_1, \ldots, s_{n+1}, \boldsymbol{P}^{(n+1)}) = \sqrt{\frac{\det(\boldsymbol{P}^{(n+1)})}{(2\pi)^{n+1}}} \exp\{-\frac{(\boldsymbol{s}^{(n+1)})^\top \boldsymbol{P}^{(n+1)} \boldsymbol{s}^{(n+1)}}{2}\},$$

with $\boldsymbol{P}^{(n+1)} = [p_{ij} : \quad i,j = 1, \ldots, n+1]$ symmetric and positive definite matrix (precision matrix).

$$\boldsymbol{C}^{(n+1)} = [c_{ij} : \quad i,j = 1, \ldots, n+1] = \{\boldsymbol{P}^{(n+1)}\}^{-1}$$

$:=$ covariance matrix.

$$\boldsymbol{C}^{(n)} = [c_{ij} : \quad i,j = 1, \ldots, n].$$

$$\boldsymbol{c}_{i*} =: (c_{i1}, \ldots, c_{in}), \quad i = 1, \ldots, n+1.$$

Gaussian model

We have

$$s(n+1, CEXP) = s(n+1, MAP) = \boldsymbol{c}_{n+1*}\{\boldsymbol{C}^{(n)}\}^{-1}\boldsymbol{s}^{(n)}$$
$$= \arg\min\{(\boldsymbol{s}^{(n+1)})^{\top}\boldsymbol{P}^{(n+1)}\boldsymbol{s}^{(n+1)} : \quad s_{n+1} \in \mathbb{R}\}.$$

$$\frac{1}{N}\sum_{i=1}^{N}|s^i_{n+1} - s^i(n+1, CEXP)|^2 \approx c_{n+1n+1} - \boldsymbol{c}_{n+1*}\{\boldsymbol{C}^{(n)}\}^{-1}\boldsymbol{c}^\top_{n+1*}$$

$$= \frac{\det(\boldsymbol{C}^{(n+1)})}{\det(\boldsymbol{C}^{(n)})}.$$

Popular Gaussian model and cubic splines

$$\int_1^{n+1} |s''(t)|^2 dt = (\boldsymbol{s}^{(n+1)})^\top \boldsymbol{Q}^{(n+1)} \boldsymbol{s}^{(n+1)},$$

$$\boldsymbol{C}^{(n+1)} = \sigma_s^2 (\boldsymbol{Q}^{(n+1)})^{ginv} + \sigma_w^2 \boldsymbol{I}_{n+1},$$

Parameters : σ_s^2, σ_w^2 and $(\boldsymbol{Q}^{(n+1)})^{ginv}$.
Dermoune, Preda, JMVA 2016.
Dermoune, Rahmania, Wei, JMVA 2013.

Popular Gaussian model and Gaussian kernel

$$k(i,j) = \exp(-\frac{|i-j|^2}{\sigma^2}),$$
$$\boldsymbol{K}^{(n+1)} = [k(i,j) : \quad i,j = 1,\dots,n+1],$$
$$\boldsymbol{C}^{(n+1)} = \sigma_s^2 \boldsymbol{K}^{(n+1)} + \sigma_w^2 \boldsymbol{I}_{n+1},$$

three parameters σ^2, σ_s^2 and σ_w^2.

Gaussian model and interpolation

Let $\boldsymbol{c}_j$ be the j-th column of the covariance matrix $\boldsymbol{C}^{(n+1)}$, and let us consider the set of functions $span(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$:

$$g_\beta = \sum_{j=1}^{n} \beta_j \boldsymbol{c}_j : i \in [1, n+1] \to \sum_{j=1}^{n} \beta_j c_{ij} = g_\beta(i).$$

The unique element of $span(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ such that

$$g_\beta(i) = s_i, \quad i = 1, \ldots, n,$$

is given by

$$(\beta_1^*, \ldots, \beta_n^*)^\top = \{\boldsymbol{C}^{(n)}\}^{-1} \boldsymbol{s}^{(n)},$$

and then

$$g_{\beta^*}(n+1) = s(n+1, CEXP) \quad \text{Gaussian predictor.}$$

Model mimicking Gaussian model :Dermoune, Es-sbaye, Es-sbaye, Moustaid, ArXiv 2018

Let $\boldsymbol{B}^{(n+1)} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n+1})$ be a basis of $\mathbb{R}^{n+1}$ such that the sub-matrix $\boldsymbol{B}^{(n)} =: [b_{ij} : \quad i, j = 1, \ldots, n]$ is invertible with Θ its inverse and θ_j denotes its j-th row. Let us consider the set of functions $span(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$:

$$g_\beta = \sum_{j=1}^{n} \beta_j \boldsymbol{b}_j : i \in [1, n+1] \to \sum_{j=1}^{n} \beta_j b_{ij} = g_\beta(i).$$

Model mimicking Gaussian model

The unique element of $span(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ such that

$$g_\beta(i) = s_i, \quad i = 1, \ldots, n,$$

is given by

$$(\beta_1^*, \ldots, \beta_n^*)^\top = \Theta \boldsymbol{s}^{(n)},$$

and then

$$g_{\beta^*}(n+1) = \sum_{j=1}^{n} \theta_j \boldsymbol{s}^{(n)} b_{n+1j} = s(n+1, CEXP)$$

when $\boldsymbol{B}^{(n+1)}$ is a covariance matrix.

We showed for the annual mean temperature that

$$g_{\beta^*}(n+1) = \sum_{j=1}^{n} \theta_j^{(n)} \boldsymbol{s}^{(n)}\, b_{n+1j}^{(n+1)}$$

is a bad predictor when $\boldsymbol{B}^{(n+1)}$ is not a covariance matrix. However with the normalization

$$\frac{\sum_{j=1}^{n} \theta_j^{(n)} \boldsymbol{s}^{(n)}\, b_{n+1j}^{(n+1)}}{\sum_{j=1}^{n} \theta_j^{(n)} \mathbf{1}^{(n)}\, b_{n+1j}^{(n+1)}} =: \sum_{i=1}^{n} w_i^{(n)} s_i$$
$$= \boldsymbol{w}^{(n)} \boldsymbol{s}^{(n)}$$

the new predictor becomes good.

Given a $n+1 \times n+1$ matrix $\boldsymbol{B}^{(n+1)} = [b_{ij} : \quad i,j = 1,\ldots,n+1]$ such that the sub-matrices $\boldsymbol{B}^{(k)} = [b_{ij} : \quad i,j = 1,\ldots,k]$ are invertible, with $k = 2,\ldots,n$. The inverse $\{\boldsymbol{B}^{(k)}\}^{-1} =: \Theta^{(k)}$. We predict s_{k+1} by

$$\hat{s}_{k+1} = \frac{\sum_{j=1}^{k} b_{k+1j}\theta_j^{(k)}\boldsymbol{s}^{(k)}}{\sum_{j=1}^{k} b_{k+1j}\theta_j^{(k)}\boldsymbol{1}^{(k)}}$$
$$=: \boldsymbol{w}^{(k)}\boldsymbol{s}^{(k)},$$

with $k = 2,\ldots,n$.

For each $k = 2, \ldots, n$, we give a $k+1 \times k+1$ matrix $\boldsymbol{B}^{(k+1)} = [b_{ij}^{(k+1)} : \quad i,j = 1, \ldots, k+1]$ such that the sub-matrix $\boldsymbol{B}^{(k)} = [b_{ij}^{(k+1)} : \quad i,j = 1, \ldots, k]$ is invertible. Its inverse $\{\boldsymbol{B}^{(k)}\}^{-1} =: \Theta^{(k)}$. We predict s_{k+1} by

$$\hat{s}_{k+1} = \frac{\sum_{j=1}^{k} b_{k+1j}^{(k+1)} \theta_j^{(k)} \boldsymbol{s}^{(k)}}{\sum_{j=1}^{k} b_{k+1j}^{(k+1)} \theta_j^{(k)} \boldsymbol{1}^{(k)}} =: \boldsymbol{w}^{(k)} \boldsymbol{s}^{(k)},$$

with $k = 2, \ldots, n$.

Linear model for prediction

A model $\mathcal{M}$ for prediction is any sequence of weights $(\boldsymbol{w}^{(k)} : \quad k = 2, \ldots, n)$, with

$$\boldsymbol{w}^{(k)} \in \mathbb{R}^k, \quad \hat{s}_{k+1} = \boldsymbol{w}^{(k)} \boldsymbol{s}^{(k)} =: \sum_{i=1}^{k} w_i^{(k)} s_i.$$

Real performance

$$MSPE =: \frac{\sum_{j=1}^{n-1} |s_{j+1} - \boldsymbol{w}^{(j)} \boldsymbol{s}^{(j)}|^2}{n-1},$$

$$MAXPE =: \max\{|s_{j+1} - \boldsymbol{w}^{(j)} \boldsymbol{s}^{(j)}| : \quad j = 1, \ldots, n-1\}.$$

Comparison of two models

Let $\{\mathcal{M}(m) : m = 1, 2\}$ be two models.
1) We say that $\mathcal{M}(1)$ is better than $\mathcal{M}(2)$ w.r.t. the MSPE criterion if

$$MSPE(\mathcal{M}(1)) < MSPE(\mathcal{M}(2)).$$

2) We say that $\mathcal{M}(1)$ is better than $\mathcal{M}(2)$ w.r.t. the MAXPE criterion if

$$MAXPE(\mathcal{M}(1)) < MAXPE(\mathcal{M}(2)).$$

3) We say that $\mathcal{M}(1)$ is statistically better than $\mathcal{M}(2)$ if

$$\frac{\sum_{k=1}^{n-1} \mathbf{1}_{[|s_{k+1} - \boldsymbol{w}^{(k)}(1)\boldsymbol{s}^{(k)}| < |s_{k+1} - \boldsymbol{w}^{(k)}(2)\boldsymbol{s}^{(k)}|]}}{n-1} > 1/2.$$

Models using cubic splines

$$\mathbb{R}^k = \mathcal{S}_3(1, \dots, k) \quad \text{the set of natural cubic splines,}$$

$$\int_1^k |s(t)|^2 dt = \{\boldsymbol{s}^{(k)}\}^\top \boldsymbol{M}^{(k)} \boldsymbol{s}^{(k)} \quad \text{non symmetric matrix}$$

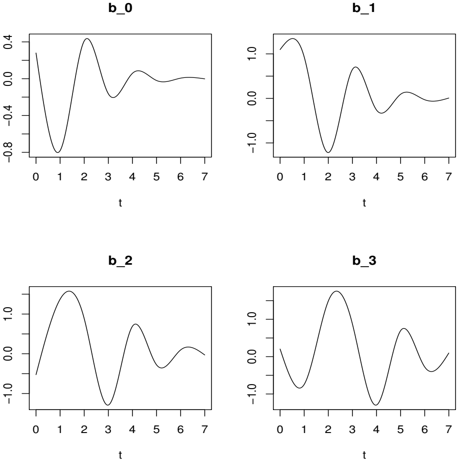
$$= \{\boldsymbol{s}^{(k)}\}^\top \boldsymbol{S}^{(k)} \boldsymbol{s}^{(k)} \quad \text{symmetric matrix.}$$

Bases

$$B^{(k)} = M^{(k)},\ \{M^{(k)}\}^{-1},\ S^{(k)}, \{S^{(k)}\}^{-1},$$

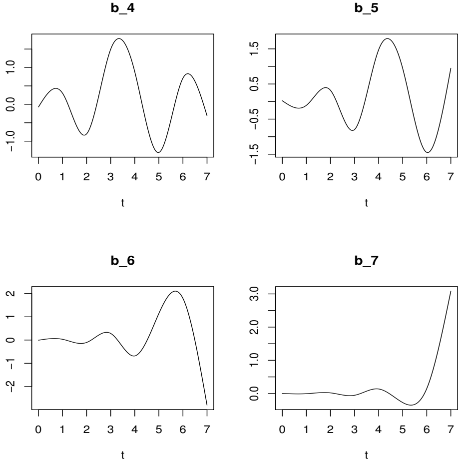
with $k = 1, \ldots, n+1$.

$B^{(k)} = M^{(k)}$, $k = 7$.



b_0
0.4
0.0
−0.4
−0.8
0 1 2 3 4 5 6 7
t
b_1
1.0
0.0
−1.0
0 1 2 3 4 5 6 7
t
b_2
1.0
0.0
−1.0
0 1 2 3 4 5 6 7
t
b_3
1.0
0.0
−1.0
0 1 2 3 4 5 6 7
t

Numerical results : $B^{(k)} = M^{(k)}$, $k = 7$



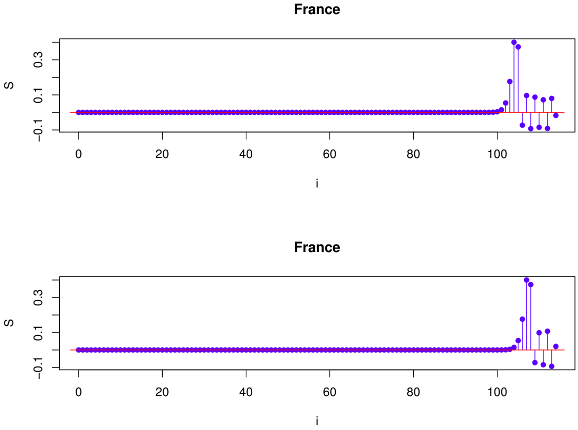
From each sequence of bases we constructed

$$8 + 4n + n(n+2) \approx 10^4 \quad \text{models},$$

with $n = 115$.

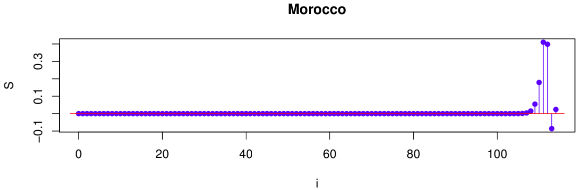
Our optimal model among $8 + 4n + n(n+2) \approx 10^4$ models

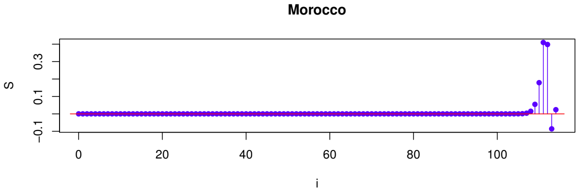
Figure : The optimal conservative row w.r.t. MAXPE, and statistical criterion in the case of France with $n = 115$.



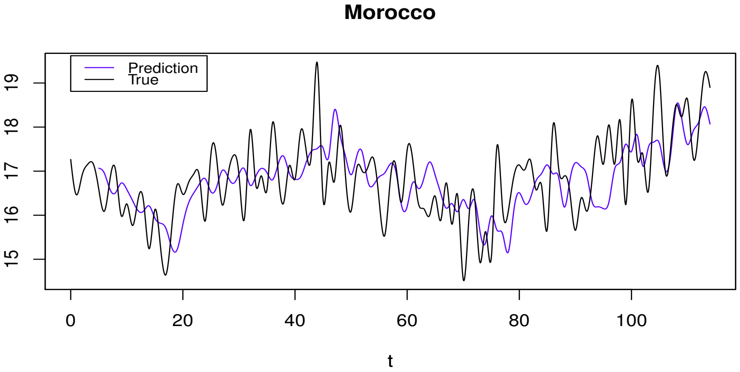
Our optimal model among $8 + 4n + n(n+2) \approx 10^4$ models

Figure : The optimal conservative row w.r.t. MAXPE, and statistical criterion in the case of Morocco with $n = 115$.

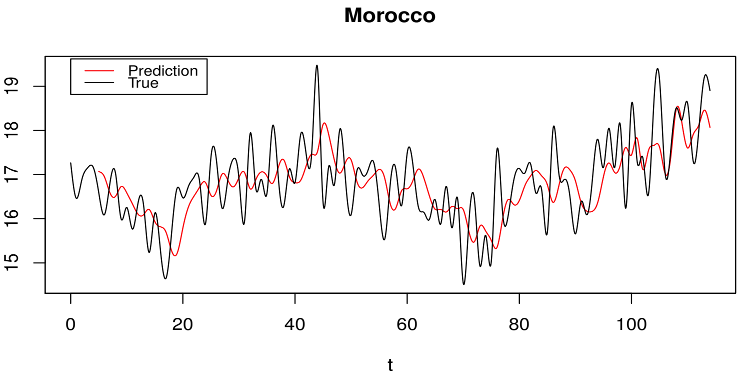




Splines of the true temperature and its optimal predictors : Morocco



Splines of the true temperature and its optimal predictors : Morocco



Comparison with AR(1) model

Table : Comparison of AR(1) model and our optimal coherent sequences w.r.t. MAXPE.

MAXPE criterion		
Country	France	Morocco
Our optimal model	1.220770	1.917094
AR(1) model	1.433288	2.490254

Comparison with AR(1) model

Statistical criterion		
Country	France	Morocco
The best	AR(1) model	Our optimal model

Conclusion

- We proposed a new method and models for prediction using only bases of the Euclidean spaces.
- We proposed comparison criteria for selecting the best model.
- We illustrated our method for the prediction of annual mean temperature of France and Morocco.

Thank you and many thanks for this perfect organization in this wonderful site.