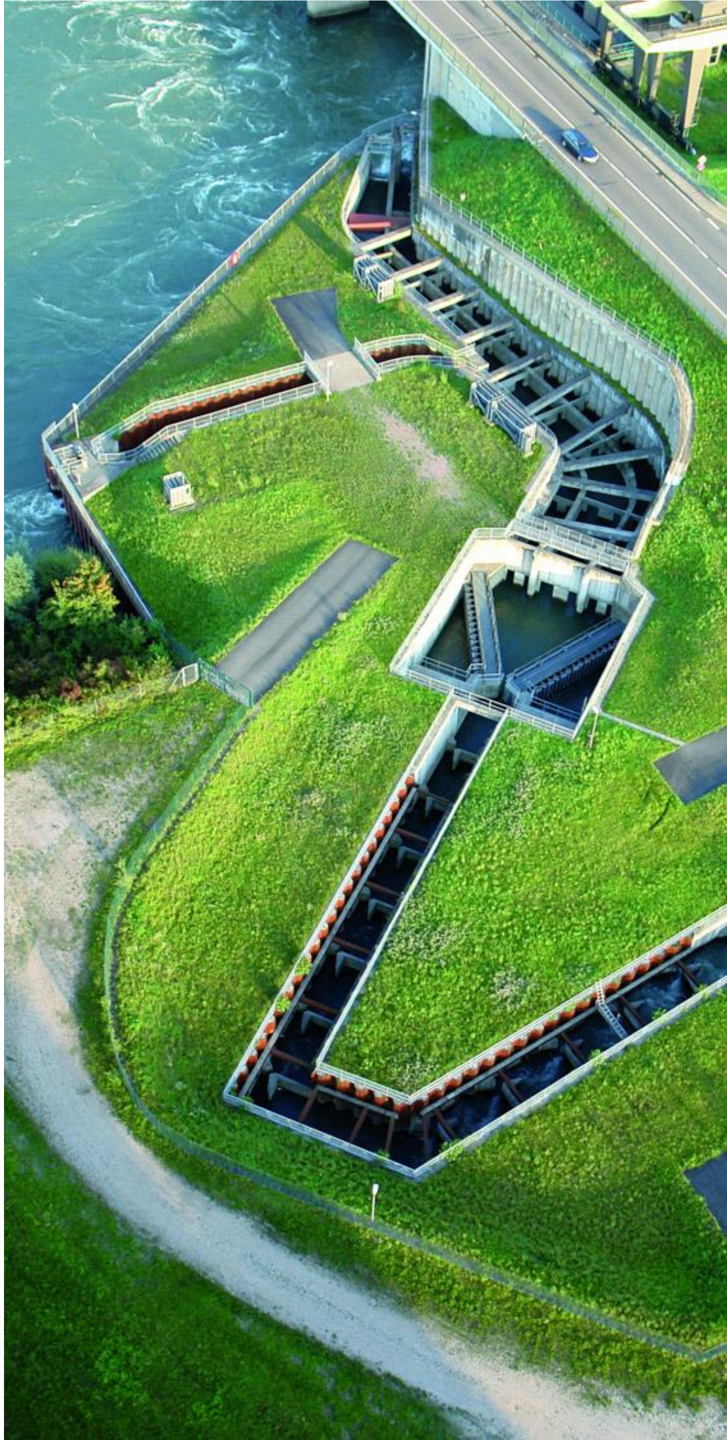
SIMULATION, RARE EVENTS AND TEMPERATURES

16/12/2013

Thi Thu huong HOANG (EDF R&D),
Sylvie PAREY(EDF R&D),
Didier DACUNHA-CASTELLE (Université Paris Sud)



OUTLINE

- Context and objectives
- Some characteristics of temperatures
- Preprocessing
- Test of cyclo-stationarity
- Simulation model for temperatures taking into account the extremes
- Validity of the model
- Some applications
- Conclusion and perspectives

CONTEXT & OBJECTIVES

Context

- EDF is interested in the impact of climate change on the energy sector
- The knowledge on those climatic evolutions and their consequences will help outline the offer/demand balance of the 21st century and envisage a decision-making procedure related to adaptation strategies.
- We concentrate on air temperature, the major climate variable influencing energy:
  - Reduce or increase the consumption of energy
  - the hot or cold waves can disturb electricity production and distribution
  - Extreme temperatures are used for design and verifications.

Objectives

- Build a simulation model (because climate models cannot easily provide results on local extremes)
  - for (maximum or minimum) daily air temperature for a fixed location
  - with good qualities for the center and the tails of distribution

- Calculate the return levels

- Use the model in combination with climate models for extremes

Some characteristics of temperature

"Theory is when you know everything and nothing works. The practice is when everything works and nobody knows why. Here we have compiled theory and practice: Nothing works ... and nobody knows why! " ***(Albert Einstein)***

- **Non stationary, non linear**

- Two **periodicities** (in mean and in variance) perhaps non constant for long periods

After the preprocessing, seasonality is still found in the correlation, sometimes in skewness or kurtosis -> cyclo-stationary (test ?)

- **Boundedness** : only a very accurate application of **extremes theory** allows to prove that every model have to take into account this feature.

It is difficult for numerical (deterministic) models to take into account extremes and even variability

- **Continuous-time process versus discrete measurements:**

The temperature is a continuous-time process but observed at discrete moments

How to apply the properties of continuous-time model on the discrete observations?

Preprocessing treatment

- This stage removes the trends in mean and in variance and the (additive and multiplicative) seasonalities to obtain an as stationary as possible reduced series

- The processing treatment uses both the nonparametric and parametric approaches

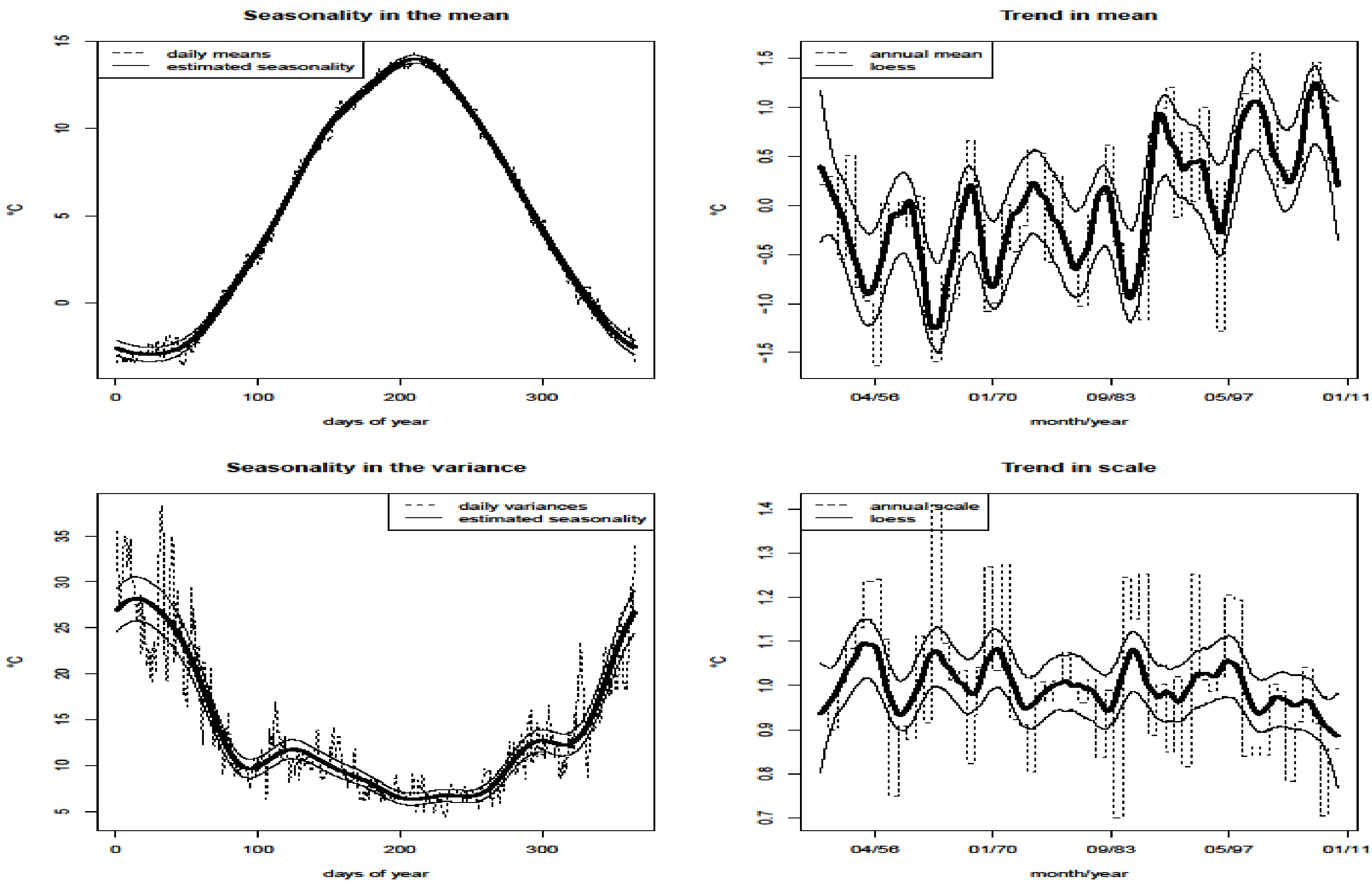
$$X(t) = m(t) + u(t) + s(t)v(t)Z(t)$$

  - $m(t), s(t)$: mean, scale function, $u(t), v^2(t)$: seasonalites in mean and in variance

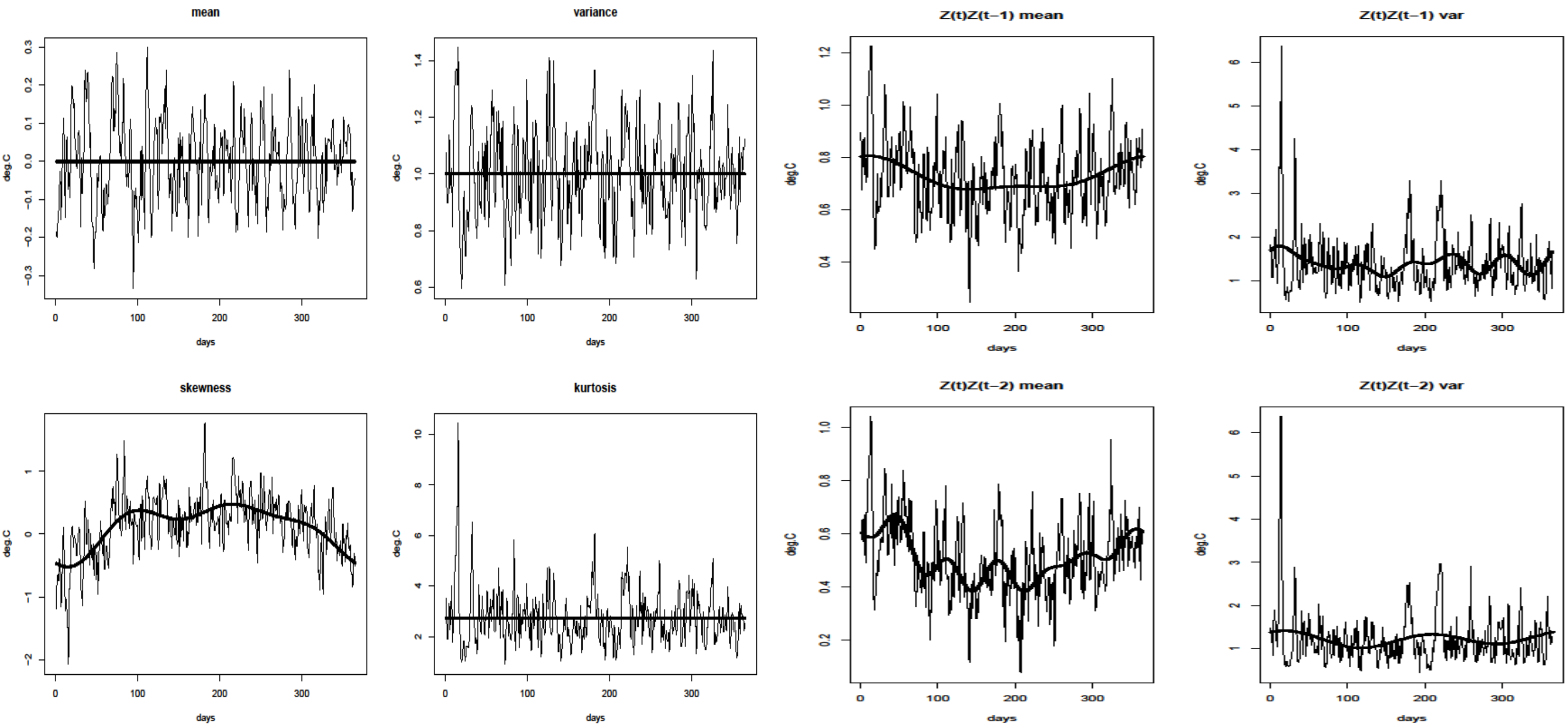
For the choice of the bandwidth for the trends, the modified partitionned cross validation (Hoang, 2010) is used

For the choice of number of Fourier terms for the seasonalities, the Akaike criterion is used

Estimated trends and seasonalities



Cyclo-stationarity of Z



Principal of our test of trend (or of stationarity)

- The considered model: ***X(t)= θ(t)+ ε(t)***, the distribution of ε is known or unknown

- **Hypothesis of test**: *θ is constant / θ is not constant*

- Let:
  - $\hat{c}_n$ the constant estimator of θ (by m.l.e if the law of ε is known, by least squares if not)
  - $\hat{\theta}_n$ the nonparametric estimator of θ (by splines if the law of ε is known, by loess if not)

- **Idea**: compare these two estimators by the L^2 distance :

$$\Delta = \left\| \hat{\theta}_n - \hat{c}_n \right\|$$

- **Asymptotics of test**: **Δ is signigficantly positive if H0 is false** (see Hoang 2010)

- **In practice**: - build a test on Δ (build the **empirical distribution of Δ under H0** hypothesis by using the simulations if the distribution of θ is known or the permutation (or the boostrap) of yearly block samples if not)

  -**tests of trend** on the **moments**: mean, variance, skewness, kurtosis, correlation

  and on the **extremes**

Discrete-time process

- First order Euler scheme of the discrete diffusion:

$$Z_n = b(Z_{n-1}) + a(Z_{n-1})\varepsilon_n, \quad \varepsilon_n \propto N(0,1)$$

with a nul outside of (r_1, r_2)

- To obtain a model of bounded support:

$$Z'_n = b(Z'_{n-1}) + a(Z'_{n-1})\varepsilon'_n(Z'_{n-1}),$$

with a nul outside of (r_1, r_2) and :

$$f'(y|x) = \frac{1}{R(x)\sqrt{2\pi}} e^{-y^2/2} 1_{\psi_1(x)<y<\psi_2(x)}$$

$$\psi_2(x) = \frac{r_2 - b(x)}{a(x)}, \psi_1(x) = \frac{b(x) - r_1}{a(x)}, \Phi(x) \text{ the Gaussian repartition}$$

$$R(x) = \Phi(\psi_2(x)) - \Phi(\psi_1(x))$$

Theorem (Dacunha Castelle et al., 2013)

Z' is geometrically ergodic with support (r_1, r_2)

The SFHAR(seasonal functional heteroscedastic autoregressive) model

- Extension: SFHAR model

$$Z(t)=\left[\theta_{0,k}+\sum_{j=1}^{p_1}\left(\theta_{1,k}^{j}\cos\frac{2j\pi t}{365}+\theta_{2,k}^{j}\sin\frac{2j\pi t}{365}\right)\right]Z(t-1)+a(t,Z_{t-1})\varepsilon'_t$$

- $a^2(t,Z_{t-1})$
  - Seasonal, depend on the state
  - Zero out of the boundaries
  - positive
  - constraints *C* on the first derivative from the continuous-time diffusion process (see thesis of Hoang, 2010):

$$\left(a^2\right)'(r_1)=\frac{2b(r_1,t)}{1-1/\xi_1}\quad et\quad \left(a^2\right)'(r_2)=\frac{2b(r_2,t)}{1-1/\xi_2}$$

  - Form of *a*:

$$\begin{cases}\hat{a}^2(t,Z_{t-1})=(\hat{r}_2-t)(t-\hat{r}_1)\sum_{k=0}^{5}\sum_{j=1}^{p_2}\left(\alpha_{1,k}^{j}\cos\frac{2j\pi t}{365}+\alpha_{2,k}^{j}\cos\frac{2j\pi t}{365}\right)Z_{t-1}^{k}\\ C(\hat{r}_1,t),C(\hat{r}_2,t)\\ \hat{a}^2(t)>0\ \forall t\end{cases}$$

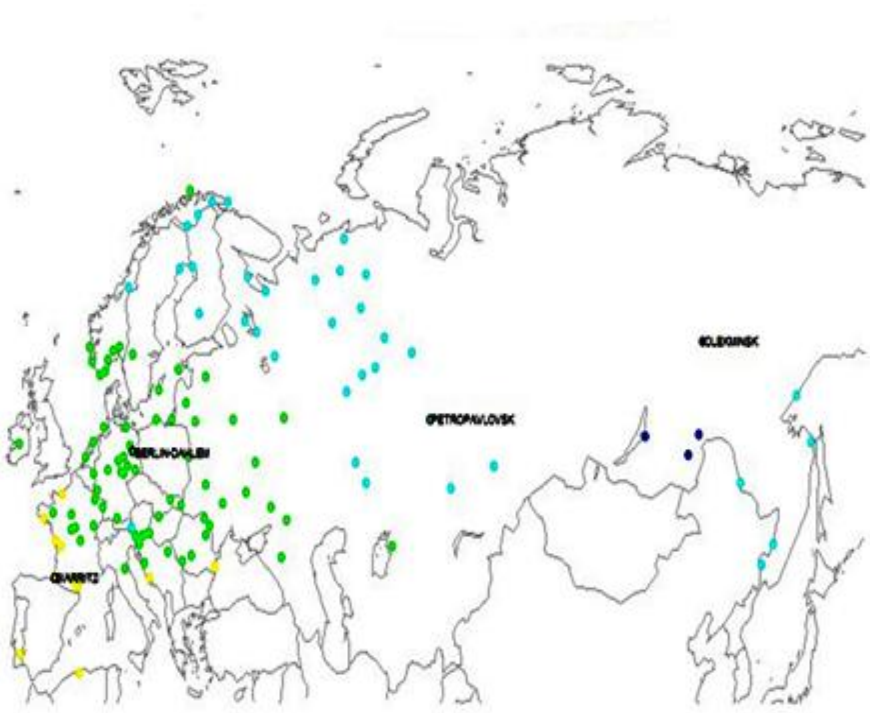
Validity of the model

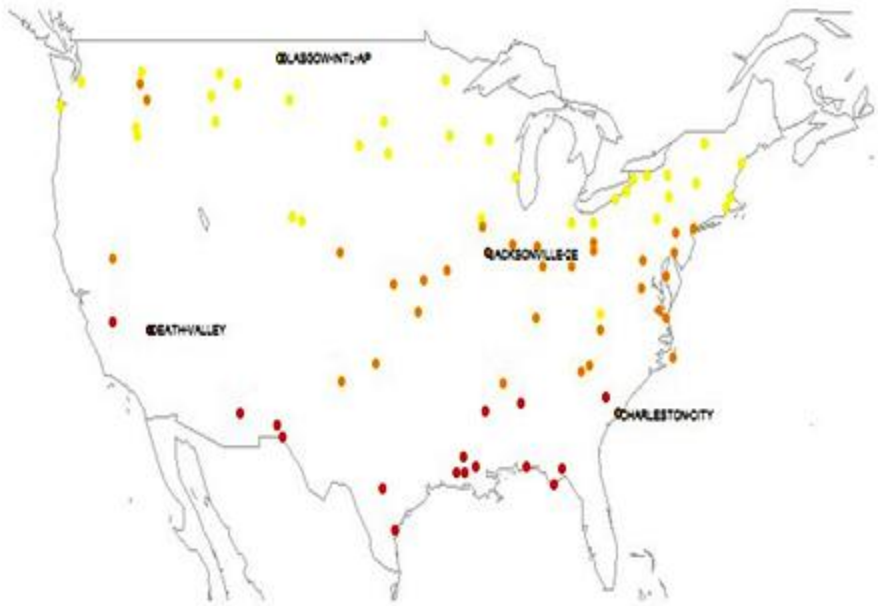
- **Residuals:**
  - **Whiteness of the residuals and the squared residuals**
  - **Tests of normality of the residuals**

- **Comparison of the observations and the simulations:**
  - **Basic statistics: mean, variance, skewness, kurtosis**
  - **Marginal distribution: density function, test of homogeneity**
  - **Quantiles**
  - **Extreme parameters for Z**
  - **Return levels**
  - **Cold waves (<=Q2%) and heat waves (>=Q98%)**

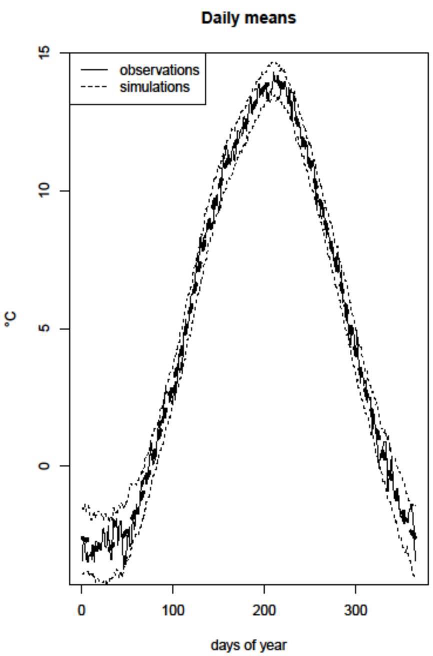
The data

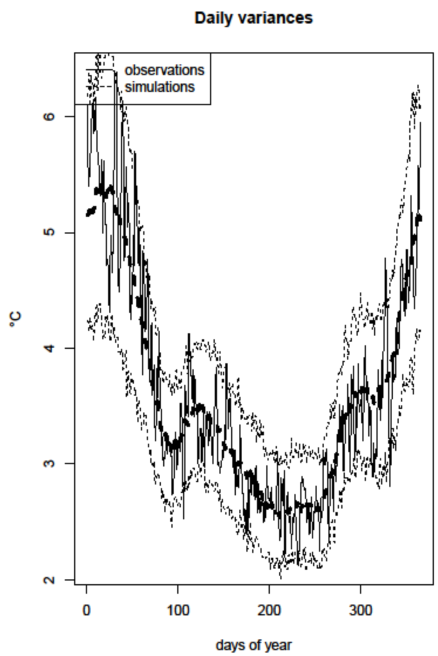
- daily maximum or minimum temperatures
- observed period: 1950-2009
- Europe
- United States



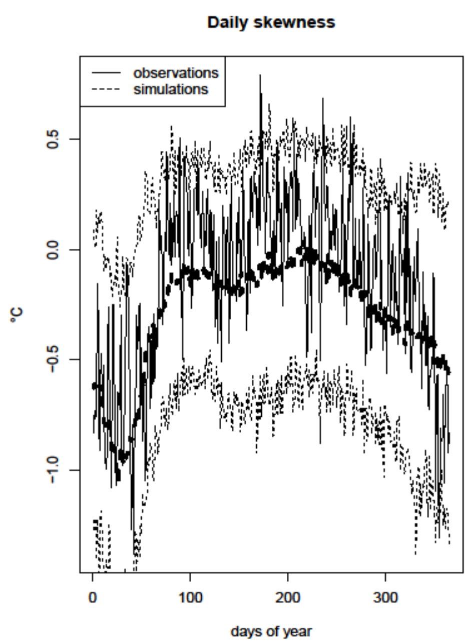


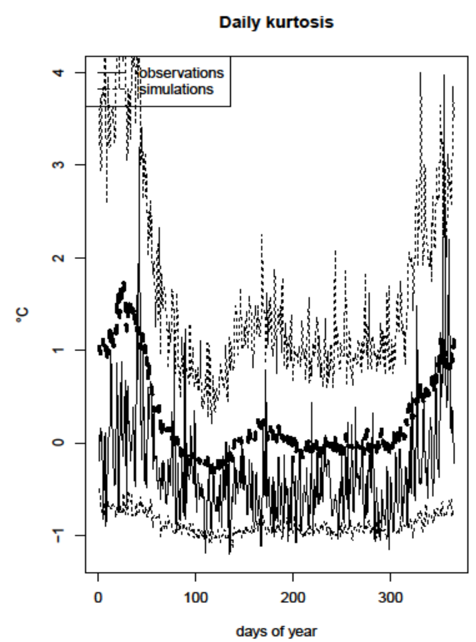
Daily characteristics





- KS tests for daily distributions -> less than 5% of reject
- very good results for daily means and variances
- Less good results for daily skewness and kurtosis, however the observed characteristics are mainly found in the confidence intervals of simulations

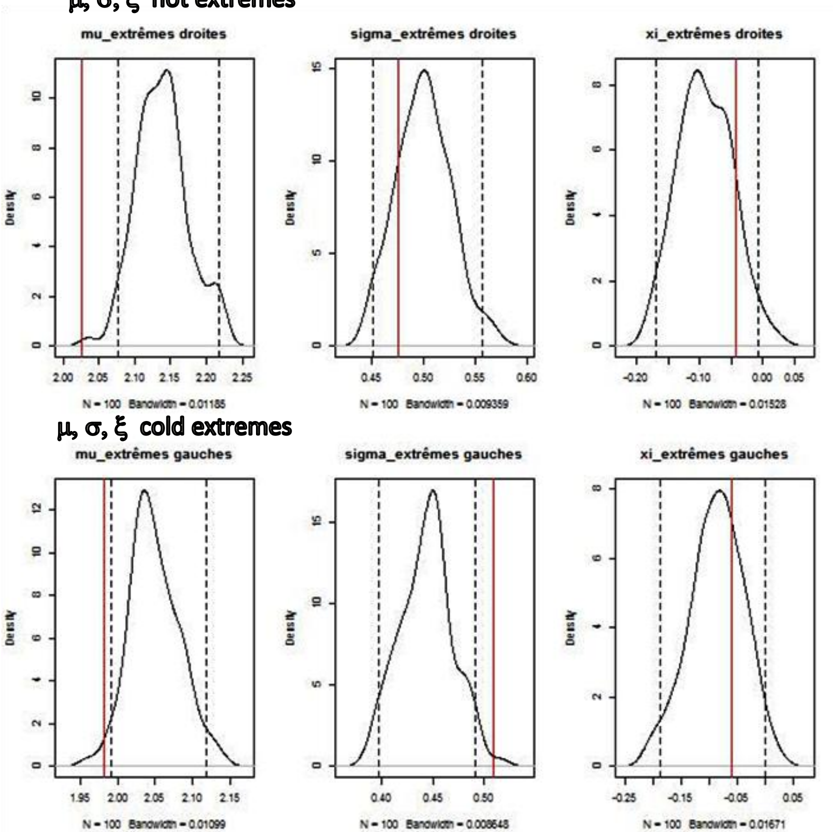




Results: extremes of the reduced series

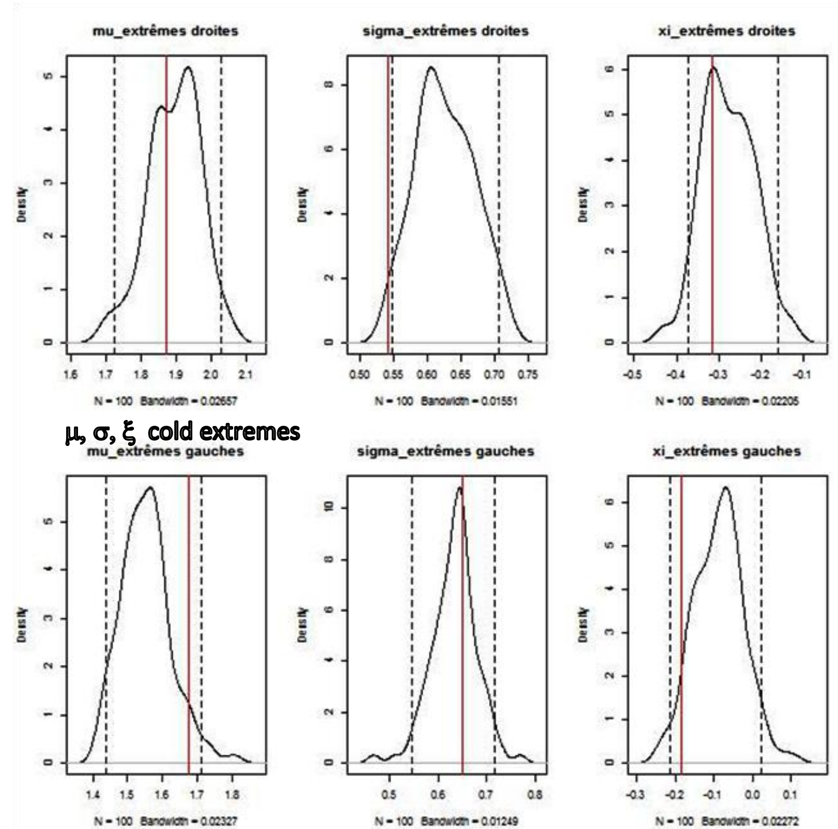
TN Death-Valley

μ, σ, ξ hot extremes



TX Berlin

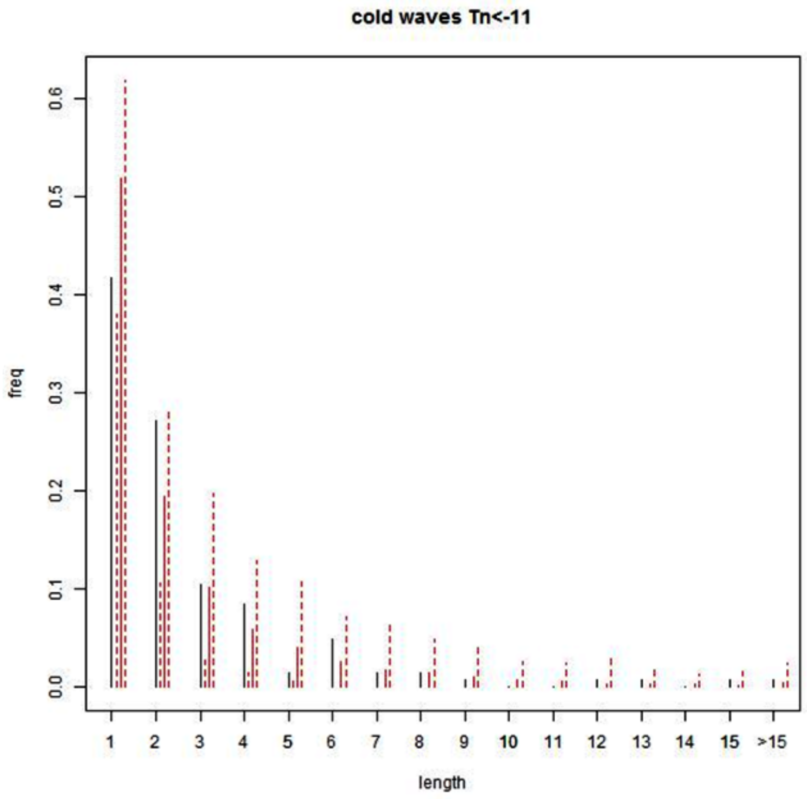
μ, σ, ξ hot extremes



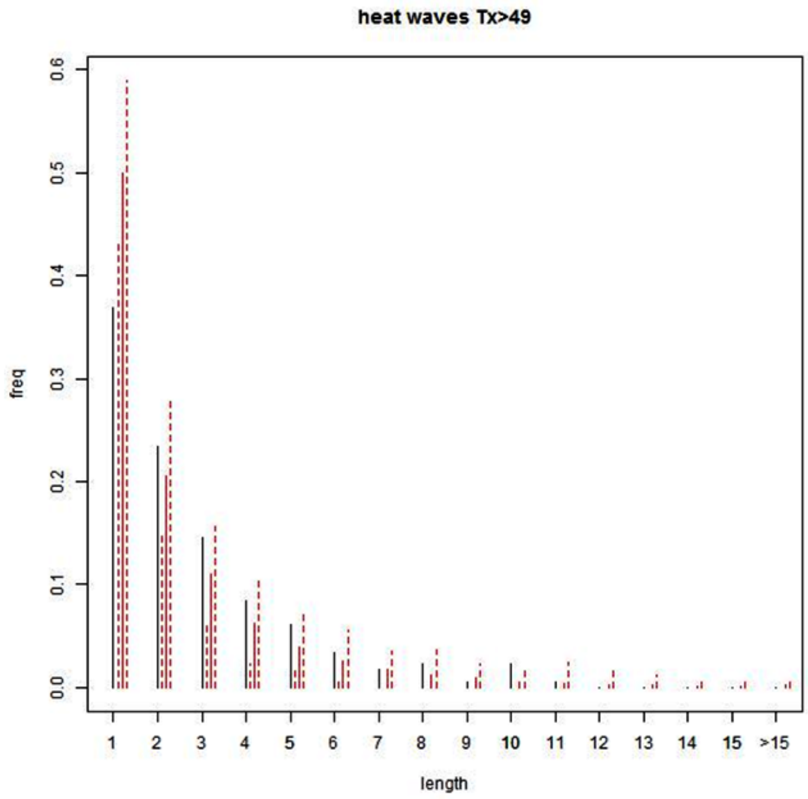
-The results are better for the shape parameter ξ (often found in the confidence interval) → important because this parameter determines the domain of max attration of the data.
-The results are less good for the location and scale parameters
- The estimation of extreme parameters is very sensible to the block size

Cold and heat waves

Cold waves in Berlin



Heat waves in Death-Valley



- The observed frequencies of the waves are always found in the confidence intervals built from the simulations
- The model tends to overestimate the proportion of 1-day cold excursions compared to the observations

→ the correlation in the extremes are maybe not completely taken into account

SOME APPLICATIONS

Freezing index

- The freezing index corresponds to the sum of daily temperatures below 0°C

- Question : 10 000-year return levels of the freezing index

- Example:
  - Agen
  - Nevers

Return levels by quantiles or GEV

- Agen: observed maximum index 113,5
  - Quantiles (1-1/(nb*A)) of simulations

NR 100	NR 1 000	NR 10 000	NR 100 000
54,0 [52,8 ; 54,5]	89,5 [86,8 ; 94,2]	130,5 [115,5 ; 146,6]	168,4 [131,7 ; 173,3]

  - Adjustment of GEV (Gumbel) on 1500 simulations

NR 1 000	NR 10 000	NR 100 000
87,3 [85,3 ; 89,2]	120,8 [117,5 ; 124,0]	153,9 [149,4 ; 158,5]

Return levels by quantiles ou GEV

- Nevers: observed maximum index 191,2
  - Quantiles (1-1/(nb*A)) of simulations

NR 100	NR 1 000	NR 10 000	NR 100 000
120,0 [118,8 ; 121,4]	180,8 [174,2 ; 184,5]	233,5 [220,7 ; 243,7]	273,3 [234,3 ; 281,1]

  - Adjustment of GEV (Gumbel) on 1500 simulations

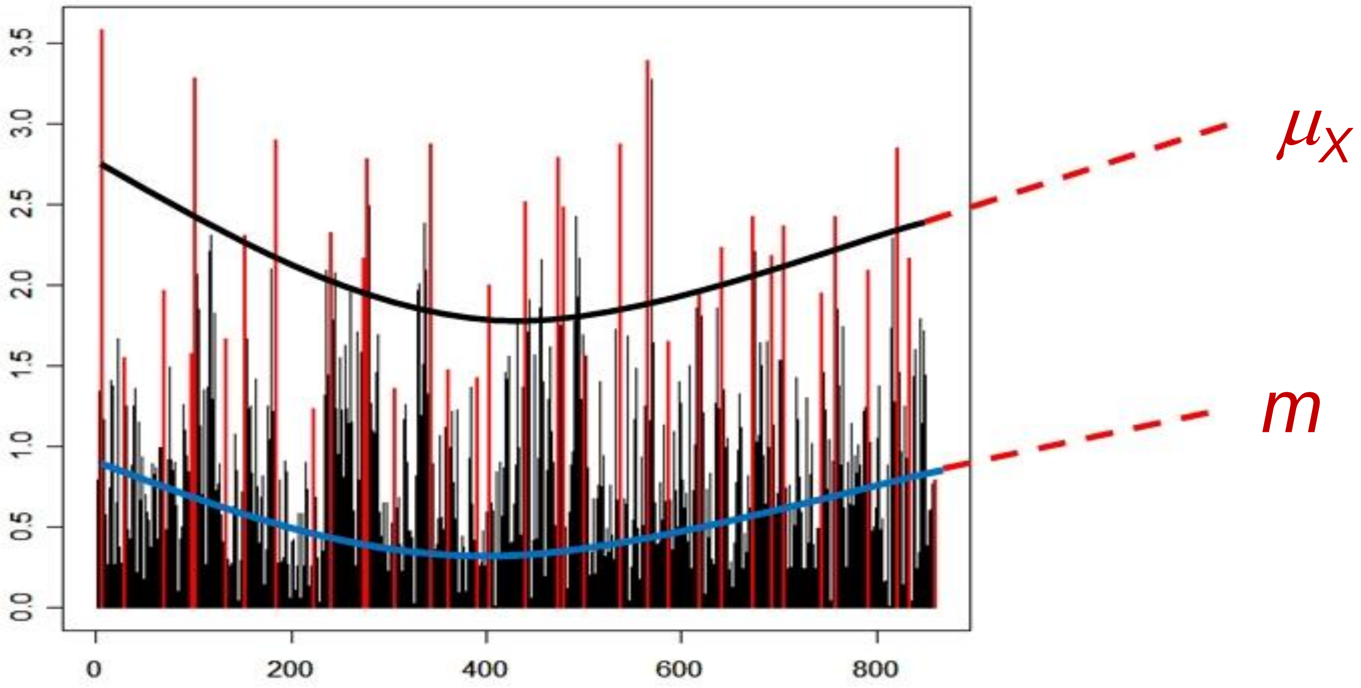
NR 1 000	NR 10 000	NR 100 000
178,4 [175,0 ; 181,9]	237,9 [232,2 ; 243,6]	296,7 [288,8 ; 304,7]

K hypothesis *(the extreme parameters of Y are constant)* and Return Levels

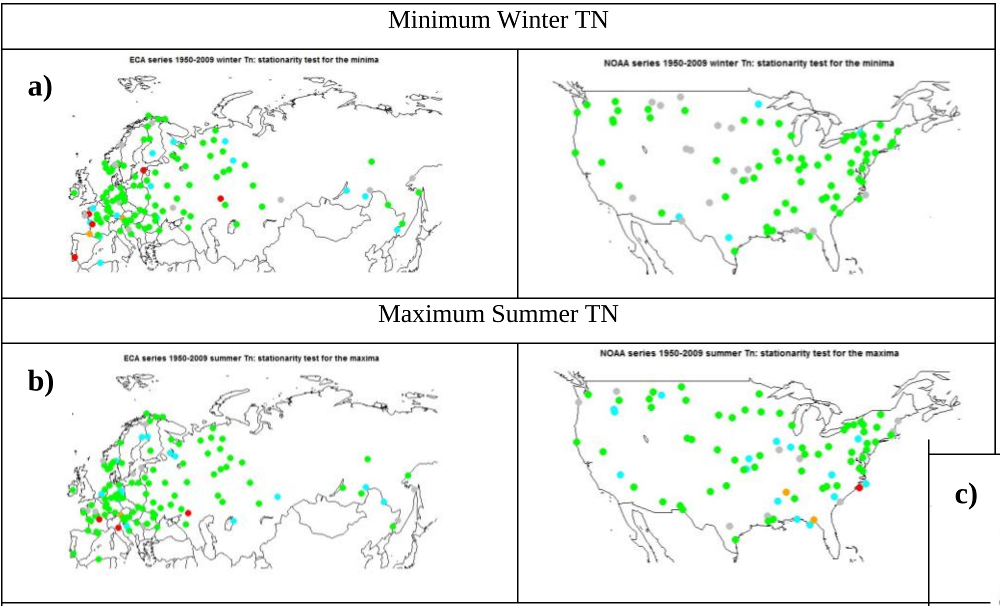
$$X_t = m(t) + s(t)Y_t$$

$$\begin{cases} \xi_X(k) = \xi_Y(k) \\ \sigma_X(k) = \sigma_Y(k)s(k) \\ \mu_X(k) = m(k) + \mu_Y(k)s(k) \end{cases}$$

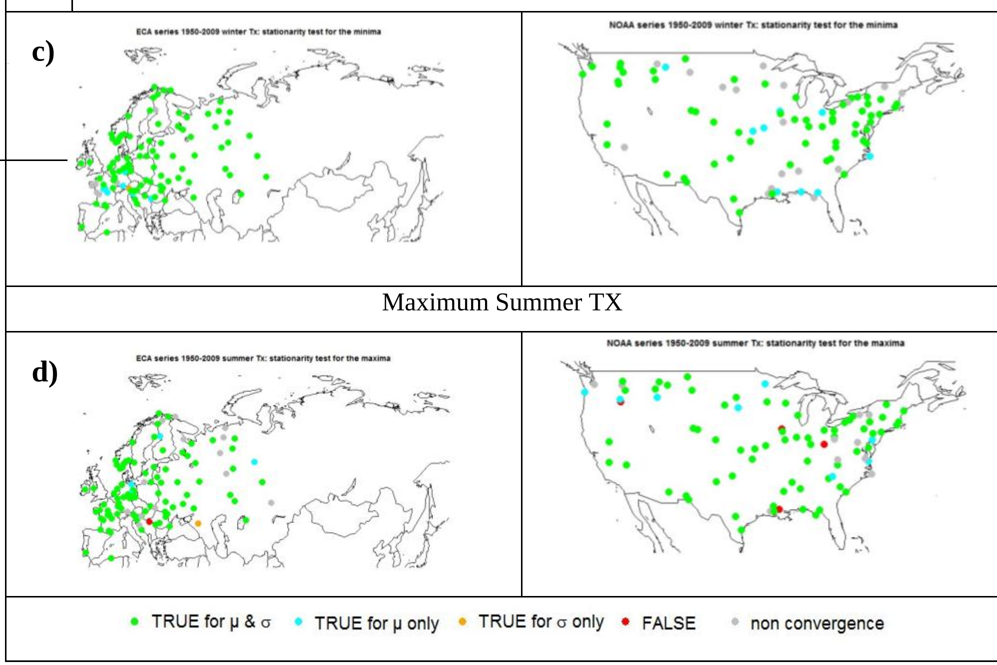
Calculate the return level



Results for different stations



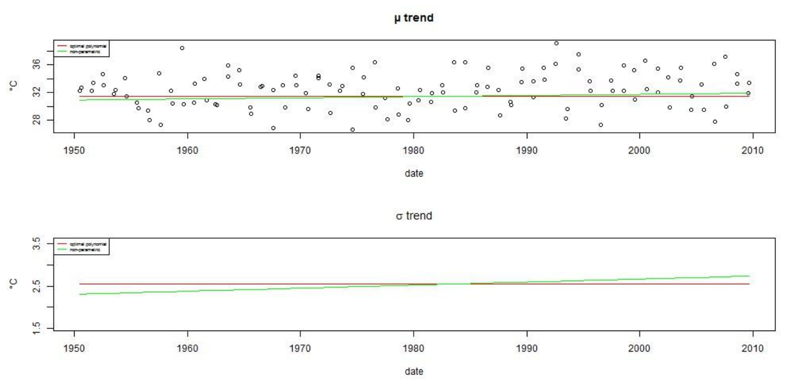
The K hypothesis is valid for almost all stations



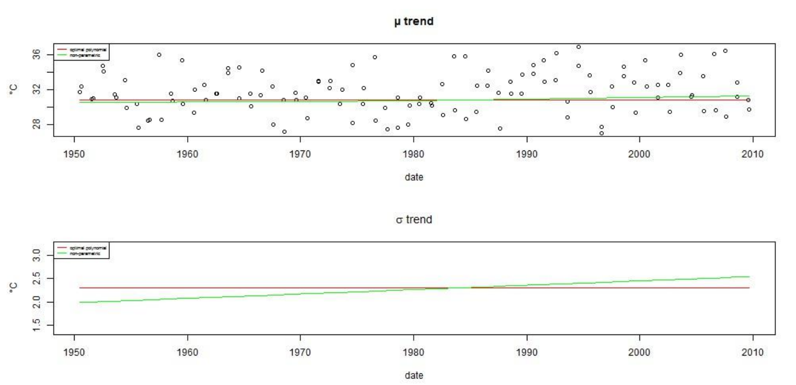
Importance of the evolution in variance

Non-parametric and optimal parametric trends in GEV parameters

BERLIN TX in summer



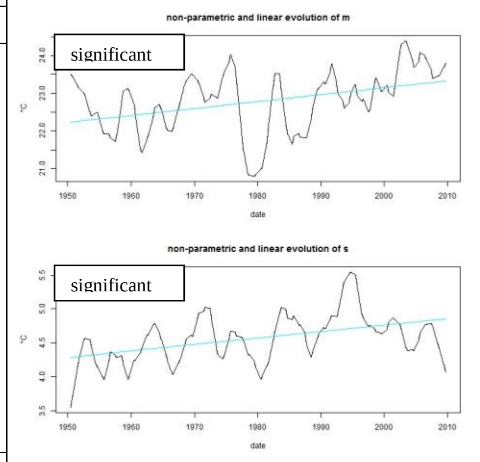
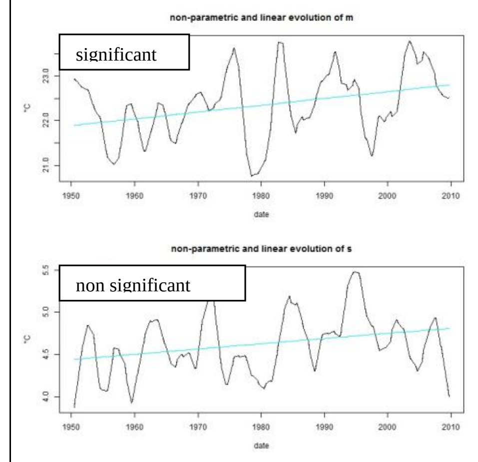
DRESDEN TX in summer



Optimal parametric

Non-parametric

Non-parametric and linear trends in mean and standard deviation

BERLIN TX in summer	DRESDEN TX in summer
	

50-year Return levels in 2030:

Dresden:

method 1: 36.9°C [35.7;38.1]

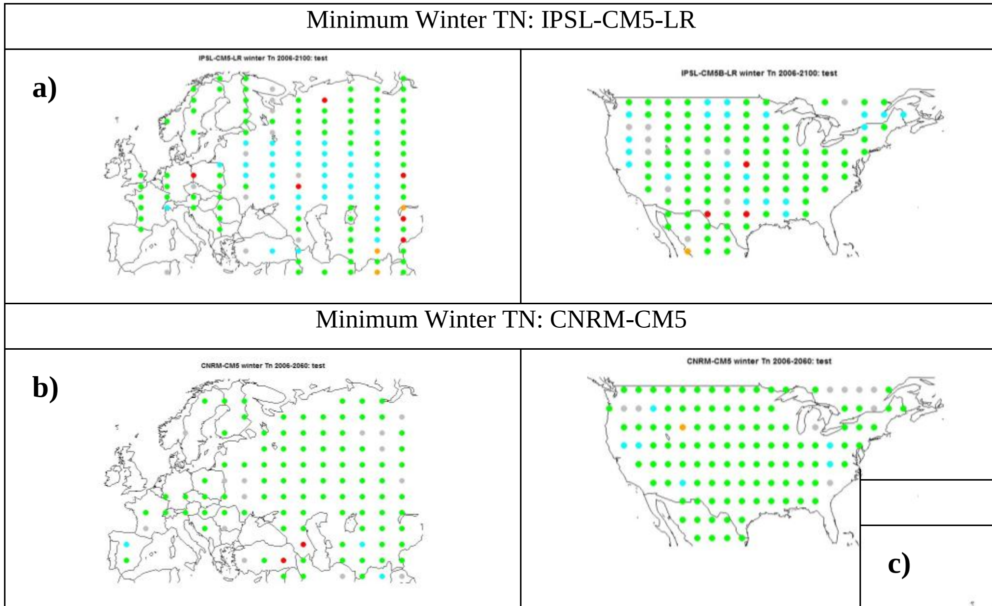
method 2: 37.8°C [36.3;38.7]

Berlin:

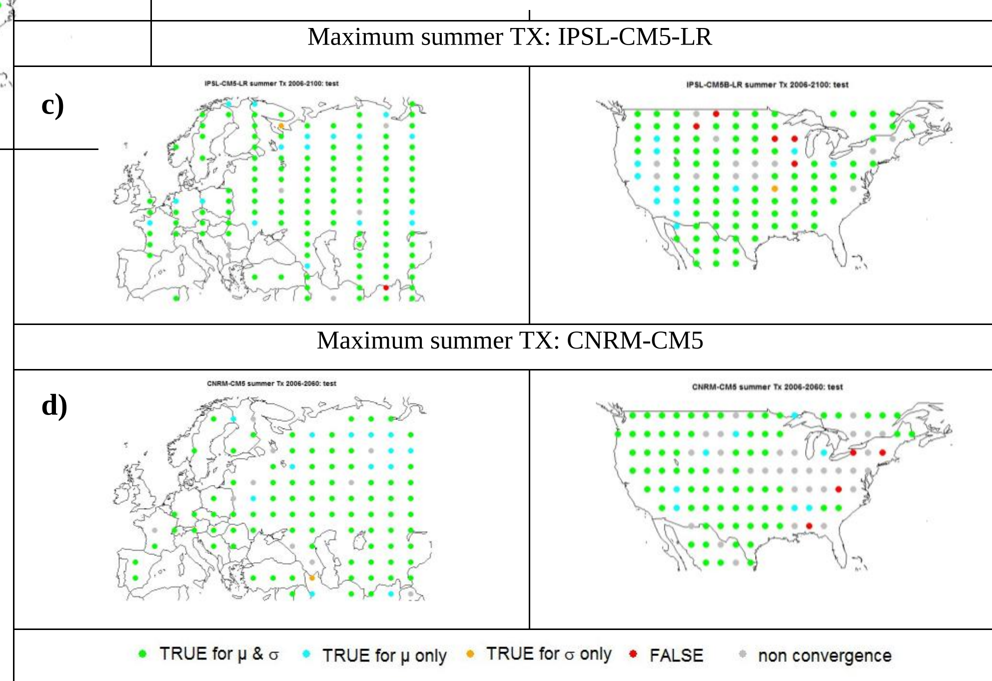
method 1: 38.2°C [37.2;39.3]

method 2: 40.9°C [39.1;42.4]

Validity for future climate



The K hypothesis is valid for almost all stations of two models

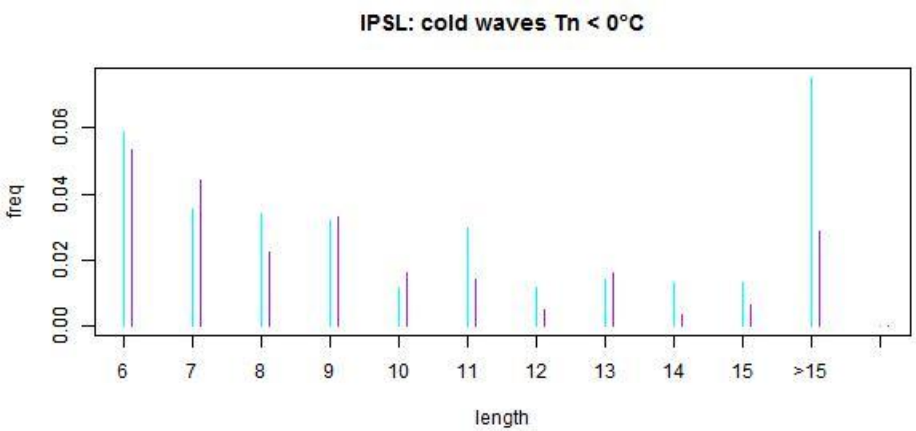


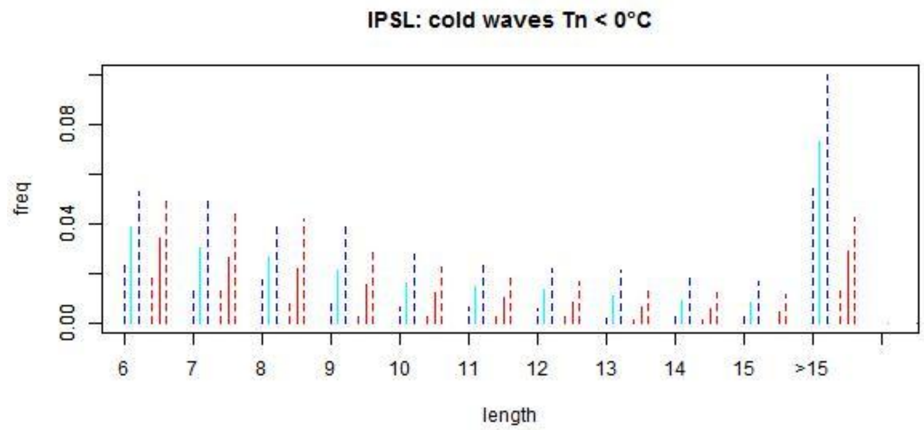
Present and future hot and cold waves

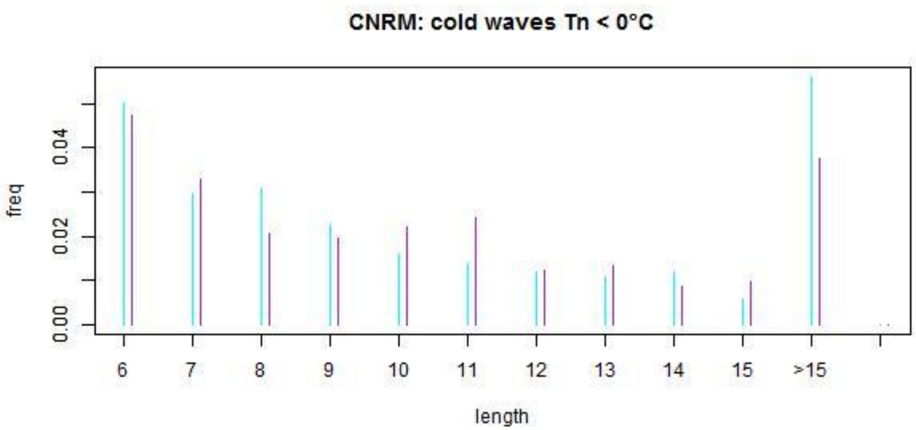
- Use of climate models
  - Estimation of parameters on an observed series
  - Simulations of the residuals Z(t)
  - Reconstruction of future temperatures with the trends and seasonalites given by the climate models
    - Past period(until 2005)
    - Future period(2006 – 2060)

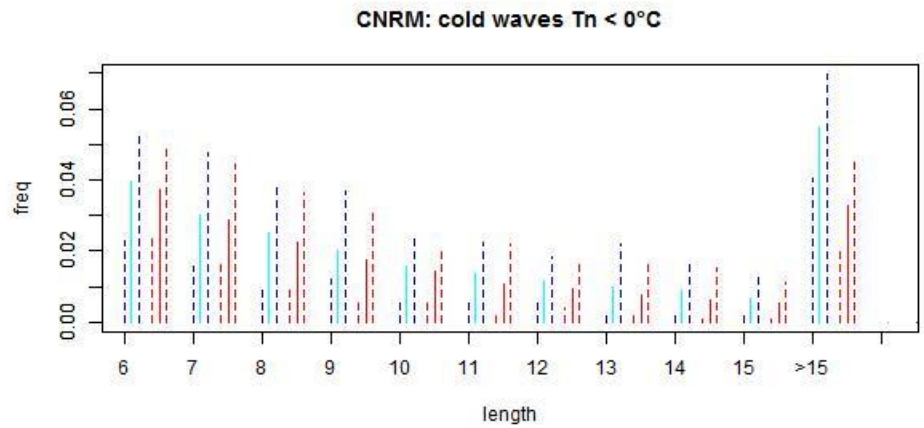
- Remark: the K hypothesis is valid, we hope that future extremes only depend on the trends and the seasonalities

Contribution of stochastic model



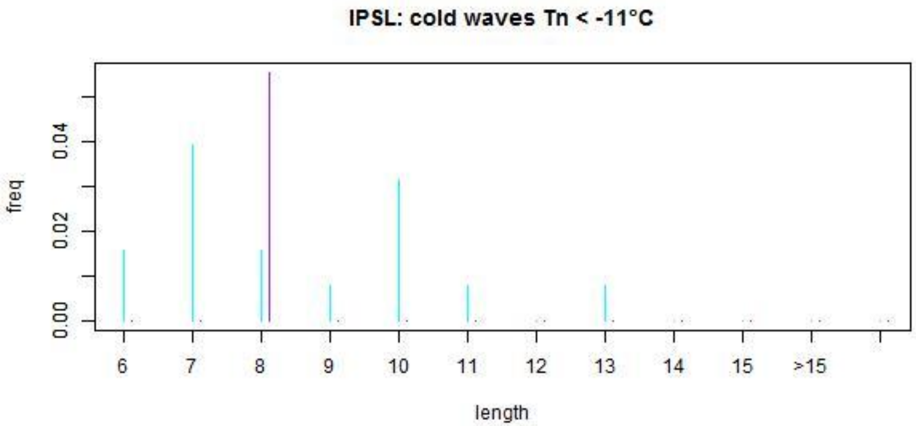


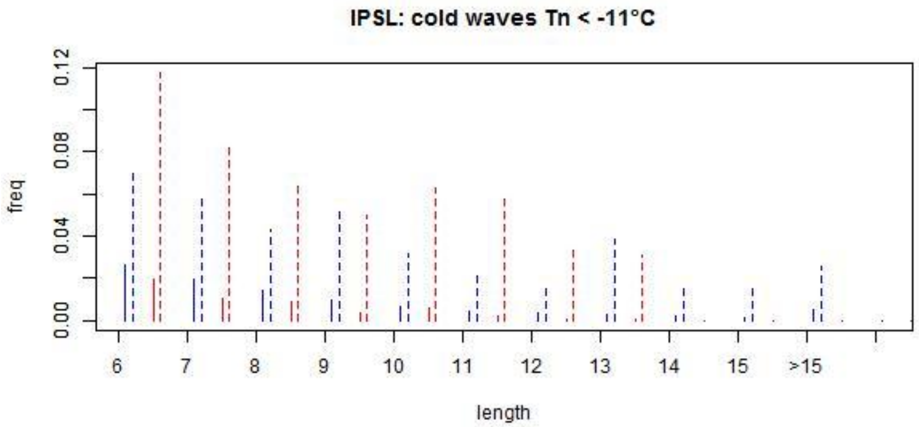


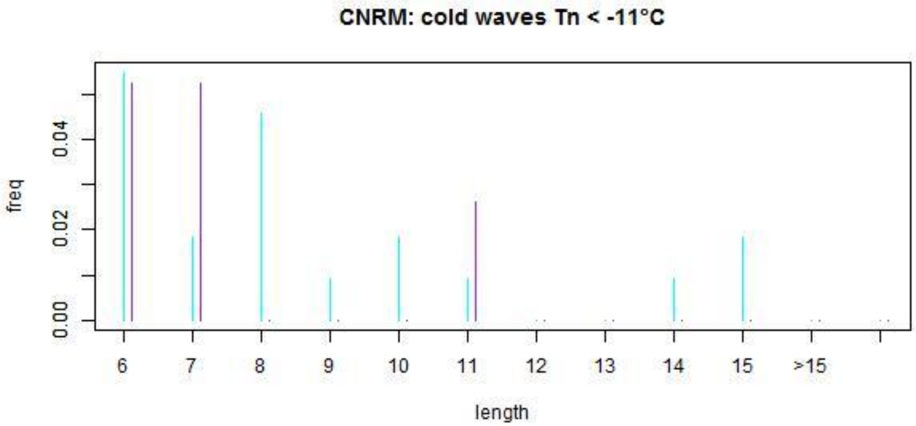


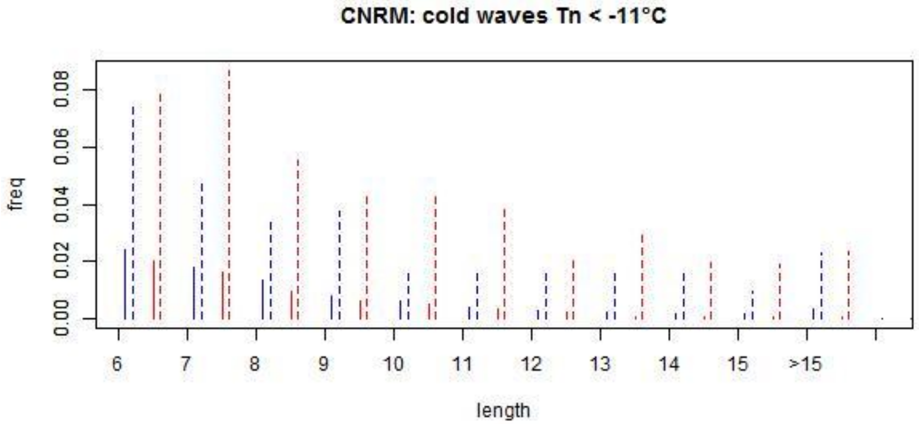
- With climate models only: difficult to say if the change of cold waves in the future is significant
- With the stochastic model + trends and seasonalities in climate models: the change is only significant for the cold waves of length greater than 15 days

Very cold waves







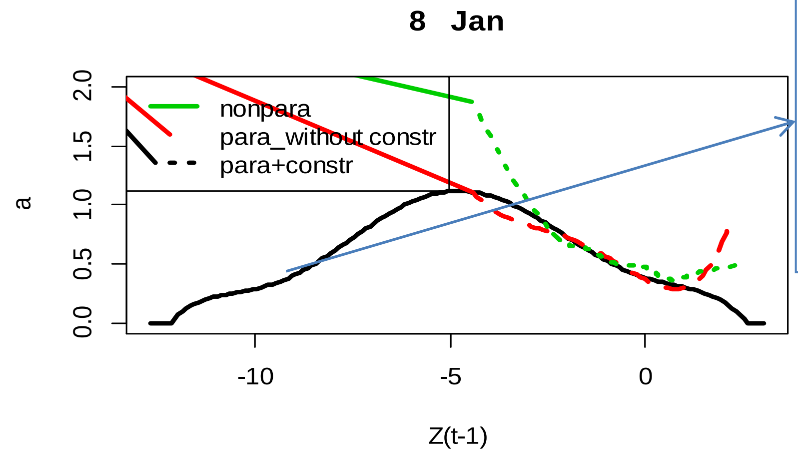


Using the stochastic model allows to simulate the distributions of long very cold waves and shows that IPSL model does not project waves longer than 14 days in the future, while CNRM model still does.

CONCLUSION & PERSPECTIVES

Conclusion on the stochastic model

- Positive: The estimation procedure is completely automatized with the optimal choice of the parameters
- Positive: Validation of the model for different climates
- Positive: give the correct results for different applications
- Negative: the likelihood estimation of the shape parameter ξ for the extremes is sensitive to the size of block maxima and so are the boundaries



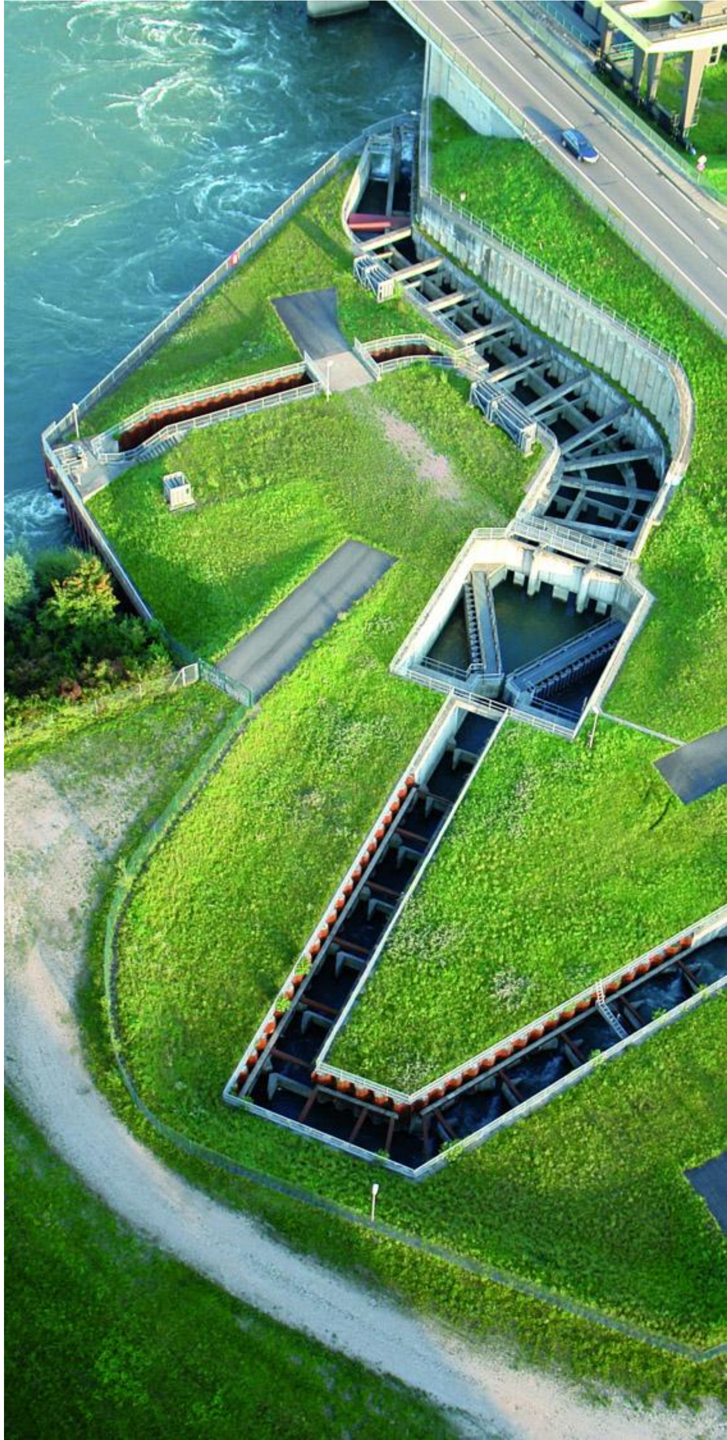
Inferior bound is irrealistic because $\xi \approx 0$

Perspectives

- Works in progress:
  - Improvement of estimation of the shape parameter ξ for the extremes with different non parametric estimators (Hill and his extensions)
  - Improvement of the quality of extremes by taking into account the more important correlation of the high quantiles
  - Direct estimation of the boundaries not using the estimation of extreme parameters

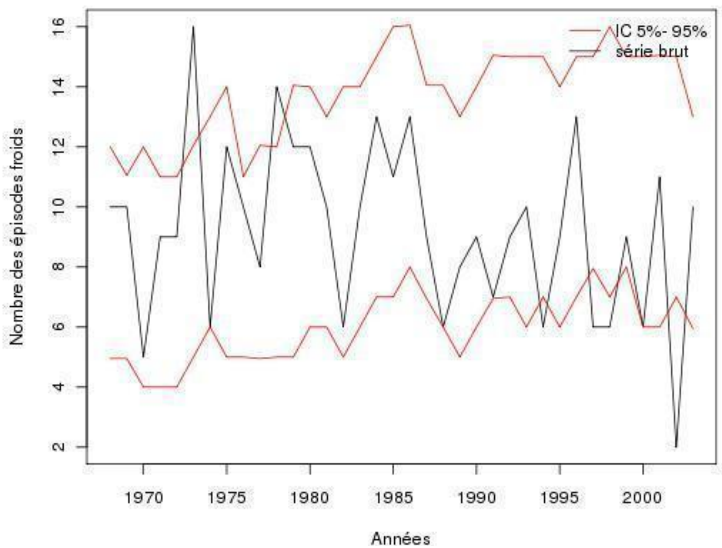
- Perspectives:
  - Build a stochastic model for two variables: temperature and precipitation taking into account the extremes

THANK YOU FOR YOUR ATTENTION !

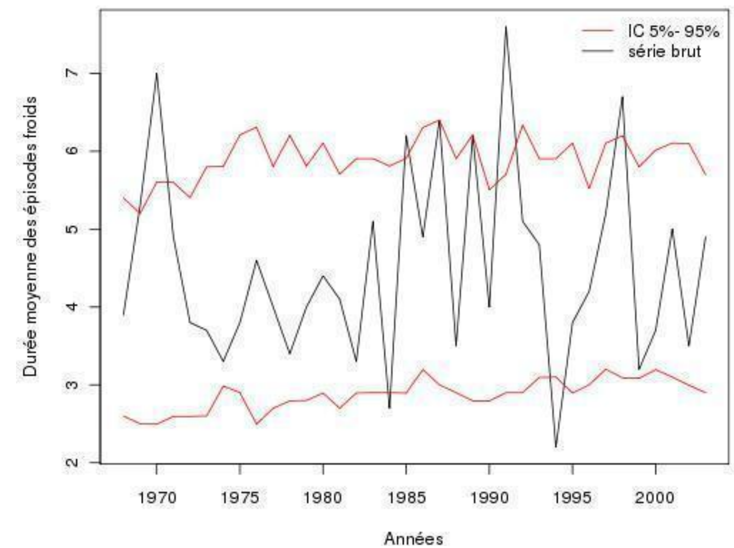


COLD WAVES (<0°C AND DURATION >2 DAYS)

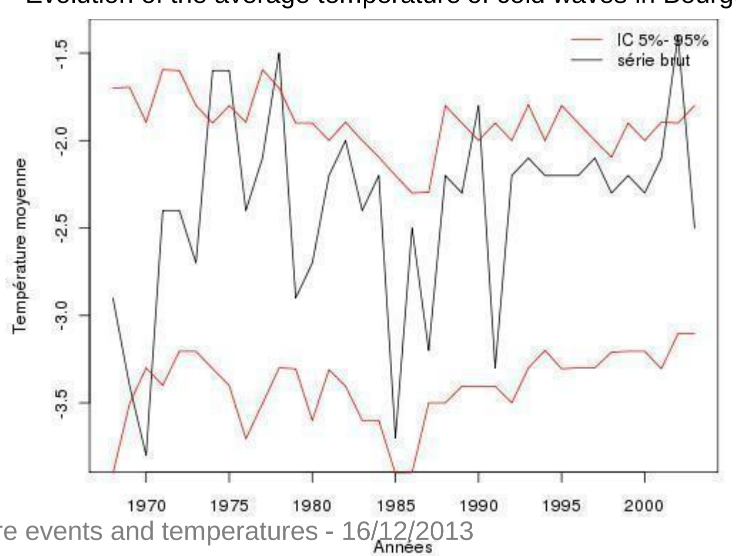
Evolution of the number of cold waves in Bourges



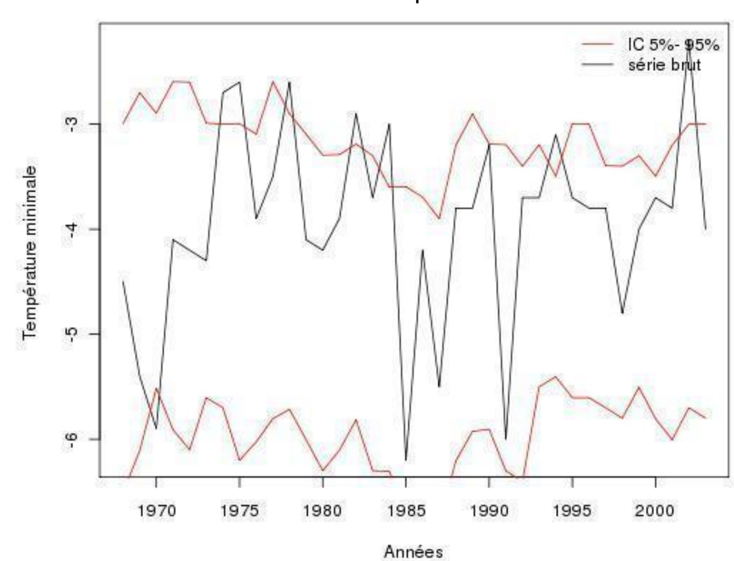
Evolution of the average duration of cold waves in Bourges



Evolution of the average temperature of cold waves in Bourges



Evolution of the minimum temperature of cold waves in Bourges



COLD WAVES

	Nombre		Durée		Moyenne		Minimum	
Nom de station	5%	95%	5%	95%	5%	95%	5%	95%
Augny	4	4	4	1	0	6	1	7
Bourges	3	2	2	4	1	4	1	6
Carcassonne	7	3	4	5	3	4	3	4
Champhol	1	3	2	2	0	6	0	6
Châteaubernard	2	5	0	2	3	5	2	5
Danne-et-Quatre-Vents	4	4	2	2	1	3	1	3
Istres	2	2	1	2	0	5	1	4
La Ciotat	0	3	0	1	6	0	4	0
Langres	3	2	6	2	2	5	1	7
Le Mans	5	5	4	4	0	5	0	7
Romilly-sur-Seine	3	3	3	4	1	7	1	7
Saint-Denis-d'Oléron	2	3	3	6	3	2	4	2
Sète	0	0	1	0	0	0	0	0
Strasbourg	3	0	3	2	2	2	2	4
Toulon	0	2	0	2	2	0	3	0
Max	**7**	**5**	**6**	**5**	**6**	**7**	**4**	**7**

The results are rather good: the proportion of simulated waves above Q95% and below Q5% is about 5%

RETURN LEVELS (50 YEARS)

	TX		TN	
	observations	simulations	observations	simulations
Berlin	38.2 [37.1;39.2]	39.8 [38.8;41.0]	-23.4 [-25.5;-21.0]	-26.5 [-31.5;-22.9]
Biarritz	39.6 [38.8;40.4]	41.0 [39.0;43.5]	-9.4 [-12.2;-6.6]	-11.0 [-12.6;-9.7]
Petropavlovsk	38.5 [37.6;39.5]	41.5 [39.3;44.8]	-43.7 [-45.2;-42.1]	-48.7 [-52.5;-45.3]
Olekminsk	-	-	-56.3 [-57.8;-54.8]	-58.8 [-61.4;-56.2]
Death Valley	54.3 [53.5;55.1]	55.2 [54.3;56.1]	-6.4 [-7.5;-5.3]	-7.4 [-8.8;-6.0]
Jacksonville	41.8 [40.3;43.3]	43.1 [41.5;44.5]	-29.5 [-31.3;-27.7]	-33.8 [-38.5;-30.6]
Glasgow	42.0 [41.1;42.8]	45.5 [44.3;46.9]	-42.9 [-44.4;-41.4]	-46.9 [-50.4;-44.0]
Charleston	39.5 [38.6;40.4]	40.3 [39.5;41.2]	-11.3 [-13.7;-9.0]	-8.8 [-10.0;-7.5]

The return levels are in general more important for the simulations → the simulations are less bounded (with 100 possibilities)
→ the model is not only able to reproduce extremes, but also to produce larger extremes than observed.