A stochastic model for daily air temperature reliably reproducing extreme values

Sylvie Parey, T.T.H. Hoang and D. Dacunha-Castelle
EDF/R&D Paris 11 university

Outline

- Introduction
- Pre-processing
- Model and estimation procedure
- Application to air temperature
- Conclusion and perspectives

Context

- EDF is interested in the impact of climate change on energy

Temperature is a key parameter influencing energy:

  - Reduce or increase the demand
  - hot or cold waves can affect overhead lines and production

Objectives

- Build a simulation model
    - for (maximum or minimum) daily air temperature for a fixed location
    - with good qualities for the bulk and the tails of distribution

able to easily produce a great number of realistic trajectories

Difficulties in the stochastic modelling of temperature

- **Non stationary, non linear**

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

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

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

  Temperature is a continuous-time process but observed at the discrete time steps

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

Pre-processing

- The aim is to remove the trends in mean and in variance and the (additive et multiplicative) seasonalities to obtain reduced series as stationary as possible

- The processing treatment uses both the nonparametric and parametric approaches

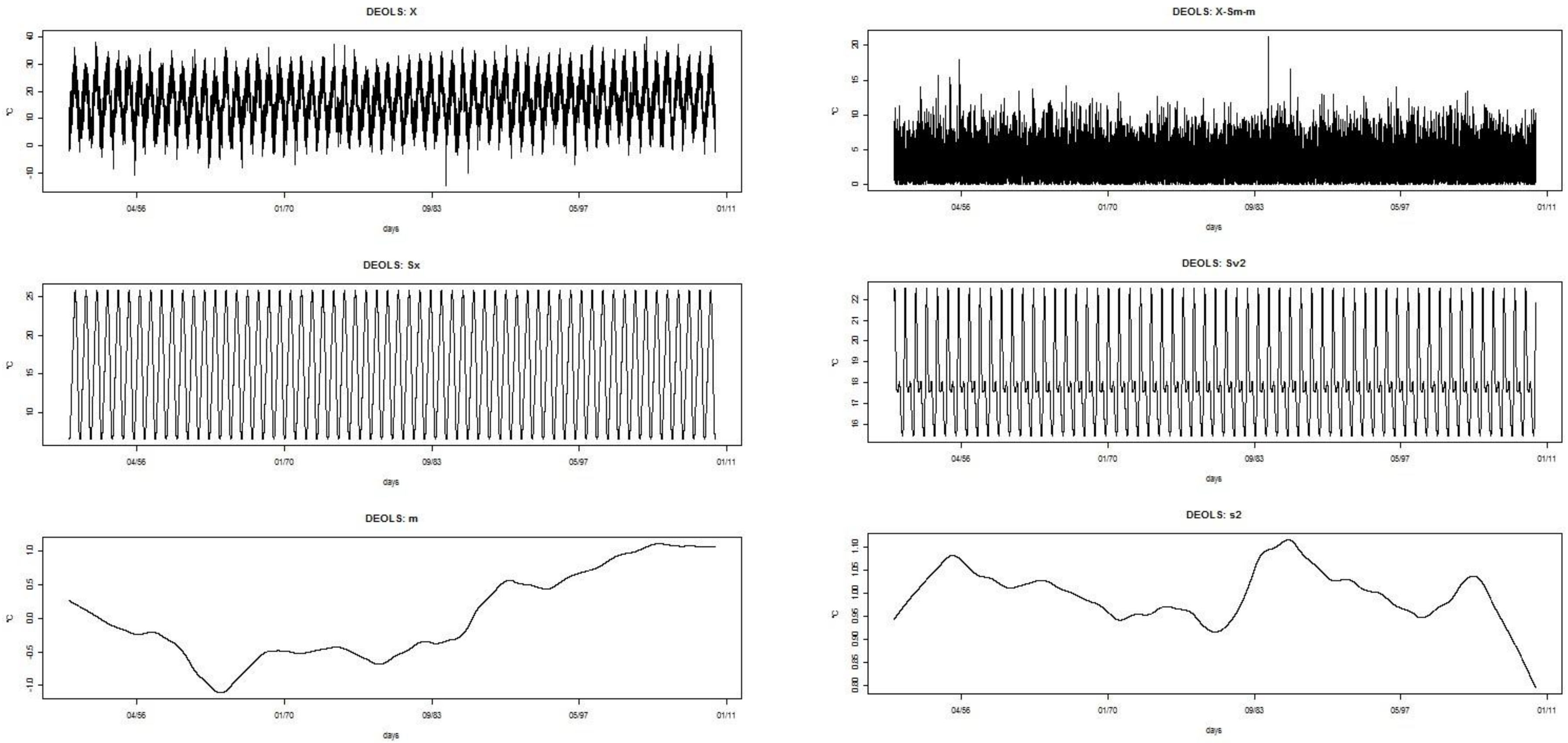
$$X(t) = m(t) + S(t) + s(t) S_V(t) Z(t)$$

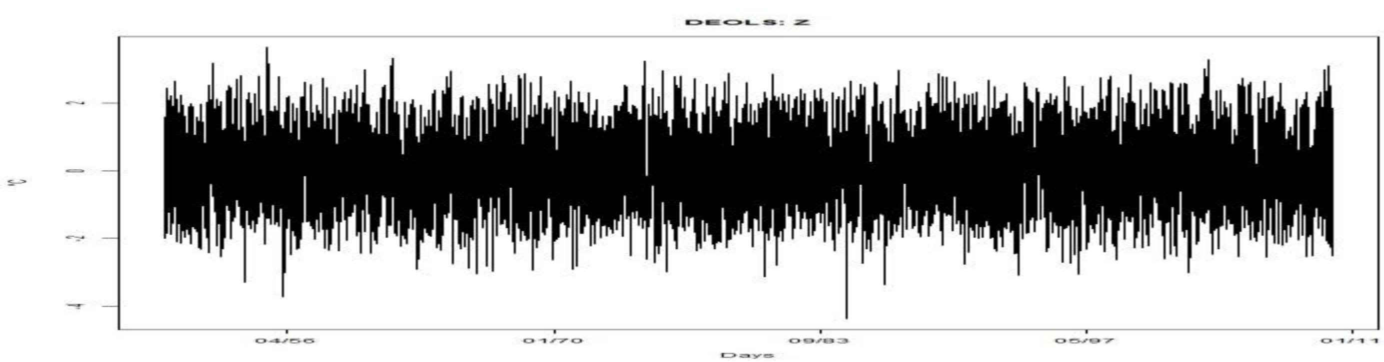
  - $m(t), s(t)$: mean, scale function, $S(t), S_V(t)$: seasonalities in mean and in scale

  - **Estimation procedure:**
    - estimate $m(t)$ by loess, $S(t)$ by a trigonometric function from the series $X(t)$, then $s(t)$ by loess and $S_V^2(t)$ by a trigonometric function from the series $\left[X(t) - \hat{m}(t) - \hat{S}(t)\right]^2$
    - For $m(t), s(t)$, the modified partitioned cross validation[1] is used , for $S(t), S_V^2(t)$, the Akaike criteria are used
    - The reduced series: $Z_t = \left(X_t - \hat{m}_t - \hat{S}_t\right) / (\hat{s}_t \hat{S}_{Vt})$

[1] *Modified partitioned CV: new algorithm for correlated data (thesis of Hoang, 2010)*

Illustration





The SFHAR(seasonal functional heteroscedastic autoregressive) model

- First order Euler scheme of a discrete diffusion: $Z_t = b(Z_{t-1}) + a(Z_{t-1})\varepsilon_t, \quad \varepsilon_t \propto N(0,1)$

***Rq**: This is a discrete approximation of the bounded diffusion. The gaussian noise is in fact a truncated gaussian*

- Extension: SFHAR model $Z(t) = \left[\theta_{0,k} + \sum_{j=1}^{p_1}\left(\theta_{1,k}^{j} \cos\frac{2j\pi t}{365} + \theta_{2,k}^{j} \sin\frac{2j\pi t}{365}\right)\right] Z(t-1) + a(t, Z_{t-1})\varepsilon_t$

  $\varepsilon_t \propto N(0,1)$

- Estimate $a^2(t, Z_{t-1})$ with constraints:
  - Zero outside 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:

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

Estimation procedure and optimisation

- **Estimation of the autoregressive part (AR(1))**

- **Choose the number of cosine and sine terms by a Akaike criterion**

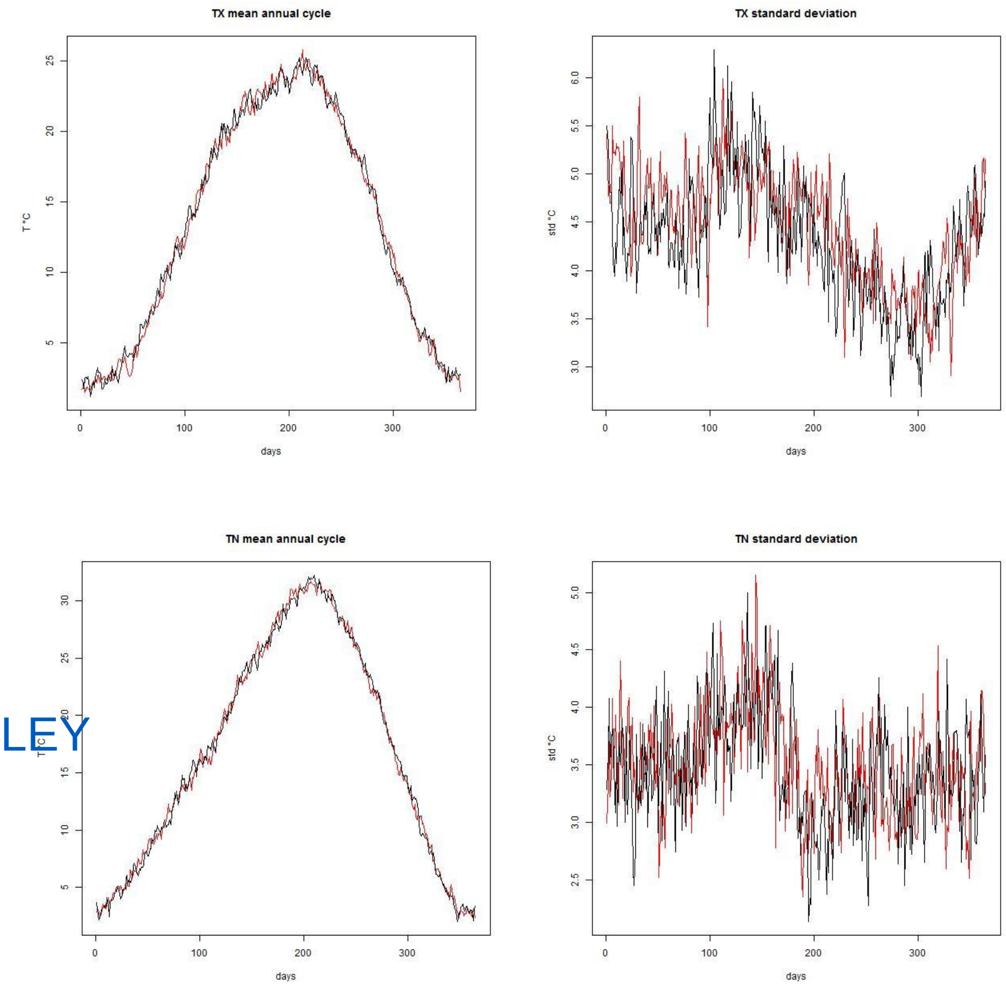
- **Estimation of the volatility through maximum likelihood with constraints**
  - **Find the initial values using least squares estimation:** problem of least squares with equality and inequality constraints → transform to the quadratic programming problem and use the algorithm of Goldfarb and Idnani (1982,1983)

  - **Maximum likelihood estimation with constraints:** use the results of least square estimation as the initial values and use the Nelder and Mead algorithm (1965)

Results: bulk of the distribution

TX BERLIN

TN DEATH VALLEY

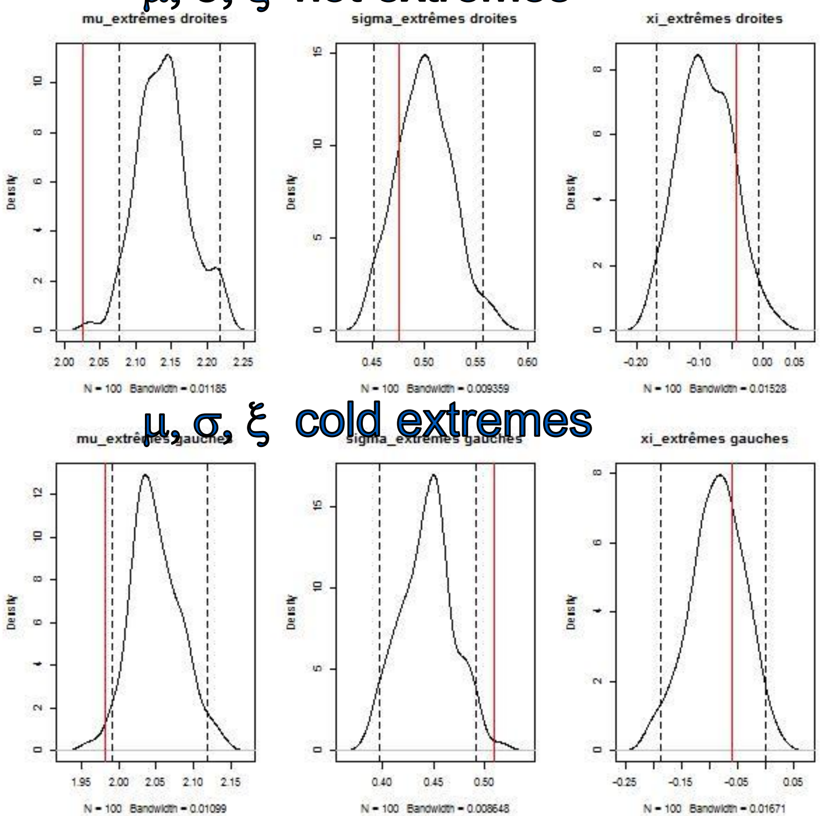


Statistical test: Comparison of daily distributions ➔ equality of the distributions accepted at 5% significance level

Results: extremes of the standardized residuals

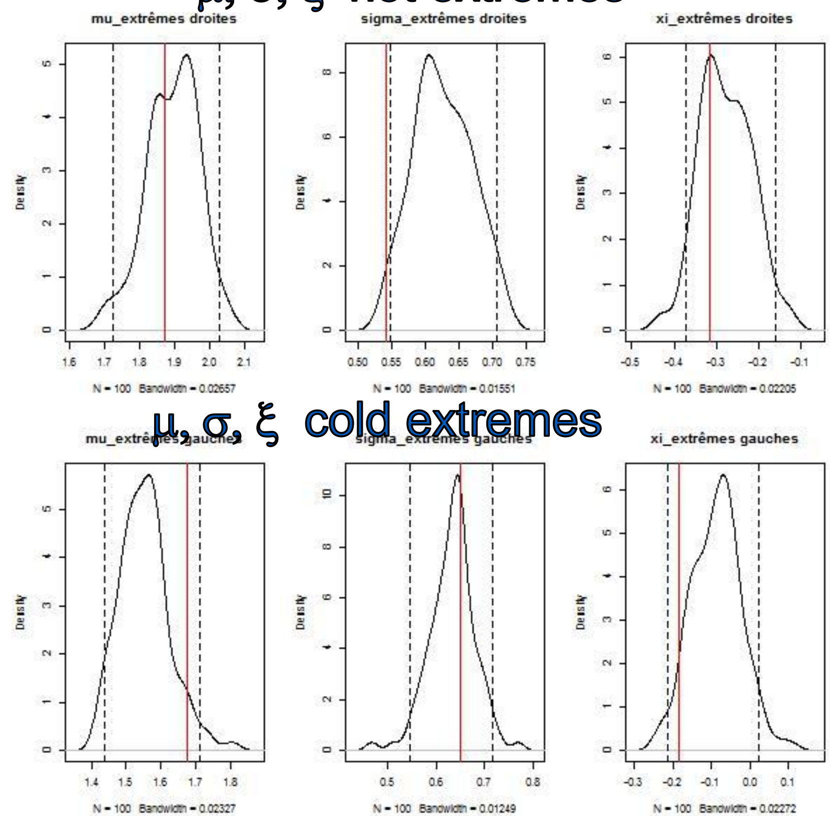
TN Death-Valley

μ, σ, ξ hot extremes



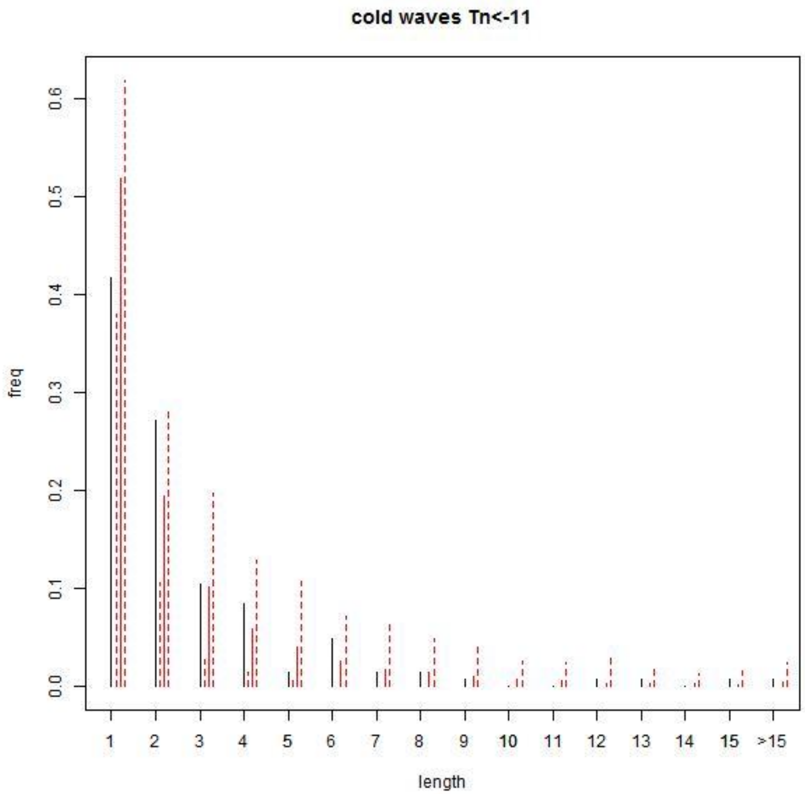
TX Berlin

μ, σ, ξ hot extremes

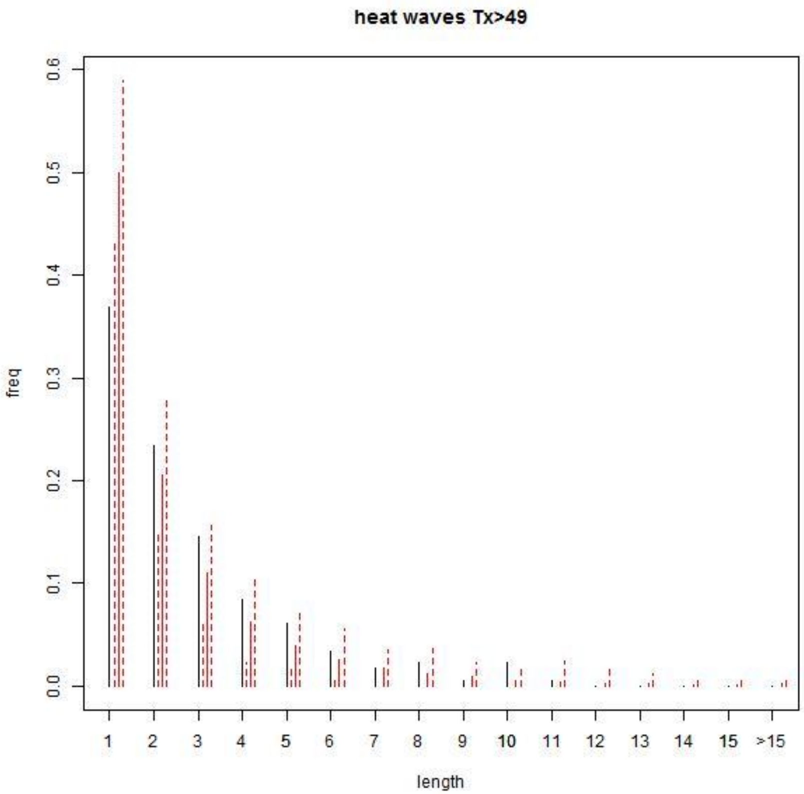


Results: cold and heat waves

Cold waves in Berlin

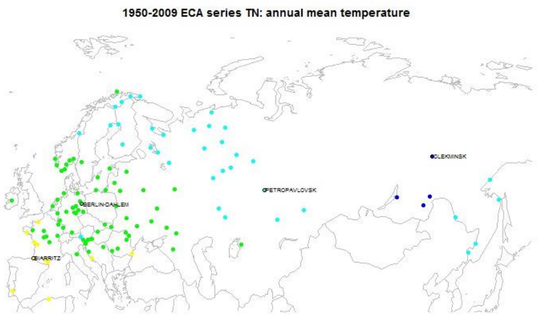


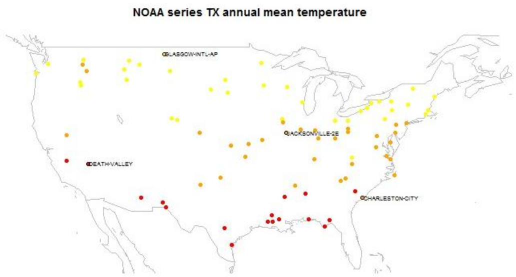
Heat waves in Death-Valley



Conclusion and perspectives

- Validation of the model for different climates





- Calibration on the 1rst period of the T time series / validation on the 2nd
  - Stepwise mean and variance change
  - Use of non parametric mean and variance of the second period
- Use to study future extremes and future evolutions of cold and heat waves
- Statistical downscaling tool for T extremes?

Thanks for your attention