Analyzing drought periods using NHPP and Bayesian inference

A case study in Campinas, Brazil

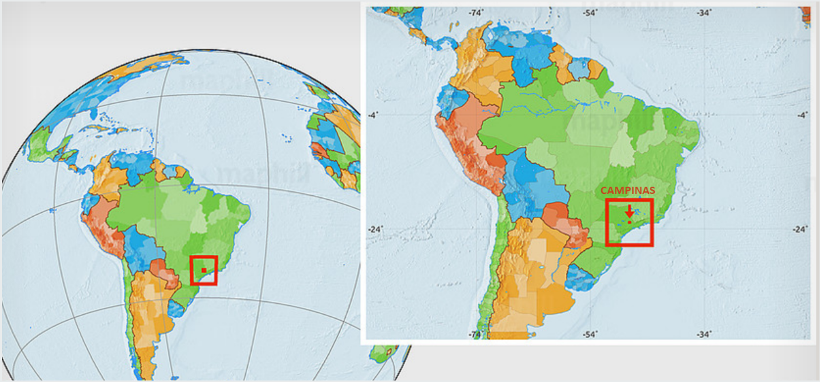
Presented by
Marcel Braasch - ist1104495
Marco Hövekenmeier - ist1104241
Nicola Dal Cin - ist1104444

Instituto Superior Técnico

Introduction

Overview

We analyze the precipitation data in Campinas (Brazil) which manifest a big change in their behaviour due to global warming.



Overview

- The data are extracted from the *Standard Precipitation Index* (SPI) from Jan 1947 to May 2011.
- We use a specific NHPP, i.e. the *power law process* (PLP), also considering the presence of change points.
- A Bayesian analysis is performed using *Markov Chain Monte Carlo Methods* (MCMC).

Overview

- The data are extracted from the *Standard Precipitation Index* (SPI) from Jan 1947 to May 2011.
- We use a specific NHPP, i.e. the *power law process* (PLP), also considering the presence of change points.
- A Bayesian analysis is performed using *Markov Chain Monte Carlo Methods* (MCMC).

Overview

- The data are extracted from the *Standard Precipitation Index* (SPI) from Jan 1947 to May 2011.
- We use a specific NHPP, i.e. the *power law process* (PLP), also considering the presence of change points.
- A Bayesian analysis is performed using *Markov Chain Monte Carlo Methods* (MCMC).

The SPI is an **precipitation indicator**, based on a large set of historic rainfall data, and also used to classify wet and dry periods [1]:

2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
−0.99 to 0.99	Near normal
−1.0 to −1.49	Moderately dry
−1.5 to −1.99	Severely dry
−2 and less	Extremely dry

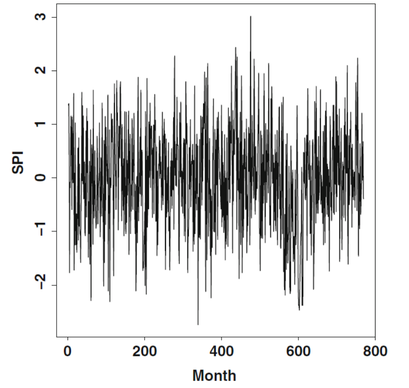


Figure: Monthly SPI – 1 values in Campinas from Jan 1947 to May 2011.

SPI values

For Campinas and different timescales

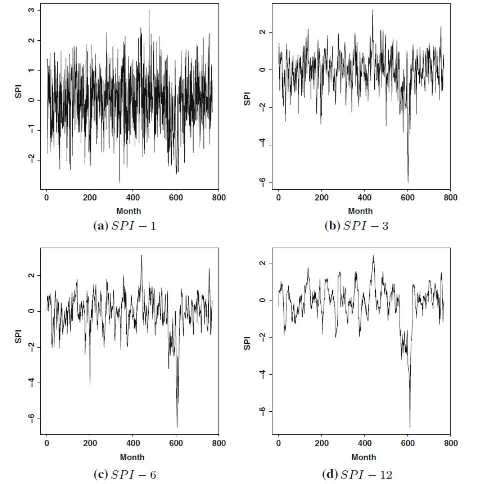


Figure: SPI values for Campinas with timescale of 1,3,6,12 months.

TÉCNICO LISBOA

Definition

We consider a **drought event** to occur any time the SPI is below the threshold $L = -1.0$.

2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2 and less	Extremely dry

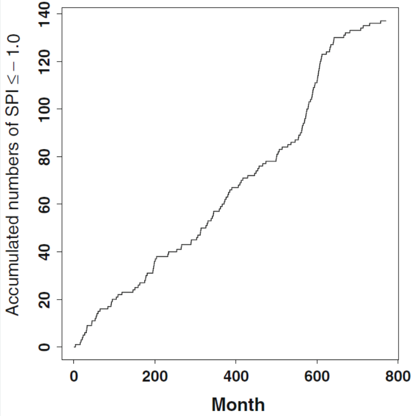


Figure: Cumulative number of drought months (using SPI – 1).

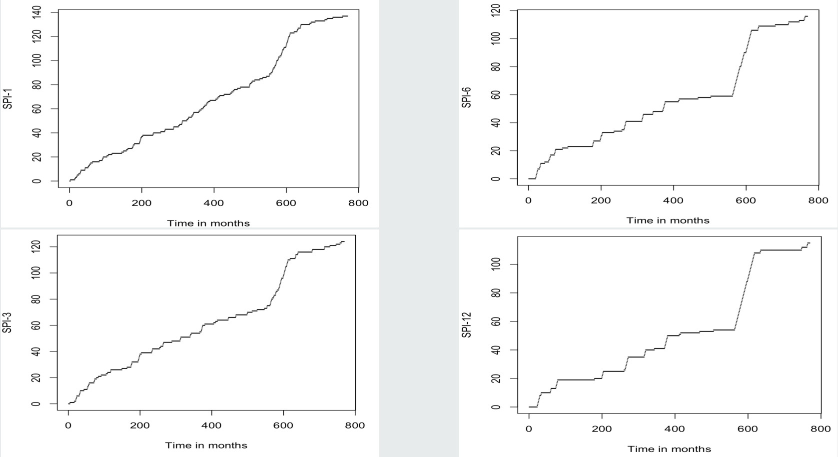


Figure: Cumulated $SPI\text{-}i$ values for $i = 1, 3, 6, 12$.

Definition

A counting process $\{N(t),\ t \geq 0\}$ is said **non–homogeneous Poisson process** (NHPP) with mean value function $m(t) = \int_0^t \lambda(x)dx$ if

i) $N(0) = 0$,

ii) $\{N(t), t \geq 0\}$ has independent increments,

iii) $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$,

iv) $P(N(t+h) - N(t) \geq 2) = o(h)$,

where $\lambda(t) > 0$ is the *intensity function*.

Theorem

A **non–homogeneous Poisson process** satisfy

$$N(t+s) - N(t) \sim \text{Poisson}(m(t+s) - m(t)),$$

for all $t, s \geq 0$.

Definition

An NHPP $\{N(t),\ t \geq 0\}$ is a **power law process** (PLP) with parameter $\boldsymbol{\theta} := (\alpha, \sigma)$ if its mean function is

$$m_{\mathsf{PLP}}(t; \boldsymbol{\theta}) = \left(\frac{t}{\sigma}\right)^{\alpha}, \quad \alpha, \sigma > 0.$$

And so the intensity is

$$\lambda_{PLP}(t; \boldsymbol{\theta}) = \left(\frac{\alpha}{\sigma}\right)\left(\frac{t}{\sigma}\right)^{\alpha - 1}.$$

The posterior estimates of the parameters are obtained through Markov Chain Monte Carlo (MCMC) sampling.

Definition

The joint posterior distribution for $\boldsymbol{\theta}$ given the data D_T is

$$p(\boldsymbol{\theta}|D_t) \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}|D_t)$$

Model and Implementation

TÉCNICO LISBOA

The SPI is an **precipitation indicator**, based on a large set of historic rainfall data, and also used to classify wet and dry periods [1]:

2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
−0.99 to 0.99	Near normal
−1.0 to −1.49	Moderately dry
−1.5 to −1.99	Severely dry
−2 and less	Extremely dry

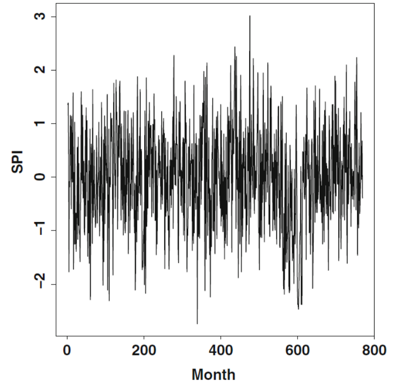


Figure: Monthly SPI values in Campinas from Jan 1947 to May 2011.

Data set

The Standard Precipitation Index

Definition

We consider a **drought event** to occur any time the SPI is below the threshold $L = -1.0$.

2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2 and less	Extremely dry

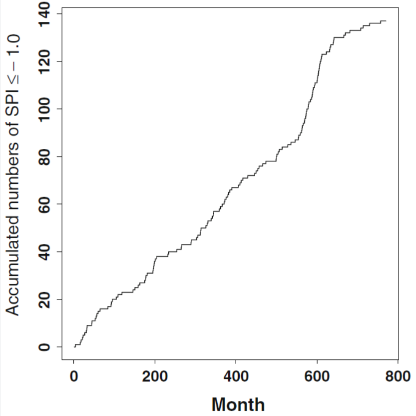


Figure: Cumulative number of drought months (using SPI – 1).

- Let $\{N(t),\ t \geq 0\}$ be the process which counts the *number of drought months to time* t with mean $m(t)$.

- The Bayesian analysis is used to determine the parameters $\boldsymbol{\theta} = (\alpha, \sigma)$ to fit the process $N(t)$ to the PLP having mean

$$m_{\mathsf{PLP}}(t; \boldsymbol{\theta}) = (t/\sigma)^{\alpha}.$$

- The likelihood function for $\boldsymbol{\theta}$ of the truncated model at time T is

$$L(\boldsymbol{\theta}; D_T) = \prod_{i=1}^{n} \lambda(t_i) \cdot \exp\left[-m(T)\right],$$

where $D_T = \{n; t_1, \ldots, t_n; T\}$ is the data set.

- Let $\{N(t),\ t \geq 0\}$ be the process which counts the *number of drought months to time* t with mean $m(t)$.

- The Bayesian analysis is used to determine the parameters $\boldsymbol{\theta} = (\alpha, \sigma)$ to fit the process $N(t)$ to the PLP having mean

$$m_{\text{PLP}}(t; \boldsymbol{\theta}) = (t/\sigma)^{\alpha}.$$

- The likelihood function for θ of the truncated model at time T is

$$L(\boldsymbol{\theta}; D_T) = \prod_{i=1}^{n} \lambda(t_i) \cdot \exp\left[-m(T)\right],$$

where $D_T = \{n; t_1, \ldots, t_n; T\}$ is the data set.

Likelihood function

No change points

- Let $\{N(t),\ t \geq 0\}$ be the process which counts the *number of drought months to time* t with mean $m(t)$.

- The Bayesian analysis is used to determine the parameters $\boldsymbol{\theta} = (\alpha, \sigma)$ to fit the process $N(t)$ to the PLP having mean

$$m_{\mathsf{PLP}}(t; \boldsymbol{\theta}) = (t/\sigma)^{\alpha}.$$

- The likelihood function for $\boldsymbol{\theta}$ of the truncated model at time T is

$$L(\boldsymbol{\theta}; D_T) = \prod_{i=1}^{n} \lambda(t_i) \cdot \exp\left[-m(T)\right],$$

where $D_T = \{n; t_1, \ldots, t_n; T\}$ is the data set.

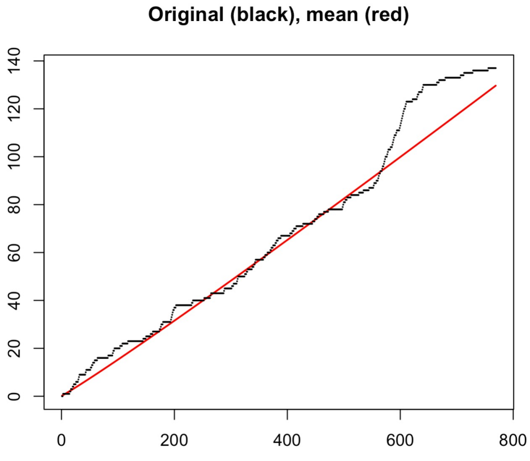


Figure: Estimated mean and original data.

Simulation

No change point

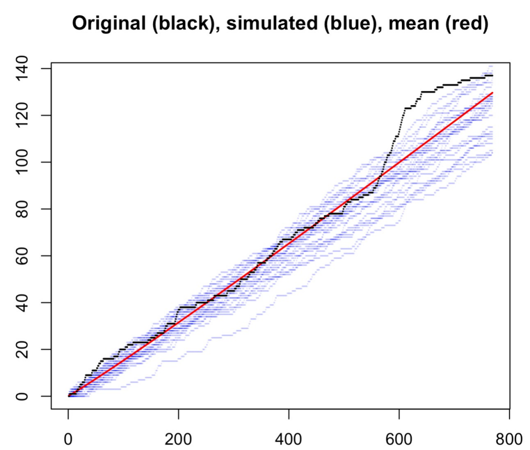


Figure: Some (25) simulations of the process, ran using the model developed.

Change points

To better represent a changing behaviour in the data, we can let our model be more flexible introducing new parameters and a **change point** $\tau \in (0, T)$ such that

$$\lambda(t;\boldsymbol{\theta}) = \begin{cases} \left(\frac{\alpha_1}{\sigma_1}\right)\left(\frac{t}{\sigma_1}\right)^{\alpha_1 - 1}, & 0 \leq t \leq \tau \\ \\ \left(\frac{\alpha_2}{\sigma_2}\right)\left(\frac{t}{\sigma_2}\right)^{\alpha_2 - 1}, & t > \tau. \end{cases}$$

So now the parameters' vector is $\boldsymbol{\theta} = (\alpha_1, \sigma_1, \alpha_2, \sigma_2, \tau)$.

Considering the data

$$D_T = \{n; t_1, \dots, t_{N(\tau)}, t_{N(\tau)+1}, \dots, t_n; T\}$$

for the previous model, the likelihood function is given by

$$L(\boldsymbol{\theta}; D_T) = \prod_{i=1}^{N(\tau)} \lambda_1(t_i) \cdot \exp[-m_1(\tau)]$$

$$\prod_{i=N(\tau)+1}^{N(T)} \lambda_2(t_i) \cdot \exp[-m_2(T) + m_2(\tau)]$$

Simulation
One change point

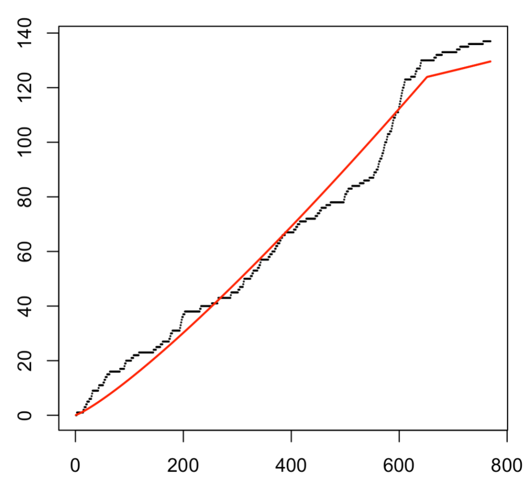


Figure: Estimated mean (red) and original data (black) of a PLP with one change point on SPI-1.

Simulation

Two change points

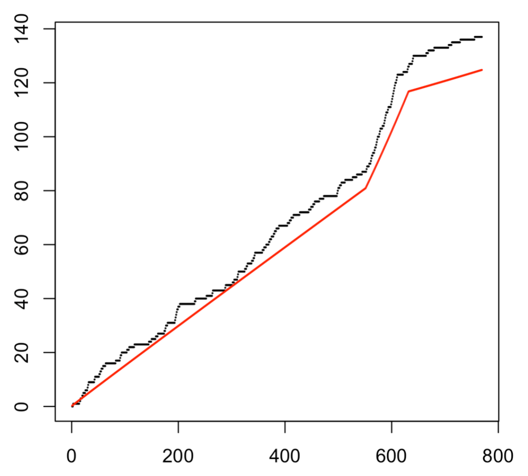


Figure: Estimated mean (red) and original data (black) of a PLP with two change points on SPI-1.

Linear (Homogeneous Poisson Process)

$$m_i(t) = a_i t, \quad t \leq \tau_i$$

Quadratic

$$m_i(t) = a_i t + b_i t^2, \quad t \leq \tau_i$$

8-th degree polynomial

$$m(t) = a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5 + a_6 t^6 + a_7 t^7 + a_8 t^8$$

SPI	$PLP_{2,0}$	$PLP_{4,1}$	$PLP_{6,2}$	$HP_{3,2}$	$QP_{6,2}$	$PP_{8,0}$
1	751.9	739.8	718.8	**714.5**	**714.5**	738.0
3	705.2	704.8	659.9	662.0	**656.8**	695.8
6	676.0	655.7	572.7	578.0	**565.5**	663.4
12	670.3	649.6	540.3	541.9	**536.0**	656.7

Table: DIC comparison between the power law (PLP), the homogeneous (HP), the quadratic (QP) and the 8th degree polynomial process (PP). $P_{a,b}$ denotes having a parameters with b change points.

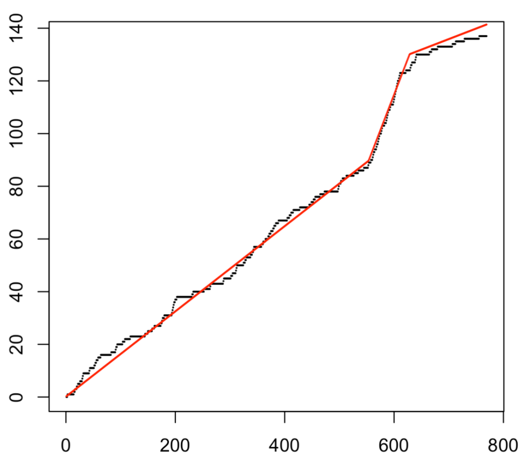


Figure: Original SPI-1 data (black) and estimated mean (red) of a homogeneous point process.

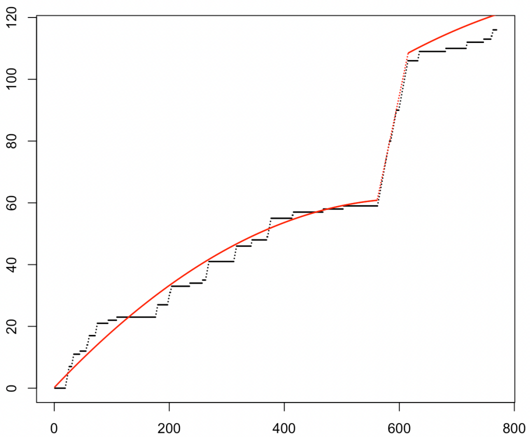


Figure: Original SPI-6 data (black) and estimated mean (red) of a quadratic point process.

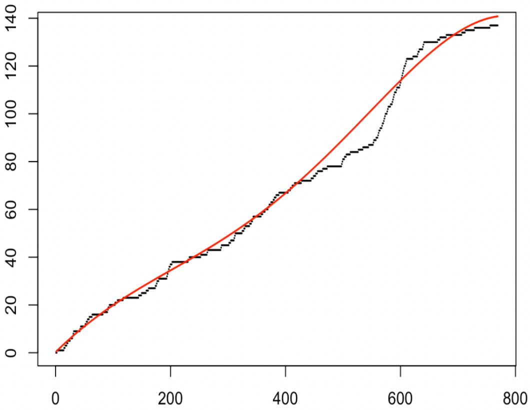


Figure: Original SPI-1 data (black) and estimated mean (red) of an 8-th degree polynomial process.

Conclusion

Critique

- Authors fit model to data and do not use it for prediction
- Power law extensively used in drought and rain forecast [2, 3], but authors do not really motivate it.
- From a practical perspective, the fitted model does not allow for interpretation
- Time series models or neural networks seem like an appropriate tools to perform prediction [4, 5]

Critique

- Authors fit model to data and do not use it for prediction
- Power law extensively used in drought and rain forecast [2, 3], but authors do not really motivate it.
- From a practical perspective, the fitted model does not allow for interpretation
- Time series models or neural networks seem like an appropriate tools to perform prediction [4, 5]

Critique

- Authors fit model to data and do not use it for prediction
- Power law extensively used in drought and rain forecast [2, 3], but authors do not really motivate it.
- From a practical perspective, the fitted model does not allow for interpretation
- Time series models or neural networks seem like an appropriate tools to perform prediction [4, 5]

Critique

- Authors fit model to data and do not use it for prediction
- Power law extensively used in drought and rain forecast [2, 3], but authors do not really motivate it.
- From a practical perspective, the fitted model does not allow for interpretation
- Time series models or neural networks seem like an appropriate tools to perform prediction [4, 5]

Review

- Modeling of a series of *Standard Precipitation Index* (SPI) values using a Power Law Process
- Reproduction of the papers results
- Extension with different mean functions and quantitative comparison

Review

- Modeling of a series of *Standard Precipitation Index* (SPI) values using a Power Law Process
- Reproduction of the papers results
- Extension with different mean functions and quantitative comparison

Review

- Modeling of a series of *Standard Precipitation Index* (SPI) values using a Power Law Process
- Reproduction of the papers results
- Extension with different mean functions and quantitative comparison

Code

If you're interested, check out our repository:

`https://github.com/marcoHoev/nhpp_drought_modeling/`

Thank you!

References I

[1] T. B. McKee, N. J. Doesken, J. Kleist, *et al.*, "The relationship of drought frequency and duration to time scales," in *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, California, 1993.

[2] R. Dickman, "Rain, power laws, and advection," *Physical review letters*, vol. 90, no. 10, p. 108701, 2003.

[3] J. D. Pelletier and D. L. Turcotte, "Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment," *Journal of hydrology*, vol. 203, no. 1-4, pp. 198–208, 1997.

[4] S. Morid, V. Smakhtin, and K. Bagherzadeh, "Drought forecasting using artificial neural networks and time series of drought indices," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 15, pp. 2103–2111, 2007.

[5] R. Modarres, "Streamflow drought time series forecasting," *Stochastic Environmental Research and Risk Assessment*, vol. 21, no. 3, pp. 223–233, 2007.