Modelling Daily Rain with Multisite Measures using Latent Gaussian Fields

P. Bulla, O. Cappé, E. Parent, J.M. Marin, C. Robert, J. Rousseau

Paris, 25/01/2006

The Problem

The Data:

1. 105 rainfall stations in the Seine basin;
2. daily observations during 27 years from 1975 to 2001 with many missing values:
   - only 14 stations are complete,
   - in 72 stations missing values $< 10\%$,
   - but in 13 they are $> 50\%$.

The Problem

The Data:

1. 105 rainfall stations in the Seine basin;
2. daily observations during 27 years from 1975 to 2001 with many missing values:
   - only 14 stations are complete,
   - in 72 stations missing values $< 10\%$,
   - but in 13 they are $> 50\%$.

The Problem

The Data:

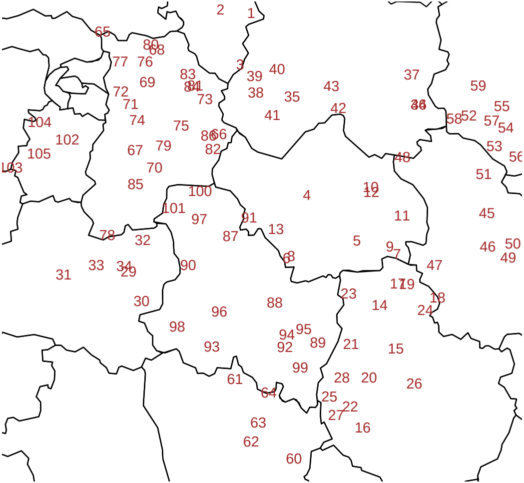
1. 105 rainfall stations in the Seine basin;
2. daily observations during 27 years from 1975 to 2001 with many missing values:
   - only 14 stations are complete,
   - in 72 stations missing values $< 10\%$,
   - but in 13 they are $> 50\%$.

The Problem

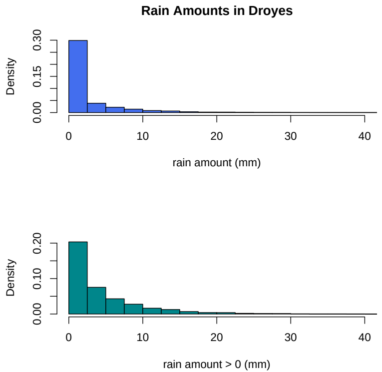
The Data:

1. 105 rainfall stations in the Seine basin;
2. daily observations during 27 years from 1975 to 2001 with many missing values:
   - only 14 stations are complete,
   - in 72 stations missing values $< 10\%$,
   - but in 13 they are $> 50\%$.

The Seine Basin



Some Histograms of Rain Amounts



The Aim

Our aim id to build a model for rainfall introducing spatial dependence between different stations.

Modelling daily rainfalls involves:

- description of the precipitation occurrences;
- specification of the distributions of nonzero rainfall amounts.

The Aim

Our aim id to build a model for rainfall introducing spatial dependence between different stations.

Modelling daily rainfalls involves:

- description of the precipitation occurrences;
- specification of the distributions of nonzero rainfall amounts.

The Aim

Our aim id to build a model for rainfall introducing spatial dependence between different stations.

Modelling daily rainfalls involves:

- description of the precipitation occurrences;
- specification of the distributions of nonzero rainfall amounts.

The Aim

Our aim id to build a model for rainfall introducing spatial dependence between different stations.

Modelling daily rainfalls involves:

- description of the precipitation occurrences;
- specification of the distributions of nonzero rainfall amounts.

Wilks' Local Precipitation Model

Precipitation occurrence process

At site k, a two state Markov chain $s_{tk},\ t \geq 0$ governs daily precipitation occurrence so that

$$s_{tk} = \begin{cases} 0 & \text{day } t \text{ is dry at } k \\ 1 & \text{day } t \text{ is wet at } k \end{cases} \tag{1}$$

The transition probabilities are stationary with respect to time.

$$P_t(k) = P(k) = \begin{bmatrix} p_0(k) & 1 - p_0(k) \\ p_1(k) & 1 - p_1(k) \end{bmatrix} \tag{2}$$

Wilks' Local Precipitation Model

Precipitation occurrence process

At site k, a two state Markov chain $s_{tk},\ t \geq 0$ governs daily precipitation occurrence so that

$$s_{tk} = \begin{cases} 0 & \text{day } t \text{ is dry at } k \\ 1 & \text{day } t \text{ is wet at } k \end{cases} \tag{1}$$

The transition probabilities are stationary with respect to time.

$$P_t(k) = P(k) = \begin{bmatrix} p_0(k) & 1 - p_0(k) \\ p_1(k) & 1 - p_1(k) \end{bmatrix} \tag{2}$$

Wilks' Local Precipitation Model

Precipitation amounts process

The time series of precipitation amounts at location k is

$$R_{tk} = r_{tk} s_{tk} \tag{3}$$

where r_{tk} represents the nonzero precipitation amounts and has density independent of t

$$f(r_k) = \frac{\alpha_k}{\beta_{k1}} \exp\left[-\frac{r_k}{\beta_{k1}}\right] + \frac{1-\alpha_k}{\beta_{k2}} \exp\left[-\frac{r_k}{\beta_{k2}}\right] \tag{4}$$

with $\beta_{k1} \geq \beta_{k2} > 0, \; 0 < \alpha_k \leq 1.$

Idea:

- Two components mixture exponential model: light and heavy rain, "continuity" property of the precipitation.

Wilks' Local Precipitation Model

Precipitation amounts process

The time series of precipitation amounts at location k is

$$R_{tk} = r_{tk} s_{tk} \tag{3}$$

where r_{tk} represents the nonzero precipitation amounts and has density independent of t

$$f(r_k) = \frac{\alpha_k}{\beta_{k1}} \exp\left[-\frac{r_k}{\beta_{k1}}\right] + \frac{1-\alpha_k}{\beta_{k2}} \exp\left[-\frac{r_k}{\beta_{k2}}\right] \tag{4}$$

with $\beta_{k1} \geq \beta_{k2} > 0,\ 0 < \alpha_k \leq 1.$

Idea:

- Two components mixture exponential model: light and heavy rain, "continuity" property of the precipitation.

Wilks' Local Precipitation Model

Precipitation amounts process

The time series of precipitation amounts at location k is

$$R_{tk} = r_{tk} s_{tk} \tag{3}$$

where r_{tk} represents the nonzero precipitation amounts and has density independent of t

$$f(r_k) = \frac{\alpha_k}{\beta_{k1}} \exp\left[-\frac{r_k}{\beta_{k1}}\right] + \frac{1-\alpha_k}{\beta_{k2}} \exp\left[-\frac{r_k}{\beta_{k2}}\right] \tag{4}$$

with $\beta_{k1} \geq \beta_{k2} > 0,\ 0 < \alpha_k \leq 1$.

Idea:

- Two components mixture exponential model: light and heavy rain, "continuity" property of the precipitation.

Multisite Precipitation Model

Consider K sites simultaneously, let $\boldsymbol{s}_t = (s_{t1}, \ldots, s_{tK})'$ and $\boldsymbol{r}_t = (r_{t1}, \ldots, r_{tK})'$, $\forall t$.

Given $\boldsymbol{s}_{t-1}$, to introduce spatial dependence between K sites, Wilks(1998) suggests to take

- $\boldsymbol{u}_t = (u_{t1}, \ldots, u_{tK})' \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Sigma}_u)$ ruling temporal state transition,
- $\boldsymbol{v}_t = (v_{t1}, \ldots, v_{tK})' \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Sigma}_v)$ ruling the amount of rain

where $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_v$ are two correlation matrices and $\boldsymbol{u}_t \perp \boldsymbol{v}_t \, \forall t$.

Multisite Precipitation Model

Consider K sites simultaneously, let $\boldsymbol{s}_t = (s_{t1}, \ldots, s_{tK})'$ and $\boldsymbol{r}_t = (r_{t1}, \ldots, r_{tK})'$, $\forall t$.

Given $\boldsymbol{s}_{t-1}$, to introduce spatial dependence between K sites, Wilks(1998) suggests to take

- $\boldsymbol{u}_t = (u_{t1}, \ldots, u_{tK})' \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ ruling temporal state transition,
- $\boldsymbol{v}_t = (v_{t1}, \ldots, v_{tK})' \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{\Sigma}_v)$ ruling the amount of rain

where $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_v$ are two correlation matrices and $\boldsymbol{u}_t \perp \boldsymbol{v}_t \, \forall t$.

Multisite Precipitation Model

Let Φ be the univariate normal df: $\forall t, k$, given $s_{t-1\,k} = i$, put

- for the precipitation occurrence process

$$s_{tk} = \begin{cases} 0 & \Phi(u_{tk}) \leq p_i(k) \\ 1 & \Phi(u_{tk}) > p_i(k) \end{cases} \tag{5}$$

- for the precipitation amount process

$$\beta_{s_{tk}}(k) = \begin{cases} \beta_{k1} & \frac{\Phi(u_{tk})}{1-p_i(k)} \leq \alpha_k \\ \beta_{k2} & \frac{\Phi(u_{tk})}{1-p_i(k)} > \alpha_k \end{cases} \tag{6}$$

$$r_{tk} = -\beta_{k\, s_{tk}} \ln[\Phi(v_{tk})] \tag{7}$$

Marginally, $\boldsymbol{s}_t$ and $\boldsymbol{r}_t$ have the same properties as before.

Multisite Precipitation Model

Let Φ be the univariate normal df: $\forall t, k$, given $s_{t-1\,k} = i$, put

- for the precipitation occurrence process

$$s_{tk} = \begin{cases} 0 & \Phi(u_{tk}) \leq p_i(k) \\ 1 & \Phi(u_{tk}) > p_i(k) \end{cases} \tag{5}$$

- for the precipitation amount process

$$\beta_{s_{tk}}(k) = \begin{cases} \beta_{k1} & \frac{\Phi(u_{tk})}{1-p_i(k)} \leq \alpha_k \\ \beta_{k2} & \frac{\Phi(u_{tk})}{1-p_i(k)} > \alpha_k \end{cases} \tag{6}$$

$$r_{tk} = -\beta_{k\,s_{tk}} \ln[\Phi(v_{tk})] \tag{7}$$

Marginally, $\boldsymbol{s}_t$ and $\boldsymbol{r}_t$ have the same properties as before.

Multisite Precipitation Model

Let Φ be the univariate normal df: $\forall t, k$, given $s_{t-1\,k} = i$, put

- for the precipitation occurrence process

$$s_{tk} = \begin{cases} 0 & \Phi(u_{tk}) \leq p_i(k) \\ 1 & \Phi(u_{tk}) > p_i(k) \end{cases} \tag{5}$$

- for the precipitation amount process

$$\beta_{s_{tk}}(k) = \begin{cases} \beta_{k1} & \frac{\Phi(u_{tk})}{1-p_i(k)} \leq \alpha_k \\ \beta_{k2} & \frac{\Phi(u_{tk})}{1-p_i(k)} > \alpha_k \end{cases} \tag{6}$$

$$r_{tk} = -\beta_{k\,s_{tk}} \ln[\Phi(v_{tk})] \tag{7}$$

Marginally, $\boldsymbol{s}_t$ and $\boldsymbol{r}_t$ have the same properties as before.

Multisite Precipitation Model

Let Φ be the univariate normal df: $\forall t, k$, given $s_{t-1\,k} = i$, put

- for the precipitation occurrence process

$$s_{tk} = \begin{cases} 0 & \Phi(u_{tk}) \leq p_i(k) \\ 1 & \Phi(u_{tk}) > p_i(k) \end{cases} \tag{5}$$

- for the precipitation amount process

$$\beta_{s_{tk}}(k) = \begin{cases} \beta_{k1} & \frac{\Phi(u_{tk})}{1-p_i(k)} \leq \alpha_k \\ \beta_{k2} & \frac{\Phi(u_{tk})}{1-p_i(k)} > \alpha_k \end{cases} \tag{6}$$

$$r_{tk} = -\beta_{k\,s_{tk}} \ln[\Phi(v_{tk})] \tag{7}$$

Marginally, $\boldsymbol{s}_t$ and $\boldsymbol{r}_t$ have the same properties as before.

A Simpler Model

Let specify and simplify the Wilks' model.

- For each t,

$$\boldsymbol{u}_t \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{\Lambda}), \qquad \Lambda_{ij} = e^{-\lambda d_{ij}} \tag{8}$$

$$\boldsymbol{v}_t \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{\Gamma}), \qquad \Gamma_{ij} = e^{-\gamma d_{ij}} \tag{9}$$

  where $D = \{d_{ij}, i, j = 1, \ldots, K\}$ is a distance matrix;
- substitute the mixture with a simple exponential distribution;
- for each site k, we have the same transition matrix

$$P(k) = \begin{bmatrix} p_0 & 1 - p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \tag{10}$$

- and the same average rainfall amount

$$\beta(k) = \beta. \tag{11}$$

A Simpler Model

Let specify and simplify the Wilks' model.

- For each t,

$$\boldsymbol{u}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Lambda}), \qquad \Lambda_{ij} = e^{-\lambda d_{ij}} \tag{8}$$

$$\boldsymbol{v}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Gamma}), \qquad \Gamma_{ij} = e^{-\gamma d_{ij}} \tag{9}$$

  where $D = \{d_{ij}, i, j = 1, \ldots, K\}$ is a distance matrix;
- substitute the mixture with a simple exponential distribution;
- for each site k, we have the same transition matrix

$$P(k) = \begin{bmatrix} p_0 & 1 - p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \tag{10}$$

- and the same average rainfall amount

$$\beta(k) = \beta. \tag{11}$$

A Simpler Model

Let specify and simplify the Wilks' model.

- For each t,

$$\boldsymbol{u}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Lambda}), \qquad \Lambda_{ij} = e^{-\lambda d_{ij}} \tag{8}$$

$$\boldsymbol{v}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Gamma}), \qquad \Gamma_{ij} = e^{-\gamma d_{ij}} \tag{9}$$

  where $D = \{d_{ij}, i, j = 1, \ldots, K\}$ is a distance matrix;
- substitute the mixture with a simple exponential distribution;
- for each site k, we have the same transition matrix

$$P(k) = \begin{bmatrix} p_0 & 1 - p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \tag{10}$$

- and the same average rainfall amount

$$\beta(k) = \beta. \tag{11}$$

A Simpler Model

Let specify and simplify the Wilks' model.

- For each t,

$$\boldsymbol{u}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Lambda}), \qquad \Lambda_{ij} = e^{-\lambda d_{ij}} \tag{8}$$

$$\boldsymbol{v}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Gamma}), \qquad \Gamma_{ij} = e^{-\gamma d_{ij}} \tag{9}$$

  where $D = \{d_{ij}, i, j = 1, \dots, K\}$ is a distance matrix;
- substitute the mixture with a simple exponential distribution;
- for each site k, we have the same transition matrix

$$P(k) = \begin{bmatrix} p_0 & 1 - p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \tag{10}$$

- and the same average rainfall amount

$$\beta(k) = \beta. \tag{11}$$

A Simpler Model

Let specify and simplify the Wilks' model.

- For each t,

$$\boldsymbol{u}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Lambda}), \qquad \Lambda_{ij} = e^{-\lambda d_{ij}} \tag{8}$$

$$\boldsymbol{v}_t \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Gamma}), \qquad \Gamma_{ij} = e^{-\gamma d_{ij}} \tag{9}$$

  where $D = \{d_{ij}, i, j = 1, \ldots, K\}$ is a distance matrix;
- substitute the mixture with a simple exponential distribution;
- for each site k, we have the same transition matrix

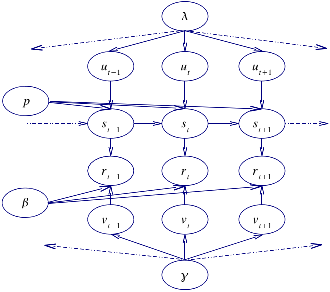
$$P(k) = \begin{bmatrix} p_0 & 1 - p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \tag{10}$$

- and the same average rainfall amount

$$\beta(k) = \beta. \tag{11}$$

Bayesian Analysis

▸ The DAG



- Parameters:

$$\lambda, \gamma, \beta, \mathbf{p} = (p_0, p_1)$$

- Latent variables:

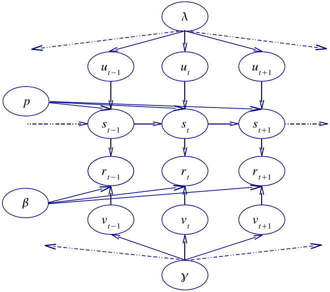
$$\mathbf{u}_t, \mathbf{v}_t,$$

- Observations:

$$\mathbf{r}_t, \mathbf{s}_t$$

Bayesian Analysis

The DAG



- Parameters:

$$\lambda, \gamma, \beta, \mathbf{p} = (p_0, p_1)$$

- Latent variables:

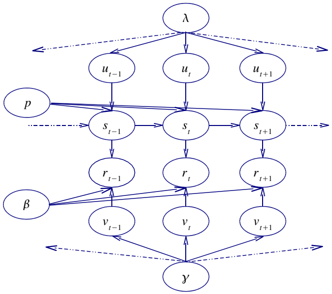
$$\mathbf{u}_t, \mathbf{v}_t,$$

- Observations:

$$\mathbf{r}_t, \mathbf{s}_t$$

Bayesian Analysis

The DAG



- Parameters:

$$\lambda, \gamma, \beta, \boldsymbol{p} = (p_0, p_1)$$

- Latent variables:

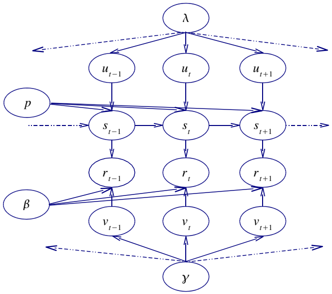
$$\boldsymbol{u}_t, \boldsymbol{v}_t,$$

- Observations:

$$\boldsymbol{r}_t, \boldsymbol{s}_t$$

Bayesian Analysis

The DAG



- Parameters:

$$\lambda, \gamma, \beta, \boldsymbol{p} = (p_0, p_1)$$

- Latent variables:

$$\boldsymbol{u}_t, \boldsymbol{v}_t,$$

- Observations:

$$\boldsymbol{r}_t, \boldsymbol{s}_t$$

Bayesian Analysis

Prior distributions:
Use noninformative priors for

$$[\lambda] \propto \frac{1}{\lambda} \tag{12}$$

$$[\gamma] \propto \frac{1}{\gamma} \tag{13}$$

$$[p_s] = \mathcal{U}\left[0, 1\right], \quad s = 0, 1 \tag{14}$$

$$P\left[\, s_{0k} = 1 \,|\, \theta\right] = \theta \quad \forall k \tag{15}$$

$$[\theta] = \mathcal{U}[0, 1] \tag{16}$$

$$\tag{17}$$

and

$$[\beta] = \mathcal{U}\left[\underline{\beta}, \overline{\beta}\right] \tag{18}$$

Bayesian Analysis

Prior distributions:
Use noninformative priors for

$$[\lambda] \propto \frac{1}{\lambda} \tag{12}$$
$$[\gamma] \propto \frac{1}{\gamma} \tag{13}$$
$$[p_s] = \mathcal{U}\,[0,1]\,, \quad s = 0,1 \tag{14}$$
$$P\,[\,s_{0k} = 1|\,\theta] = \theta \quad \forall k \tag{15}$$
$$[\theta] = \mathcal{U}[0,1] \tag{16}$$
$$\tag{17}$$

and

$$[\beta] = \mathcal{U}\left[\underline{\beta}, \overline{\beta}\right] \tag{18}$$

Bayesian Analysis

Transforming each component of $\boldsymbol{v}_t$ according to

$$r_{tk} = \begin{cases} -\beta \log(\Phi(v_{tk})) & s_{tk} = 1 \\ 0 & s_{tk} = 0 \end{cases} \qquad (19)$$

we get the conditional distribution of

$$[\boldsymbol{r}_t | \boldsymbol{s}_t(\boldsymbol{u}_t, \boldsymbol{r}_{t-1}), \boldsymbol{p}, \beta, \gamma] \propto \frac{1}{|\Gamma|^{1/2}} e^{-\frac{1}{2}\boldsymbol{v}_t' \Gamma^{-1} \boldsymbol{v}_t} \prod_{k:s_{tk}>0} e^{\frac{1}{2}\Phi^{-1}\left(e^{-\frac{r_{tk}}{\beta}}\right)^2} \frac{1}{\beta} e^{-\frac{r_{tk}}{\beta}} \qquad (20)$$

Bayesian Analysis

Inference can be performed via MCMC methods.
Through the latent variables we can complete the dataset simulating the missing values given the available as follows

$$\left[s^*_{tk} \middle| \boldsymbol{s}_{t-1}, \boldsymbol{s}_{t\,(-k)}, \boldsymbol{s}_{t+1}, \boldsymbol{u}_{t\,(-k)}, \boldsymbol{u}_{t+1}, \boldsymbol{p}, \Lambda \right] \propto \\ \left[s^*_{tk} \middle| \boldsymbol{s}_{t-1}, \boldsymbol{u}_{t\,(-k)}, \boldsymbol{p}, \Lambda \right] \\ \left[s_{t+1\,k} \middle| s^*_{tk}, \boldsymbol{s}_{t\,(-k)}, \boldsymbol{u}_{t+1\,(-k)}, \boldsymbol{p}, \Lambda \right] \\ \left[u_{t+1\,k} \middle| s_{t+1\,k}, s^*_{tk}, \boldsymbol{u}_{t+1\,(-k)}, \boldsymbol{p}, \Lambda \right] \tag{21}$$

$$P\left[s^*_{tk} = 0 \middle| \boldsymbol{s}_{t-1}, \boldsymbol{u}_{t\,(-k)}, \boldsymbol{p}, \Lambda \right] = \\ = \int_{-\infty}^{\Phi^{-1}(p_{s_{t-1\,k}})} \phi(u; \boldsymbol{\lambda}'_{k\,(-k)} \Lambda^{-1}_{-k} \boldsymbol{u}_{t\,(-k)}, 1 - \boldsymbol{\lambda}'_{k\,(-k)} \Lambda^{-1}_{-k} \boldsymbol{\lambda}_{k\,(-k)}) du \tag{22}$$

Bayesian Analysis

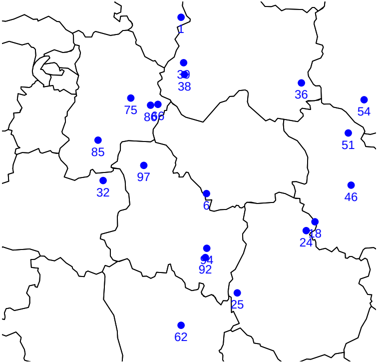
$$[u_{tk} | \ldots] \propto \mathcal{N}\left(\boldsymbol{\lambda}'_{k\,(-k)} \Lambda_{-k}^{-1} \boldsymbol{u}_{t\,(-k)}, \lambda_{kk} - \boldsymbol{\lambda}'_{k\,(-k)} \Lambda_{-k}^{-1} \boldsymbol{\lambda}_{k\,(-k)}\right) \mathbb{1}_{\left\{(-1)^{s_{tk}}\left(\Phi(u_{tk}) - p_{s_{t-1\,k}}\right) < 0\right\}} \tag{23}$$

$$[\boldsymbol{v}_t^* | \boldsymbol{v}_t^a, \ldots] = \mathcal{MN}\left(\Gamma_{*a}\Gamma_a^{-1}\boldsymbol{v}_t^a,\ \Gamma_* - \Gamma_{*a}\Gamma_a^{-1}\Gamma_{a*}\right) \tag{24}$$

and then compute $\boldsymbol{r}_t^*$.

An example

We considered twenty stations during the 27 months of April ('75-'01).



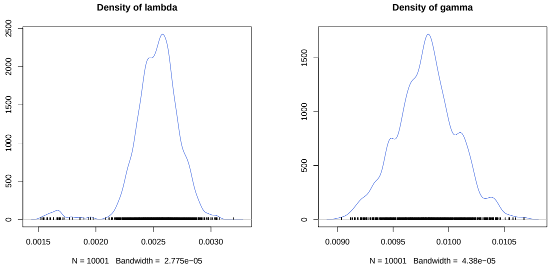
Estimate Results

	Mean	SD	Naive SE	Time-series SE
λ	0.002525	0.0002060	2.060e-06	2.890e-05
γ	0.009804	0.0002729	2.729e-06	2.852e-05
β	3.928176	0.1022724	1.023e-03	8.009e-03
p_0	0.653934	0.0061804	6.180e-05	6.497e-04
p_1	0.527557	0.0027752	2.775e-05	3.605e-04

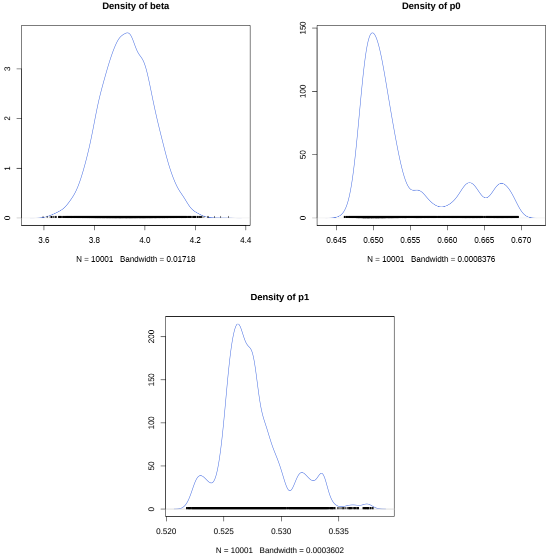
	2.5%	25%	50%	75%	97.5%
λ	0.002116	0.002424	0.002542	0.002645	0.002863
γ	0.009247	0.009622	0.009801	0.009971	0.010380
β	3.731738	3.856482	3.927303	3.998626	4.130867
p_0	0.647833	0.649566	0.651281	0.656248	0.668298
p_1	0.522774	0.525827	0.527030	0.528700	0.533796

An example

The densities



An example


Density of beta
0
1
2
3
3.6
3.8
4.0
4.2
4.4
N = 10001 Bandwidth = 0.01718
Density of p0
0
50
100
150
0.645
0.650
0.655
0.660
0.665
0.670
N = 10001 Bandwidth = 0.0008376
Density of p1
0
50
100
150
200
0.520
0.525
0.530
0.535
N = 10001 Bandwidth = 0.0003602

Perspectives

- Better and complete implementation of the Wilks' model according to hydrological knowledge:
  - parameters specific to each different station?
  - Is an exponential mixture necessary?
  - Is there an *a posteriori* dependence between the $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ and is not sufficient to use just one latent field ruling both occurrences and rainfall amounts?

Perspectives

- Better and complete implementation of the Wilks' model according to hydrological knowledge:
  - parameters specific to each different station?
  - Is an exponential mixture necessary?
  - Is there an *a posteriori* dependence between the $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ and is not sufficient to use just one latent field ruling both occurrences and rainfall amounts?

Perspectives

- Better and complete implementation of the Wilks' model according to hydrological knowledge:
  - parameters specific to each different station?
  - Is an exponential mixture necessary?
  - Is there an *a posteriori* dependence between the $\mathbf{u}_t$ and $\mathbf{v}_t$ and is not sufficient to use just one latent field ruling both occurrences and rainfall amounts?

Perspectives

- Better and complete implementation of the Wilks' model according to hydrological knowledge:
  - parameters specific to each different station?
  - Is an exponential mixture necessary?
  - Is there an *a posteriori* dependence between the $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ and is not sufficient to use just one latent field ruling both occurrences and rainfall amounts?

Perspectives

- Embedding the model in a generic decision problem aiming the optimal shrinkage of the network of the rainfall stations, possibly integrating information about the flows of the rivers in the basin.

  Let $\theta = (\lambda, \gamma, \beta, \boldsymbol{p})$ the vector of paramters of the model and let indicate the joint distribution of rainfalls and parameters as

$$\pi\left(R, \theta\right) = \pi\left(\left. R \right| \theta\right) \pi\left(\theta\right) \tag{25}$$

Perspectives

Let $d = (d_1, \ldots, d_n)$ be a vector s.t. $d_i = 0$ means the station i is removed from the network, $d_i = 1$ the station remains and $D = \{i : d_i = 1\}$.
The aim is making prediction about the streamflow of some river in the basin

$$q_t = \sum_k \omega_k \sum_j \rho_k^j r_{t-t_k^0-j\,k} \tag{26}$$

The loss function considered is

$$L\left(q_t, \hat{q}_t^d\right) = a\left(q_t - \hat{q}_t^d\right)^{\alpha} \mathbb{1}_{\{q_t > \hat{q}_t^d\}} + b\left(\hat{q}_t^d - q_t\right)^{\beta} \mathbb{1}_{\{q_t < \hat{q}_t^d\}} + \sum_{i \in D} c_i \tag{27}$$

$$L\left(\boldsymbol{q}, \hat{\boldsymbol{q}}^d\right) = \sum_t L\left(q_t, \hat{q}_t^d\right) \tag{28}$$

Perspectives

We are looking for an optimal decision d^* such that

$$d^* = \arg\min_d W(d) \tag{29}$$

$$\begin{aligned} W(d) &= E_\theta E_{R|\theta}\left[L\left(\boldsymbol{q}, \hat{\boldsymbol{q}}^d\right)\right] \\ &= \int L\left(\boldsymbol{q}, \hat{\boldsymbol{q}}^d\right) d\pi\left(R|\,\theta\right) d\pi\left(\theta\right) \end{aligned} \tag{30}$$

The particle algirithm developed in Amzal et al. (2003) can be employed.

Bibliography

- Amzal, B., Bois, F., Parent, E. and Robert, C. (2003), Bayesian optimal design via interacting MCMC. Technical Report, *Les Cahiers du CEREMADE - Université Paris Dauphine*
- Wilks, D.S. (1998), Multisite generalization of a daily stochastic precipitation generation model, *Journal of Hydrology 210(1998), 178-191.*

The DAG