Spatial Kernel Interpolation for Annual Rainfall Maxima

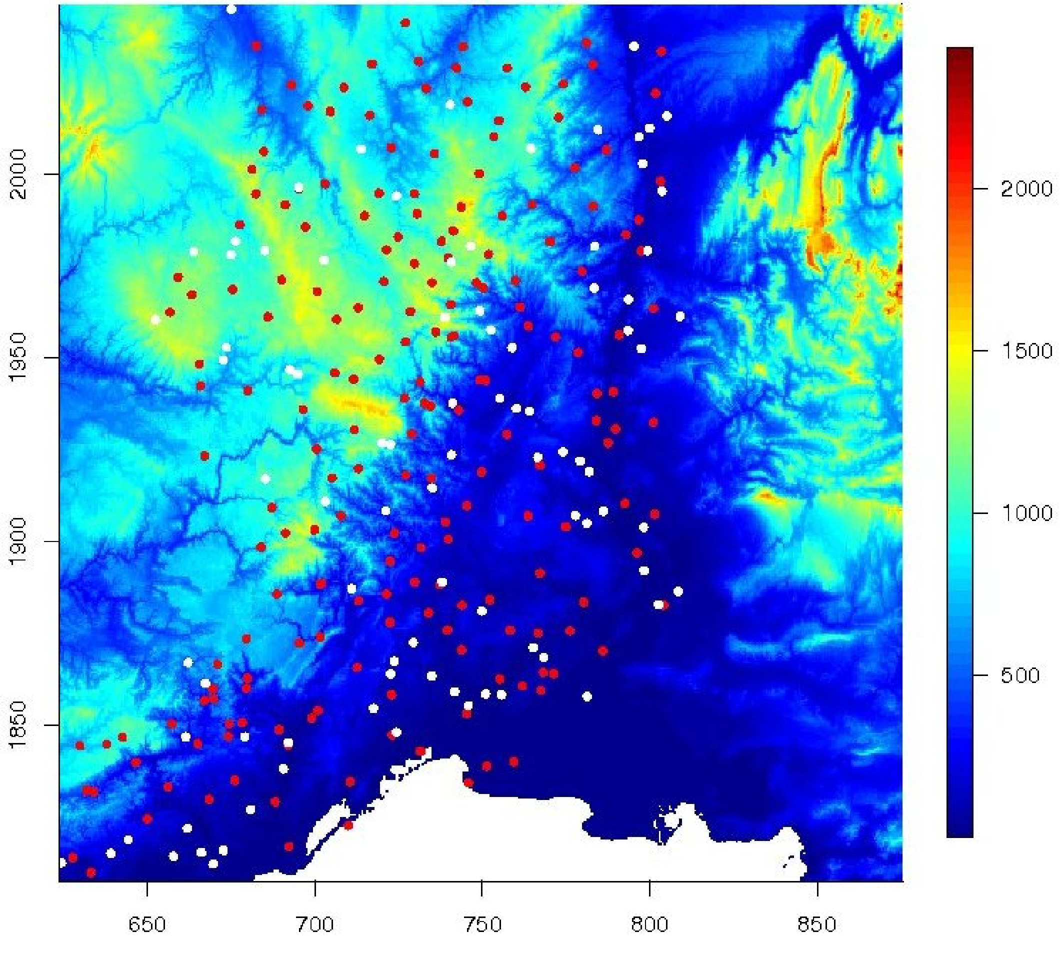
Julie Carreau & **Stephane Girard** & **Eugen Ursu**

julie.carreau@inrialpes.fr stephane.girard@inrialpes.fr eugen.ursu@inrialpes.fr

INRIA Rhone-Alpes, FRANCE

Cevennes-Vivarais Precipitation Data

Analyse spatially the impact of heavy rainfall



- daily rainfall measurements
- 43 annual maxima
- 198 (92) stations

Modelling Maxima

Generalized Extreme Value Distribution (GEV)

$$F(y;\mu,\sigma,\xi) = \begin{cases} \exp\{-(1+\frac{\xi}{\sigma}(y-\mu))^{-1/\xi}\} & \text{if } \xi > 0 \\ \exp\{-\exp\{-\left(\frac{y-\mu}{\sigma}\right)\}\} & \text{if } \xi = 0 \end{cases}$$

Modelling Maxima

Generalized Extreme Value Distribution (GEV)

$$F(y;\mu,\sigma,\xi) = \begin{cases} \exp\{-(1+\frac{\xi}{\sigma}(y-\mu))^{-1/\xi}\} & \text{if } \xi > 0 \\ \exp\{-\exp\{-\left(\frac{y-\mu}{\sigma}\right)\}\} & \text{if } \xi = 0 \end{cases}$$

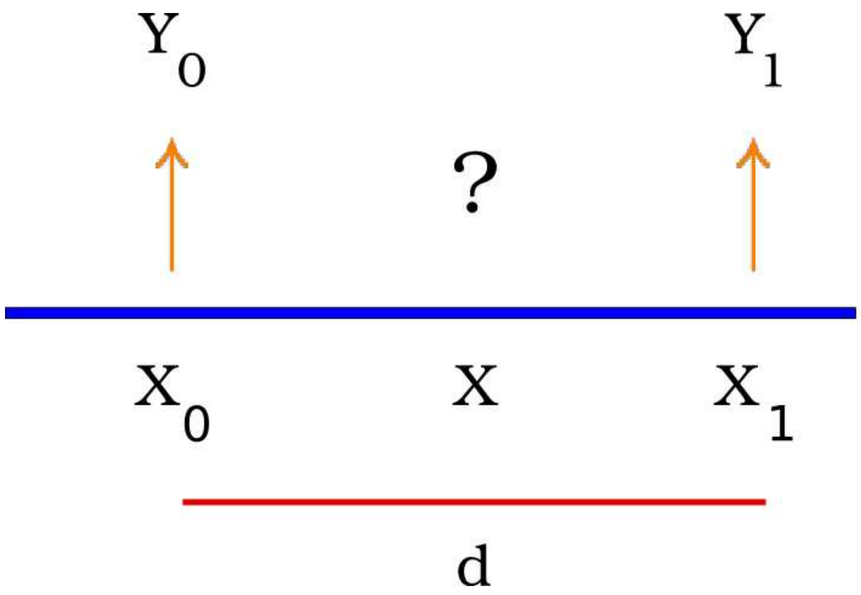
Estimate the GEV parameters at a new station X

$$l(\mu,\sigma,\xi) = -\sum_{i=1}^{n} w_i \log\left(f(Y_i;\mu,\sigma,\xi)\right)$$

Y_i are observations at neighboring stations X_{s_i}

$w_i = K\left(\frac{||X-X_{s_i}||}{h}\right)$ where $K(\cdot)$ is a kernel function, h is the bandwidth

1D Interpolation Study



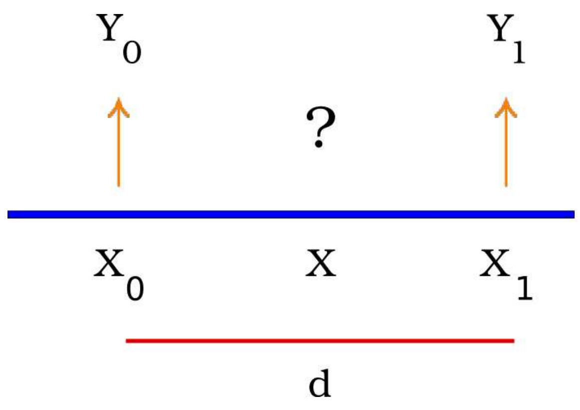
Observations

Stations

$$Y_0 \sim GEV(\mu_0, \sigma_0, \xi)$$

$$Y_1 \sim GEV(\mu_0 + d, \sigma_0 + d/20, \xi)$$

1D Interpolation Study



Observations

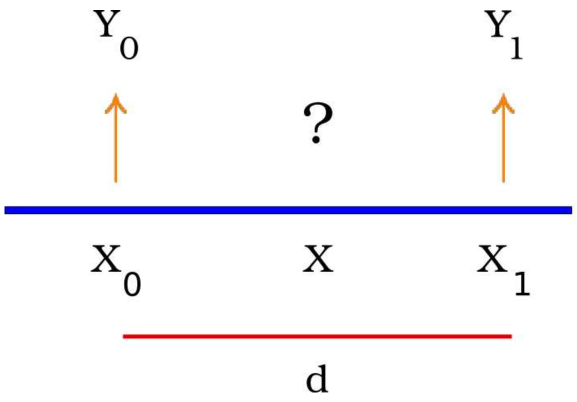
Stations

$$Y_0 \sim GEV(\mu_0, \sigma_0, \xi)$$

$$Y_1 \sim GEV(\mu_0 + d, \sigma_0 + d/20, \xi)$$

$$l(\mu, \sigma, \xi) = (1-\lambda)\sum_{i=1}^{n} \log f(y_i^{(0)}; \mu, \sigma, \xi) + \lambda \sum_{j=1}^{n} \log f(y_j^{(1)}; \mu, \sigma, \xi)$$

1D Interpolation Study



Observations

Stations

$$Y_0 \sim GEV(\mu_0, \sigma_0, \xi)$$

$$Y_1 \sim GEV(\mu_0 + d, \sigma_0 + d/20, \xi)$$

$$l(\mu,\sigma,\xi) = (1-\lambda)\sum_{i=1}^{n} \log f(y_i^{(0)};\mu,\sigma,\xi) + \lambda \sum_{j=1}^{n} \log f(y_j^{(1)};\mu,\sigma,\xi)$$

How to choose λ?

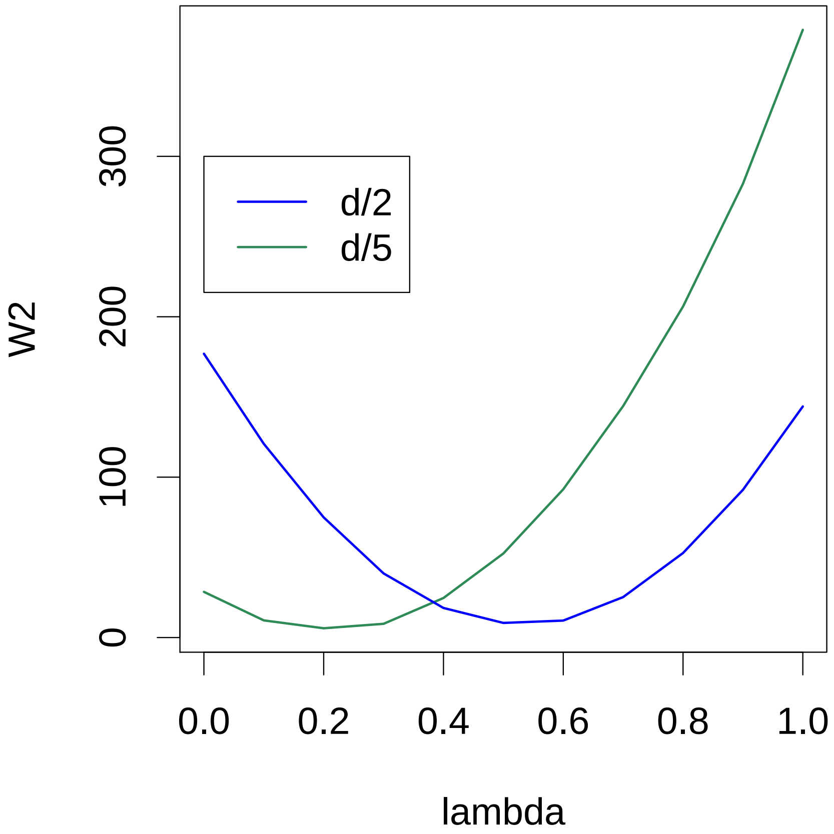
How does the optimal λ varies with d and with X?

Goodness-of-Fit

- Estimate $\mu^i_\lambda, \sigma^i_\lambda, \xi^i_\lambda,\ \lambda \in [0,1]$ and $i = 1, \ldots, 1000$
- Evaluate the fit with the **Cramer-Von Mises statistic at site X**

Goodness-of-Fit

- Estimate $\mu^i_\lambda, \sigma^i_\lambda, \xi^i_\lambda$, $\lambda \in [0,1]$ and $i = 1, \ldots, 1000$
- Evaluate the fit with the **Cramer-Von Mises statistic at site X**

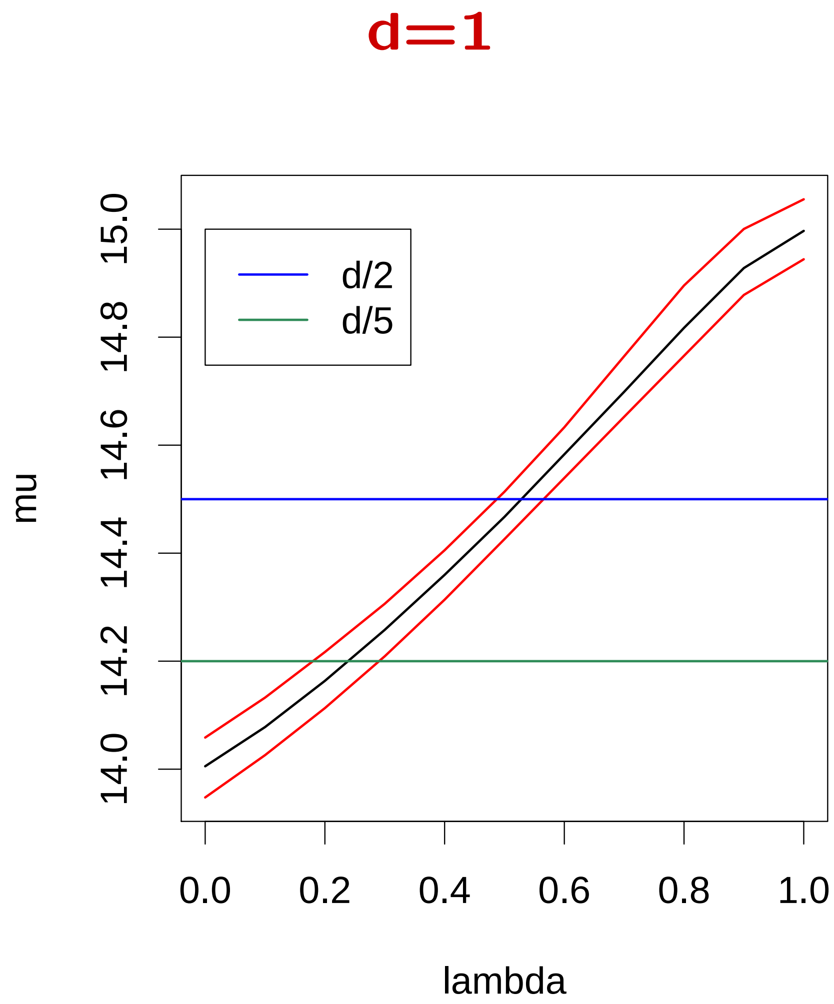


d=1

$$X = d/2 \Longrightarrow \lambda_{\text{opt}} = 1/2$$

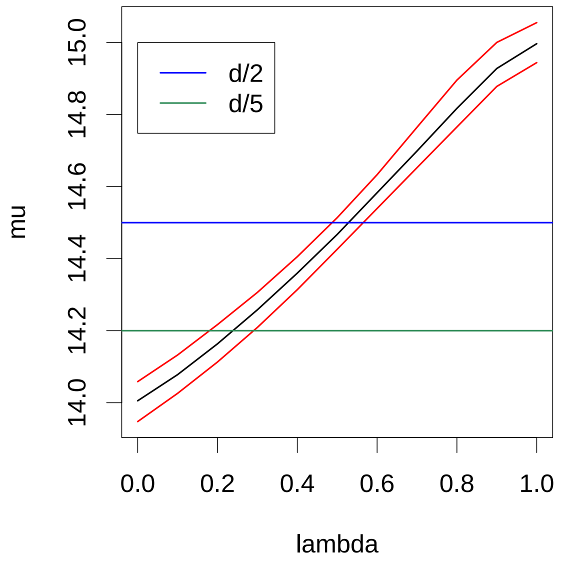
$$X = d/5 \Longrightarrow \lambda_{\text{opt}} = 1/5$$

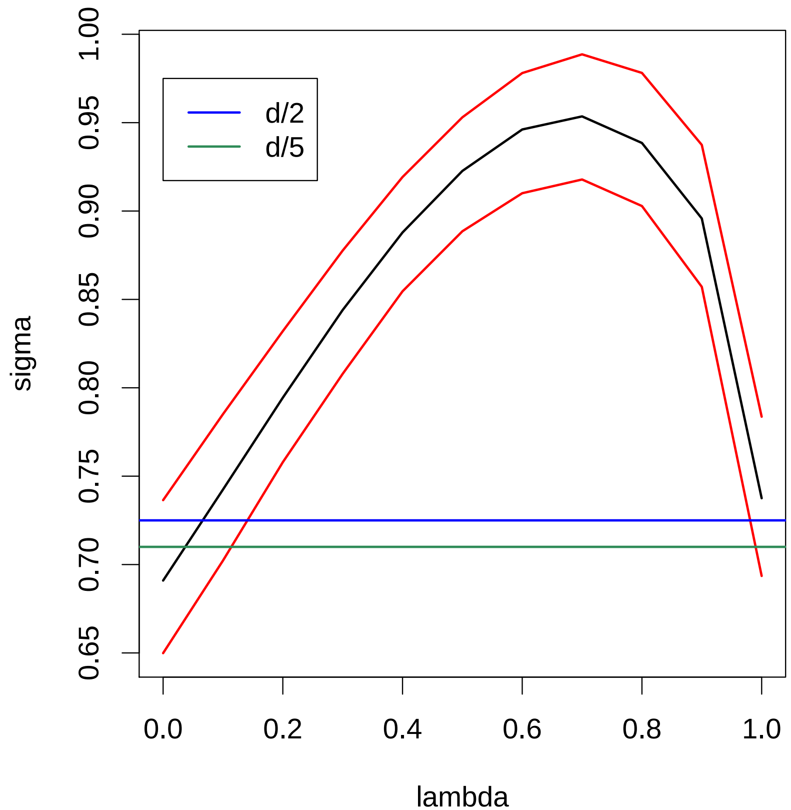
GEV parameters



GEV parameters

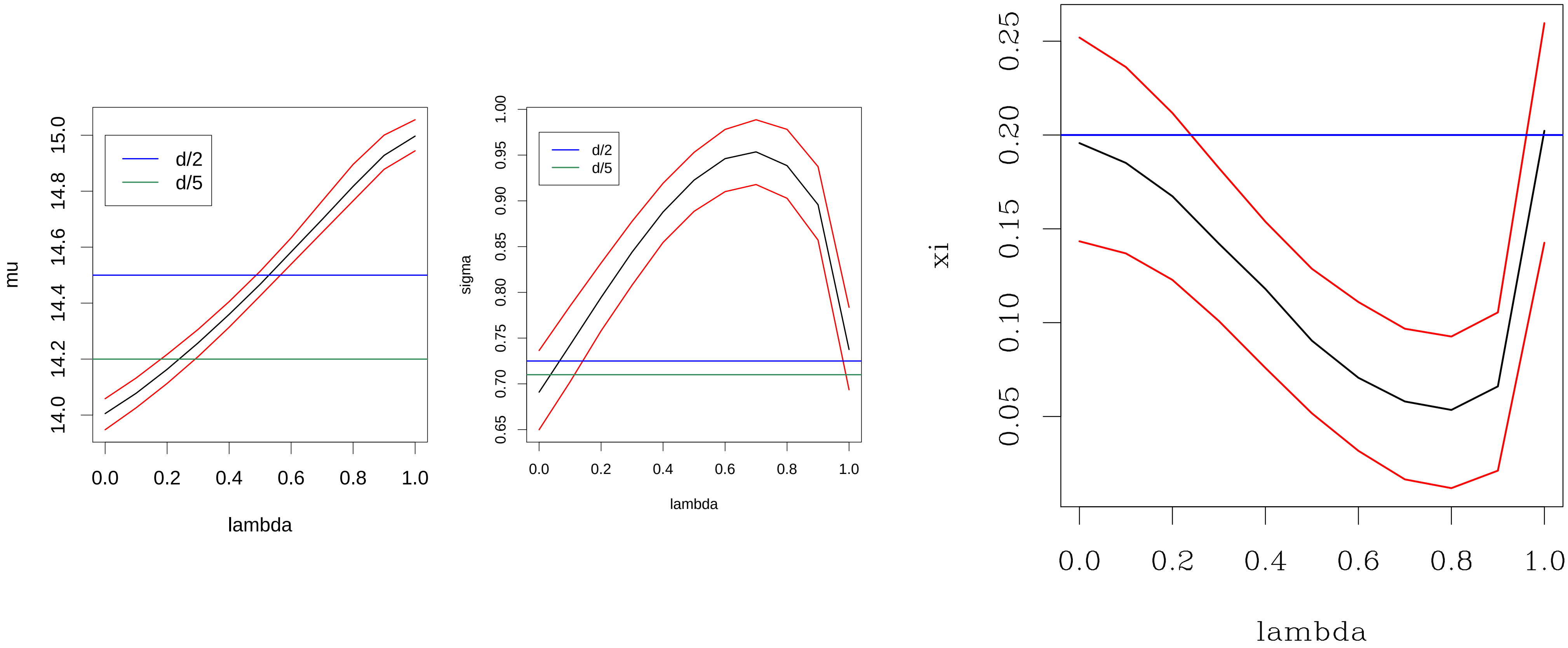
d=1



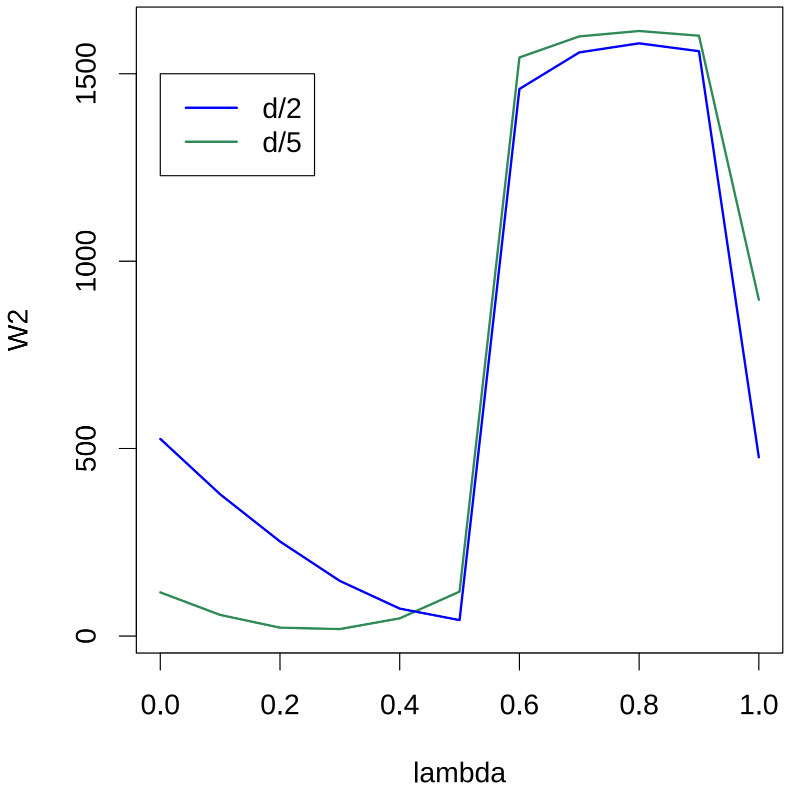


GEV parameters

d=1



Goodness-of-Fit



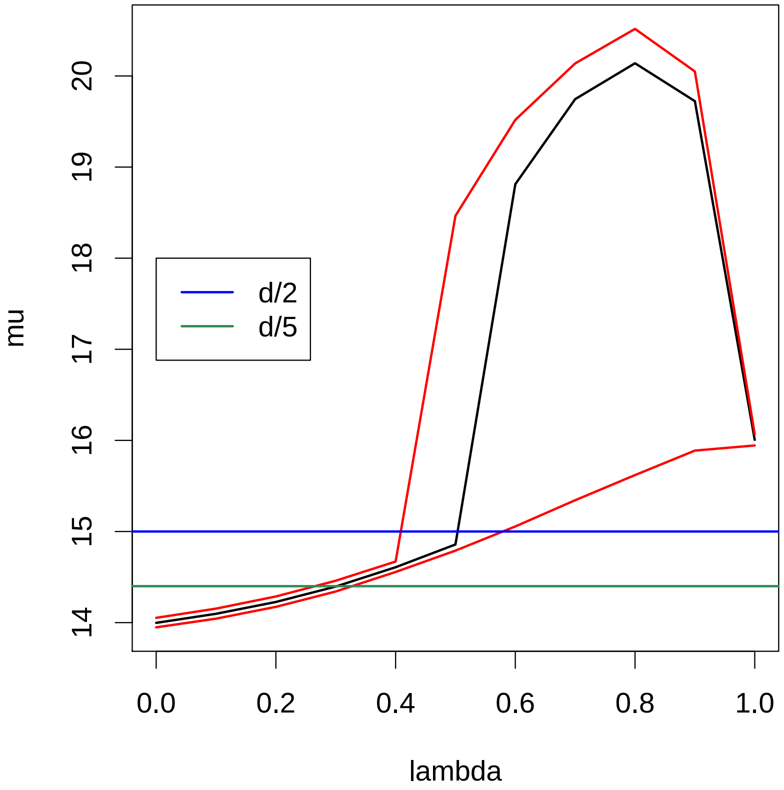
d=2

$$X = d/2 \Longrightarrow \lambda_{\text{opt}} = 1/2$$

$$X = d/5 \Longrightarrow \lambda_{\text{opt}} = 1/5$$

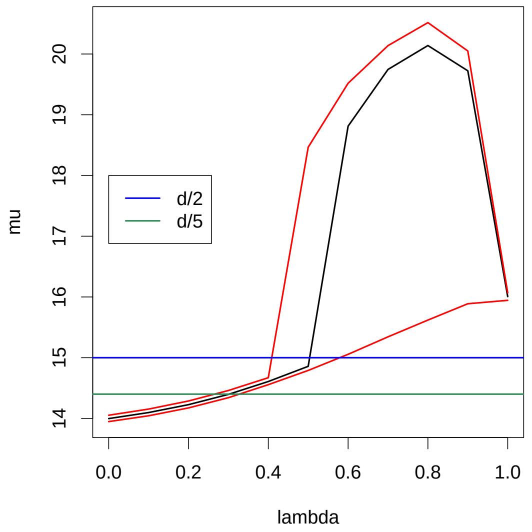
GEV parameters

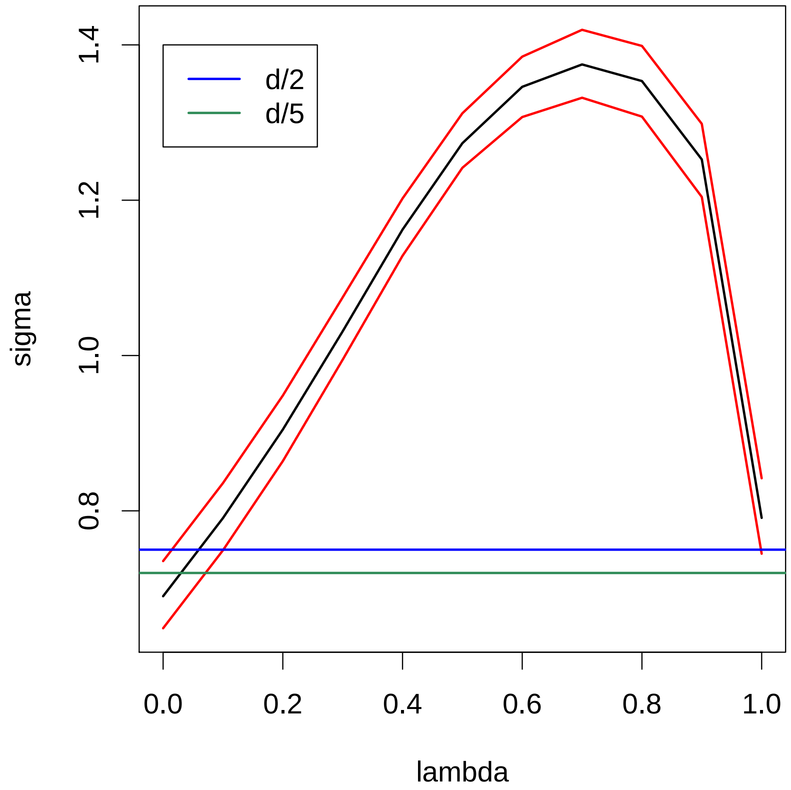
d=2



GEV parameters

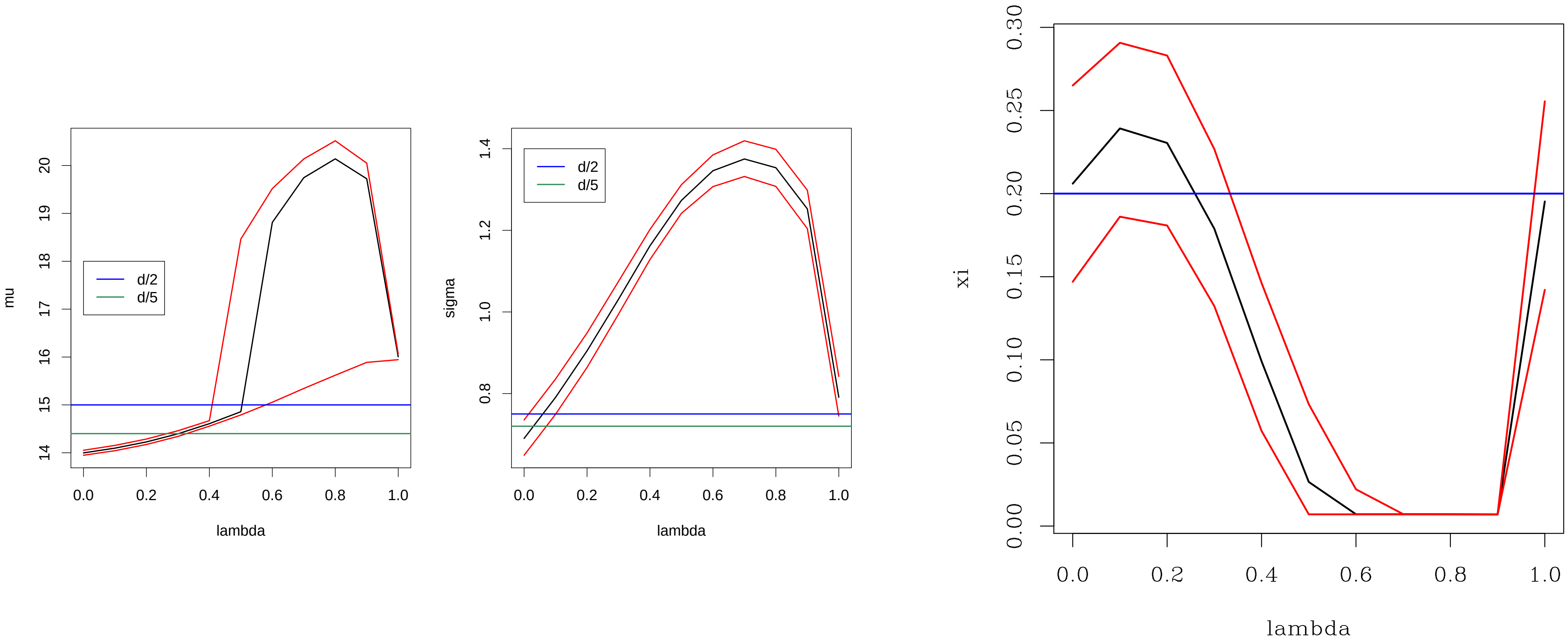
d=2





GEV parameters

d=2



Distance in Distributions

✷ Sensitivity of the Kernel GEV estimates to the notion of neighbourhood

✷ Define the neighbourhood in terms of closeness in distributions

Distance in Distributions

- Sensitivity of the Kernel GEV estimates to the notion of neighbourhood

- Define the neighbourhood in terms of closeness in distributions

- **Kolmogorov-Smirnov statistic (KS)** as a measure of dissimilarity

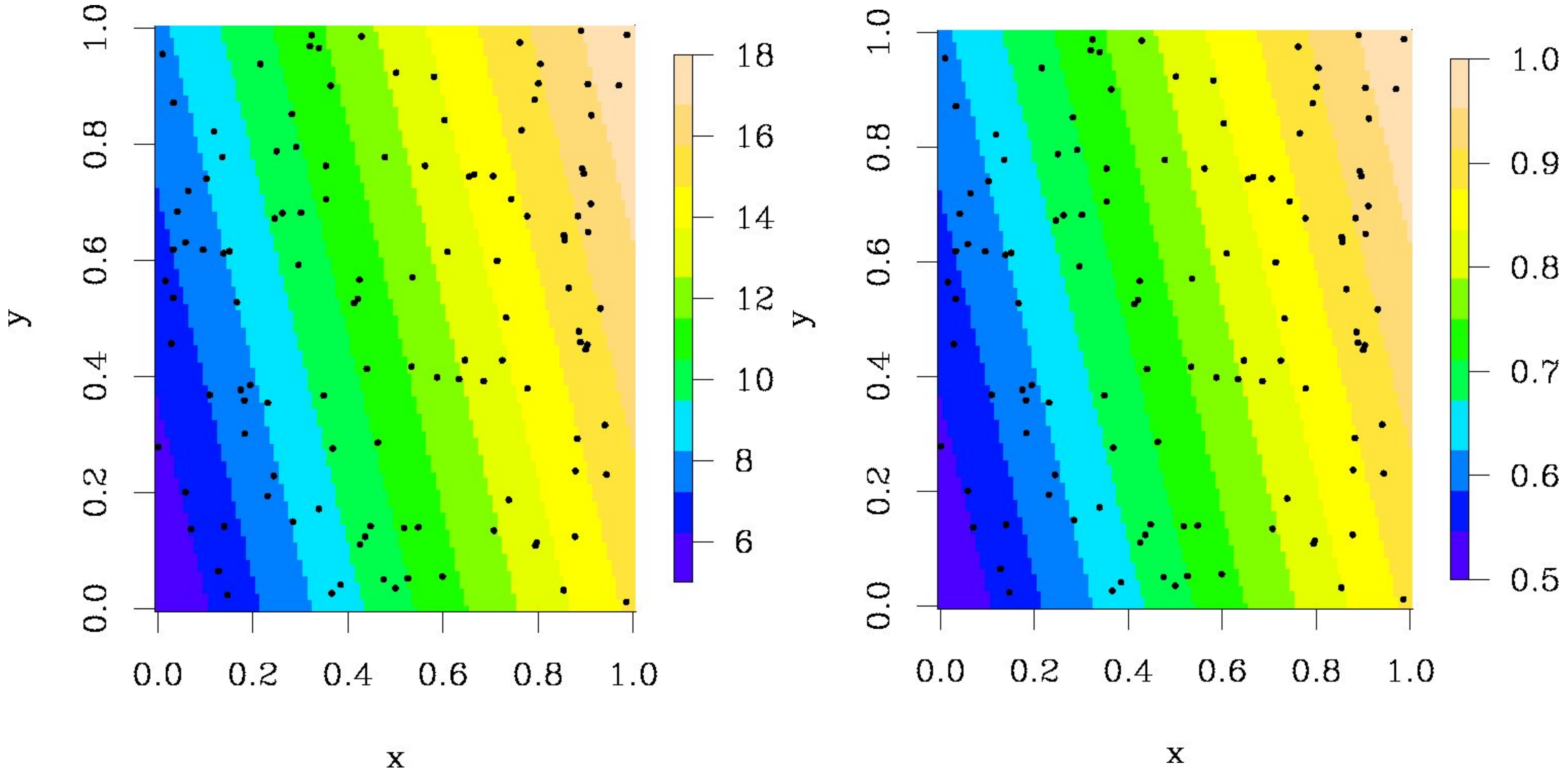
- **Multidimensional Scaling (MDS)** to build an embedding space

- **Map :** Euclidean space to Embedding space

2D Uniform Data

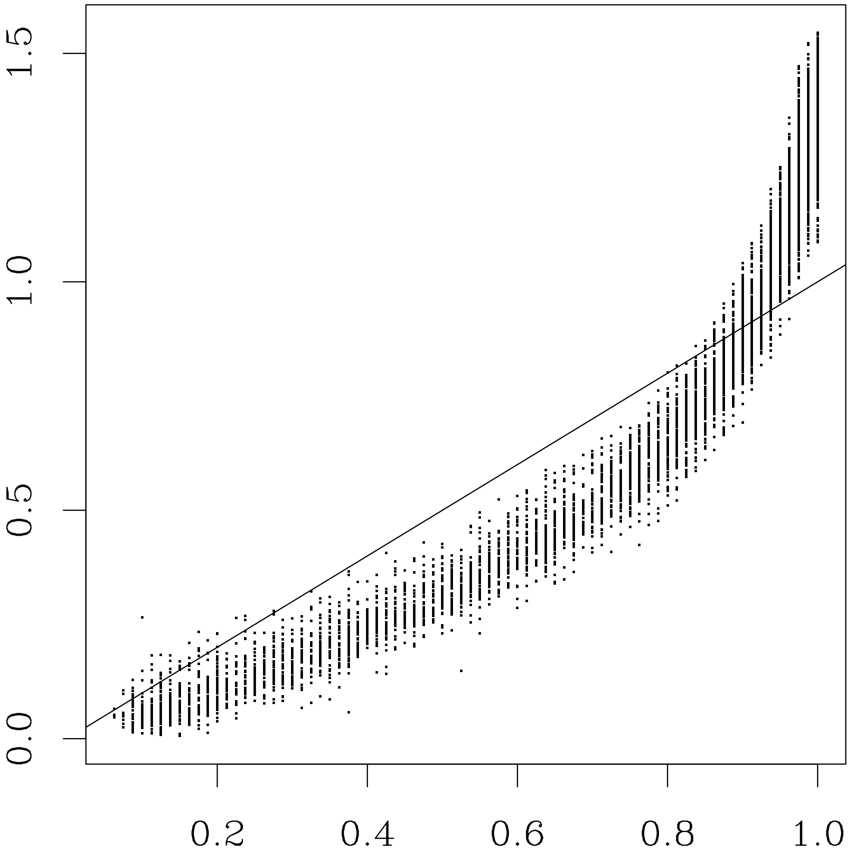
120 stations drawn uniformly across the unit square

80 maxima sampled at each station from a GEV with $\xi = 0.2$

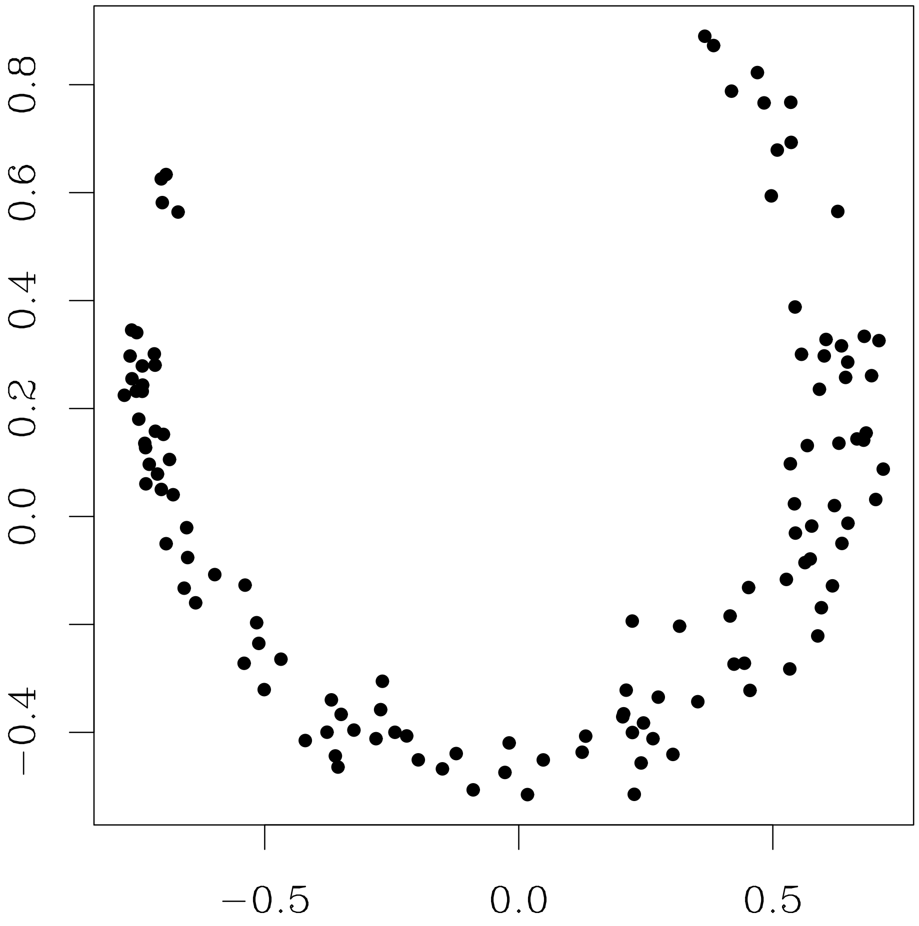


MDS Embedding

MDS distances vs KS dissimilarities

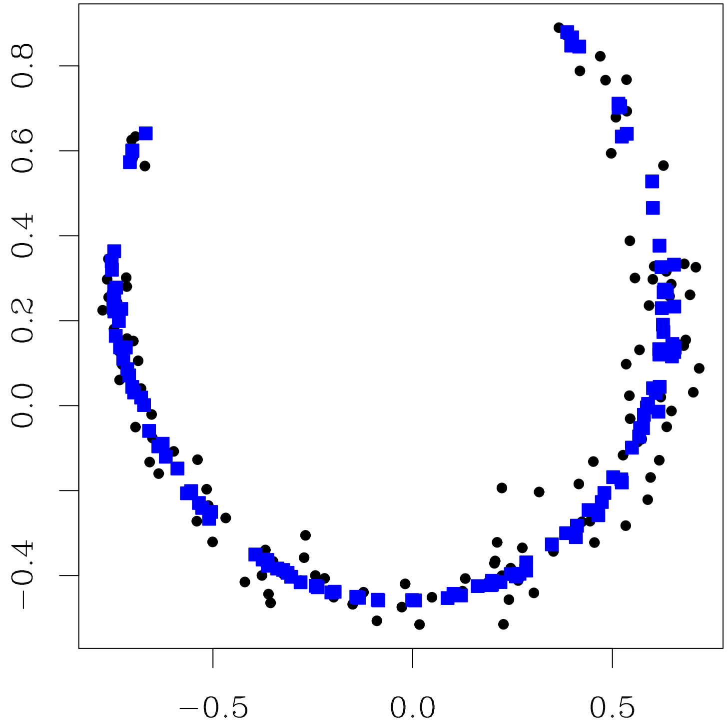


Embedding



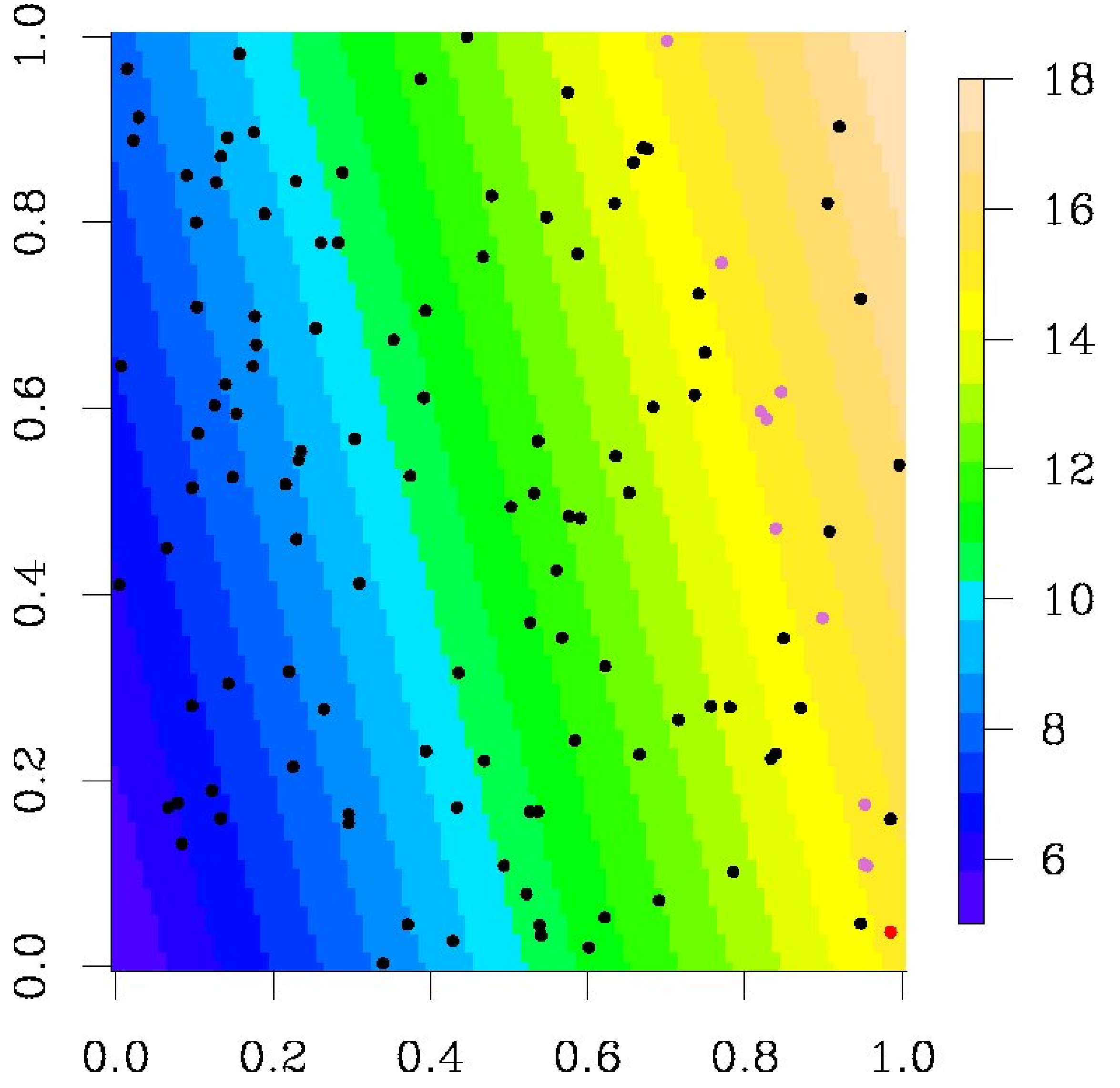
MDS Mapping

Neural network : Euclidean space $\longmapsto$ MDS embedding space

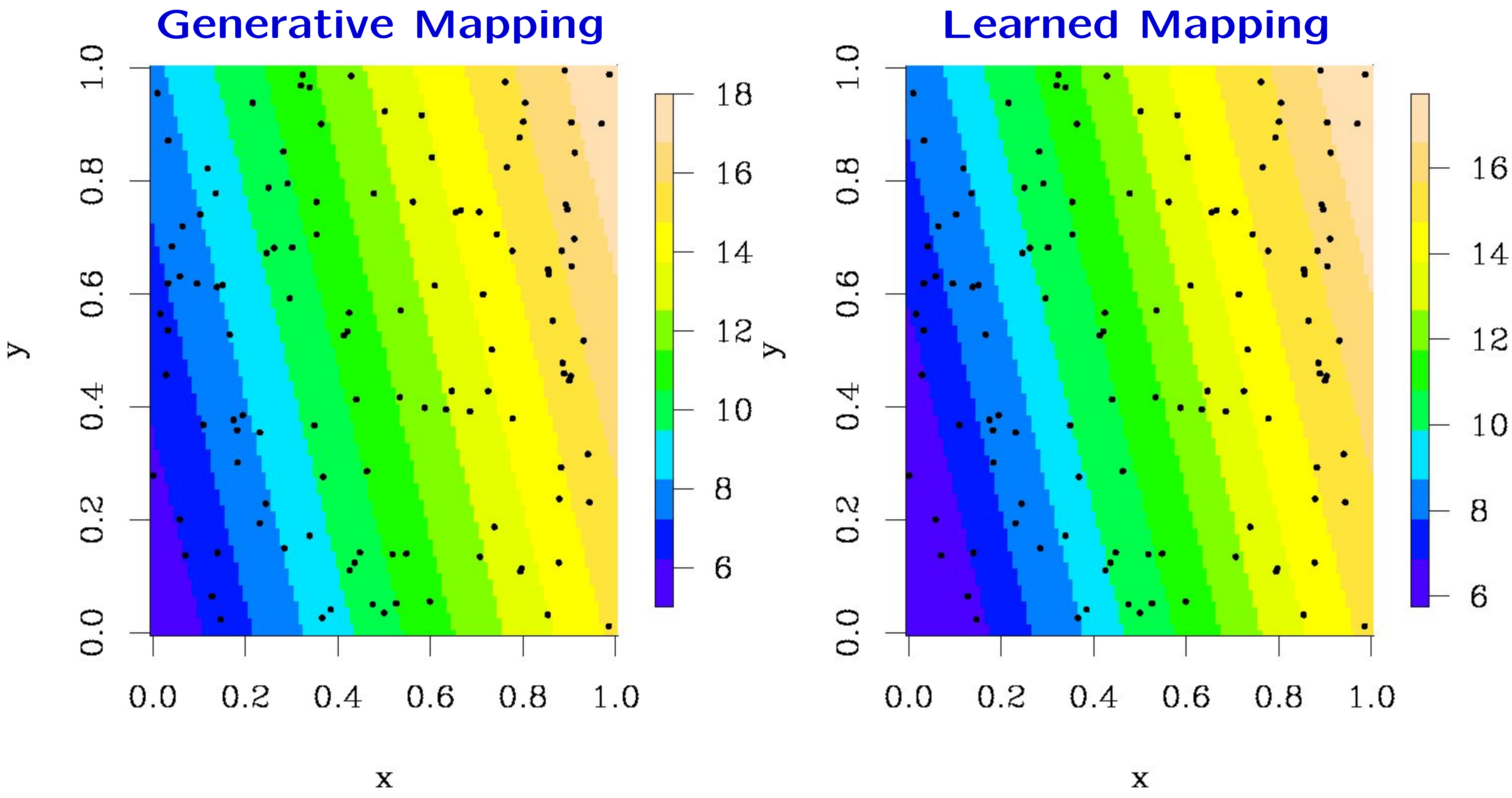


To evaluate the neighbourhood of station with no observations

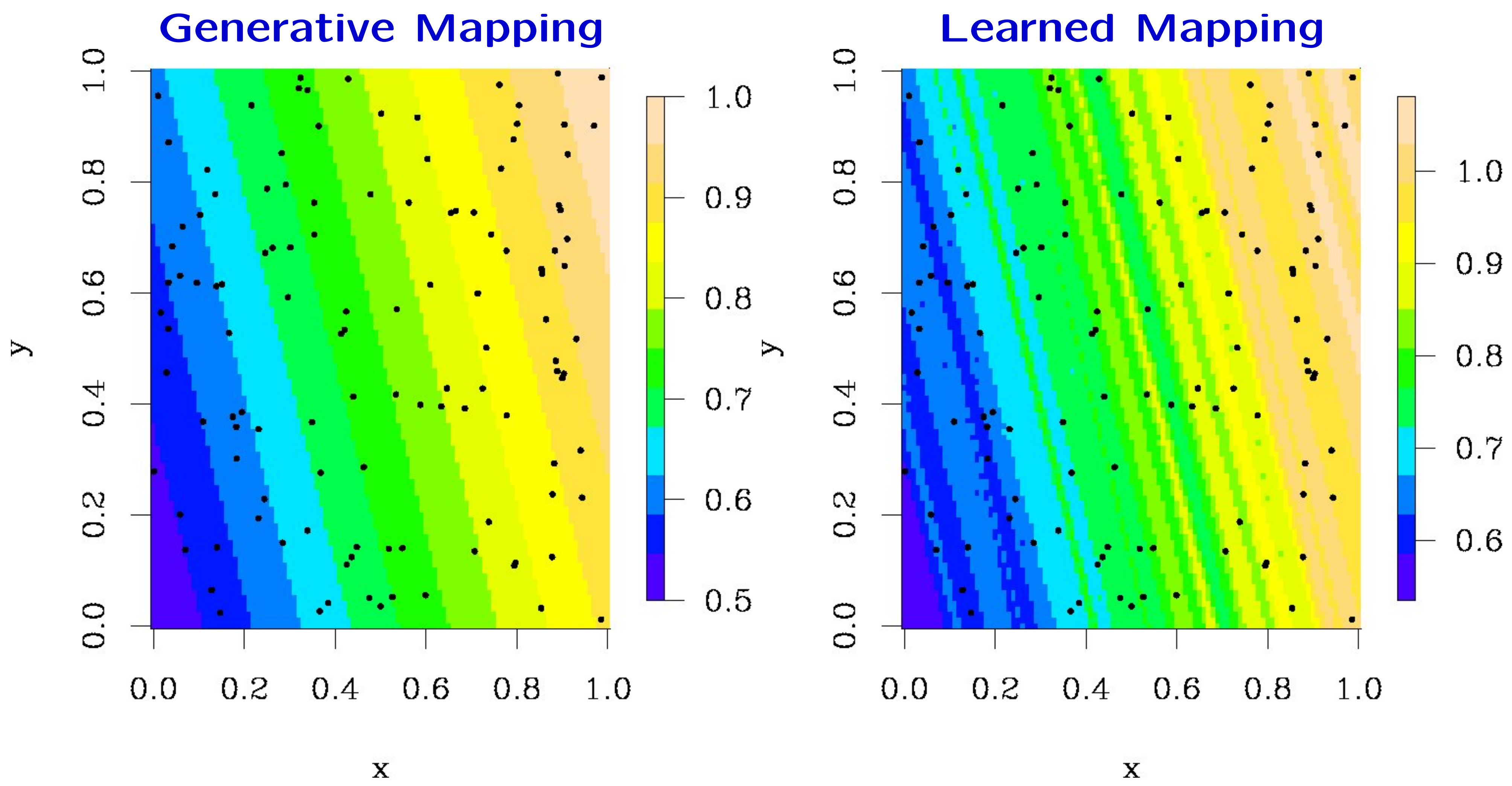
Distributional Neighbourhood



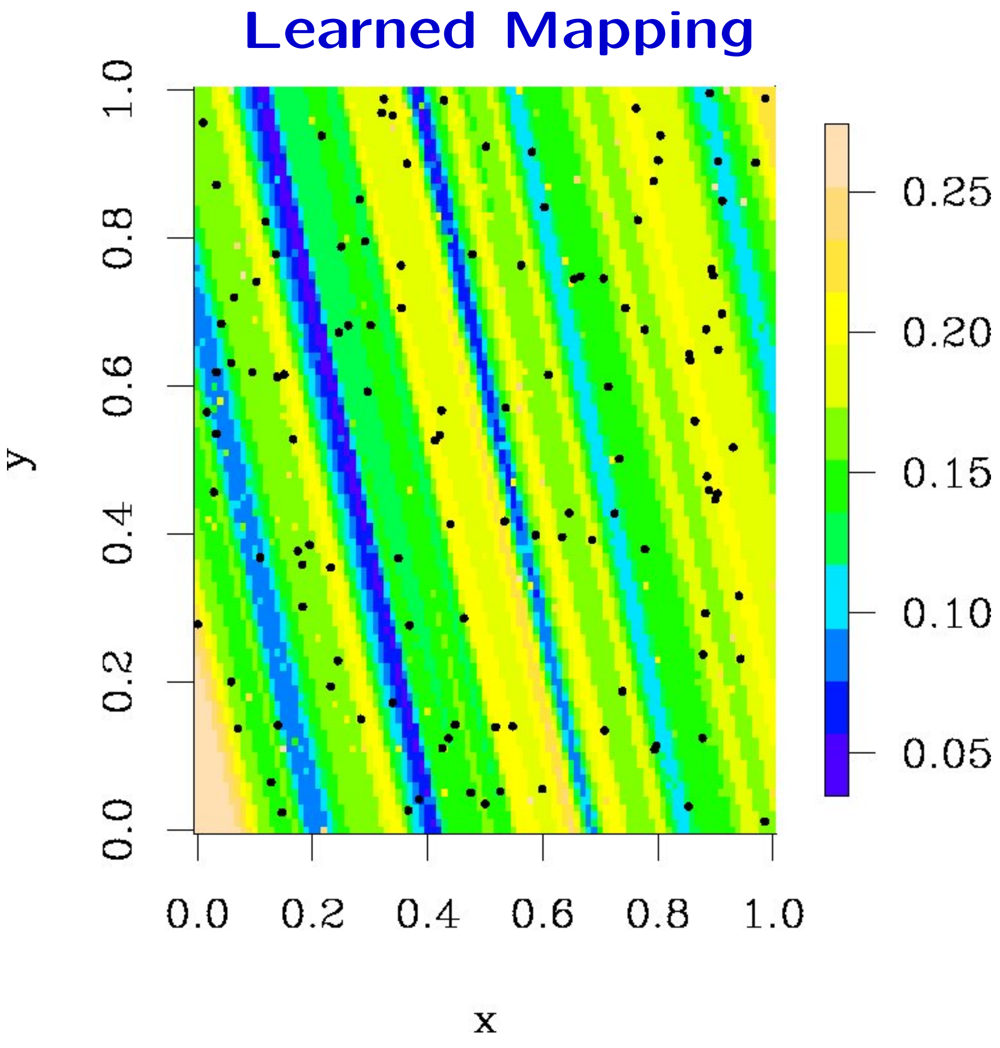
GEV Location Parameter



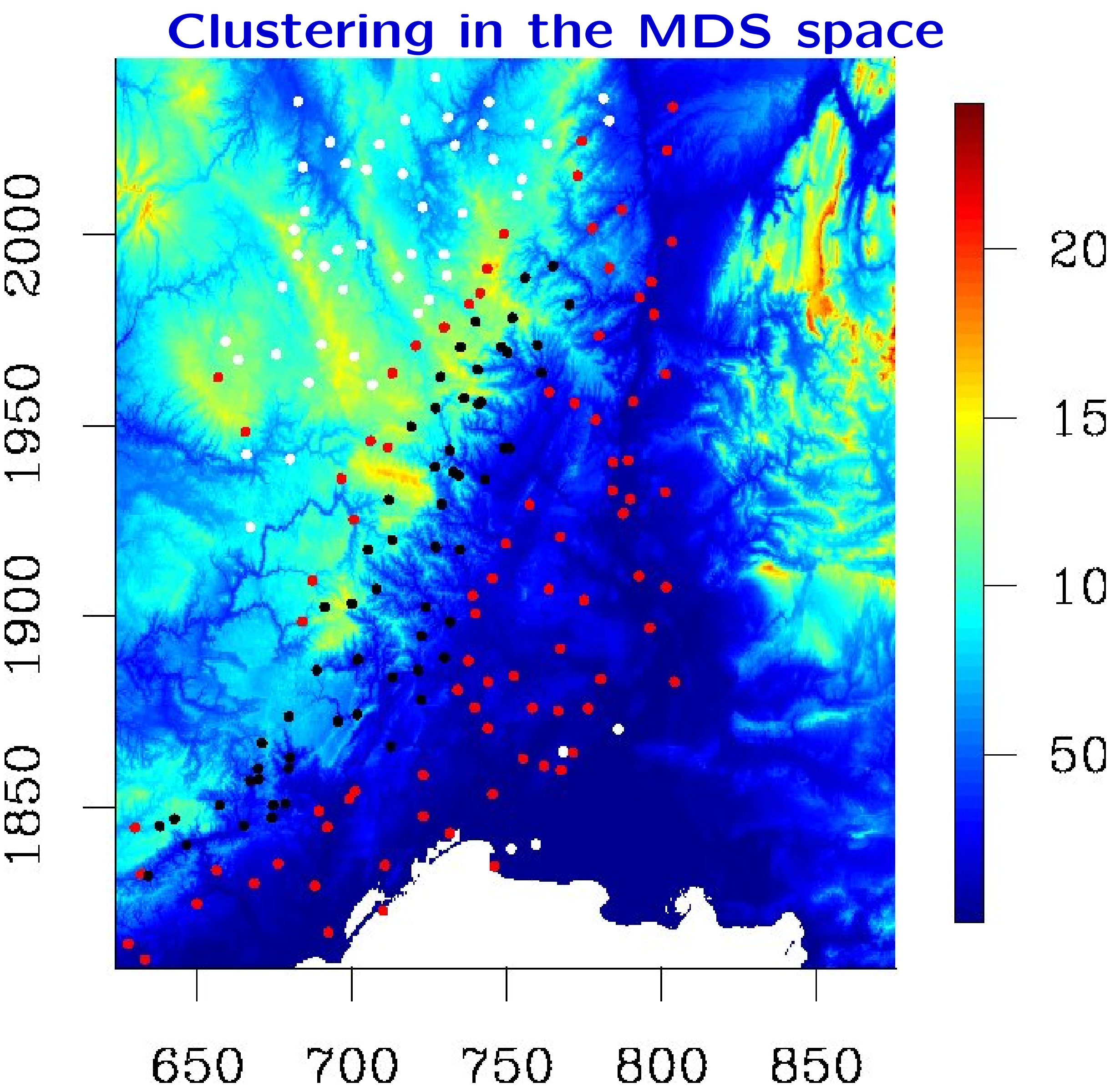
GEV Spread Parameter



GEV Shape Parameter



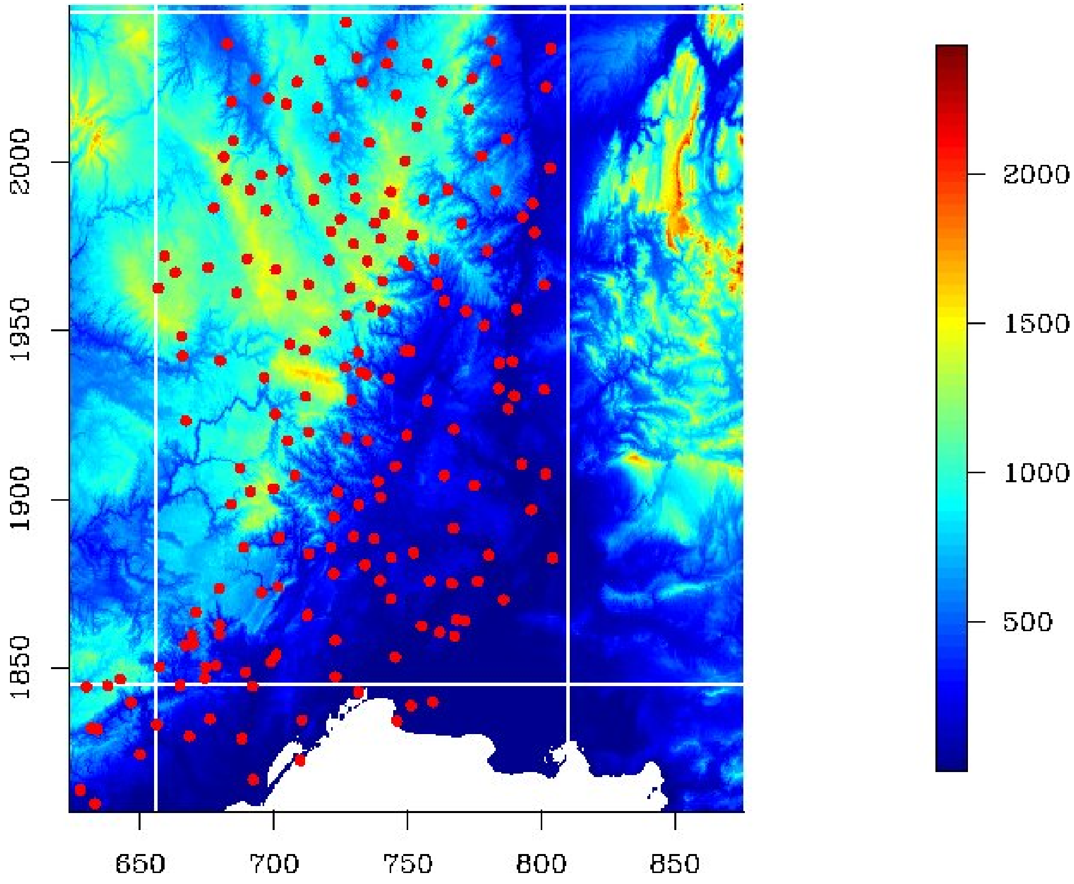
Cevennes-Vivarais Data



- White : smallest μ σ
- Black : largest μ σ
- Red : highest ξ

Cevennes-Vivarais Data

Grid on the area covered by the stations



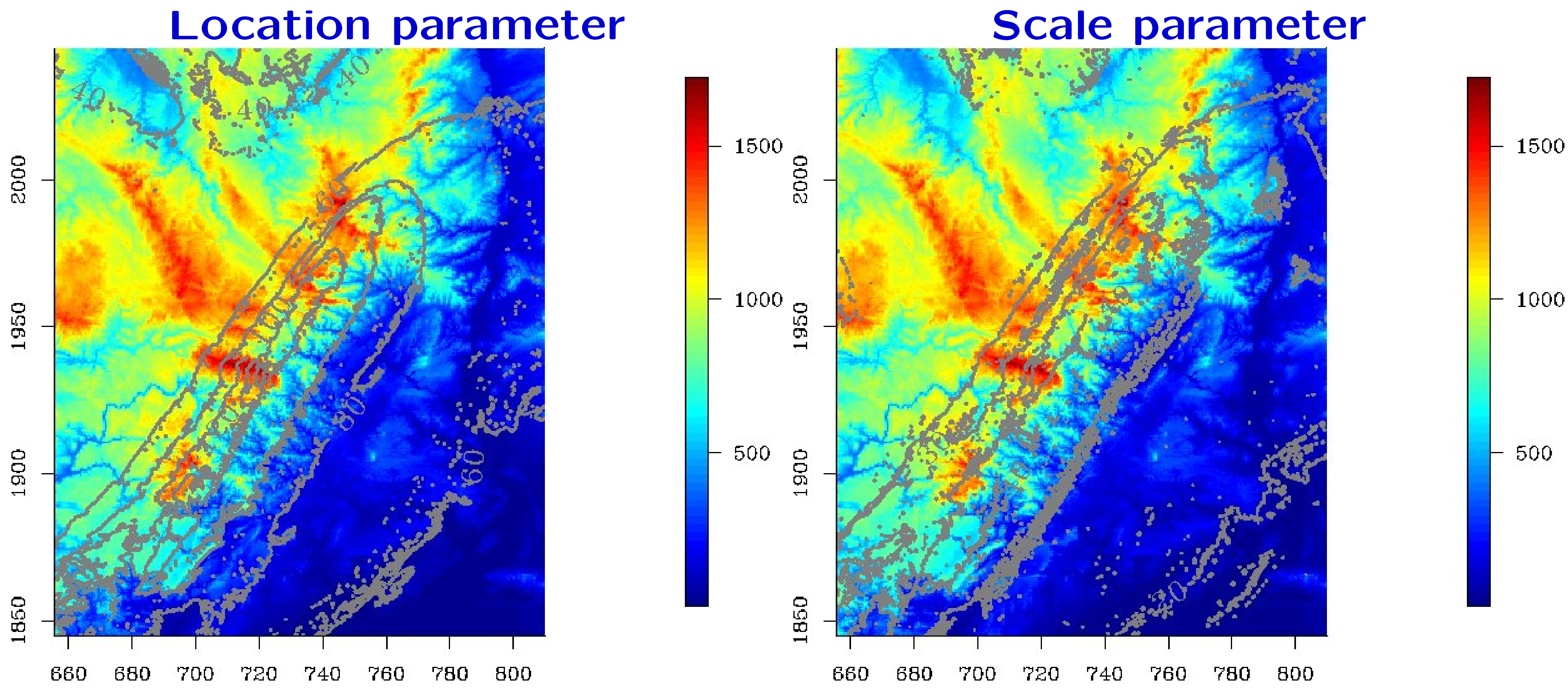
Training data : stations in RED

$K_{\mathrm{MDS}} = 3$

hidden units = 8

Kernel $(1 - u^2)^2$

Estimated GEV Parameters



Estimated GEV Parameters