# Localized/Shrinkage Kriging Predictors

Zeytu Gashaw Asfaw and Henning Omre

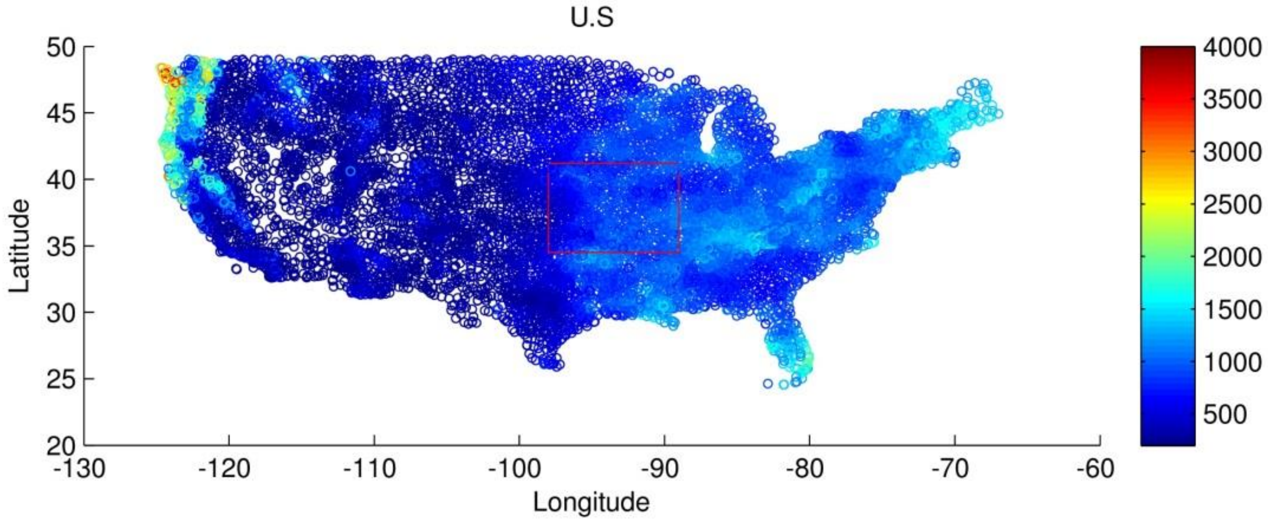
Department of Mathematical Sciences
NTNU Trondheim Norway

Gløshaugen March 2021

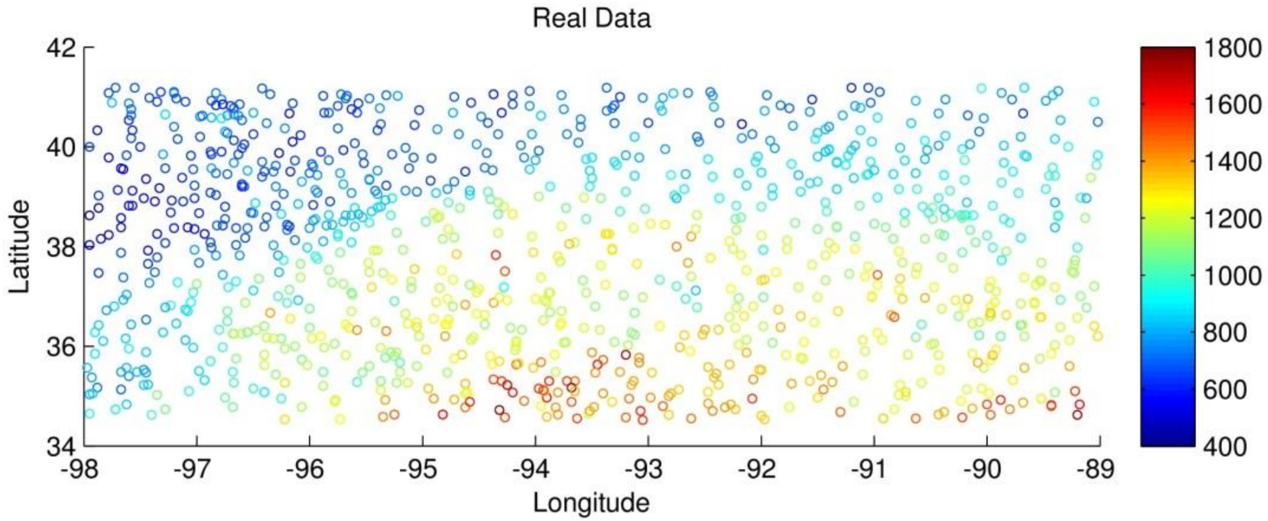
Zeytu & Omre (2016) - Math Geosc
Røislien & Omre (2006) – Math Geosc
Efron & Morris (1973) – JASA
+ Charles Stein – of course !

# Data Set

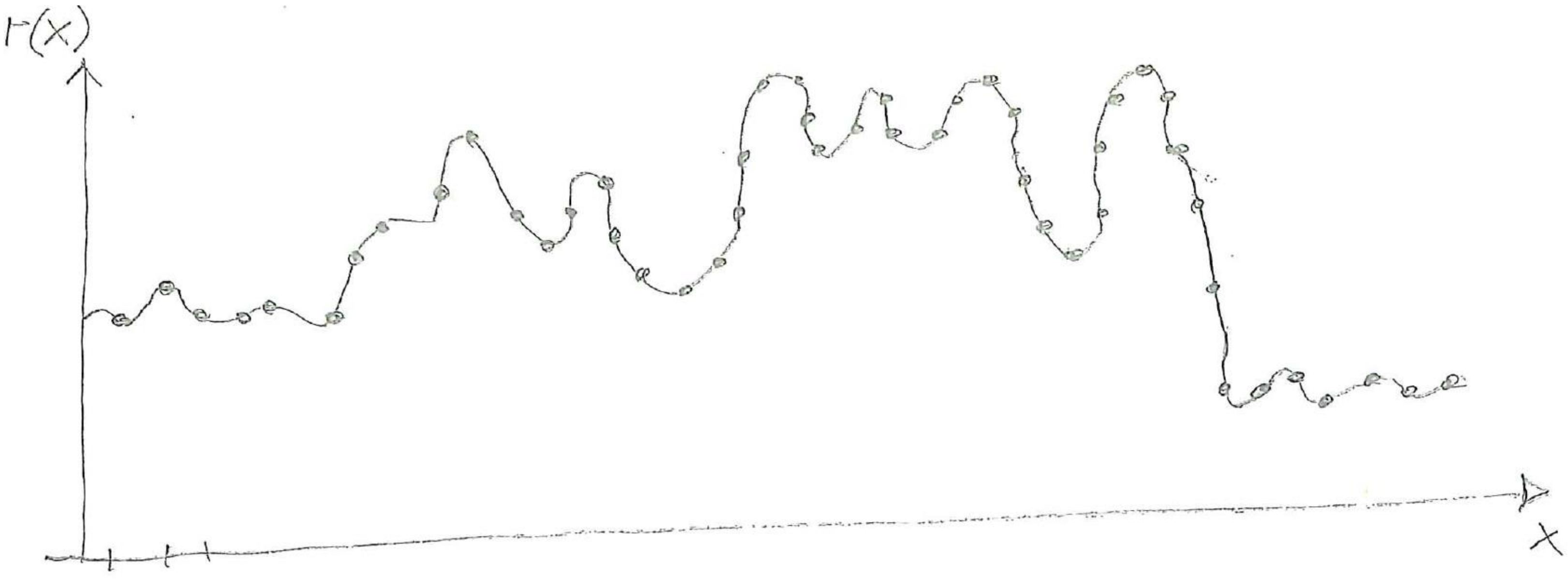
Precipitation in 1997
- spatial problem only




Subset n=1001



# Data Acquisition




Likelihood – Dirac – linear

$$\left[\mathbf{r}_0 \mid \{r(\mathbf{x});\ \mathbf{x}\in\mathcal{D}\}\right] = \mathcal{H}\left(r(\mathbf{x});\ \mathbf{x}\in\mathcal{D}\right) = \begin{bmatrix} r(\mathbf{x}_1) \\ \vdots \\ r(\mathbf{x}_n) \end{bmatrix} + \cancel{\varepsilon} = \begin{bmatrix} r_1 \\ \\ r_n \end{bmatrix}$$

# TRAD KRIGING

**Prior** – $\mathrm{GaussRF}(\mu, \sigma^2; \theta)$

$$\left\{ r(x) = \mu + \cancel{\sum_\ell \mu_\ell g_\ell(x)} + \sigma\, r_s(x) \right\}$$

$r_s(x) \leftarrow \mathrm{GaussRF}(0, 1, \theta)$

$$E\{r(x)\} = \mu$$

$$\mathrm{Cov}\{r(x'), r(x'')\} = \sigma^2 \rho(\tau; \theta) \longrightarrow \exp\left\{ -\left[\frac{\tau}{\tau_0}\right]^\nu \right\}$$

$\tau = |x' - x''|$

$\downarrow$

$\omega_{0t}, \Omega_{00}$

<u>Predictor</u> $r(\mathbf{x}_+) = r_+$

$$\hat{r}_{+|0} = \hat{\mu}_{+|0} = \mu + \sigma^2 \omega_{+0}^T \left[\sigma^2 \Omega_{00}\right]^{-1} \left[\mathbb{r}_0 - \mu \mathring{\mathbb{1}}_n\right]$$

$$= \mu + \underset{1\times n}{\omega_{+0}^T} \underset{n\times n}{\Omega_{00}^{-1}} \underset{n\times 1}{\left[\mathbb{r}_0 - \mu \mathring{\mathbb{1}}_n\right]} \quad - \text{ indep } \sigma^2$$

$$\hat{\sigma}^2_{+|0} = \sigma^2 - \sigma^2 \omega_{+0}^T \left[\sigma^2 \Omega_{00}\right]^{-1} \sigma^2 \omega_{+0}$$

$$= \sigma^2 \left[1 - \underset{1\times n}{\omega_{+0}^T} \underset{n\times n}{\Omega_{00}^{-1}} \underset{n\times 1}{\omega_{+0}}\right] \quad - \quad \text{indep } \overset{\mathbb{r}_0}{\mu}$$

→ Localization:

binary selection matrix

$$\underset{n\times 1}{\mathbb{r}_0} \longrightarrow \underset{n_+\times 1}{\mathbb{r}_0^+} = \underset{n_+\times n}{\mathbb{G}_+^{n_+}} \underset{n\times 1}{\mathbb{r}_0}$$

$$\longrightarrow \underset{n_+\times 1}{\omega_{+0}^+} = \mathbb{G}_+^{n_+} \omega_{+0}$$

$$\longrightarrow \underset{n_+\times n_+}{\Omega_{00}^+} = \mathbb{G}_+^{n_+} \Omega_{00} \mathbb{G}_+^{n_+ T}$$

Good approx!

Model parameters: $\begin{cases} \mu \in \mathbb{R} \\ \sigma^2 \in \mathbb{R}_+ \\ \theta \longrightarrow \exp\left\{-\left[\frac{\tau}{\tau_0}\right]^{\nu}\right\} \end{cases}$

## Parameter inference

Assume known: $\rho(\tau) = \exp\left\{-\left|\frac{\tau}{\tau_0}\right|^{\nu}\right\} \rightarrow \underset{n \times n}{\Omega_{oo}}$ $[\nu \in [0,2]]$

ML estimators:

$$\hat{\mu} = [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{i}_n]^{-1} [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{r}_o],$$

$$\hat{\sigma}^2 = \frac{1}{n} [\mathbf{r}_o - \hat{\mu}\mathbf{i}_n]^T \Omega_{oo}^{-1} [\mathbf{r}_o - \hat{\mu}\mathbf{i}_n],$$

Localization more difficult!

# GEN KRIGING

$$\underline{\text{Prior}} \;-\; \text{GaussRF}\left(\begin{bmatrix}\mu(x)\\ \sigma^2(x)\end{bmatrix}; x\in\mathcal{D}, \theta\right)$$

$$\left\{ r(x) = \mu(x) + \cancel{\sum_i \mu^i g_i(x)} + \sigma(x)\, r_s(x) \right\}$$

$$E\{r(x)\} = \mu(x)$$

$$\operatorname{Cov}\{r(x'), r(x'')\} = \sigma(x')\,\sigma(x'')\,\rho(\tau;\theta)$$

$$\hookrightarrow \exp\left\{-\left[\frac{\tau}{\tau_0}\right]^{\nu}\right\}$$

$$\downarrow$$

$$\omega_{+0}\ \Omega_{00}$$

# Predictor: $r(x_+) = r_+$

$$\hat{r}_{+|0} = \mu_{+|0} = \mu_+ + \sigma_+ \omega_{+0}^T \Gamma_0 \left[\Gamma_0 \Omega_{00} \Gamma_0\right]^{-1} \left[r_0 - \mu_0\right] \text{ - indep?}$$

$$\Gamma_0 = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}_{n \times n} \qquad \mu_0 = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}_{n \times 1}$$

$$\sigma_{+|0}^2 = \sigma_+^2 - \sigma_+ \omega_{+0}^T \Gamma_0 \left[\Gamma_0 \Omega_{00} \Gamma_0\right]^{-1} \Gamma_0 \omega_{+0} \text{ - indep } \frac{r_0}{\mu_0}$$

Localization:

$\llcorner$ Good approx

Model parameters: $\mu_0 \in \mathbb{R}^n$

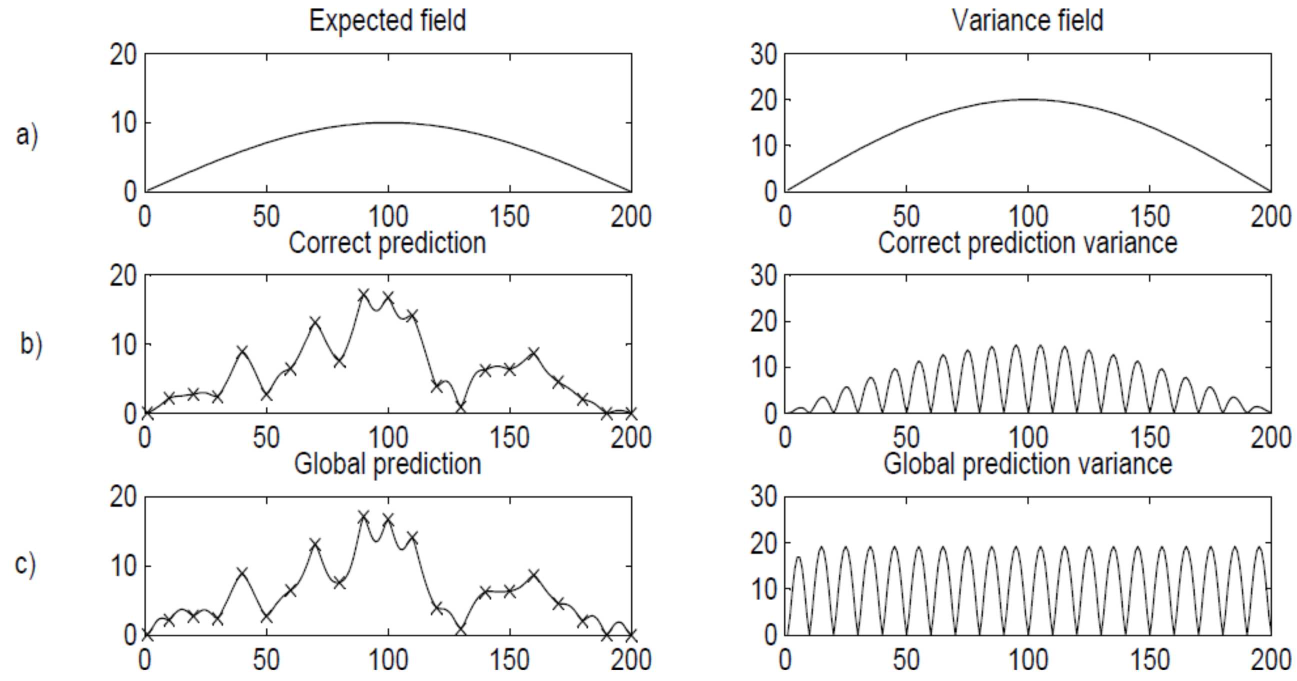
$\Gamma_0 \in \mathbb{R}_+^n$

$\Theta \rightarrow \exp\left\{-\left[\frac{z}{z_0}\right]^{\nu}\right\}$

$\mu_+ \in \mathbb{R}$

$\sigma_+^2 \in \mathbb{R}_+$

# Synthetic Example




$$\rho(\tau) = exp\{-0.2\tau^{1.5}\}$$

**Fig. 10** General Gaussian random field. Expectation and variance field (a), Predictions and prediction variances for one realization (b-e)

# Parameter inference

Assume known: $\rho(\tau) = \exp\{-[\frac{\tau}{\tau_o}]^{\nu}\} \rightarrow \underset{n \times n}{\Omega_{oo}}$

Focus on location $x_o \longrightarrow D_o^k \subset D$ ; $\underset{k \times 1}{r_o^o} = \underset{k \times n}{G_o^k} \underset{n \times 1}{r_o}$

$\rightarrow$ Inference model in $D_o^k \subset D$:

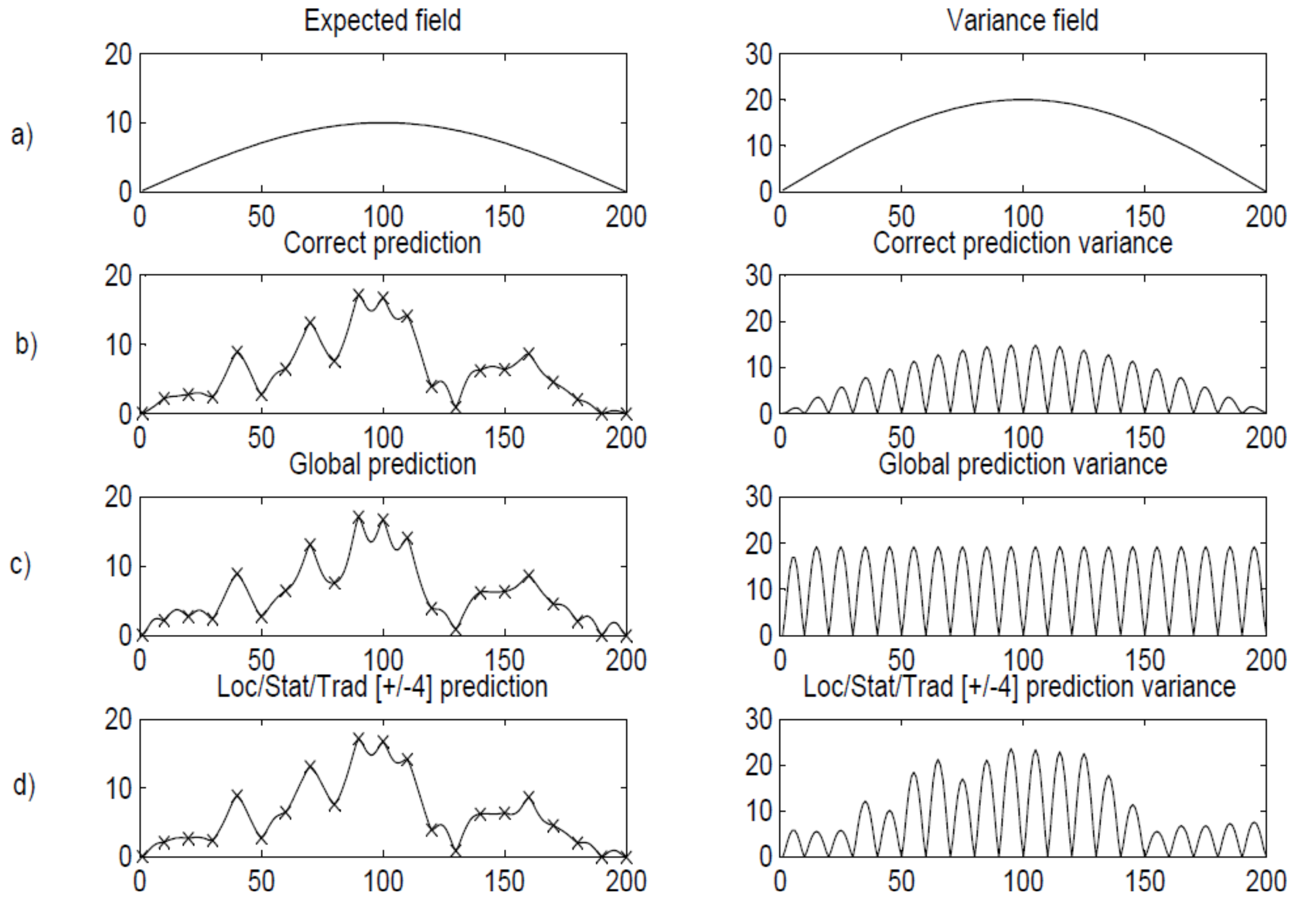
$\text{GaussRF}(\mu_o, \sigma_o^2) \longrightarrow \hat{\mu}_o, \hat{\sigma}_o^2$

$$\hat{\mu} = [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{i}_n]^{-1} [\mathbf{i}_n^T \Omega_{oo}^{-1} \mathbf{r}_o],$$

$$\hat{\sigma}^2 = \frac{1}{n} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n]^T \Omega_{oo}^{-1} [\mathbf{r}_o - \hat{\mu} \mathbf{i}_n],$$

Challenge: Bias/Variance trade-off

# Synthetic Example




$$\rho(\tau) = exp\{-0.2\tau^{1.5}\}$$

Bayesian Inference

**Fig. 10** General Gaussian random field. Expectation and variance field (a), Predictions and prediction variances for one realization (b-e)

Bayesian Inference

Inference model in $\mathcal{D}_0^k \subset \mathcal{D}$; $\underset{k \times 1}{r_0^\circ} = G_0^k r_0$

$T\text{-}RF(\mu_m, \tau_m, \xi_s, \gamma_s)$ . Røislien & Omre (2006)

Defined by:

$[r(x) \mid m, s^2] \rightarrow GaussRF(m, s^2)$

conjugate priors:

$[m \mid s^2] \rightarrow Gauss(\mu_m, \tau_m s^2)$

$$= [2\pi]^{\frac{-1}{2}} [\tau_m s^2]^{-\frac{1}{2}} exp \left\{ \frac{-1}{2} \left[\tau_m s^2\right]^{-1} [m - \mu_m]^2 \right\}$$

$s^2 \rightarrow InvGam(\xi_s, \gamma_s)$

$$= [\Gamma(\xi_s)]^{-1} \gamma_s^{\xi_s} [s^2]^{-[\xi_s + 1]} exp \left\{ -\gamma_s [s^2]^{-1} \right\}$$

Posterior E-estimators:

$\hat{\mu}_o = \mu_{m|o} = E\{m | s^2; \mathbf{r}_o\} =$

$$\mu_{m|o} = \mu_m + \tau_m \mathbf{i}_n^T \left[\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}\right]^{-1} \left[\mathbf{r}_o - \mu_m \mathbf{i}_n\right]$$

$\hat{\sigma}_o^2 = [\xi_{s|o} - 1]^{-1} \gamma_{s|o} = E\{s^2 | \mathbf{r}_o\} =$

$$\xi_{s|o} = \xi_s + \frac{n}{2}$$

$$\gamma_{s|o} = \gamma_s + \frac{1}{2}\left[\left[\mathbf{r}_o - \mu_m \mathbf{i}_n\right]^T \left[\tau_m \mathbf{i}_n \mathbf{i}_n^T + \Omega_{oo}\right]^{-1} \left[\mathbf{r}_o - \mu_m \mathbf{i}_n\right]\right]$$

Challenge: Elisitation of

$\mu_m, \tau_m, \xi_s, \gamma_s$

Empirical

Bayesian Inference

→ Empirical Bayes elisitation

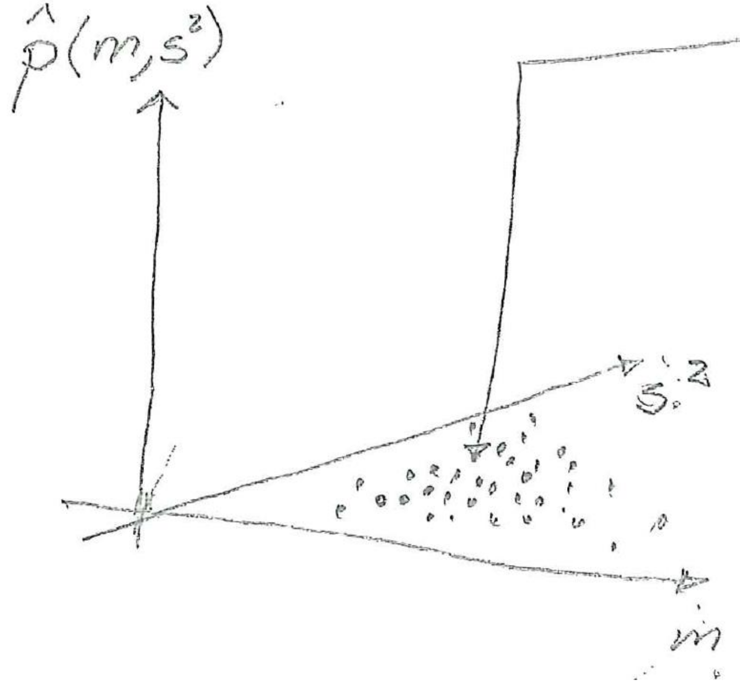
Efron&Morris (1973)

Hyper-population: ← obs. locations

$\mathcal{D}^{k}_{\circ}$ with $\circ \in \{x_1, \dots, x_n\}$

ML-estimates $\hat{\mu}_\circ, \hat{\sigma}^2_\circ \longrightarrow m, s^2$

Empirical Bayesian Inference




$\longrightarrow$ $\hat{p}(m|s^2)\,\hat{p}(s^2)$

$\tilde{\mu}_m, \tilde{J}_m, \tilde{\xi}_s, \tilde{\gamma}_s$

# Localized Predictors

Challenge: No global anchoring!

NOTE:
Cross-validation (CV) under correct model

$$e_i = \left[\frac{r_i - \hat{r}_{i\,10-i}}{\hat{\sigma}_{i\,10-i}}\right] \sim \text{Gauss}(0,1)$$

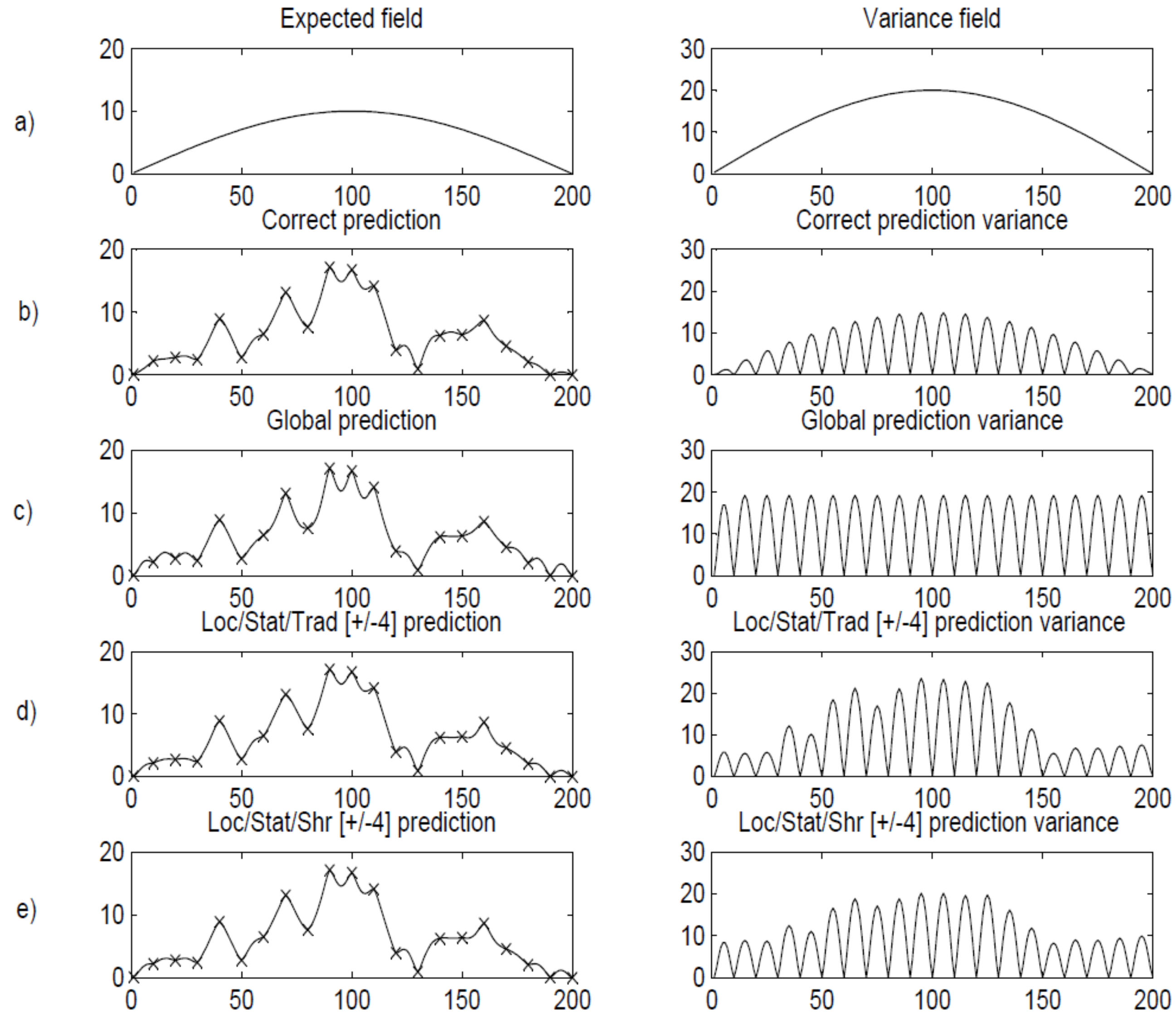
In Practice use:
CV-corrected (CVC) Predictor:

$$\hat{\hat{r}}_{+10} = \hat{r}_{+10} + \hat{\sigma}_{+10}\,\hat{\mu}_e \longrightarrow \sum_i \hat{\hat{e}}_i = 0$$

$$\hat{\hat{\sigma}}^2_{+10} = \hat{\sigma}_e^2\,\hat{\sigma}^2_{+10} \qquad \sum_i \hat{\hat{e}}_i^2 = 1$$

NOTE:
Use CV as criterion!

# Synthetic Example




$$\rho(\tau) = exp\left\{-0.2\tau^{1.5}\right\}$$

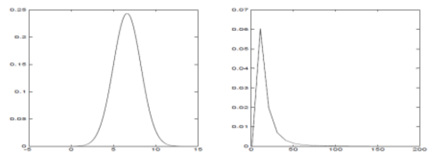
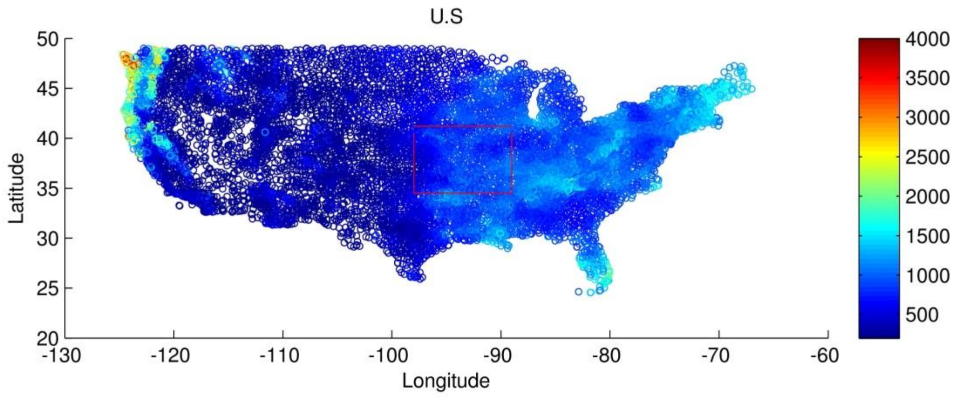
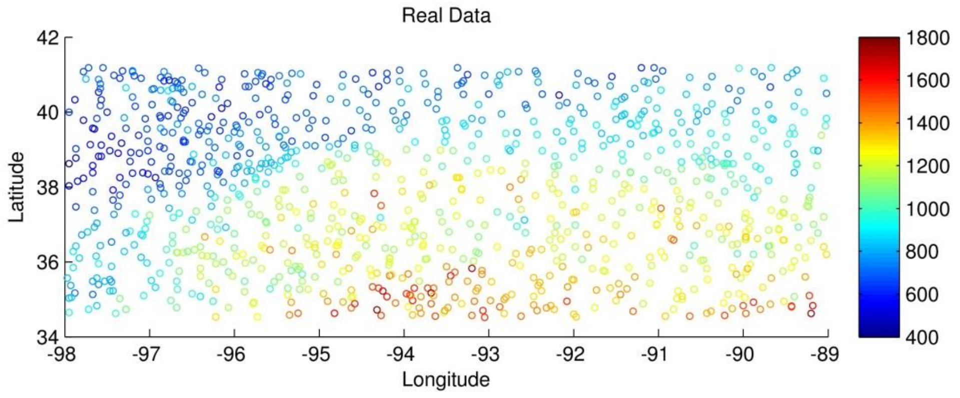


Fig. 11 General Gaussian random field. Prior model for expectation and variance for one realization

**Fig. 10** General Gaussian random field. Expectation and variance field (a), Predictions and prediction variances for one realization (b-e)

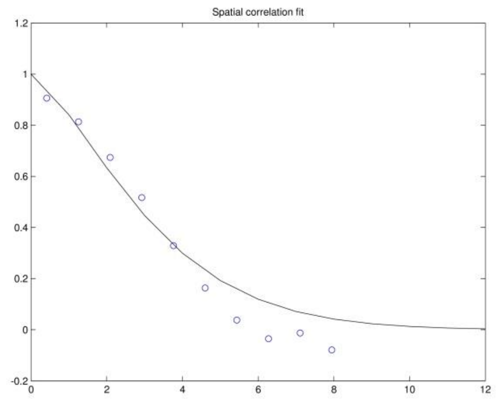
# Evaluation Data - US Precipitation 1997 – subset n=1001



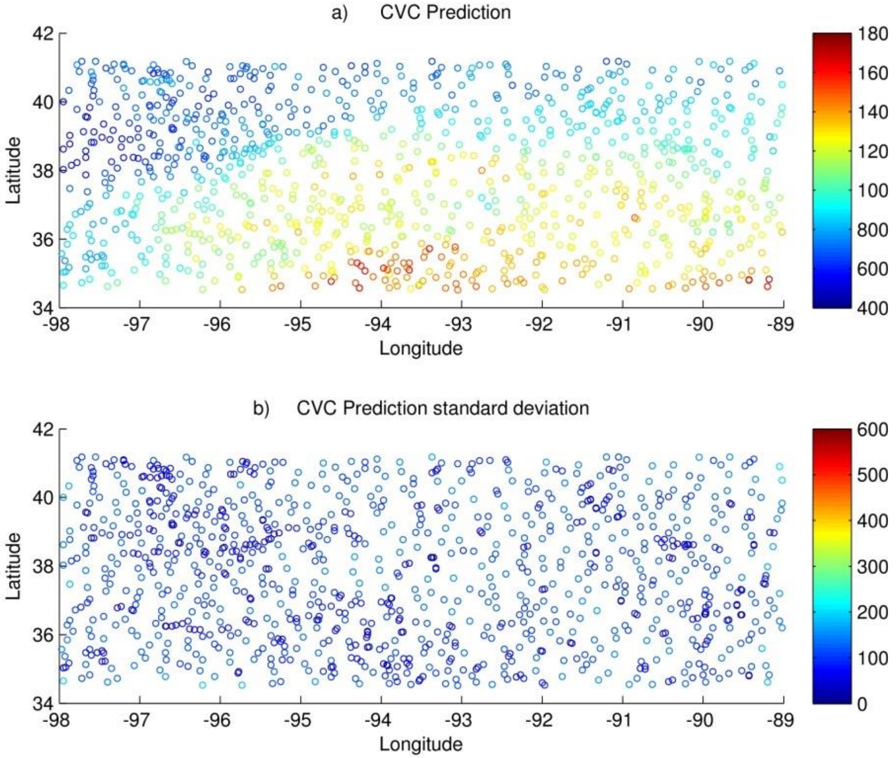



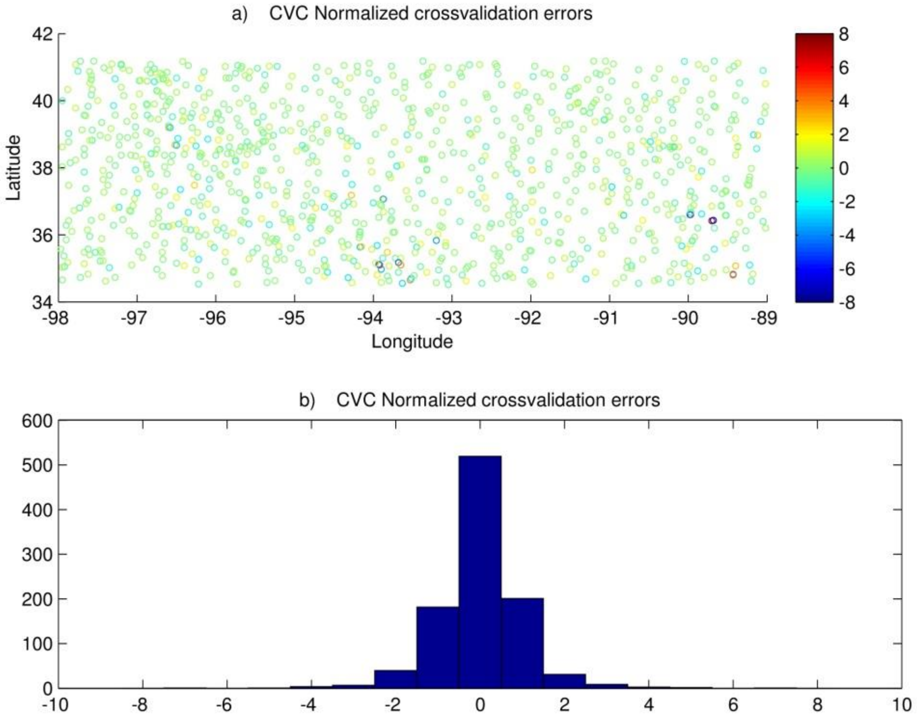

Estimated global spatial correlation function:

$$Corr\left\{r\left(x'\right),r\left(x''\right)\right\}=\rho\left(x'-x''\right)$$

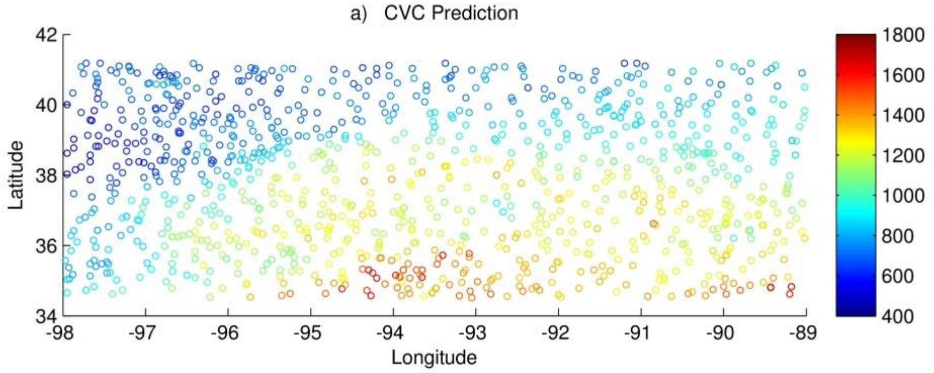


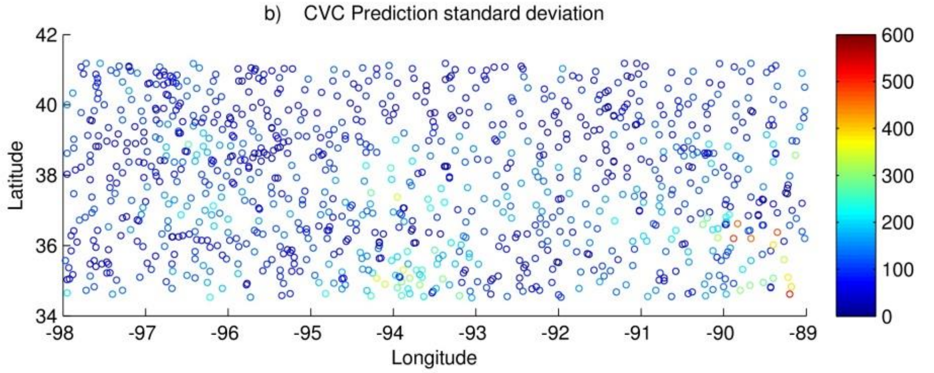
# Glob/Stat/Trad k=1000 CVC Kriging Predictor:

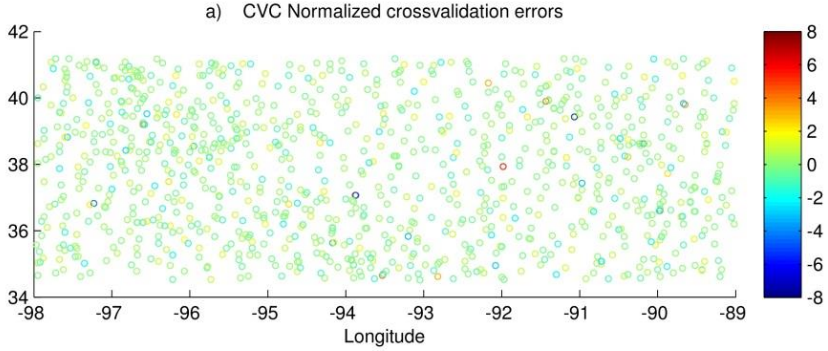


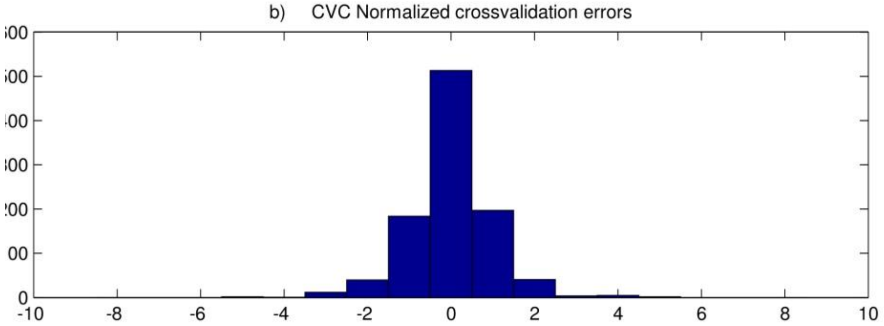



# Loc/Stat/Trad k=10 CVC Kriging Predictor:

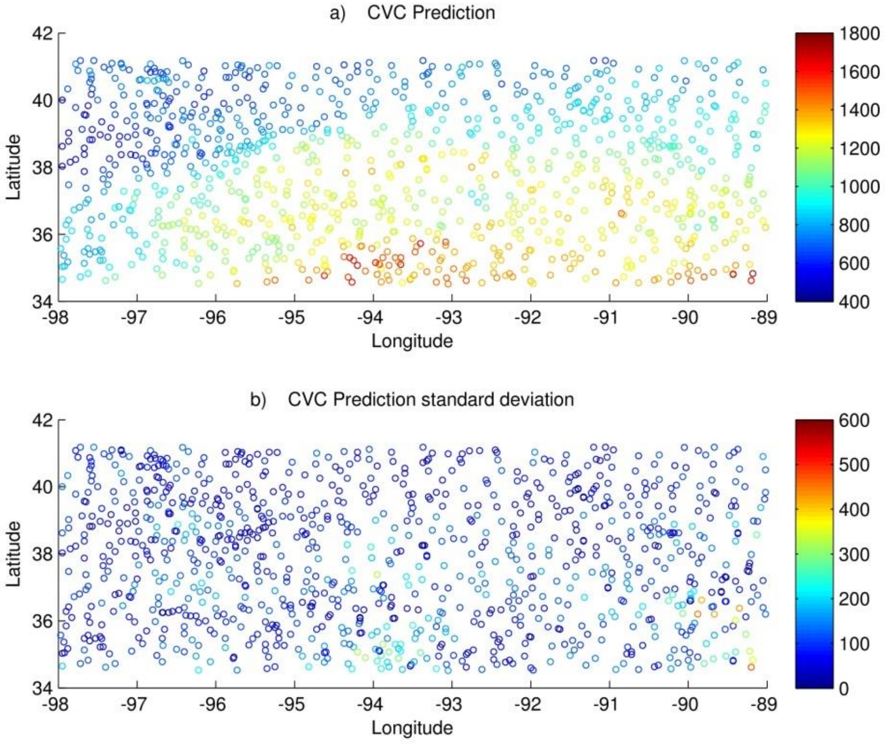


# Loc/Stat/Shr k=10 CVC Kriging Predictor:



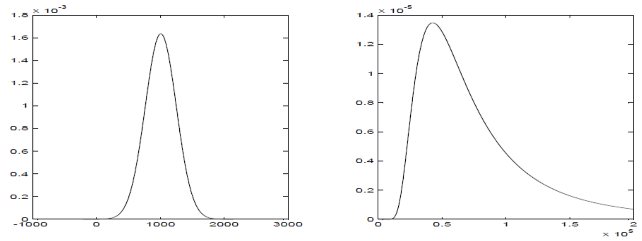
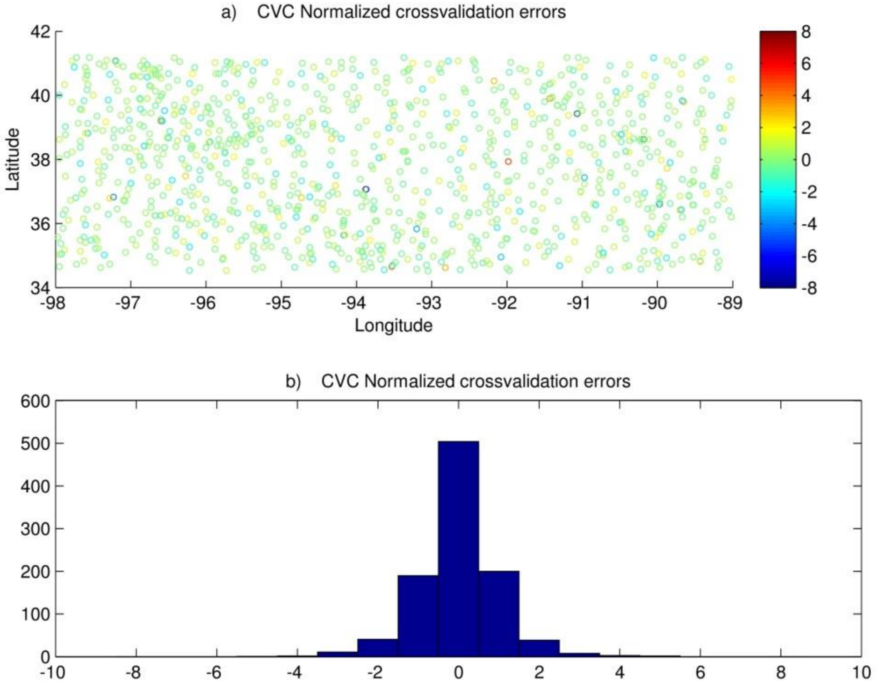



**Fig. 9** US Precipitation study. Priors model for expectation and variance



## Empirical study – Summary Table:

Empercal Evaluation

Based on CV:

$$PMSE = \frac{1}{n}\sum_{i=1}^{n}\left[r_i - \hat{r}_{i|o-i}\right]^2 \text{ – prediction accuracy}$$

$$VMSE = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left[r_i - \hat{r}_{i|o-i}\right]^2}{\hat{\sigma}^2_{i|o-i}} - 1\right]^2 \text{ – variance accuracy}$$

**Table 1** US Precipitation study. Evaluation criteria: Mean normalized error (MNE), Mean square normalized error (MSNE), Prediction mean squared error (PMSE)and Variance mean squared error (VMSE)

| Model | Localized/Stationary | | | Localized/Non-stationary | |
|---|---|---|---|---|---|
| | Traditional | | Shrinkage | Traditional | Shrinkage |
| Test D. | $k = 1,000$ | $k = 10$ | $k = 10$ | $k = 10$ | $k = 10$ |
| MNE | 1.0148e-17 | -5.3792e-18 | 5.9892e-17 | 1.5084e-16 | -1.1579e-16 |
| MSNE | 1.5399 | 2.9740 | 3.3928 | 9.2487 | 5.1702 |
| PMSE | $6.8758e+03$ | $6.8745e+03$ | $6.8654e+03$ | $2.2181e+04$ | $9.3555e+03$ |
| VMSE | 9.8749 | 5.9746 | 5.2027 | 4.0787 | 4.3475 |

**Conclusion:**

- Hierarchical non-stationary Gaussian RF – fully analytically tractable
- Localized model parameter inference – expectation & variance
- Robustified parameter inference by empirical Bayes approach
- Prediction by Localized/Shrinkage Kriging – analytically tractable
- Computer demands by Spatial Predictor is linear in no of observations.
- Spatial Prediction by ‘ non-stationary spatial Covariance function’
- Encouraging results from empirical study – actually more than so !!!

EXTRA !

# KRIGING - Alternative views

Spatial variable $\{r(x); x \in D \subset \mathbb{R}^1\}$

Likelihood:

$$\mathbb{r}_o = \mathbb{H}(r(x); x \in D) = \begin{bmatrix} r(x_1) \\ \vdots \\ r(x_n) \end{bmatrix} = \begin{bmatrix} r_1 \\ \\ r_n \end{bmatrix}$$

$$\text{Prior} - \text{GaussRF}(0, 1, \theta_o) \longrightarrow \exp\{-\tau^2\}$$

$$\tau \quad |x' - x''|$$

$$E\{r(x)\} = 0$$

$$\text{cov}\{r(x'), r(x'')\} = \exp\{-\tau^2\} \longrightarrow \omega_{o+} \; \Omega_{oo}$$

Prior $-$ GaussRF$(0,1,\theta_0) \longrightarrow \exp\{-\tau^2\}$

$\tau = |x' - x''|$

$E\{r(x)\} = 0$

$\mathrm{cov}\{r(x'), r(x'')\} = \exp\{-\tau^2\} \longrightarrow \omega_{0+} \quad \Omega_{00}$

Predictor $r(x_+) = r_+$

$$\underset{}{\hat{r}_{+|0}} = \mu_{+|0} = \underset{1\times n}{\omega_{+0}^T}\ \underset{n\times n}{\Omega_{00}^{-1}}\ \underset{n\times 1}{r_0} \qquad \text{- simple Kriging}$$

$$\hat{\sigma}^2_{+|0} = \left[1 - \omega_{+0}^T\, \Omega_{00}^{-1}\, \omega_{+0}\right]$$

## Closer look at Predictor

$$\hat{r}_{+10} = \omega_{+0}^{T} \Omega_{00}^{-1} r_0 \qquad \text{- simple Kriging}$$

## Alt A-trad Kriging

$$= \underbrace{\alpha_{+0}^{T}}_{1\times n} \cdot \underbrace{r_0}_{n\times 1} = \sum_{i=1}^{n} \alpha_i r_i$$

Lin in obs

$\alpha_{+0}$ - dep $x_+$

dep $\rho(\cdot)$

dep $X_0$

indep $r_0$

Closer look at Predictor

$$\hat{r}_{t|0} = w_{t0}^{T} \Omega_{00}^{-1} r_0$$ - simple Kriging

A&B - dual Kriging

$$= \underset{1\times n}{w_{t0}^{T}} \cdot \underset{n\times 1}{\beta_{t0}} = \sum_{i=1}^{n} \beta_i \rho(x_t - x_i)$$

Lin in $\rho(\cdot)$;

$\beta_{t0}$ - dep $r_0$

dep $\rho(\cdot)$

dep $x_0$

indep $x_t$

NOTE:

$$\left\{ \hat{r}(x) = \sum_{i=1}^{n} \beta_i \cdot \rho(x - x_i);\ x \in D \right\}$$ - cont in $x \in D$

NOTE:

$$\left\{ \hat{F}(x) = \sum_{i=1}^{n} \beta_i \cdot \rho(x - x_i);\ x \in \mathcal{D} \right\} - \text{cont in } x \in \mathcal{D}$$

hence

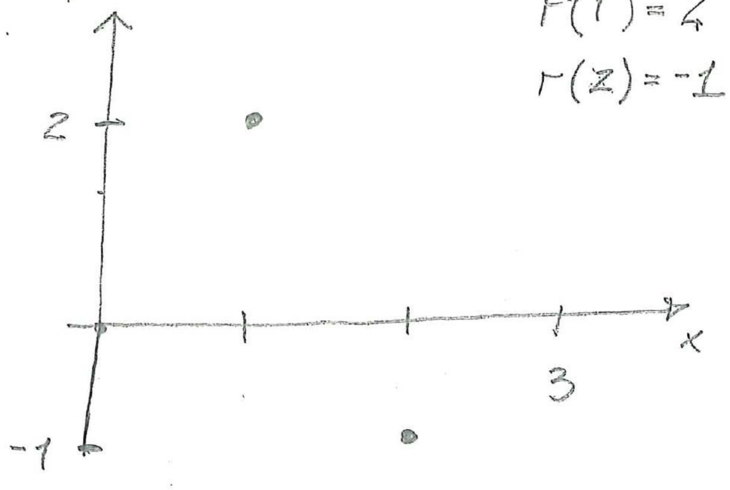
$$\frac{d\hat{F}(x)}{dx} \; -$$

analytically tractable!

$$\int_A \hat{F}(x)\,dx \; -$$

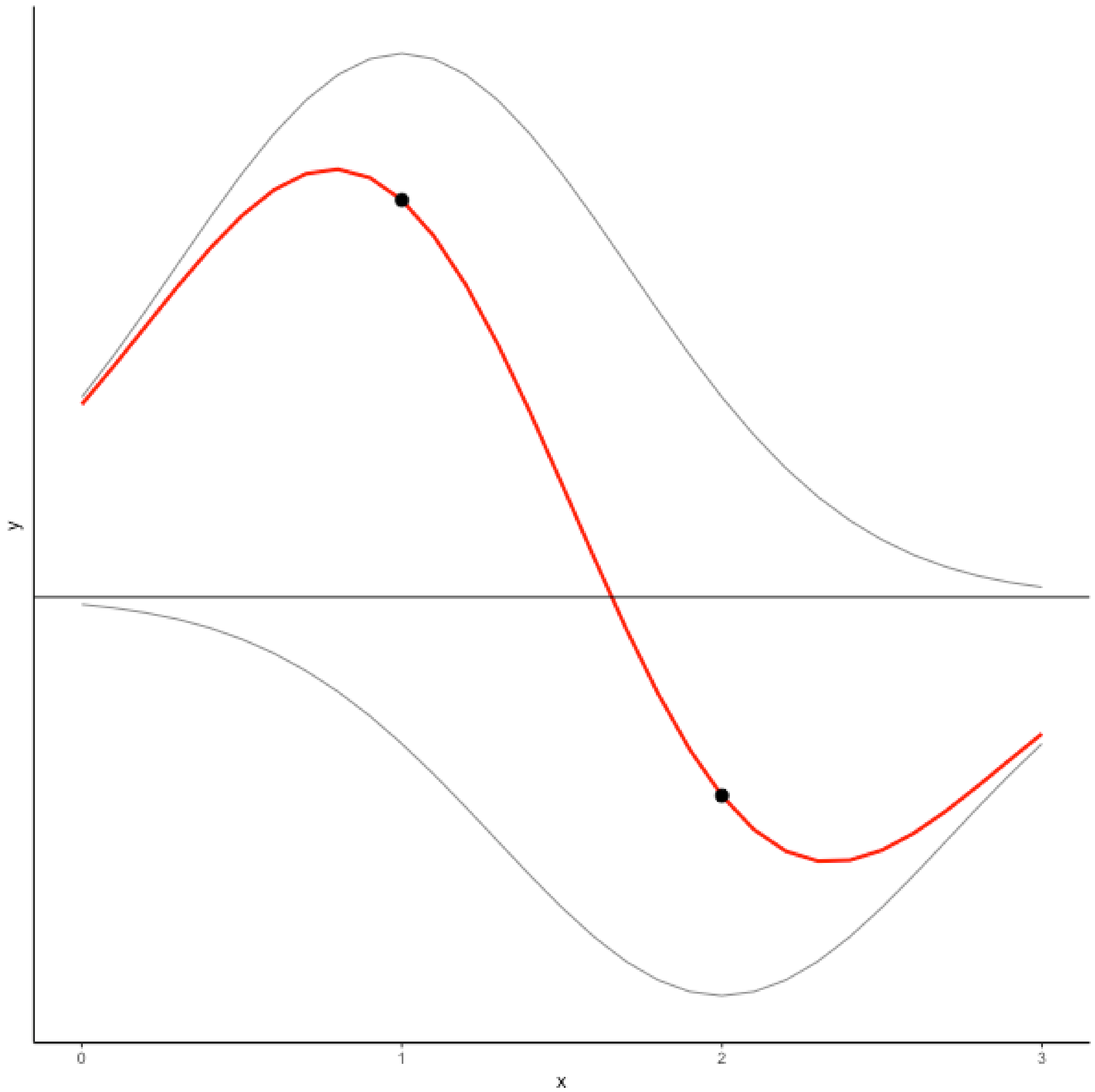
## Examples:

(A)

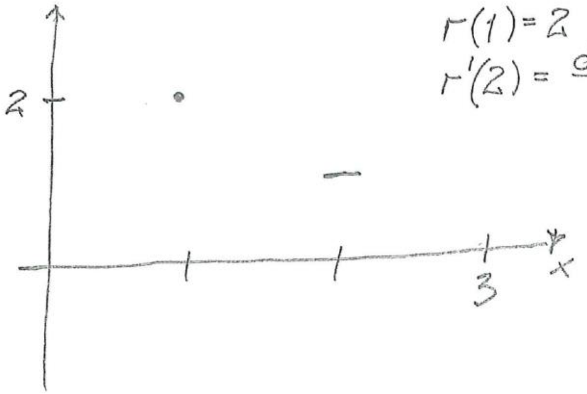



$r(1) = 2$

$r(2) = -1$



B




$r(1) = 2$

$r'(2) = \left.\frac{dr(x)}{dx}\right|_{x=2} = 0$