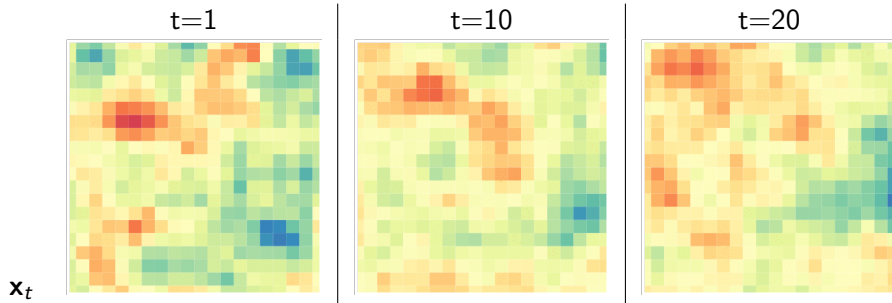


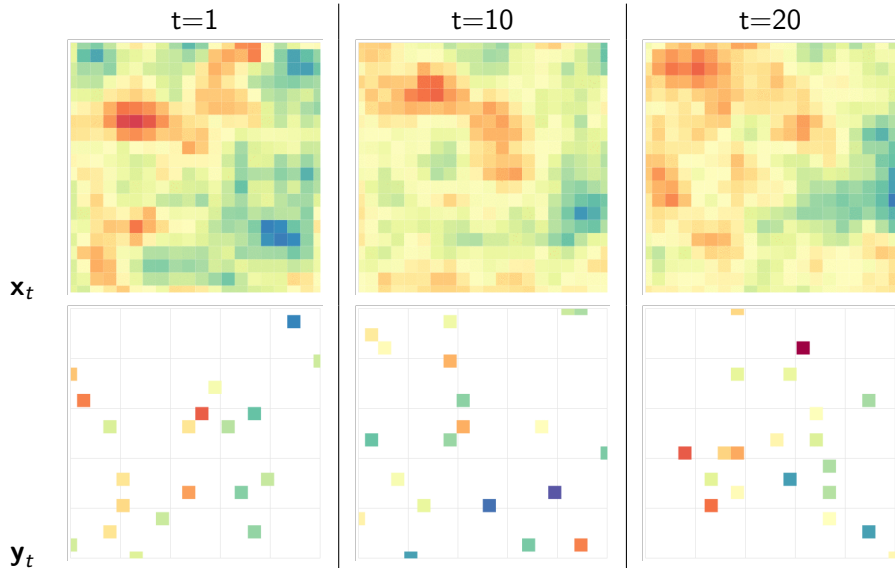
Scalable filters for massive spatio-temporal data

Marcin Jurek

Texas A&M University

May 26, 2020





First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \quad \text{where } \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \quad \text{where} \quad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

then

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$$

First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \quad \text{where} \quad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

then

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$$

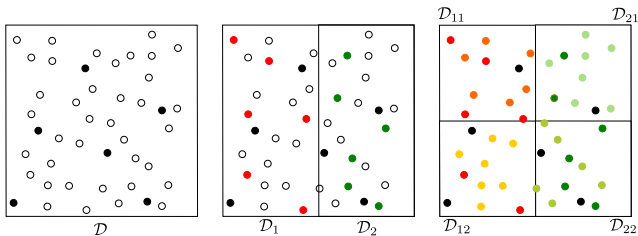
where

$$\boldsymbol{\Lambda}^{-1} = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^\top + \mathbf{R}$$

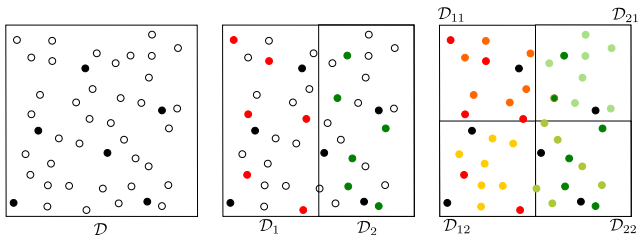
$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{H}^\top \boldsymbol{\Lambda}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{H}^\top \boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Sigma}$$

The hierarchical Vecchia (HV) approximation

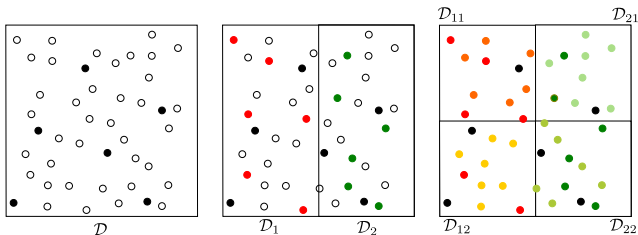


The hierarchical Vecchia (HV) approximation

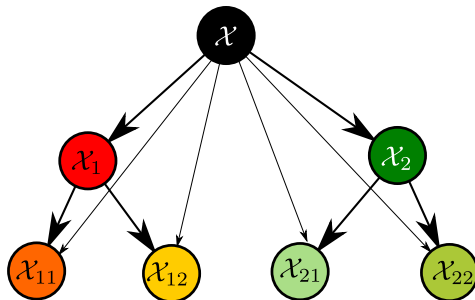


$$\mathbf{x} = \begin{bmatrix} \mathcal{X} \\ \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_{11} \\ \mathcal{X}_{12} \\ \mathcal{X}_{21} \\ \mathcal{X}_{22} \end{bmatrix}$$

The hierarchical Vecchia (HV) approximation



$$\mathbf{x} = \begin{bmatrix} \mathcal{X} \\ \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_{11} \\ \mathcal{X}_{12} \\ \mathcal{X}_{21} \\ \mathcal{X}_{22} \end{bmatrix}$$



Incomplete Cholesky decomposition

Input: positive-definite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, sparsity matrix $\mathbf{S} \in \{0, 1\}^{n \times n}$

Result: lower-triangular $n \times n$ matrix \mathbf{L}

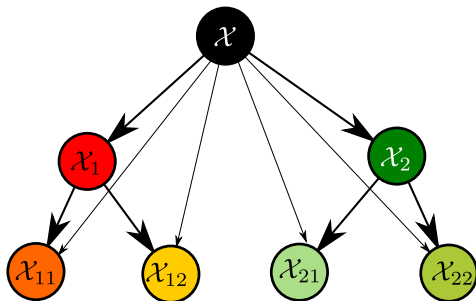
```

1: for  $i = 1$  to  $n$  do
2:   for  $j = 1$  to  $i - 1$  do
3:     if  $\mathbf{S}_{i,j} = 1$  then
4:        $\mathbf{L}_{i,j} = \frac{1}{\mathbf{L}_{j,j}} \left( \mathbf{A}_{i,j} - \sum_{k=1}^{j-1} \mathbf{L}_{i,k} \mathbf{L}_{j,k} \right)$ 
5:     end if
6:   end for
7:    $\mathbf{L}_{i,i} = \sqrt{\mathbf{A}_{i,i} - \sum_{k=1}^{i-1} \mathbf{L}_{k,k}}$ 
8: end for

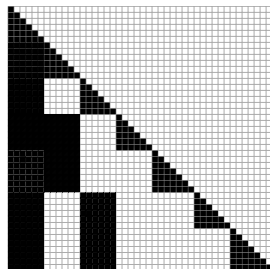
```

back to the HV approximation

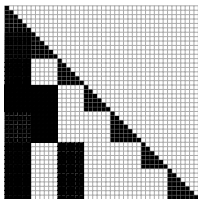
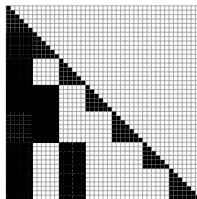
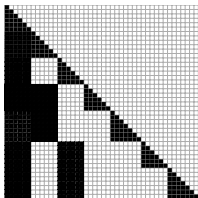
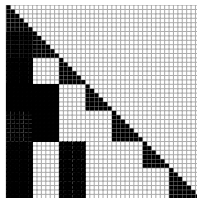
The graph



corresponding S



Sparsity patterns (we proved it)


 S

 $L := \text{ichol}(\Sigma, S)$

 L^{-1}

 $\tilde{L} := (\mathbf{P} \text{chol}(\mathbf{P} \Lambda \mathbf{P}) \mathbf{P})^{-\top}$

Some matrix algebra

The goal:

$$\tilde{\mu} = \mu + \Sigma \mathbf{H}^\top \Lambda (\mathbf{y} - \mathbf{H}\mu)$$

$$\tilde{\Sigma} = \Sigma + \Sigma \mathbf{H}^\top \Lambda \mathbf{H} \Sigma$$

where $\Lambda^{-1} = \mathbf{H}\Sigma\mathbf{H}^\top + \mathbf{R}$.

Some matrix algebra

The goal:

$$\tilde{\mu} = \mu + \Sigma \mathbf{H}^\top \Lambda (\mathbf{y} - \mathbf{H}\mu)$$

$$\tilde{\Sigma} = \Sigma + \Sigma \mathbf{H}^\top \Lambda \mathbf{H} \Sigma$$

where $\Lambda^{-1} = \mathbf{H}\Sigma\mathbf{H}^\top + \mathbf{R}$.

$$\Sigma \approx \mathbf{L}\mathbf{L}^\top \implies \Lambda \approx \mathbf{L}^{-\top}\mathbf{L}^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}$$

$$\left. \begin{array}{l} \mathbf{P} - \text{reverses order} \\ \tilde{\mathbf{L}} = (\mathbf{P} \text{chol}(\mathbf{P}\Lambda\mathbf{P})\mathbf{P})^{-\top} \end{array} \right\} \implies \tilde{\Sigma} \approx \tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top$$

$$\tilde{\mu} \approx \mu + \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{-\top}\mathbf{H}^\top\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mu)$$

Posterior inference using HV

Input: $\mathbf{y}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{H}, \mathbf{R}$

Result: $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{L}}$

1: $\mathbf{L} = \text{ichol}(\boldsymbol{\Sigma}, \mathbf{S})$

2: $\mathbf{U} = \mathbf{L}^{-\top}$

3: $\boldsymbol{\Lambda} = \mathbf{U}\mathbf{U}^{\top} + \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}$

4: $\tilde{\mathbf{L}} = (\mathbf{P}(\text{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}))\mathbf{P})^{-\top}$, where \mathbf{P} is the order-reversing permutation matrix

5: $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\top}\mathbf{H}^{\top}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$

Posterior inference using HV

Input: $\mathbf{y}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{H}, \mathbf{R}$

Result: $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{L}}$

1: $\mathbf{L} = \text{ichol}(\boldsymbol{\Sigma}, \mathbf{S})$

2: $\mathbf{U} = \mathbf{L}^{-\top}$

3: $\boldsymbol{\Lambda} = \mathbf{U}\mathbf{U}^{\top} + \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}$

4: $\tilde{\mathbf{L}} = (\mathbf{P}(\text{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}))\mathbf{P})^{-\top}$, where \mathbf{P} is the order-reversing permutation matrix

5: $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\top}\mathbf{H}^{\top}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$

Can be easily extended to non-Gaussian data using the Vecchia-Laplace algorithm (Zilber & Katzfuss, 2019)

Adding time

Assume that

$$\mathbf{x}_t = \mathcal{E}_t(\mathbf{x}_{t-1}) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$$

At each time t we want to know the *filtering distribution*:

$$\mathbf{x}_t | \mathbf{y}_{1:t} \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$$

which is Gaussian, provided that $\mathbf{x}_0 | \mathbf{y}_0$ is Gaussian.

Extended Kalman filter

Idea: recursion

Want $\mathbf{x}_t | \mathbf{y}_{1:t}$

Assume we know $\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1})$

Best guess (forecast):

$$\mathbb{E}(\mathbf{x}_t | \mathbf{y}_{t-1}) \quad := \quad \boldsymbol{\mu}_{t|t-1} \quad = \quad \mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1})$$

$$\begin{aligned} \mathbf{E}_t &:= \nabla \mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1}) \\ \text{var}(\mathbf{x}_t | \mathbf{y}_{t-1}) &:= \boldsymbol{\Sigma}_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{E}_t^\top + \mathbf{Q}_t \end{aligned}$$

Update current data (\mathbf{y}_t)

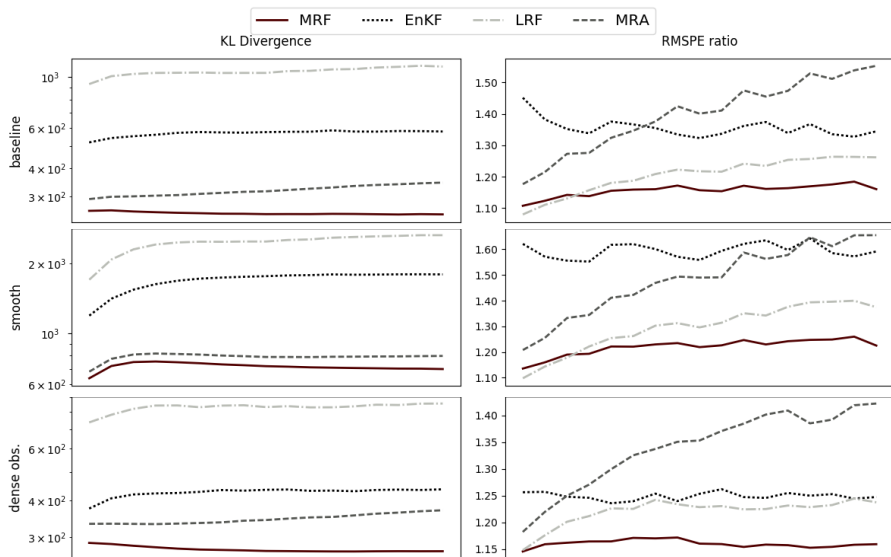
(posterior inference using $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$ as prior moments)

Extended Kalman-Vecchia filter

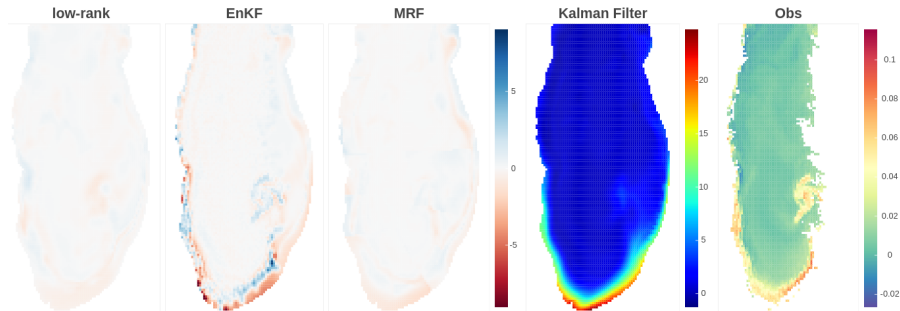
Input: \mathbf{S} , $\mu_{0|0}$, $\Sigma_{0|0}$, $\{(\mathbf{y}_t, \mathcal{E}_t, \mathbf{Q}_t, \mathbf{H}_t, \mathbf{R}_t)\}_{t=1,2,\dots}$

Result: $\mu_{t|t}$ and $\mathbf{L}_{t|t}$

- 1: Compute $\mathbf{L}_{0|0} = (\text{ichol}(\Sigma_{0|0}, \mathbf{S}))^{-\top}$
 - 2: **for** $t = 1, 2, \dots$ **do**
 - 3: Calculate $\mathbf{E}_t = \nabla \mathcal{E}_t(\mu_{t-1|t-1})$
 - 4: Forecast: $\mu_{t|t-1} = \mathcal{E}_t(\mu_{t-1|t-1})$
 $\mathbf{L}_{t|t-1} = \mathbf{E}_t \mathbf{L}_{t-1|t-1}$
 - 5: Calculate $\Sigma_{t|t-1;ij} = \mathbf{L}_{t|t-1;i,:} \mathbf{L}_{t|t-1;j,:}^{\top} + \mathbf{Q}_{t;ij}$ ($i, j : , \mathbf{S}_{ij} = 1$):
 - 6: Update: $[\mu_t, \mathbf{L}_{t|t}] = \text{HV}(\mathbf{y}_t, \mathbf{S}, \mu_{t|t-1}, \Sigma_{t|t-1}, \mathbf{H}_t, \mathbf{R}_t)$
 - 7: **return** $\mu_{t|t}, \mathbf{L}_{t|t}$
 - 8: **end for**
-

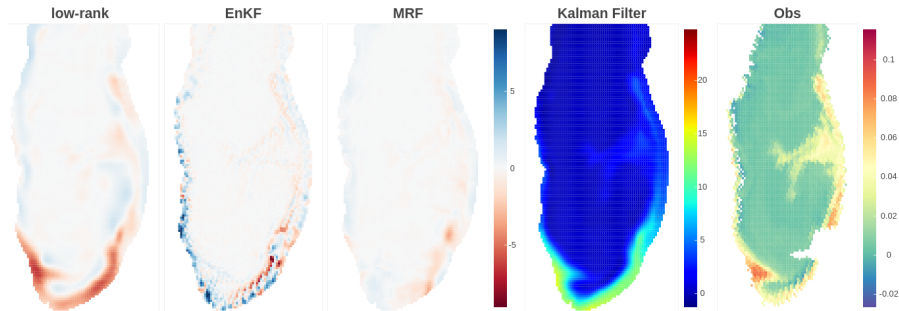


Illustration



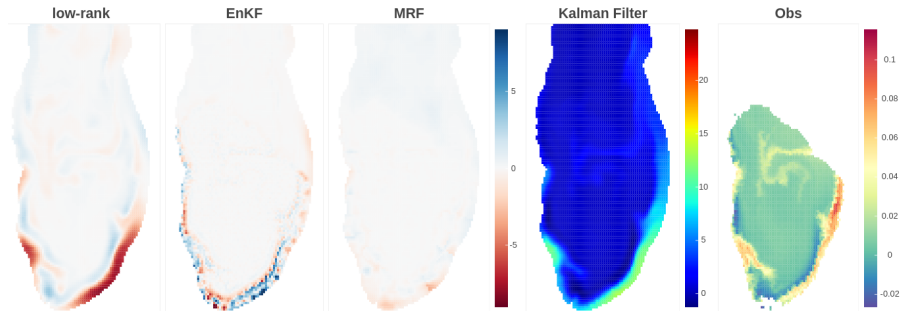
$t = 2$

Illustration



$t = 267$

Illustration



$t = 409$

Thanks Texas A&M statistics!