Geophysical Statistics Project
National Center for Atmospheric Research

Presentation for NSF, March 2003

Ensemble Kalman Filter
Tapering and Kriging
Estimating CO_2 Fluxes

Reinhard Furrer, GSP-CGD, NCAR

Outline of the Talk

- **Ensemble Kalman Filter:**
  Quantitaive description of the effects of forecast covariance estimation in the Kalman filter.
- **Tapering and Kriging:**
  Introduce a sparseness structure in the covariance via tapering.
- **Estimating CO_2 Fluxes:**
  Modeling the CO_2 cycle with a state-space approach to gain insight in the uncertainty of the error.

Ensemble Kalman Filter

Qualitative and quantitative description of the effects on the forecast and analysis covariance for different ensemble Kalman filters.

Prerequisites

Given the state-space model

$$\begin{aligned} \mathbf{y}_t &= \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t & \mathbf{w}_t &\sim \mathcal{N}_m(\mathbf{0}, \mathbf{R}_t) \\ \mathbf{x}_t &= \mathbf{G}_t\mathbf{x}_{t-1} + \mathbf{v}_t & \mathbf{v}_t &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t) \end{aligned}$$

the Kalman filter uses the iterative quantities

$$\begin{aligned} \mathbf{P}_t^f &= \mathbf{G}_t\mathbf{P}_{t-1}^a\mathbf{G}_t^\mathsf{T} + \mathbf{Q}_t \\ \mathbf{K}_t &= \mathbf{P}_t^f\mathbf{H}_t^\mathsf{T}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^\mathsf{T} + \mathbf{R}_t)^{-1} \\ \mathbf{P}_t^a &= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t^\mathsf{T})\mathbf{P}_t^f \end{aligned}$$

to filter the state $\mathbf{x}_t$ given the observations $\mathbf{y}_t, \ldots, \mathbf{y}_0$.

Approximations to Covariance matrices

The forecast covariance matrix $\mathbf{P}_t^f$ is often approximated with $\widetilde{\mathbf{P}}_t^f$:

1. $\mathbf{P}_t^f$ is estimated with N ensembles
   $\rightsquigarrow$ $\widetilde{\mathbf{P}}_t^f = \widehat{\mathbf{P}}_t^f$: the ensemble Kalman filter (EnKF)
2. $\mathbf{P}_t^f$ is estimated with N ensembles and tapered with $\mathbf{C}$
   $\rightsquigarrow$ $\widetilde{\mathbf{P}}_t^f = \widehat{\mathbf{P}}_t^f \circ \mathbf{C}$

We want to quantify the effect of the approximation with

$$\left\|\mathbf{P}_t^f - \widetilde{\mathbf{P}}_t^f\right\| \qquad \left\|\mathbf{P}_t^a - \widetilde{\mathbf{P}}_t^a\right\|$$

forecast analysis difference

We suppose $\mathbf{H} = \mathbf{I}$ and $\mathbf{R} = \mathbf{I}$ ($\rightsquigarrow$ $\widetilde{\mathbf{P}}_t^a = (\widetilde{\mathbf{P}}_t^f + \mathbf{I})^{-1}$) and we consider only one-step forecasts ($\rightsquigarrow$ drop subscript t).

Goal of the Study

What is the dependence of

$$\left\|\mathbf{P}^f - \widetilde{\mathbf{P}}^f\right\| \qquad \left\|(\mathbf{P}^f + \mathbf{I})^{-1} - (\widetilde{\mathbf{P}}^f + \mathbf{I})^{-1}\right\|$$

with

$$\widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f \qquad \widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f \circ \mathbf{C}$$

on ensemble size N, state dimension n and eigenvalues of λ_i of $\mathbf{P}^f$?

Supplementary questions:

- can we generalize to other matrices $\mathbf{H}$ and $\mathbf{R}$?
- what is an optimal taper matrix $\mathbf{C}$?

EnKF 1 2 3: Forecast Covariance

Straightforward analysis leads to

$$\mathsf{E}\left\|\mathbf{P}^f - \widehat{\mathbf{P}}^f\right\|^2 = \frac{2}{N}\sum_i \lambda_i^2 + \frac{2}{N}\sum_{i<j}\lambda_i\lambda_j$$

Remarks:

- we do not need Gaussianity,
- off-diagonal terms dominate,
- for polynomial spectra $\lambda_i \sim i^{-\theta}$ we get simple asymptotic results ($n \to \infty$).

EnKF 1 2 3: Analysis Covariance

We cannot compute the inverses.

If there is a matrix norm such that $\|\mathbf{\Lambda D} - \widehat{\mathbf{\Lambda}}\mathbf{D}\| < 1$, then the following equation holds

$$\left\|\mathbf{P}^a - \widehat{\mathbf{P}}^a\right\|^2 = \sum_{i=2}^{\infty}(i-1)\,\mathrm{tr}\Big((\mathbf{\Lambda D} - \widehat{\mathbf{\Lambda}}\mathbf{D})^i\mathbf{D}^2\Big)$$

To use the approximation

$$\left\|\mathbf{P}^a - \widehat{\mathbf{P}}^a\right\|^2 \approx \mathrm{tr}\Big((\mathbf{\Lambda D} - \widehat{\mathbf{\Lambda}}\mathbf{D})^2\mathbf{D}^2\Big) - 2\,\mathrm{tr}\Big((\mathbf{\Lambda D} - \widehat{\mathbf{\Lambda}}\mathbf{D})^3\mathbf{D}^2\Big)$$

we need to calculate

$$\mathsf{E}\Big(\widehat{\lambda}_{ij}^2\Big) \qquad \text{and} \qquad \mathsf{E}\Big(\widehat{\lambda}_{ij}\widehat{\lambda}_{jk}\widehat{\lambda}_{ki}\Big)$$

EnKF 1 2 3: Analysis Covariance

Evaluating the expressions, we have

$$\begin{aligned}\mathsf{E}\left\|\mathbf{P}^a - \widehat{\mathbf{P}}^a\right\|^2 \approx{}& \frac{1}{N}\Big(\sum_i \frac{\lambda_i^2}{(\lambda_i+1)^4} + \sum_{i,j}\frac{\lambda_i\lambda_j}{(\lambda_i+1)^3(\lambda_j+1)}\Big)\\ &+ \frac{1}{N^2}\Big(\sum_i \frac{\lambda_i^3}{(\lambda_i+1)^5} + \sum_{i,j}\frac{\lambda_i\lambda_j^2}{(\lambda_i+1)^3(\lambda_j+1)^2}\\ &+ \sum_{i,j}\frac{\lambda_i^2\lambda_j}{(\lambda_i+1)^4(\lambda_j+1)}\\ &+ \sum_{i,j,\ell}\frac{\lambda_i\lambda_j\lambda_\ell}{(\lambda_i+1)^3(\lambda_j+1)(\lambda_\ell+1)}\Big)\end{aligned}$$

For small N the approximation is useless.

EnKF with tapering 1 2

The Schur product induced by the tapering implies

$$\mathsf{E}\left\|\mathbf{P}^f - \widehat{\mathbf{P}}^f \circ \mathbf{C}\right\|^2 = \mathsf{function}\Big(\{\lambda_i\},\{\gamma_{ij}\}\Big)$$

If $\mathbf{P}^f$ is diagonal we have

$$\mathsf{E}\left\|\mathbf{P}^f - \widetilde{\mathbf{P}}^f\right\|^2 = \sum_i \lambda_i^2(c_{ii}-1)^2 + \frac{1}{N}\sum_i c_{ii}^2\lambda_i^2 + \frac{1}{N}\sum_{i,j}\lambda_i\lambda_j c_{ij}^2$$

For the ‘analysis’ difference we encounter the same problems as above.

EnKF with tapering 1 2

With $\mathbf{P}^f = (p_{ij})$, the optimal taper matrix $\mathbf{C} = (c_{ij})$ satisfies

$$\sum_{i,j}\Big(-2c_{ij}p_{ij}^2 + c_{ij}^2\Big(p_{ij}^2 + \frac{1}{N}(p_{ij}^2 + p_{ii}p_{jj})\Big)\Big)$$

Without further constraints it is impossible to find the optimum.
A component-wise minimization leads to a taper matrix, but $\mathbf{C}$ is:

- not a correlation matrix
- not always positive definite

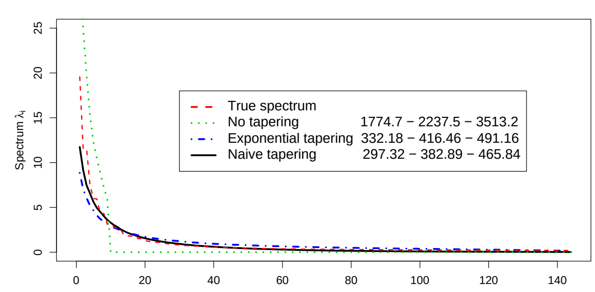
For parameterized $\mathbf{P}^f$ we could find optimal parameters.

Numerical Example

Suppose a regular 12×12 grid in $[0,1]^2$ and let

$$(\mathbf{P}^f)_{ij} = \exp\Big(-0.2|i-j|\Big)$$

For 100 MC samples we calculated the spectrum.



Further research

- Find optimal $\mathbf{C}$ for specific cases.
- Understand the 'analysis' case for small N.

Tapering and Kriging

Introduce a sparseness structure in the covariance via tapering to gain computational advantages.

- Theoretical justification
- Numerical errors
- Construction of optimal tapers

Prerequisites

Suppose a zero mean second order stationary parameterized spatial process

$$\left\{ Z(\mathbf{x}), \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d \right\}$$

with observations $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))^\mathsf{T}$.

The kriging estimator (BLUP) is

$$\widehat{Z}(\mathbf{x}_0) = \mathbf{c}^\mathsf{T}\mathbf{C}^{-1}\mathbf{Z}$$

with $\mathbf{c}_i = \mathrm{Cov}\big(Z(\mathbf{x}_0), Z(\mathbf{x}_i)\big)$, $\mathbf{c}_{ij} = \mathrm{Cov}\big(Z(\mathbf{x}_i), Z(\mathbf{x}_j)\big)$.

The mean squared error is

$$\mathrm{MSE}(\mathbf{x}_0) = c_{00} - \mathbf{c}^\mathsf{T}\mathbf{C}^{-1}\mathbf{c}$$

Matérn Covariogram 1 2

We need a broad, flexible class of covariograms to describe spatial processes.

We recommend the Matérn class given by

$$c(\mathbf{h}) \propto \big(\alpha\|\mathbf{h}\|\big)^\nu \mathcal{K}_\nu\big(\alpha\|\mathbf{h}\|\big)$$

and with spectral density

$$f(\boldsymbol{\omega}) \propto \frac{1}{\big(\alpha^2 + \|\boldsymbol{\omega}\|^2\big)^{\nu+d/2}}$$

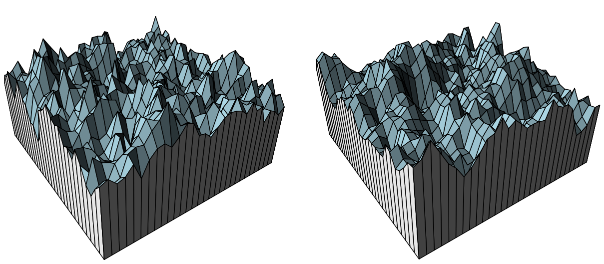
Differentiability at the origin of the covariogram is related to the tail behavior of the spectrum.

The process is m times mean squared differentiable iff

$$m < \nu$$

Matérn Covariogram 1 2

Two realizations of Matérn spatial fields:



$\nu = 0.75$ $\nu = 1.25$

effective range is 0.2

Apply M. Steins Results

Stein gives asymptotic results for misspecified covariances.

Tapering is a form of misspecification.

If we choose the taper function such that the tapered covariance satisfies 'certain' conditions, then asymptotically

- the relative increase in the MSE due to tapering tends to zero
- we have bounds for this relative increase

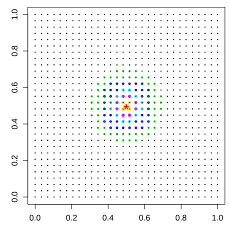
The taper has to be as differentiable at the origin as the original covariogram.

Numerical Studies 1 2 3: Setup

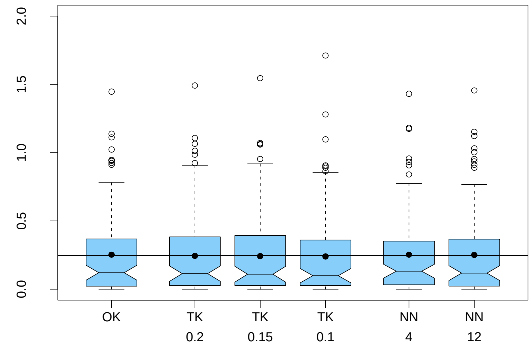
Observe processes on an equispaced 30×30 grid in $[0, 1]^2$ with a Matérn covariance structure (ν, θ).

We predict at $(0.5, 0.5)$ with:

- ordinary kriging (OK),
- ordinary kriging with tapering (TK 0.2, 0.15, 0.1),
- nearest neighbor kriging (NN 4, 16).

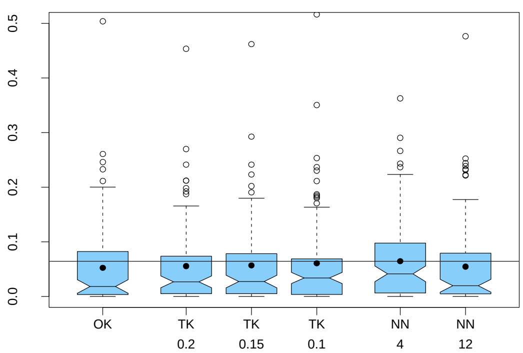


Numerical Studies 1 2 3: Results



Covariogram parameters $\nu = 0.75$, $\theta = 0.01$.

Numerical Studies 1 2 3: Results



Covariogram parameters $\nu = 1.25$, $\theta = 0.0$.

Further research

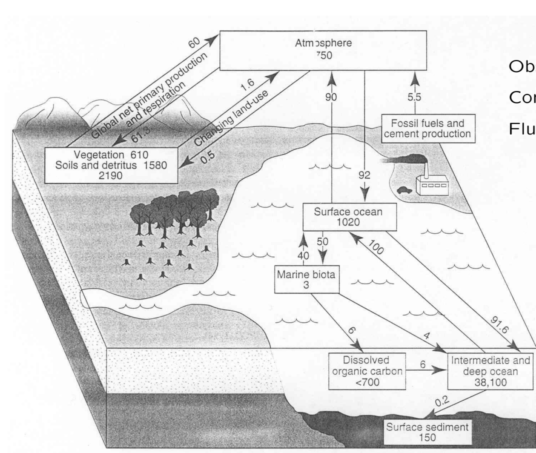
- Extend theory to covariances with a nugget,
- Computationally exploit the sparseness,
- Smooth optimal taper at origin.

Estimating CO_2 Fluxes

Examine how atmospheric carbon dioxide concentrations have changed.

Understand what and how the carbon budget affects the environment.

Schematic CO_2 Cycle



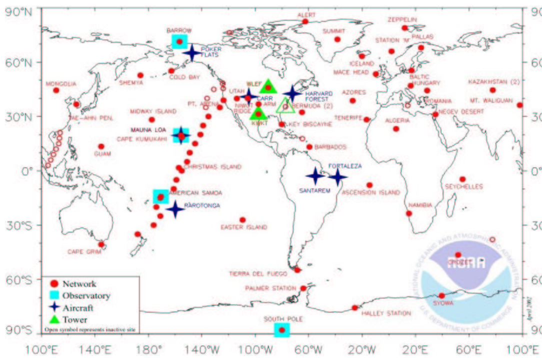
Observations

Concentrations

Fluxes

Statistical Model 1 2: Observations

We observe the CO_2 concentrations $Z(\mathbf{s}_i, t)$ at time t at location $\mathbf{s}_i$ with different measurement techniques.



Statistical Model 1 2: Concentrations and Fluxes

Further, we model the CO_2 cycle with:

$X(\mathbf{s}, t)$ actual concentrations (mixing ratios),
are dynamically constrained to $X(\mathbf{s}, t - \delta)$ and fluxes,

$U(\mathbf{s}, t)$ surface CO_2 fluxes,
represent the sources and sinks in the model.

We solve for the unknown fluxes $U(\cdot, \cdot)$

Inverse Method 1 2

The CO_2 surface flux problem is formulated as an optimality problem (simplified vector notation)

$$\min_{\{\mathbf{u}\}} \sum_t (h(\mathbf{x}_t) - \mathbf{z}_t)^\mathsf{T} \mathbf{W}_1 (h(\mathbf{x}_t) - \mathbf{z}_t) + \sum_t (\mathbf{u}_t - \mathbf{u}_t^{\mathsf{con}})^\mathsf{T} \mathbf{W}_2 (\mathbf{u}_t - \mathbf{u}_t^{\mathsf{con}}) + (\mathbf{x}_0 - \mathbf{x}_0^{\mathsf{con}})^\mathsf{T} \mathbf{W}_3 (\mathbf{x}_0 - \mathbf{x}_0^{\mathsf{con}})$$

subject to dynamical constraints

where:

$h(\cdot)$ is the measurement function,
$\mathbf{W}_i$ are weight matrices,
superscript ‘con’ are constraints,
subscript ‘0’ are initial conditions.

Inverse Method 1 2

Run model forward:

with initial concentrations and forced by a priori fluxes
⇝ obtain modeled measurement history.

Run (adjoint) model backward:

forced by the weighted measurement differences
⇝ get the adjoint to the concentrations,
⇝ get new estimates for the fluxes and initial concentrations.

Repeat until convergence is achieved.

Statistical Approach to CO_2 Cycle

Questions of Analysis:

- How does a observation network influence the space-time resolution of the uncertainty?
- Can we improve the knowledge of the state with inclusion of spatial cofactors?
- . . .

State-Space Decomposition

We decompose the observed mixing ratios $Z(\cdot, \cdot)$ as

$$Z(\mathbf{s}_i, t) = h(X(\mathbf{s}, t), \mathbf{s}_i, t) + \varepsilon(\mathbf{s}_i, t)$$

$$X(\mathbf{s}, t + \delta) = \int_{\mathcal{D}} K(\mathbf{u}, \mathbf{s}, t) X(\mathbf{u}, t) \, \mathrm{d}\mathbf{u} + U(\mathbf{s}, t)$$

$$U(\mathbf{s}, t) = g(\mathbf{s}, t) + \sum_{\tau = t - \Delta}^{t - \delta} \kappa(\mathbf{s}, \tau) U(\mathbf{s}, \tau) \, \mathrm{d}\tau + e(\mathbf{s}, t)$$

where

$h(\cdot, \cdot, \cdot)$ is the measurement function
$K(\cdot, \cdot, \cdot)$ is the (random) transport function
$g(\cdot, \cdot)$ and $\kappa(\cdot, \cdot)$ are (semi)-parametric functions
$\varepsilon(\cdot, \cdot)$ and $e(\cdot, \cdot)$ are colored spatio-temporal processes

State Equation 1 2: Concentration Model

Discretization of space and time leads to the equation

$$X(\mathbf{s}, t+\delta) = \sum_{\mathbf{u} \in \mathcal{U}_{\mathbf{s}}} K(\mathbf{u}, \mathbf{s}, t) X(\mathbf{u}, t) + U(\mathbf{s}, t)$$

where $\mathcal{U}_{\mathbf{s}}$ contains $\mathbf{s}$ and its first order neighbors.

We consider a random transport function.

For example add Gaussian perturbations to its elements

$$\rightsquigarrow \qquad K(\cdot,\cdot,\cdot) = k(\cdot,\cdot,\cdot) + \nu(\cdot,\cdot,\cdot)$$

State Equation 1 2: Source and Sink Model

The same discretization leads to

$$U(\mathbf{s}, t) = g(\mathbf{s}, t) + \kappa(\mathbf{s}, t) U(\mathbf{s}, t-\delta) + e(\mathbf{s})$$

where $e(\cdot)$ is a Gaussian spatial process.

The cofactors $g(\cdot,\cdot)$ are modeled as

$$g(\mathbf{s}, t) = \sum_k \beta_k v_k(\mathbf{s}, t)$$

Markov Chain Monte Carlo

The concentrations and fluxes can be simulated with a Gibbs sampler.

Write $\mathbf{U} = \big(U(\mathbf{s}_i, t_j)\big)$, $\mathbf{X} = \big(X(\mathbf{s}_i, t_j)\big)$ and
$\mathbf{X}_k$, $\mathbf{U}_k$ the respective kth column.

Then
$\mathbf{U}_k \big| \mathbf{U}_{k-1}$ is Gaussian,

$\mathbf{X}_k \big| \{\mathbf{X}_{k-1}, \mathbf{U}_k\}$ is Gaussian.

Apply a Gibbs sampler to get a sequence $\text{vec}(\mathbf{X})^{(n)}$.

Work in Progress

- Run the Gibbs sampler
- Optimize Gibbs sampler algorithm
- Answer the scientific questions

Quo Vadis

- Ensemble Kalman Filter
- Tapering and Kriging
- Estimating CO_2 Fluxes