Statistics and the DAI

Douglas Nychka
Geophysical Statistics Project
National Center for Atmospheric Research

- Participants
- Non Gaussian filters
- Multiresolution models
- Tapering covariance functions

Overview

The Geophysical Statistics Project is a statistical research group embedded within NCAR. It is supported mainly by NSF Division of Mathematical Sciences.

The Initiative motivates fundamental statistical problems that underlie DA and provides a context for new statistical research.

DART also gives statisticians, who have little or no graduate training in geophysical computing, access to large numerical models and nontrivial test cases.

Activities Past and Future

Statistics post docs

Thomas Bengtsson (NonGaussian filters)
Reinhard Furrer (Covariance Tapering)
Tomoko Matsuo (Wavelet covariances, Mesosphere DA) *50% DAI*

Undergraduates in math

Sponsored a student team at the Institute for Pure and Applied Mathematics, UCLA 6/04-8/04.

DA focus with NSF Mathematics Centers 1/05-5/05

samsi
NSF • Duke • NCSU • UNC • NISS

Institute for Pure & Applied Mathematics
I P A M
University of California, Los Angeles

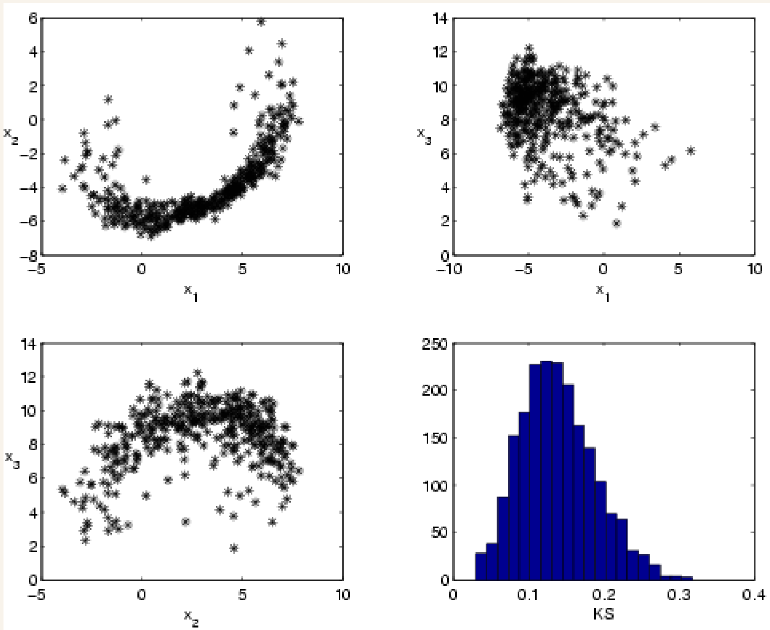
(Includes the NCAR/DAI Summer school)

NonGaussian filters

Is it possible to improve assimilation by accounting for non-Gaussian distributions?

NonGaussian filters

Is it possible to improve assimilation by accounting for non-Gaussian distributions?

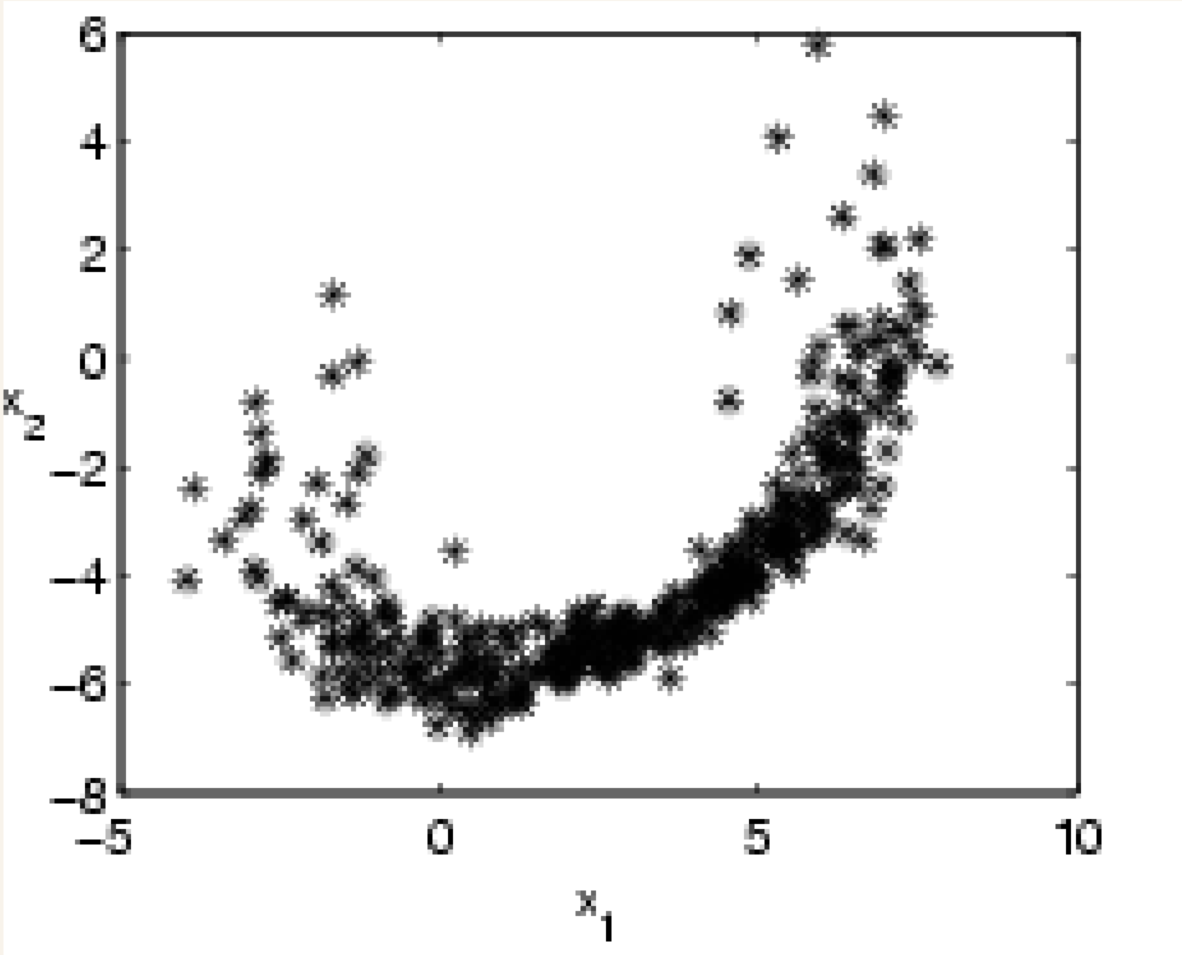


Components 1,2,3 from an ensemble Kalman Filter for "Lorenz 40".

Clear departures from a multivariate normal.

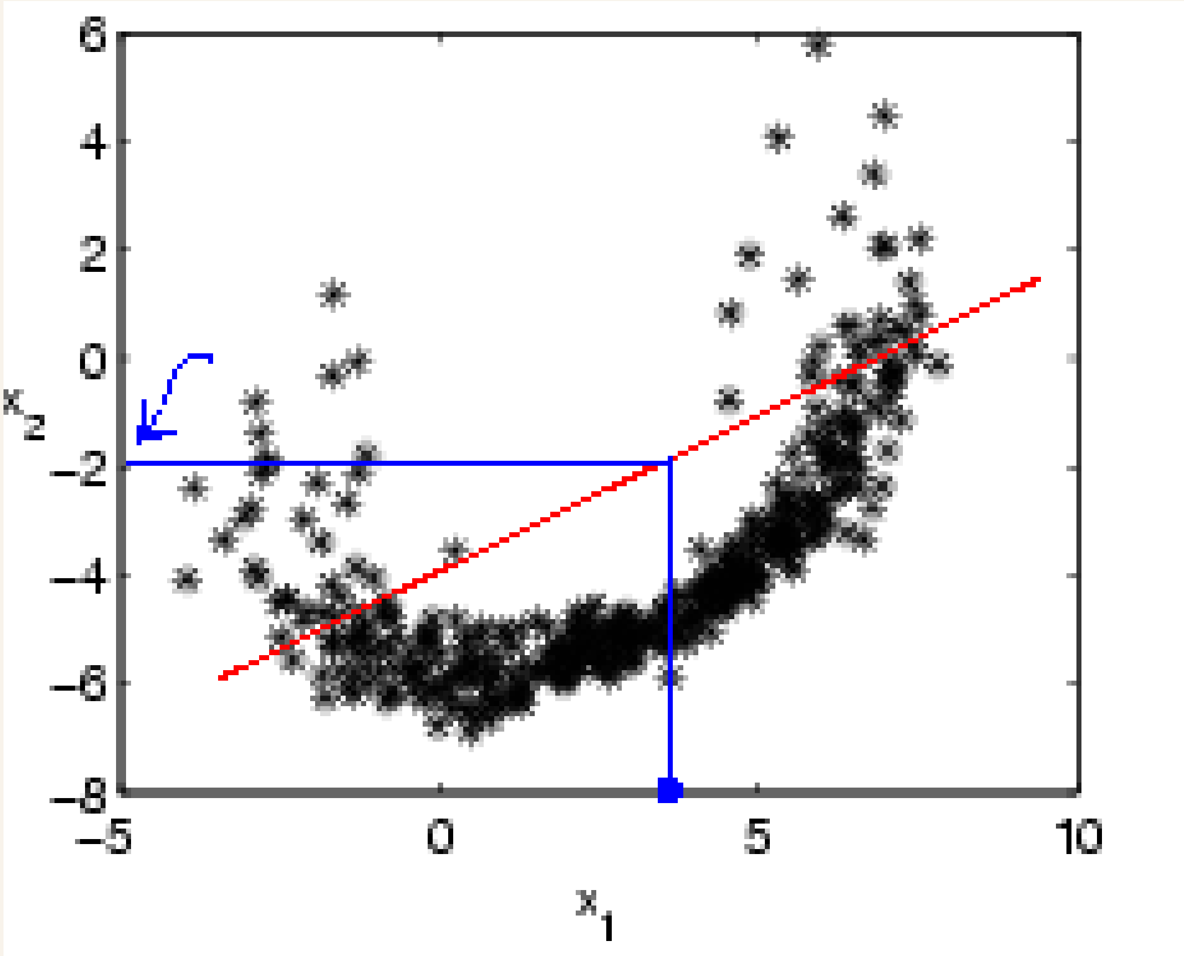
This is not a straight line world!

Focus on (X1, X2) ensemble members



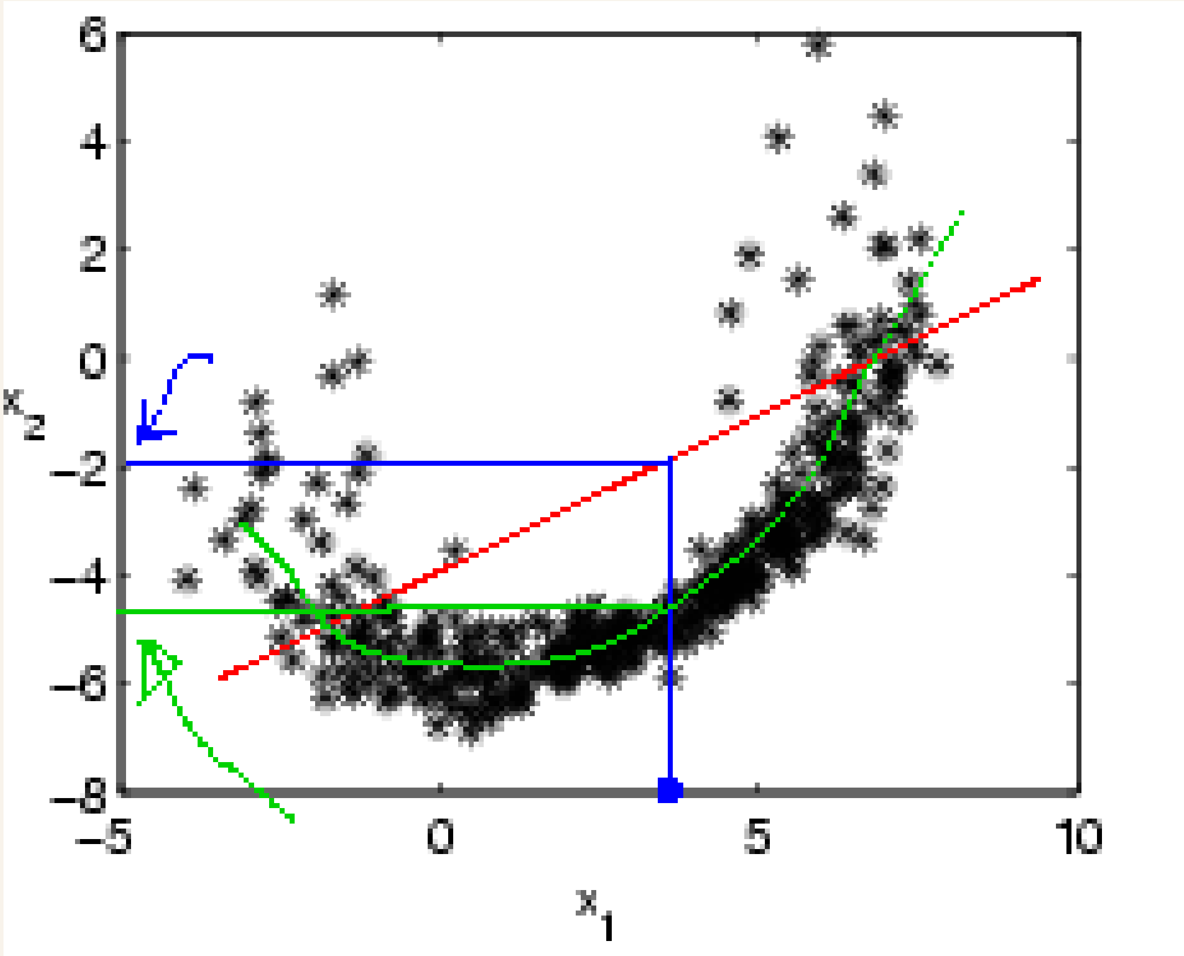
This is not a straight line world!

Focus on (X1, X2) ensemble members

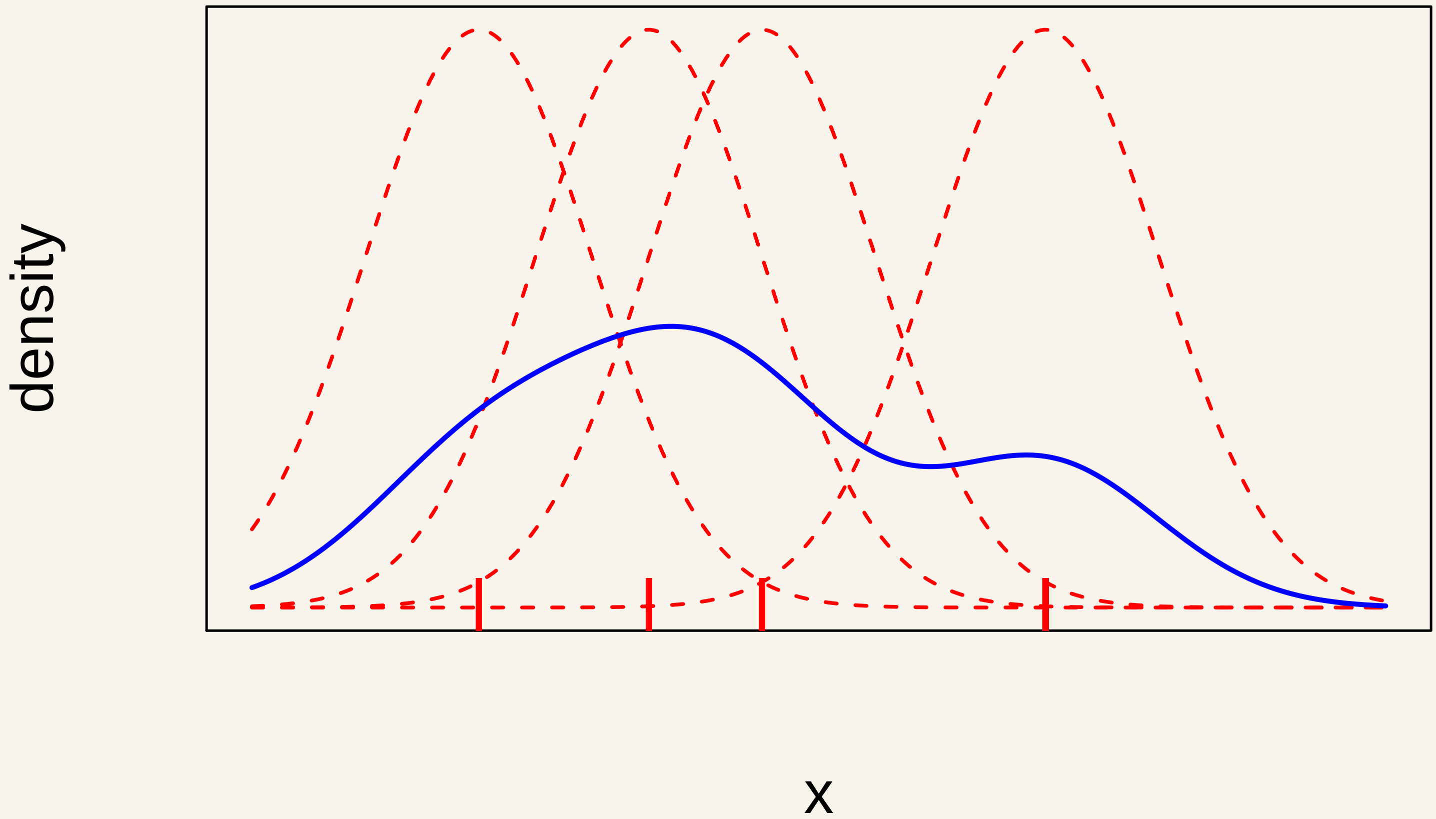


This is not a straight line world!

Focus on (X1, X2) ensemble members



Mixtures of Normals



Mixture Filters

Ensemble Kalman Filter can be extended in an efficient way from a Gaussian to mixtures of Gaussians.

Basic ingredients for a hybrid local filter.

- Sequentially assimilate observations.
- Apply mixture filter to a low dimensional set of state components "close" to the observations.
- Update other components of the state vector using the Gaussian EKF but *conditional* on the close updates.

Mixture Filters

Ensemble Kalman Filter can be extended in an efficient way from a Gaussian to mixtures of Gaussians.

Basic ingredients for a hybrid local filter.

- Sequentially assimilate observations.
- Apply mixture filter to a low dimensional set of state components "close" to the observations.
- Update other components of the state vector using the Gaussian EKF but *conditional* on the close updates.

Performance (in Lorenz 1996) improves over usual EnKF.
Multi-modal ensemble forecasts can be preserved in the update.

Multiresolution and nonstationary spatial covariances

Problem:

Traditional methods of representing spatial covariances break down for nonstationary fields.

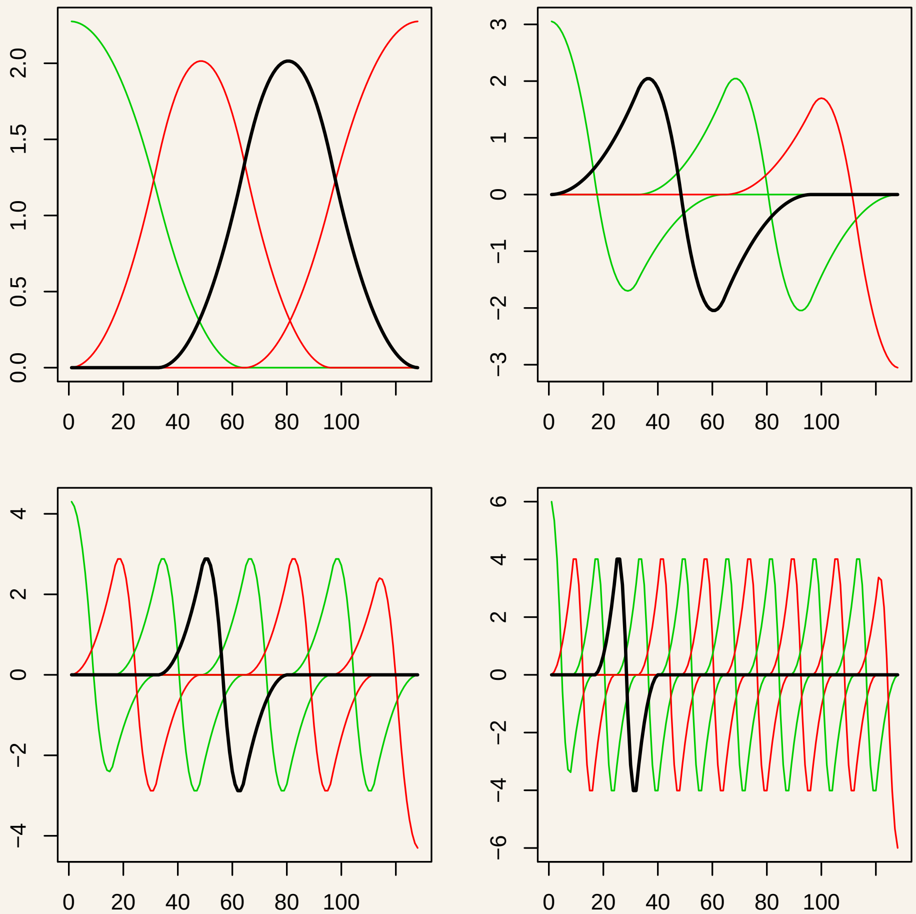
Also, traditional models are computationally infeasible applied to large numbers of locations.

Goal:

Use wavelet based representations to capture both nonstationary structure and to facilitate computing.

Creating a random function or field.

A (wavelet) Basis



Basis functions $\{\psi_j\}$,

$Random$ coefficients, $\{a_j\}$

$$z(x) = \sum_j a_j \psi_j(x)$$

Why wavelets?

Reduction in complexity

Many random functions and fields may be succinctly described by just a few correlations among the coefficients.

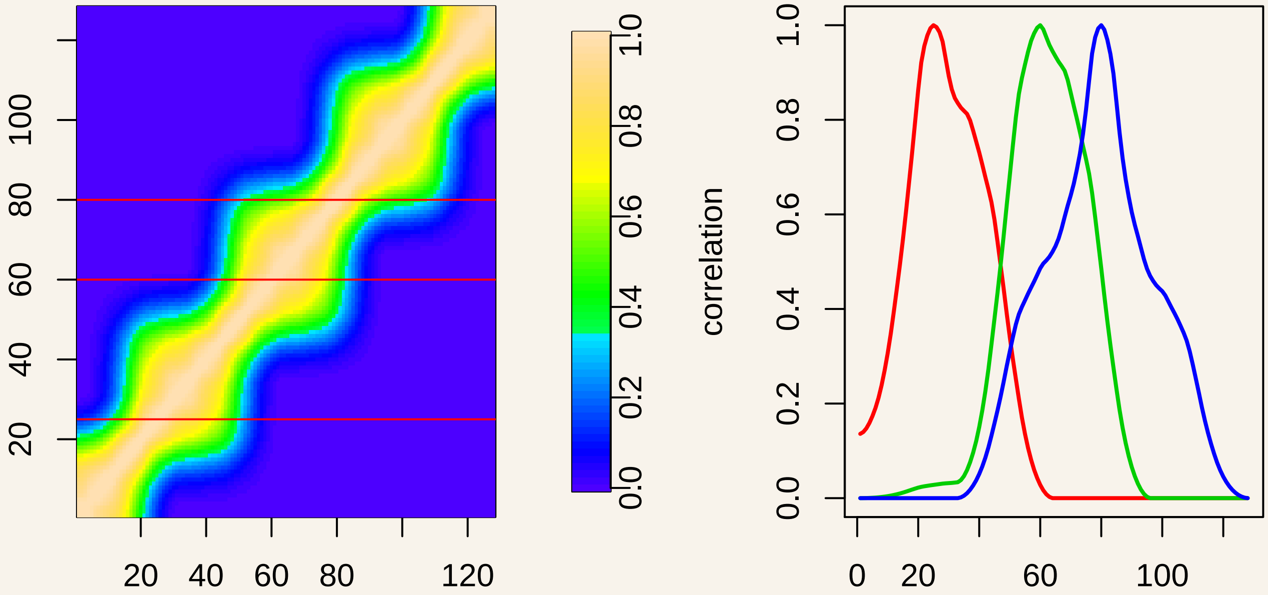
Local support:

Wavelets can represent nonstationary covariances easily.

Another good thing:

If the covariances among the $\{a_j\}$ are sparse, this leads to efficient numerical algorithms. e.g. generating *ensembles*, updating

BAU: a naive correlation matrix.



Assuming the coefficients are uncorrelated. *Strange!*

The goal is to craft wavelet models to reproduce reasonable covariances.

Some Results

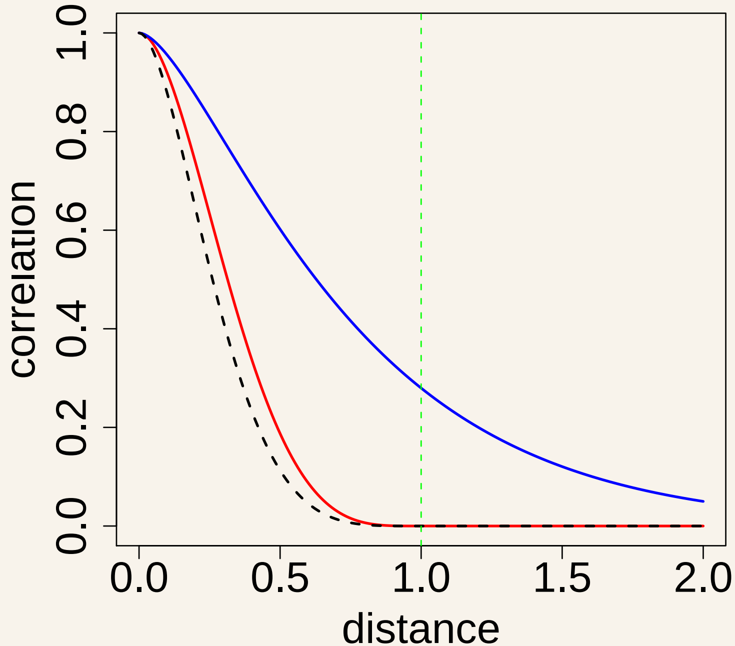
The key is to add some off diagonal correlations among the wavelet coefficients ... but not too many!

- It is possible to approximate a wide range of 2-d isotropic covariance models using wavelets. The number of nonzero covariance elements appears to scale linear in grid size.
- The EM algorithm for "missing" data can be adapted to handle estimating covariances for irregularly spaced locations.

Covariance Tapering

Tapering a covariance matrix, (Σ), introduces zeroes by multiplying the covariance elements directly by a tapering function (T).

$$\Sigma_{i,j}^{tapered} = \Sigma_{i,j} * T_{i,j}$$

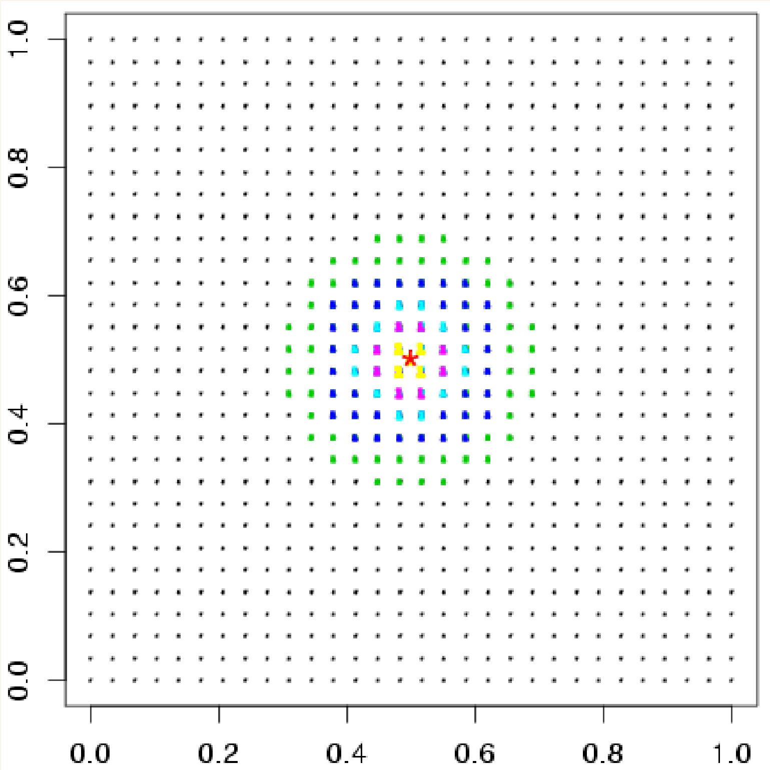


Results:

Judicious tapering does not greatly compromise the "optimal" properties of the Kalman Filter or Optimal Interpolation.

This is confirmed numerically with simulations and also based on asymptotic theory for spatial processes.

Updates using nearest neighbors can be as effective as the entire set of observations.



Future Work

The extensions will provide new DA methods that can be tested in the DART framework.

- Implement the nonGaussian filter in DART.
- Apply the wavelet covariance models to archived NCEP forecasts to estimate the background covariance.
- Determine the best way to taper a sample covariance matrix that is nonstationary.