Regularization of the ensemble Kalman filter with constrained non-stationary convolutions

M Tsyrulnikov, A Sotskiy, D Gayfulin

HydroMetCenter of Russia

EnKF Workshop. Online, 10 June 2021

Regularization

The *sample covariance matrix* is a good estimator of the true covariance matrix — **if** the ensemble size is much greater than the size of the matrix, which is never the case in high-dimensional problems.

The (small) ensemble simply contains too little information to reliably estimate the covariance matrix.

Hence, additional information is needed $\Rightarrow$ *regularization*.

Popular covariance regularization techniques and our proposal

1. Domain localization
2. Covariance localization
3. Mixing with climatological (time mean) covariances

These techniques are simple and effective, but

(i) they are ad hoc, there are no underlying stochastic models, no optimality criteria satisfied.

(ii) they require tuning. In a practical "heavy-weight" system tuning hundreds of interacting parameters can be problematic.

(iii) they, basically, first, allow the noise to contaminate the signal (by relying on sample covariances) and, then, apply a device to filter out the noise.

In this research, we propose a *model* for forecast-error covariances and an estimator of a square root of the covariance matrix directly from the ensemble.

Non-stationarity

The key feature of the prior distribution that allows EnKF to thrive is non-stationarity (both in time and in space).

(If there is no non-stationarity, we just have to carefully estimate the time-mean prior covariances and then use them every time.)

So, the spatial model we wish to build is to be non-stationary in space (and, of course, in time).

Approach

We aim to build

- a non-stationary stochastic model for the spatial forecast error field
- an affordable in high dimensions estimator of the model given the prior ensemble.

The ultimate goal is to use the new technique for practical data assimilation in meteorology.

The non-stationary spatial model

Process convolution model

Let the forecast-error field $\xi(x)$ be the general linear Gaussian process:

$$\xi(x) = \int w(x,y)\,\alpha(y)\,\mathrm{d}y$$

(α is the white noise). Its space-discrete counterpart is

$$\boldsymbol{\xi} = \mathbf{W}\boldsymbol{\alpha}$$

(with $\boldsymbol{\alpha} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$), so that

$$\mathbf{B} = \mathbf{W}\,\mathbf{W}^\top$$

The model is overcomplete (too general): with any orthogonal matrix $\mathbf{Q}$, $\mathbf{W}' = \mathbf{W}\mathbf{Q}$ is another "square root" of $\mathbf{B}$.
So, $\mathbf{W}$ needs to be *constrained* to become unique.
Another required feature of $\mathbf{W}$ is *sparsity*.

Stationary process convolution model

•$w(x, y) = u(x - y)$ corresponds to a homogeneous model on $\mathbb{R}^d$ or $\mathbb{S}^1$.
•$w(x, y) = u(\rho(x, y))$ corresponds to an isotropic model on $\mathbb{R}^d$ or $\mathbb{S}^2$.
•$u(x)$ is the *convolution kernel*.

To simplify the presentation, let us consider the process on the unit circle $\mathbb{S}^1$:

$$\xi(x) = \int u(x - y)\, \alpha(y)\, \mathrm{d}y$$

In spectral space, with $\xi(x) = \sum \widetilde{\xi}_\ell \, \mathrm{e}^{\mathrm{i}\ell \mathrm{x}}$:

$$\widetilde{\xi}_\ell \propto \widetilde{u}_\ell \, \widetilde{\alpha}_\ell$$

The *spectrum* of $\xi(x)$ is then

$$f_\ell := \mathbb{E}\, |\widetilde{\xi}_\ell|^2 \propto |\widetilde{u}_\ell|^2$$

Given the spectrum f_ℓ, the ambiguity in $u(x)$ comes here from the *modulus* of $\widetilde{u}_\ell$.

Selecting a unique stationary model

For computational reasons, we need the kernel $u(x)$ to be as *localized* as possible. Therefore, we require that $u(x)$ is the narrowest kernel among all that satisfy $|\widetilde{u}_\ell|^2 \propto f_\ell$ with the fixed $\{f_\ell\}$.

It can be shown that the unique narrowest kernel has the real and non-negative Fourier transform. Correspondingly, the narrowest kernel $u(x)$ is a positive-definite function.

We postulate this feature for the non-stationary model as well.

From stationary to non-stationary model

Stationary:

$$\xi(x) = \int u(\rho(x,y))\,\alpha(y)\,\mathrm{d}y$$

Let the kernel u also depend on the location x (Higdon 2002):

$$\boxed{\xi(x) = \int u(x, \rho(x,y))\,\alpha(y)\,\mathrm{d}y}$$

The non-stationary model

$$\text{On } \mathbb{S}^1 \text{ and } \mathbb{S}^2\text{: } \xi(x) = \int u(x, \rho(x,y))\, \alpha(y)\, \mathrm{d}y$$

where $u(x, \rho)$ is a positive-definite function of its 2nd argument ρ (this is our **first constraint**). Fourier transforming $u(x, \rho)$ w.r.t. ρ yields the equivalent definition of the model:

$$\text{On } \mathbb{S}^1\text{:} \qquad \xi(x) = \sum_{\ell=-\ell_{\max}+1}^{\ell_{\max}} \sigma_\ell(x)\, \widetilde{\alpha}_\ell\, \mathrm{e}^{\mathrm{i}\ell \mathrm{x}}$$

$$\text{On } \mathbb{S}^2\text{:} \qquad \xi(x) = \sum_{\ell=0}^{\ell_{\max}} \sigma_\ell(x) \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m}\, Y_{\ell m}(x)$$

where $\widetilde{\alpha}_\ell, \widetilde{\alpha}_{\ell m}$ are uncorrelated zero-mean unit-variance random variables, $Y_{\ell m}$ are spherical harmonics, and, in both cases, $\sigma_\ell(x) \geq 0$.

When $\sigma_\ell(x) = \sigma_\ell$ (i.e. independent of x), the model Eq.(*) becomes *stationary*.

Local spectrum

On the circle (similarly, on the sphere), having the model

$$\boxed{\xi(x) = \sum \sigma_\ell(x)\, \widetilde{\alpha}_\ell\, \mathrm{e}^{\mathrm{i}\ell\mathrm{x}}} \qquad (*)$$

with *independent* $\widetilde{\alpha}_\ell$, we have

$$\mathrm{Var}\,\xi(x) = \sum \sigma_\ell^2(x) = \sum f_\ell(x)$$

so that we can call $f_\ell(x) = \sigma_\ell^2(x)$ the local spectrum and the model Eq.(*) the Local Spectrum Model (LSM).

Weakly non-stationary model

The model

$$\xi(x) = \sum \sigma_\ell(x)\, \widetilde{\alpha}_\ell\, \mathrm{e}^{\mathrm{i}\ell \mathrm{x}}$$

is unique but it still has too many degrees of freedom (the functions $\sigma_\ell(x)$) to be reliably estimated from the ensemble.

Therefore, our **second constraint** is the assumption that the **structure** of the process (i.e. the local spectra $\{\sigma_\ell^2(x)\}$ or the kernel $u(x,\rho)$) varies in space (i.e. with x) on a scale **significantly larger than the length scale of the process itself**.

This greatly reduces the number of degrees of freedom, makes the process weakly non-stationary or locally stationary (*cf.* evolutionary spectrum by Priestley 1965).

Smooth spectra

Finally, we assume that $\{f_\ell = \sigma_\ell^2(x)\}$ varies smoothly in wavenumber space space (i.e. with ℓ). This is our **third constraint**, which further reduces the number of degrees of freedom to be estimated from the ensemble.

Summary of the spatial model

The Local Spectrum Model is a process convolution model with the spatially variable kernel $u(x, \rho)$:

$$\xi(x) = \int u(x, \rho(x, y))\, \alpha(y)\, \mathrm{d}y$$

satisfying:

1. $u(x, \rho)$ is a positive definite function of the distance ρ
2. $u(x, \rho)$ a smooth function of the physical space location x
3. The Fourier image of $u(x, \rho)$ w.r.t. ρ, i.e. $\sigma_\ell(x)$ is a smooth function of the wavenumber ℓ.

Estimation of the spatial model from the ensemble

Multi-scale bandpass filter

Motivation: as $f_\ell = \sigma_\ell^2(x)$ are constrained to be smooth functions of the wavenumber ℓ, measuring the spectrum f_ℓ averaged over a few wavenumber bands would suffice to recover the whole spectrum.

We introduce $J = 5...10$ linear bandpass filters with the spectral transfer functions $H_j(\ell)$ and impulse response functions $h_j(\rho)$ $(j = 1, \dots, J)$. With the weakly non-stationary process

$$\xi(x) = \sum_{\ell=0}^{\ell_{\max}} \sigma_\ell(x) \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m} \, Y_{\ell m}(x),$$

if $\sigma_\ell(x)$ only slightly change within the effective support of h_j (which is a more specific formulation of our **second constraint**), then the bandpass filtered processes satisfy

$$\xi_{(j)}(x) \approx \sum_{\ell=0}^{\ell_{\max}} H_j(\ell) \, \sigma_\ell(x) \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m} \, Y_{\ell m}(x)$$

With this equation and the standard (complex) Gaussian $\widetilde{\alpha}_{\ell m}$, we can write down the pointwise likelihood of $\sigma_\ell(x)$ given the filtered data, i.e. $p(\xi_{(1:J)}(x) \,|\, \sigma_\ell(x))$.

Specification of the bandpass filters

Requirements:

1. The bandpass filters should have narrow enough impulse response functions – to resolve non-stationary structures in physical space.

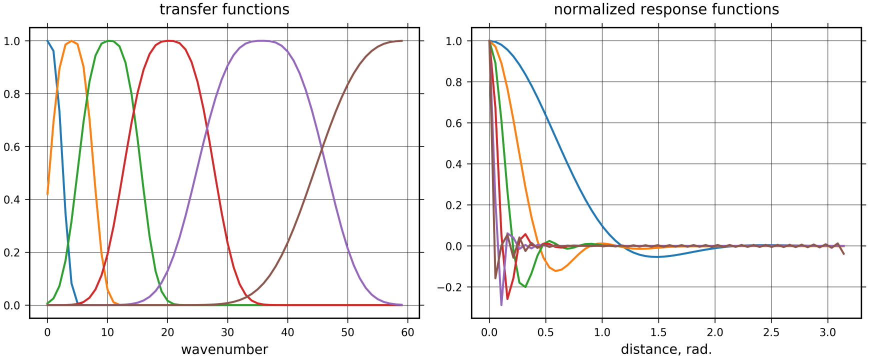
2. The bandpass filters should have narrow enough spectral transfer functions – to have good resolution in spectral space.

As a compromise between items 1 and 2, we found that the transfer functions

$$H_j(\ell) = \mathrm{e}^{-\left|\frac{\ell - \ell_j^c}{d_j}\right|^q}$$

(with $q = 2...3$) work well.

Spectral transfer and impulse response functions of the multi-scale filter



Estimator

Working at each grid point x independently, we aim to estimate $\{\sigma_\ell(x)\}$.

Employing a Bayesian approach, we may estimate/specify a prior distribution for $p(\sigma_\ell(x))$ and use the likelihood $p(\xi_{(1:J)}(x) \,|\, \sigma_\ell(x))$ to obtain the posterior

$$p_{\rm post}(\sigma_\ell(x)) \propto p(\sigma_\ell(x))\; p(\xi_{(1:J)}(x) \,|\, \sigma_\ell(x))$$

and an optimal estimate.

We did this in a parametric setting and it worked (despite the posterior density is not convex), but the following simpler approach proved to be more effective (and more appropriate with real-world high-dimensional problems).

Estimator: a simplified approach

$$\operatorname{Var}\xi_{(j)}(x) \approx \sum_{\ell=0}^{\ell_{\max}} H_j^2(\ell)\, f_\ell(x)$$

$$\boxed{\mathbf{\Omega}\,\mathbf{f} = \mathbf{v}}$$

where $(\mathbf{\Omega})_{j\ell} = H_j^2(\ell)$, $\mathbf{f}$ is the variance spectrum vector, and $(\mathbf{v})_j = \operatorname{Var}\xi_{(j)}(x)$.
We use the pseudo-inverse solution

$$\mathbf{\Omega} = \mathbf{U\Sigma V}^\top \quad \Rightarrow \quad \boxed{\widehat{\mathbf{f}} = \mathbf{\Omega}^+\,\mathbf{v} = \mathbf{V\Sigma}^+\mathbf{U}^\top\,\mathbf{v},}$$

and fit a two-parameter model $\boxed{A \cdot g(\ell/a)}$ to $\widehat{\mathbf{f}}$.

Finally, the local spectra are transformed (again, at each x independently) to the kernels $u(x,\rho)$, which, in turn, yield the $\mathbf{W}$ matrix (the square root of $\mathbf{B}$).

The problem is **low-dimensional** if solved for each x independently, and this can be done **in parallel**.

Implementation in the filter

Summary of the analysis algorithm

1. Apply the multi-scale bandpass filter to the prior ensemble and compute the band variances at all spatial grid points (we employed spectral filtering).
2. From the band variances, compute the local spectra (using the pseudo-inversion).
3. From the local spectra, compute the kernels, i.e. the $\mathbf{W}$ matrix (using the inverse spectral transform).
4. Perform a kind of thresholding of $\mathbf{W}$, getting a sparse matrix.
5. Use $\mathbf{W}$ to compute the gain matrix:

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{R}^{-1} = \mathbf{W}(\mathbf{I} + \mathbf{W}^\top \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{H}^\top \mathbf{R}^{-1}.$$

   This numerical scheme enjoys sparsity, provides efficient pre-conditioning, and has no rank deficiency problem.

We call the resulting filter the Local Spectrum Ensemble Filter (LSEF).

The posterior ensemble is computed as in the classical stochastic EnKF (at this stage of development).

Numerical experiments

Three non-stationary models of truth

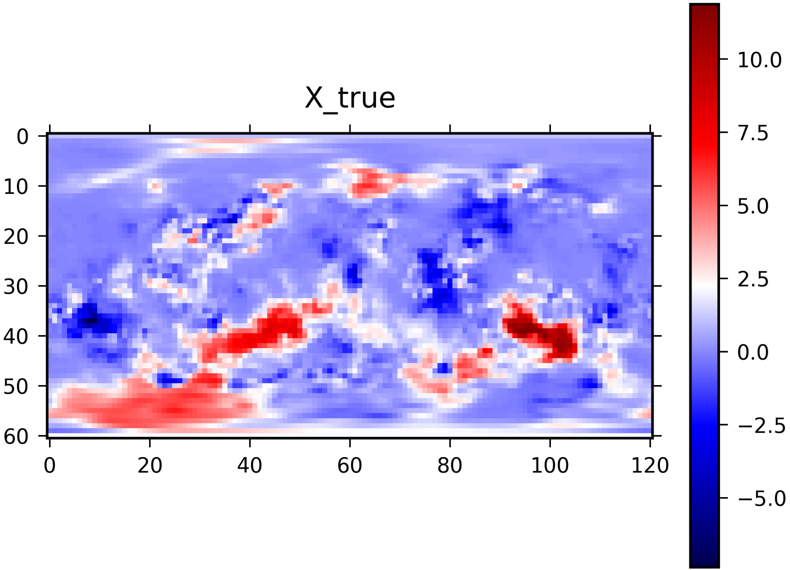
1 static model on the circle + 1 static model on the sphere + 1 dynamic model on the circle.

All three models are hierarchical (doubly stochastic), with random *parameter fields* and conditionally Gaussian true fields.

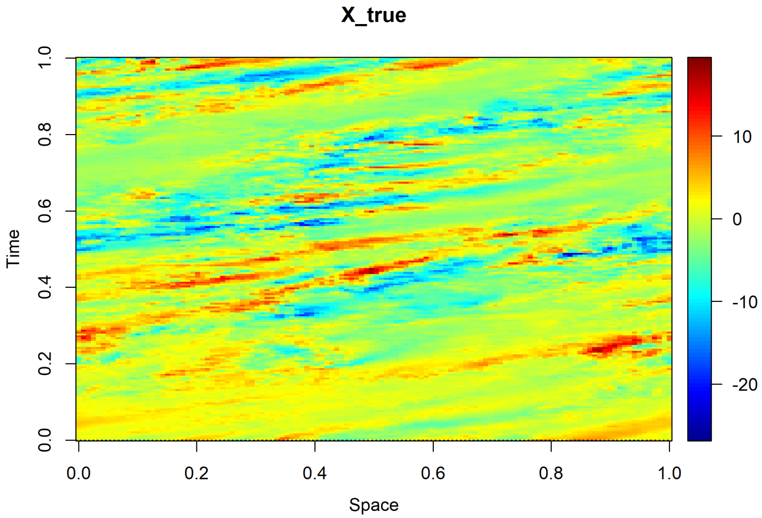
1. $\mathbb{S}^1$ static.
   Kernel: $u(x, \rho) = S(x) \cdot u_0(\rho / L(x))$
   Two *parameter fields* : $L(x)$ and $S(x)$
2. $\mathbb{S}^2$ static.
   Local spectrum: $f_\ell(x) = \frac{c(x)}{1+(\lambda(x)\ell)^{\gamma(x)}}$
   Three *parameter fields* : $c(x), \lambda(x), \gamma(x)$
3. $\mathbb{S}^1$ dynamic.
   A doubly stochastic advection-diffusion-decay model (Tsyrulnikov and Rakitko, 2019, QJRMS).
   Three spatio-temporal *parameter fields*: $U(x,t), \nu(x,t), \delta(x,t)$.

The *parameter fields* are logit-Gaussian.

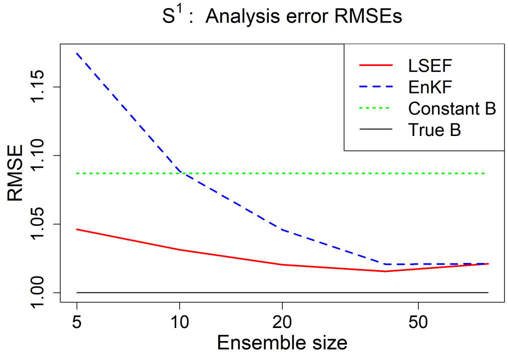
An example of the non-stationary spatial field on $\mathbb{S}^2$

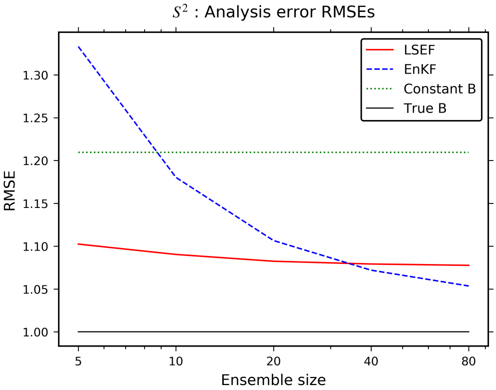


An example of the non-stationary spatio-temporal field on $\mathbb{S}^1$

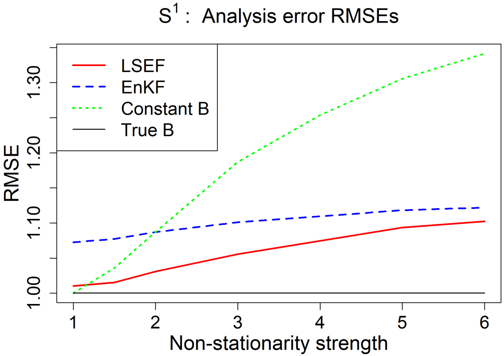


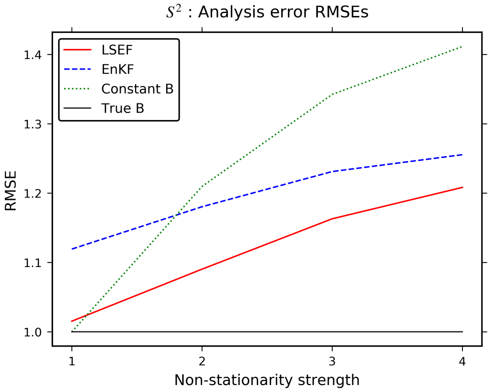
Filters' performance: static setup, dependence on ensemble size



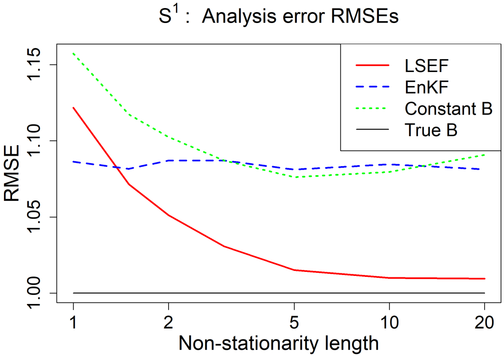


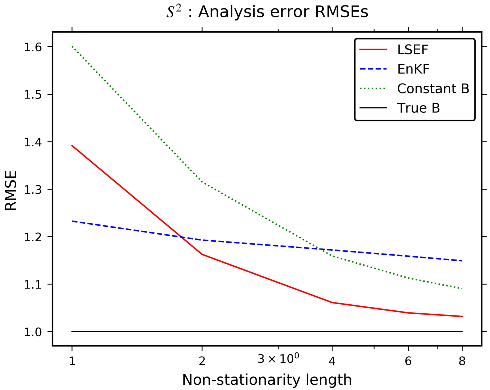
Filters' performance: static setup, dependence on Non-Stationarity Strenth



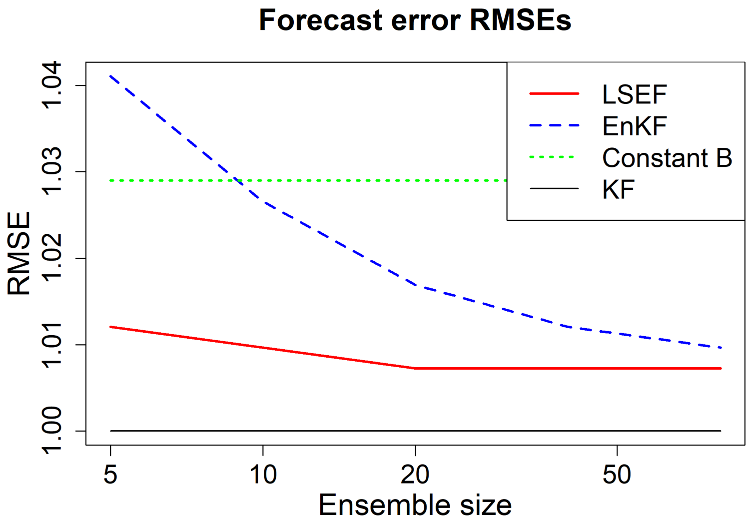


Filters' performance: static setup, dependence on Non-Stationarity Length

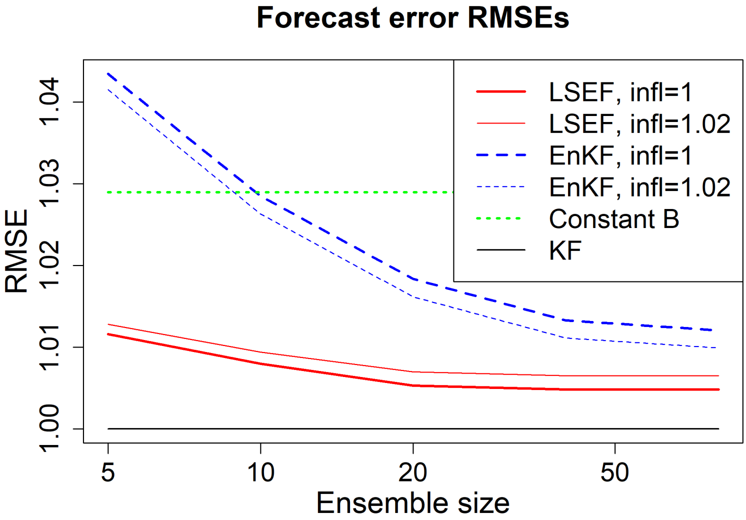




Filters' performance: $\mathbb{S}^1$, dynamic setup



LSEF needs no covariance inflation



Conclusions

- The Local Spectrum Ensemble Filter (LSEF) estimates the gain matrix directly from the prior ensemble using a constrained non-stationary spatial convolution model.
- The constraints imposed on the convolution model include slow variation of the kernel in physical space and smoothness of the (parametric or non-parametric) kernel in spectral space.
- The estimation of the convolution model is performed in spectral space: the local spectrum is estimated gridpoint by gridpoint from the output of a multi-scale bandpass filter.
- In numerical experiments with two spatial models of truth (on the circle and on the sphere) and a spatio-temporal model on the circle, the LSEF outperformed the standard stochastic EnKF for small to moderate ensembles and under weak to moderate non-stationarities.
- The technique is computationally tractable: its *non-ensemble* version has been used at our center for operational meteorological data assimilation for several years.

The End