Advantages and disadvantages with finite-ensemble sizes

→ When the observation time frequency is short enough, the prior is close to Gaussian.
→ When the nonlinearity is too strong the prior is no longer Gaussian and the update becomes suboptimal.

But what is the right ensemble size?

→ In meteorology we hope for a number of 100.
→ Particle filters are prohibited with this ensemble size.
→ The sampling errors in the covariance underestimate the covariance even in the perfect model.
   This is because of the strong nonlinearity that causes big ensembles to collapse into a single direction.
→ Moreover, when the ensemble size is small, there are significant sampling errors due to spurious correlations from distant grid points.
   → These can be prevented using various physical space localized updates – This kind of localized updates can be expensive.

→ Don’t forget that L63 model is very special – is weakly mixing with one unstable direction on an attractor with high symmetry.
→ Realistic systems have a very large phase space with a high d unstable manifold on the attractor.
Systematic strategies for real time filtering of turbulent signals in complex systems

Filtering Turbulent Nonlinear Dynamical Systems by Finite Ensemble Methods
The Lorenz 96 model

- Introduced by Lorenz to represent an atmospheric variable at 2N equally spaced points around a circle of constant latitude (periodic boundary conditions).

\[
\frac{du_j}{dt} = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F
\]

Satisfies three basic principles:

1. It has linear dissipation, that decreases the total energy
2. An external forcing that can increase or decrease the total energy
3. A quadratic discrete advection-like term that conserves energy.
The Lorenz 96 model

Dynamical – statistical properties over different dynamical regimes

Table 11.1: Dynamical properties of L-96 model for regimes with $F = 6, 8, 16$. $\lambda_1$ denotes the largest Lyapunov exponent, $N^+$ denotes the dimension of the expanding subspace of the attractor, $KS$ denotes the Kolmogorov-Sinai entropy, and $T_{corr}$ denotes the correlation time of the energy-rescaled time correlation function.

<table>
<thead>
<tr>
<th>$F$</th>
<th>$\lambda_1$</th>
<th>$N^+$</th>
<th>$KS$</th>
<th>$T_{corr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1.02</td>
<td>12</td>
<td>5.54</td>
<td>8.23</td>
</tr>
<tr>
<td>8</td>
<td>1.74</td>
<td>13</td>
<td>10.94</td>
<td>6.70</td>
</tr>
<tr>
<td>16</td>
<td>3.94</td>
<td>16</td>
<td>27.94</td>
<td>5.59</td>
</tr>
</tbody>
</table>

$T_{corr} = \int_0^\infty |\langle (u_j(t) - \bar{u})(u_j(t + \tau) - \bar{u}) \rangle_t |d\tau,$

$N^+$ Unstable directions of the attractor

$\lambda_1$ Largest Lyapunov exponent

$KS$ Kolmogorov-Sinai entropy (sum of all positive Lyapunov exponents) which measures information loss
The Lorenz 96 model

Weakly skewed statistics
Ensemble square root filters on L96

We discuss both ensemble square root filters ETKF and EAKF

To avoid particle collapse we fix the variance inflation coefficient to \( r = 0.05 \)

\( P = N/M \) is the ratio of number of positive Fourier modes of the model to the number of positive Fourier modes of the observations.

**Plentiful observations**

The two filters are comparable

For the fully turbulent regime ETKF is much better.

**Sparse observations**

Overall, EAKF is more accurate compared to ETKF for sparse observations (\( P > 1 \)). Furthermore, EAKF rarely suffers from catastrophic filter divergence while ETKF is very prone to this problem.
Ensemble square root filters on L96

Table 11.2: \( F = 6, T_{\text{obs}} = 0.078 \) and 0.234, \( r^o = 0.01, 3 \). The quantity in bracket near \( \infty \) denotes the number of cycles the filter go through before it blows up.

<table>
<thead>
<tr>
<th>p</th>
<th>( T_{\text{obs}} = 0.078, \ r^o = 0.01 )</th>
<th>( T_{\text{obs}} = 0.234, \ r^o = 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EAKF RMS</td>
<td>corr</td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.99</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td>( T_{\text{obs}} = 0.078, \ r^o = 3 )</td>
<td>( T_{\text{obs}} = 0.234, \ r^o = 3 )</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.34</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>0.54</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.77</td>
<td>0.96</td>
</tr>
<tr>
<td>4</td>
<td>1.14</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>1.54</td>
<td>0.84</td>
</tr>
<tr>
<td>No Obs</td>
<td>2.84</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
Table 11.3: $F = 8$, $\Delta t = 0.078$ and 0.234, $r^o = 0.01, 3$. The quantity in bracket near $\infty$ denotes the number of cycles the filter go through before it blows up.

<table>
<thead>
<tr>
<th>p</th>
<th>$T_{obs} = 0.078, r^o = 0.01$</th>
<th>$T_{obs} = 0.078, r^o = 3$</th>
<th>$T_{obs} = 0.234, r^o = 0.01$</th>
<th>$T_{obs} = 0.234, r^o = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EAKF</td>
<td>RMS</td>
<td>corr</td>
<td>ETKF</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>1</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>0.03</td>
<td>1</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.99</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
<td>0.99</td>
<td></td>
<td>$\infty(23)$</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.99</td>
<td></td>
<td>$\infty(9)$</td>
</tr>
</tbody>
</table>

| No Obs | 3.66  | 0.01 |      |      |      |      |      |      |      |      |      |      |

Ensemble square root filters on L96
Ensemble square root filters on L96

Table 11.4: $F = 16$, $\Delta t = 0.078$ and 0.234, $r^o = 0.01$, 3. The quantity in bracket near $\infty$ denotes the number of cycles the filter go through before it blows up.

<table>
<thead>
<tr>
<th></th>
<th>$T_{obs} = 0.078$, $r^o = 0.01$</th>
<th>$T_{obs} = 0.234$, $r^o = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EAKF</td>
<td>ETKF</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>corr</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>0.09</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>$T_{obs} = 0.078$, $r^o = 3$</td>
<td>$T_{obs} = 0.234$, $r^o = 3$</td>
</tr>
<tr>
<td>1</td>
<td>3.28</td>
<td>0.86</td>
</tr>
<tr>
<td>2</td>
<td>6.68</td>
<td>0.56</td>
</tr>
<tr>
<td>3</td>
<td>6.43</td>
<td>0.57</td>
</tr>
<tr>
<td>4</td>
<td>7.07</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>6.20</td>
<td>0.45</td>
</tr>
<tr>
<td>No Obs</td>
<td>6.35</td>
<td>0</td>
</tr>
</tbody>
</table>
11.4 The two-layer quasi-geostrophic model

11.5 Local least-square EAKF

**Atmosphere**
Goddard Earth Observing System Model, earthobservatory.nasa.gov

**Ocean**
Ross Tulloch, Eddying Double Drake (ocean only model). This picture comes from http://ocean.mit.edu/~tulloch/Research.html
The two-layer quasi-geostrophic model

- Both mid-latitude (30~65 degree) atmospheric and oceanic turbulence dynamics

- Small Rossby number
  \[ Ro = \frac{\text{advection term}}{\text{rotational term}} = \frac{U}{fL} < 1 \]

- Comparable motion length scale relative to the radius of deformation \( L \sim L_d \)

- Small variation of the Coriolis parameter \( |\beta L| \ll f_0 \)

- Time scales advectively \( T = \frac{L}{U} \)
The two-layer quasi-geostrophic model

As a simple paradigm, we consider the two-layer quasi-geostrophic (QG) model in a double periodic domain with instability induced by mean vertical shear (Salmon, 1998). The properties of the turbulent cascade have been extensively discussed in this setting, e.g. see Salmon (1998) and the citations in Smith et al. (2002). The governing equations for the two-layer QG model with a flat bottom, rigid lid and equal depth layers $H$ can be written as

\[
\begin{align*}
\frac{\partial q_1}{\partial t} + J(\psi_1, q_1) + U \frac{\partial q_1}{\partial x} + (\beta + k_d^2 U) \frac{\partial \psi_1}{\partial x} + \nu \nabla^s q_1 &= 0, \\
\frac{\partial q_2}{\partial t} + J(\psi_2, q_2) - U \frac{\partial q_2}{\partial x} + (\beta - k_d^2 U) \frac{\partial \psi_2}{\partial x} + \kappa \nabla^2 \psi_2 + \nu \nabla^s q_2 &= 0,
\end{align*}
\]

where $k_d = \sqrt{8/L_d}$ is the wavenumber corresponding to the Rossby radius of deformation $L_d$; the coefficient $\kappa$ is the Ekman bottom drag coefficient; and the dissipative operator $\nu \nabla^s q$ is added to filter out the energy buildup on the smaller scales. Here, the hyperviscosity coefficient $\nu$ is chosen such that it only damps the smaller scale. Note

\[
J(\psi, q) = \psi_x q_y - \psi_y q_x
\]

\[
q_i = \beta y + \nabla^2 \psi_i + \frac{k_d^2}{2} (\psi_{3-i} - \psi_i), \quad i = 1, 2,
\]

where

stream function $\Psi_1 = -U_y$, $\Psi_2 = U_y$ as the background state.
The two-layer quasi-geostrophic model

In our numerical simulations, the true signal is generated by resolving Eqns (11.5) with 63 modes meridionally and zonally, which corresponds to \(128 \times 128 \times 2\) grid points. With such resolution, the numerical integration of this turbulent system only needs slightly more than 30,000 state variables since one can always compute \(\psi\) from knowing \(q\) and \(\theta\).

Recall the QG model

\[
\frac{\partial q_1}{\partial t} + J(\psi_1, q_1) + U \frac{\partial q_1}{\partial x} + (\beta + k_d^2 U) \frac{\partial \psi_1}{\partial x} + \nu \nabla^2 q_1 = 0,
\]

\[
\frac{\partial q_2}{\partial t} + J(\psi_2, q_2) - U \frac{\partial q_2}{\partial x} + (\beta - k_d^2 U) \frac{\partial \psi_2}{\partial x} + \kappa \nabla^2 \psi_2 + \nu \nabla^2 q_2 = 0,
\]

Table 11.5 Parameter values for the numerical experiments;

\((k, \ell)_{\text{max}}\) is the horizontal wavenumber associated with the maximum growth rate, \(\tilde{\sigma}_{\text{max}} = \text{Im}[c]k\), where \(c\) is the wave speed.

<table>
<thead>
<tr>
<th>Regime</th>
<th>(b)</th>
<th>(F)</th>
<th>(U)</th>
<th>(\kappa)</th>
<th>(\tilde{\sigma}_{\text{max}})</th>
<th>((k, \ell)_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere</td>
<td>2</td>
<td>4</td>
<td>0.2(U_o)</td>
<td>0.1</td>
<td>0.21</td>
<td>(3,0)</td>
</tr>
<tr>
<td>Ocean</td>
<td>2</td>
<td>40</td>
<td>0.1(U_o)</td>
<td>0.1</td>
<td>0.48</td>
<td>(9,0)</td>
</tr>
<tr>
<td>Atmosphere</td>
<td>2</td>
<td>4</td>
<td>0.2(U_o)</td>
<td>0.2</td>
<td>0.18</td>
<td>(3,0)</td>
</tr>
</tbody>
</table>
The two-layer quasi-geostrophic model

(see Salmon (1998) and fig. 1 in Kleeman and Majda (2005)). In the atmosphere case, the
barotropic mode, which provides the bulk depth-averaged component of the flow with the
stream function $\psi_b = (\psi_1 + \psi_2)/2$, is dominated by a strong zonal jet (see the barotropic
velocity vector field and stream function in Fig. 11.7). On the other hand, the baroclinic
modes govern the transport of heat and have the stream function $\psi_c = (\psi_1 - \psi_2)/2$. 
Strong zonal jets

Rossby wave

ATM regime $F=4$ at time = 2500

OCN regime $F=40$ at time = 100

ATM regime $F=4$ at time = 5000

OCN regime $F=40$ at time = 200
The two-layer quasi-geostrophic model
The two-layer quasi-geostrophic model

For both the atmosphere and ocean cases, the most energetic modes are on the large-scale barotropic modes (see Fig. 11.8) in the statistical equilibrium state (see Kleeman and Majda (2005) for a discussion of the energy cascades in these two cases with topography).

The horizontal axis in Fig. 11.8 corresponds to the barotropic modes, ordered according to empirical orthogonal functions (EOFs) using the barotropic stream function variance, i.e. from the largest to the smallest variance (averaged over a long period of time). In the atmosphere case, 95\% of the energy or variability is represented by modes 1–5, which corresponds to the two-dimensional horizontal Fourier modes

\[(0,1), \quad (1,1), \quad (-1,1), \quad (0,2), \quad (1,0)\]

**Large-scale zonal jet modes**

**Rossby mode**

In the ocean case, the first five modes are ordered as follows: \((1,0), \quad (0,1), \quad (1,1), \quad (-1,1), \quad \text{and} \quad (0,2)\). Here the large-scale zonal jet modes carry

In the atmosphere with stronger bottom drag, the 5\textsuperscript{th} most energetic mode is \((2,0)\). The mode \((1,0)\) become 8\textsuperscript{th} mode
The two-layer quasi-geostrophic model

In Fig. 11.9, we show the marginal pdfs of the first five and the eighth most energetic modes for both the atmosphere with weaker and stronger bottom drag and the ocean cases. These marginal pdfs are generated through bin-counting the barotropic stream function, centered at 0 (each panel in Fig. 11.9 shows a histogram of $d\psi = \hat{\psi}_b - \langle \hat{\psi}_b \rangle$) and they encompass solutions of Eqns (11.5) up to $T = 10,000$ time units at every 0.25 time interval in the atmosphere case and up to $T = 400$ time units at every 0.01 time interval in the ocean case.
Marginal probability density distribution

Solid: real part
Dashed: imaginary part
Correlation functions of the barotropic stream function as functions of time

Solid: real part
Dashed: imaginary part
Local least-square EAKF

**Advantage**
- It produces very accurate solutions in many contexts
- It does not require singular value decomposition
  (see Anderson 2003)

A least-squares relation between prior distribution of an observation and model state variables

**Two Steps**
- Update the prior ensemble estimate of the observation variables with a scalar ensemble filter
- Perform a linear regression to the prior ensemble member of the state variables based on the increment on the observation variable

Locally in space, each observation corrects only variables within a two-dimensional rectangular box of size \((2D+1) \times (2D+1)\) centered at the corresponding observation location
Local least-square EAKF

A one-analysis step beginning with a prior ensemble of state variables within a rectangular box of radius $D$ grid points, $u_{m+1|m}^i \in \mathbb{R}^{(2D+1)^2}$, where the index $i = 1, \ldots, K$ denotes the ensemble member, is given as follows:

1. We compute $v_{m+1|m}^i = g(u_{m+1|m}^i)$; in our case $g \in \mathbb{R}^{1 \times (2D+1)^2}$ is zero everywhere and one for the component that corresponds to the location of the observation; hence $v_{m+1|m}^i$ is scalar.
2. Compute the ensemble average

$$
\bar{u}_{m+1|m} = \frac{1}{K} \sum_{i=1}^K u_{m+1|m}^i, \quad \bar{v}_{m+1|m} = \frac{1}{K} \sum_{i=1}^K v_{m+1|m}^i.
$$

3. Compute the cross-covariances

$$
\sigma_{UV} = \frac{1+r}{K-1} U V^T, \quad \sigma_{VV} = \frac{1+r}{K-1} V V^T,
$$

where $r$ is the variance inflation coefficient (in our numerical experiment, we will empirically find the optimal inflation coefficient). Each column of $U$ is $u_{m+1|m}^i - \bar{u}_{m+1|m}$ and each column of $V$ is $v_{m+1|m}^i - \bar{v}_{m+1|m}$. In our case, $\sigma_{UV} \in \mathbb{R}^{(2D+1)^2 \times 1}$ and $\sigma_{VV} \in \mathbb{R}$.
4. Correct the mean observation state

$$
\bar{v}_{m+1|m+1} = \bar{v}_{m+1|m} + (\sigma_{VV} + r^o)^{-1} \sigma_{VV} (v_{m+1} - \bar{v}_{m+1|m}),
$$

where $v_m$ is a scalar observation with noise variance $r^o$. 

First Step

Update the prior ensemble estimate of the observation variables with a scalar ensemble filter
Local least-square EAKF

Second Step

Perform a linear regression to the prior ensemble member of the state variables based on the increment on the observation variable.

5. Correct each ensemble member of the observation state

\[
v_{m+1|m+1}^i = \bar{v}_{m+1|m+1}^i + \sqrt{\frac{r^o}{r^o + \sigma_{VV}}} (v_{m+1|m}^i - \bar{v}_{m+1|m})
\]

and compute \( \Delta v_{m+1}^i = v_{m+1|m+1}^i - v_{m+1|m}^i \in \mathbb{R} \).

6. For each ensemble member, we update each ensemble member with the least-square formula

\[
u_{m+1|m+1}^i = u_{m+1|m}^i + \frac{\sigma_{UV}}{\sigma_{VV}} \Delta v_{m+1}^i.
\] (11.7)

7. Repeat steps 1–6 for the remaining observations one-by-one. If there are overlaps between local boxes, use the updated variables from earlier analysis for the overlapped regions instead of the prior forecasts.
Local least-square EAKF

We consider sparse observations of the barotropic stream function at 36 grid points, distributed uniformly in a two-dimensional $2\pi$-periodic domain.

\[ \psi^o_m(x_i, y_j) = \psi_{b,m}^t(x_i, y_j) + \sigma^o_m(x_i, y_j), \quad i, j = 1, \ldots, 6, \]

\[ \sigma^o_m(x_i, y_j) \sim \mathcal{N}(0, r^o) \] is a white noise

A sparse observation network is consistent with a typical sparse ocean observation platform (Cane et al., 1996)

Sufficient to explain >90% of the time averaged model variability, the most energetic modes are large-scale

To represent the radius of deformation

<table>
<thead>
<tr>
<th></th>
<th>atmosphere</th>
<th>ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td>D (local box radius)</td>
<td>14</td>
<td>14(5)</td>
</tr>
<tr>
<td>$\Delta t$</td>
<td>0.001</td>
<td>0.00025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00005 for data assimilation</td>
</tr>
</tbody>
</table>

F=40, stiff problem
Local least-square EAKF

Table 11.6  Average RMS errors (averaged over 10,000 assimilation cycles ignoring the first 200 cycles) with LLS-EAKF.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>$T_{obs} = 0.5$, perfect</th>
<th>$r^o = 0.0856$, model error</th>
<th>$T_{obs} = 0.25$, perfect</th>
<th>$r^o = 0.1711$, model error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{r^o}$</td>
<td>0.2925</td>
<td>0.2925</td>
<td>0.4137</td>
<td>0.4137</td>
</tr>
<tr>
<td>LLS-EAKF</td>
<td>0.1582</td>
<td>0.1471</td>
<td>0.1790</td>
<td>0.1510</td>
</tr>
</tbody>
</table>

Better accurate and numerically fast filtering strategy in Chapter 12&13

stochastic parameterization:
- comparable filter skill to LLS-EAKF in the atmospheric regime
- extremely high skill beyond LLS-EAKF in the oceanic regime