Improved proposal densities for particle filters

Data Assimilation Research Centre | University of Reading

Javier Amezcua (j.AmezcuaEspinosa@reading.ac.uk) | Peter Jan van Leeuwen (p.j.VanLeeuwen@reading.ac.uk)

Abstract

The standard implementation of particle filters is unfeasible in large dimensional geophysical systems. The ensemble tends to collapse in few particles (degeneracy of weights) when the number of independent observations is large. An alternative is to use proposal densities other than the prior. In this work we use simple nudging and 4dvar as improved proposal densities. The results suggest that the use of 4dvar with an additional equal-weights step is promising.

Particle filters

Particle filters are **Monte Carlo implementations** of **Bayes theorem**, in which the prior probability is weakly approximated by an ensemble of M **weighted particles**.

$$p(\mathbf{x}\mid\mathbf{y})=\frac{p(\mathbf{y}\mid\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

posterior; likelihood; prior: $p(\mathbf{x})=\frac{1}{M}\sum_{m=1}^{M}\delta(\mathbf{x}-\mathbf{x}_m)$

$\mathbf{x}\in\Re^{N_x}$ state variables

$\mathbf{y}\in\Re^{N_y}$ observations

marginal probability of the observations: $p(\mathbf{y})=\int_{-\infty}^{\infty}p(\mathbf{y}\mid\mathbf{x})p(\mathbf{x})d\mathbf{x}$ (common to all particles)

One can **use proposal densities other than the prior** and weight accordingly. In particular, these can **include information about (future) observations**:

Then $p(\mathbf{x}\mid\mathbf{y})=\sum_{m=1}^{M}w_m\delta(\mathbf{x}-\mathbf{x}_m^*)$ with $w_m=\frac{p(\mathbf{y}\mid\mathbf{x}_m^*)p(\mathbf{x}_m^*)}{p(\mathbf{y})q(\mathbf{x}_m^*\mid\mathbf{y})}$ ($\underbrace{\;\;}_{q(\mathbf{x}\mid\mathbf{y})}$)

A simple system

We will use the following system for our experiments, consisting of **1 forecast and 1 assimilation step**.

$$\mathbf{x}^1=\alpha\mathbf{x}^0+\boldsymbol{\beta}^1 \qquad \mathbf{x}^0\sim N(\boldsymbol{\mu}^0=\mathbf{0},\mathbf{B}) \qquad \mathbf{B}=b^2\mathbf{I}$$

$$\mathbf{y}^1=\mathbf{x}^1+\boldsymbol{\eta}^1 \qquad \boldsymbol{\beta}^1\sim N(\mathbf{0},\mathbf{Q}) \qquad \mathbf{Q}=q^2\mathbf{I}$$

$$\boldsymbol{\eta}^1\sim N(\mathbf{0},\mathbf{R}) \qquad \mathbf{R}=r^2\mathbf{I}$$

We consider a fixed ensemble size and different state space sizes.

***Given* observations** are considered as: $\mathbf{y}^{obs}=y^{obs}\mathbf{1}\in\Re^{N_x}$

The truth is $\mathbf{x}^{truth}=\mathbf{0}\in\Re^{N_x}$ and the dynamics $\alpha=1$

Conditioning on the background probability

In the **absence of model error** we can estimate the posterior $p(\mathbf{x}^0\mid\mathbf{y}^1)$ as:

$$p(\mathbf{x}^0\mid\mathbf{y}^1)=\frac{p(\mathbf{y}^1\mid\mathbf{x}^0)p(\mathbf{x}^0)}{p(\mathbf{y}^1)q(\mathbf{x}^0\mid\mathbf{y}^1)}q(\mathbf{x}^0\mid\mathbf{y}^1)$$ ← proposal

with 3 options for proposals:

- The **prior** $q(\mathbf{x}^0\mid\mathbf{y}^1)=p(\mathbf{x}^0):N(\boldsymbol{\mu}^0,\mathbf{B})$ — $\zeta=0$ Corresponds to using the prior as proposal
- **Simple nudging** $q(\mathbf{x}^0\mid\mathbf{y}^1):N(\boldsymbol{\mu}^0+\zeta(\mathbf{y}-\boldsymbol{\mu}^0),\mathbf{B})$

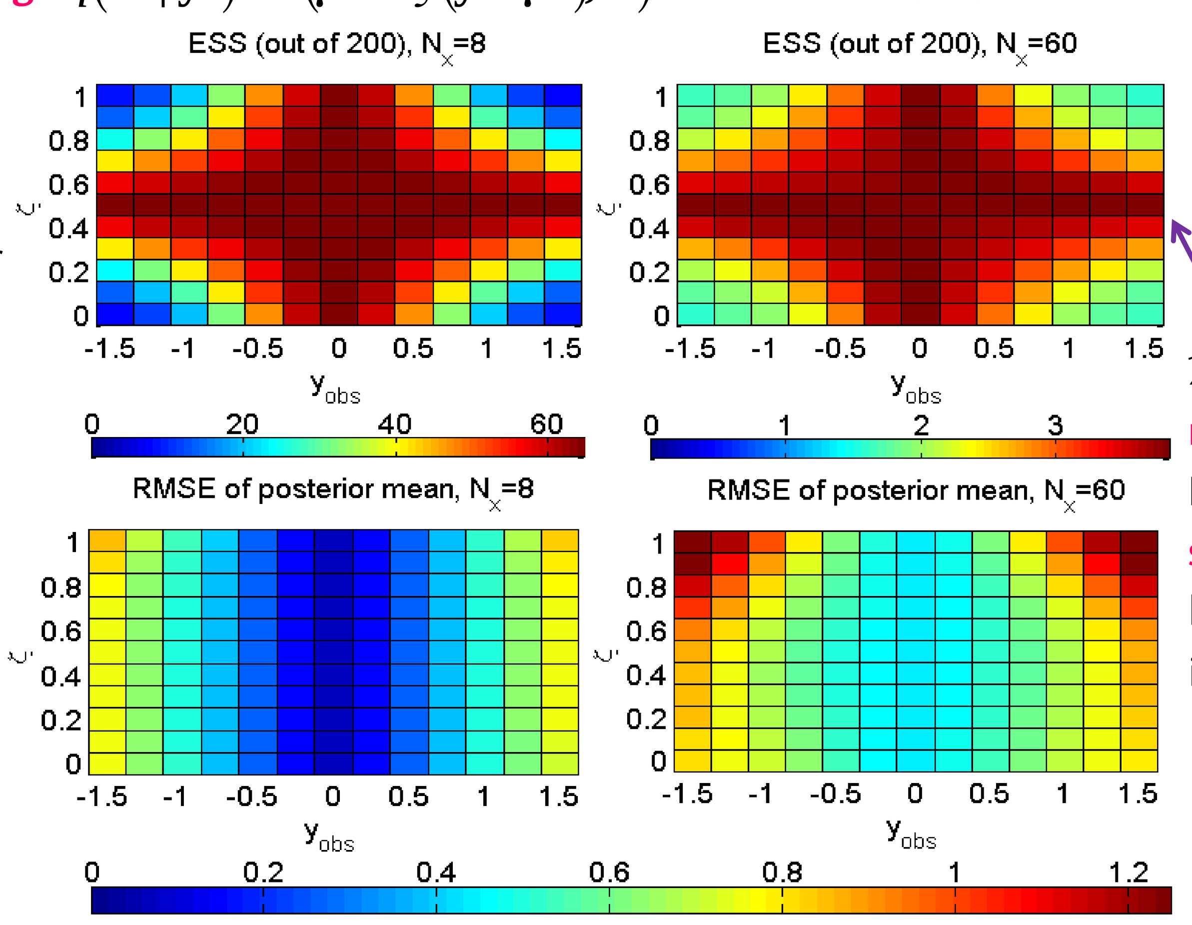

Figure 1: Effective ensemble size (**ESS**, top row) and root mean squared error (**RMSE**, bottom row) of the posterior mean for **different given observations** (horizontal axis) and **different nudging strength** (vertical axis) and 2 state space sizes (columns).

½ is the **optimal nudging** strength. In general **nudging slows degeneracy**, but does not stop it.

- **4dvar**: $q(\mathbf{x}^0\mid\mathbf{y}^1):N(\boldsymbol{\mu}^{0a},\mathbf{A})$ → Coming from **families of solutions of the type**:

$$\mathbf{x}_m^{0a}=\mathbf{x}_m^{0b}+\mathbf{K}(\mathbf{y}_m^1-\mathbf{x}_m^{0b})$$

$$\mathbf{x}_m^{0b}=\boldsymbol{\mu}^0+\boldsymbol{\varepsilon}_m^0 \qquad \mathbf{K}=\alpha\mathbf{B}^T(\alpha^2\mathbf{B}+\mathbf{R})^{-1} \qquad \mathbf{y}_m^1=\mathbf{y}^1+\boldsymbol{\eta}_m^1$$

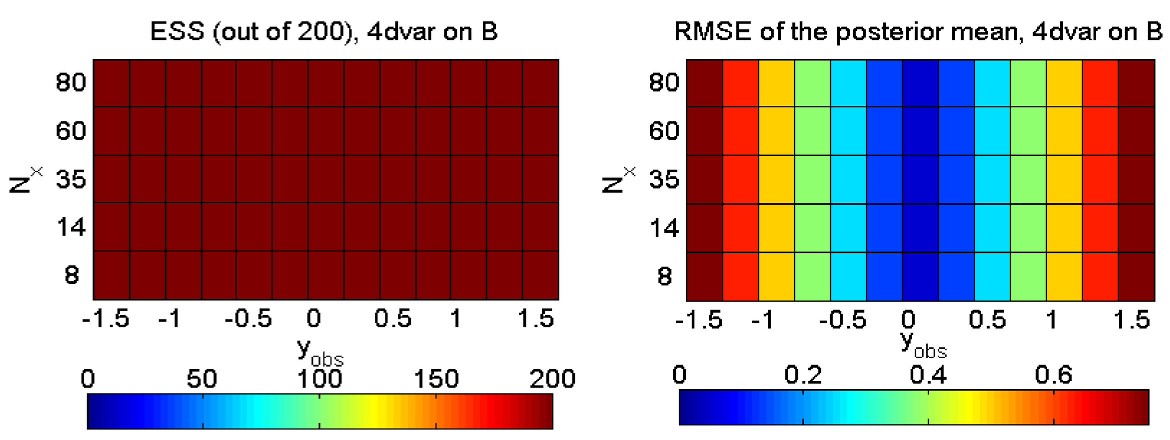

Figure 2: ESS (left) and **RMSE** (right) for **4dvar on B** used as proposal for different given observations (horizontal axis) and state space sizes (vertical axis). **By construction the weights are equal.**

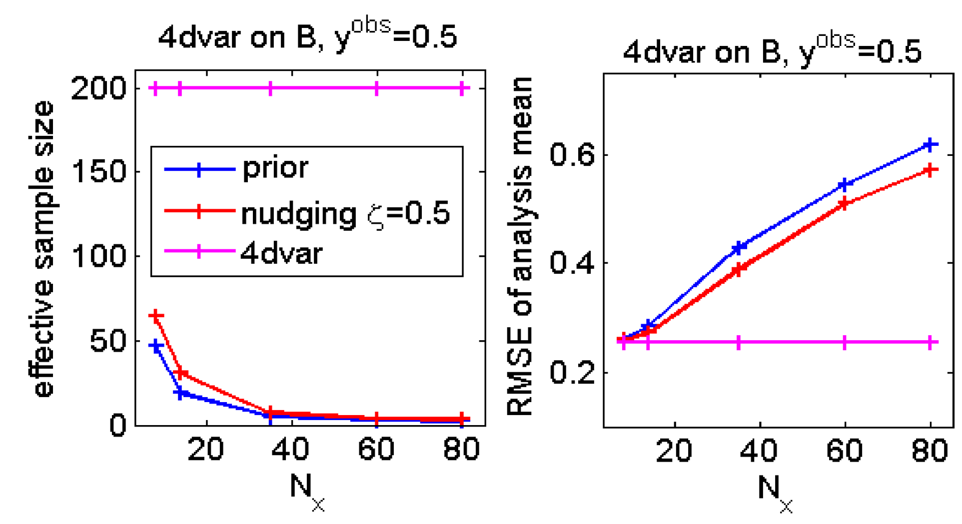

Figure 3: ESS and **RMSE** for the **3 proposal densities,** 1 given observation and different state space dimensions (horizontal axis).

Conditioning on the transition probability

In PF, usually one does **not worry about the background** and uses the **transition probability as proposal**

$$p(\mathbf{x}^1\mid\mathbf{y}^1)=\frac{p(\mathbf{y}^1\mid\mathbf{x}^1)}{p(\mathbf{y}^1)}p(\mathbf{x}^1)=\frac{p(\mathbf{y}^1\mid\mathbf{x}^1)}{p(\mathbf{y}^1)}\int_{-\infty}^{\infty}p(\mathbf{x}^1\mid\mathbf{x}^0)p(\mathbf{x}^0)d\mathbf{x}^0$$

$$=\frac{p(\mathbf{y}^1\mid\mathbf{x}^1)}{p(\mathbf{y}^1)}\frac{p(\mathbf{x}^1\mid\mathbf{x}_m^0)}{q(\mathbf{x}^1\mid\mathbf{x}_m^0,\mathbf{y}^1)}q(\mathbf{x}^1\mid\mathbf{x}_m^0,\mathbf{y}^1)$$ ← use particle representation

with 3 options for proposals:

- The **transition density** $q(\mathbf{x}^1\mid\mathbf{x}_m^0,\mathbf{y}^1)=p(\mathbf{x}^1\mid\mathbf{x}_m^0):N(\alpha\mathbf{x}_m^0,\mathbf{Q})$
- **Simple nudging** $q(\mathbf{x}^1\mid\mathbf{x}_m^0,\mathbf{y}^1):N(\alpha\mathbf{x}_m^0+\zeta(\mathbf{y}^1-\mathbf{x}_m^0),\mathbf{Q})$

$\zeta=0$ corresponds to using the transition as proposal

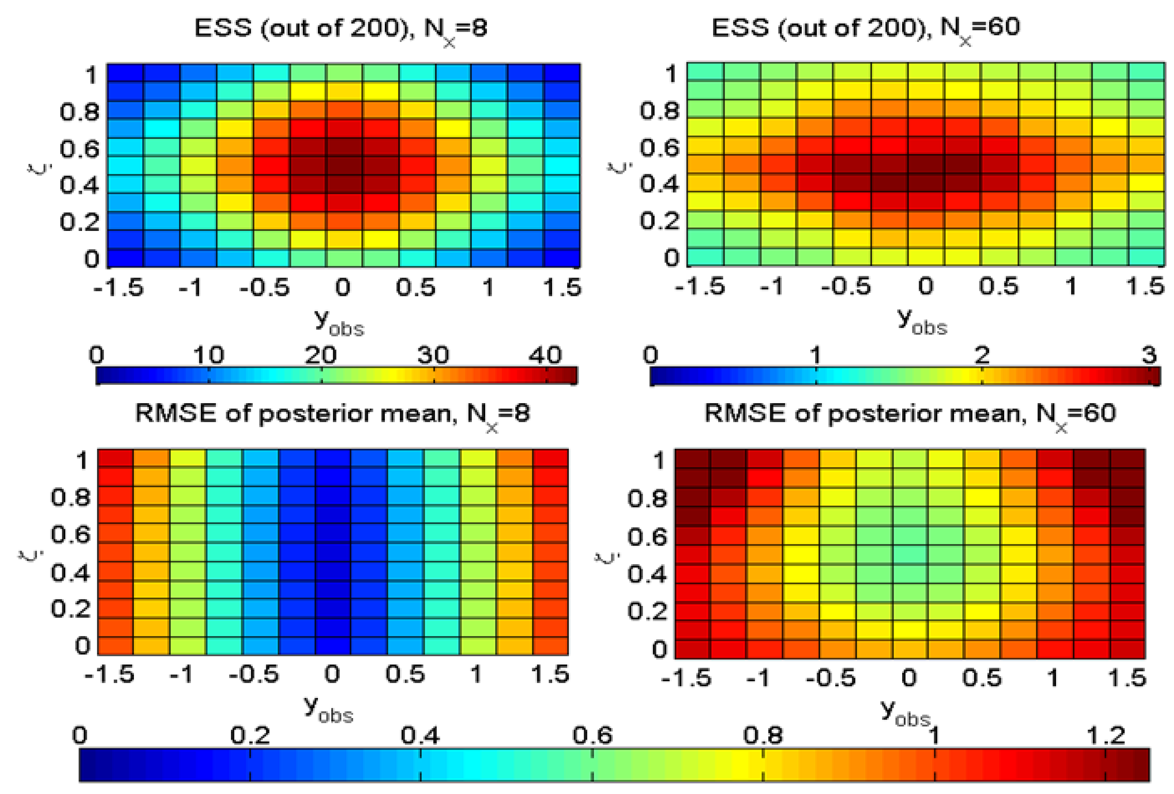

Figure 4: ESS (top row) and **RMSE** for different given observations (horizontal axis) and different nudging strength (vertical axis) and 2 state space sizes (columns). Again the **optimal nudging strength** is ½. In general, **degeneracy occurs faster than when conditioning on the background**.

- **4dvar**: $q(\mathbf{x}^1\mid\mathbf{x}_m^0,\mathbf{y}^1):N(\mathbf{x}_m^{a1},\mathbf{A})$ → Coming from **families of solutions of the type**:

$$\mathbf{x}_m^{1a}=\mathbf{x}_m^1+\mathbf{K}(\mathbf{y}_m^1-\mathbf{x}_m^1)$$

$$\mathbf{x}_m^1=\alpha\mathbf{x}_m^0+\boldsymbol{\beta}_m^1 \qquad \mathbf{K}=\mathbf{Q}^T(\mathbf{Q}+\mathbf{R})^{-1} \qquad \mathbf{y}_m^1=\mathbf{y}^1+\boldsymbol{\eta}_m^1$$

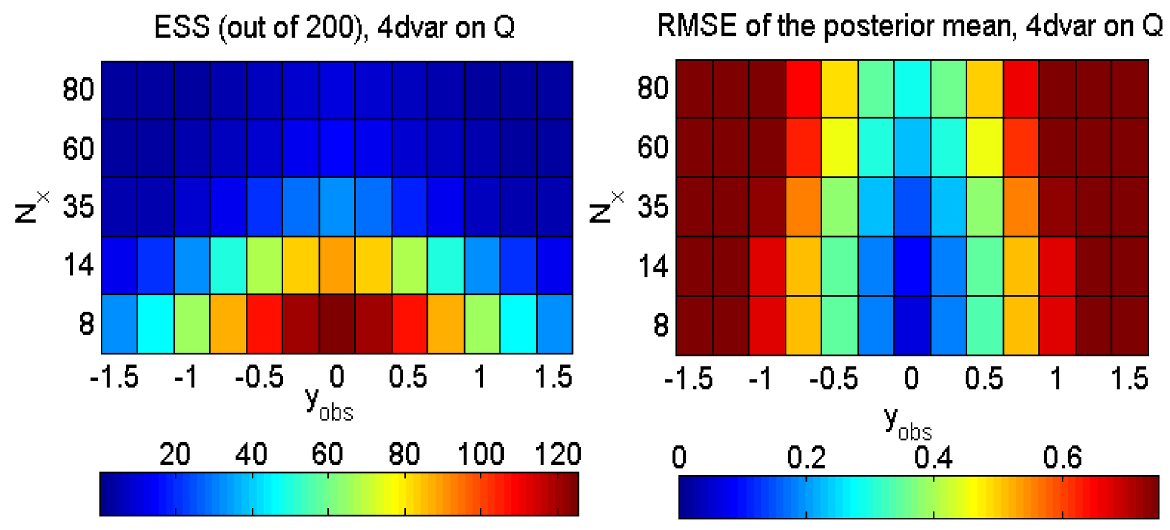

Figure 5: ESS (left) and **RMSE** (right) for **4dvar on Q** used as proposal for different given observations (horizontal axis) and state space sizes (vertical axis). In this case, **degeneracy does occur.**

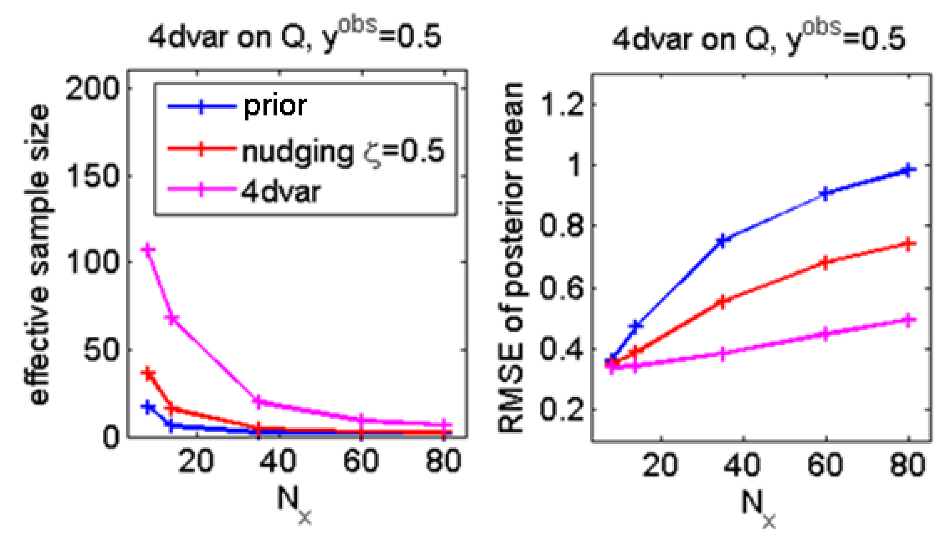

Figure 6: ESS and **RMSE** for **the 3 proposal densities**, 1 given observation and different state space dimensions (horizontal axis).

Equivalent weight steps.

Even **the optimal proposal density degenerates**. One can **set a target weight**, propose an **'incomplete' 4dvar solution** $\mathbf{x}_m^{1*}=\mathbf{x}_m^1+c\mathbf{K}(\mathbf{y}_m^1-\mathbf{x}_m^1)$ and solve for **c** from the equation:

$$-2\ln(w^{target})=(\mathbf{y}^1-\mathbf{x}_m^{1*})^T R^{-1}(\mathbf{y}^1-\mathbf{x}_m^{1*})+(\mathbf{x}_m^{1*}-\alpha\mathbf{x}_m^0)^T Q^{-1}(\mathbf{x}_m^{1*}-\alpha\mathbf{x}_m^0)$$

(See [1] for more details on how to add a stochastic component to this deterministic move).

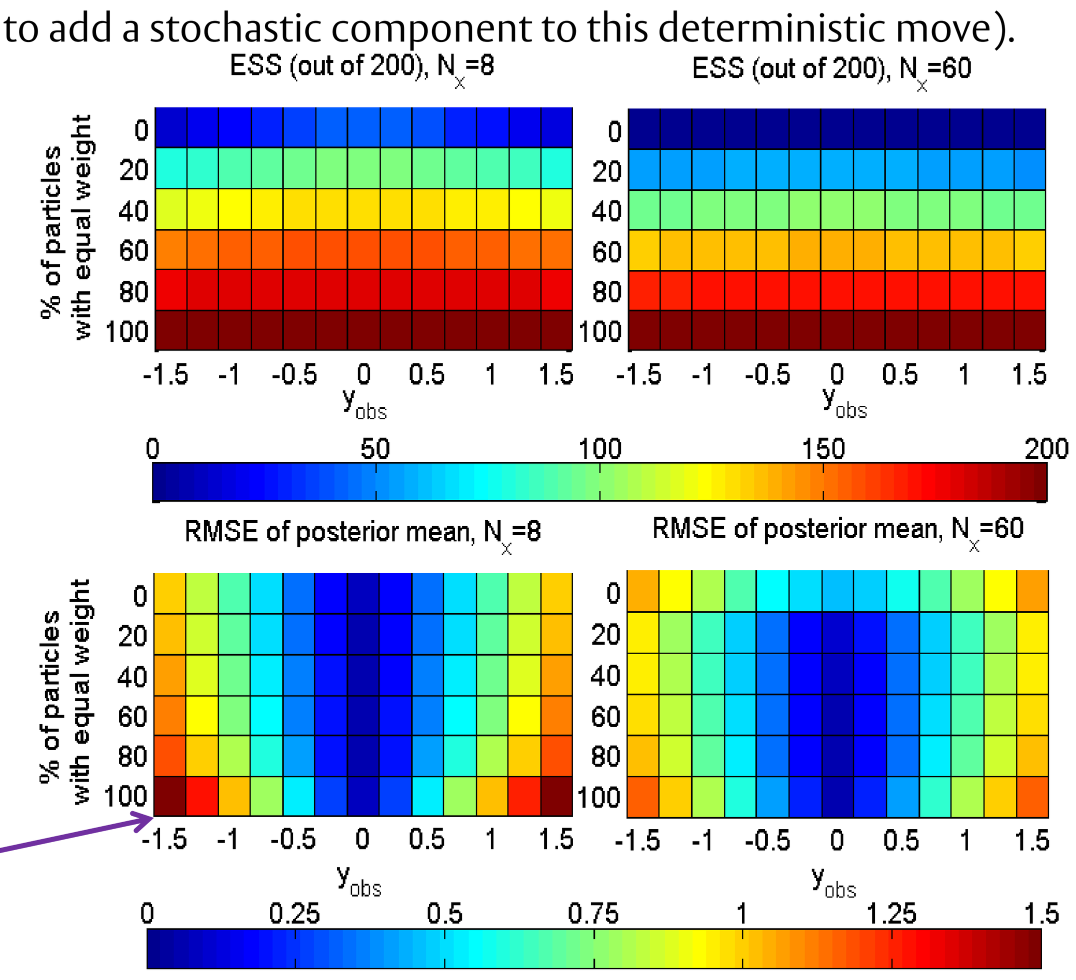

Figure 7: ESS (top row) and **RMSE** (bottom row) for **4dvar on Q** after **applying the equivalent weight step** for 2 state space dimensions (columns) and given observations (horizontal axis). The vertical axis is the **percentage of particles with equal weights.** No resampling was applied.

100% equal weights means all the particles have the weight of the worst particle. This may **not** be **convenient** (as shown), so that other **threshold** should be applied.

References

1. M. Ades and P.J. van Leeuwen, 2012. An exploration of the equivalent weights particle filter. *Q. J. R. Meteorolo. Soc.*, **139**: 820–840. 2. J. Amezcua and P.J. van Leeuwen, 2013. Ensembles of 4dvars and their use within a particle filter context. *In preparation*. 3. C. Snyder, T. Bengtsson, P. Bickel and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.*, **136**, 4629-4640. 4. P. J. van Leeuwen, 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. Q. J. R. Meteorol. Soc., 136, 1991-1999.