The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model

Michael Würsch, George C. Craig
Meteorological Institute
University of Munich

Joint SRNWP Workshop on Data Assimilation and Ensemble Prediction Systems
23-25 February 2011, Bologna, Italy

Motivation

- Introduce a minimal model
  - Key features: spatial intermittency, stochastic evolution
- Examine the performance of ETKF and particle filter
- Look at two different strategies to improve methods
  - Localization and observation averaging

Problems of radar data assimilation

- Lack of spatial correlations prevents reduction in degrees of freedom
  - Curse of dimensionality (Bellman 1957, 1961)
- Lack of temporal correlation between observation times
  - Stochastic evolution of the model field
- No balance constraints
  - No geostrophic or comparable balance constraints applicable

Data Assimilation Methods

- Ensemble Transform Kalman Filter (ETKF)
  - ETKF (Bishop et al. 2001; Hunt et al. 2007)
- Sequential Importance Resampling (SIR)
  - SIR (Gordon et al. 1993, Van Leeuwen 2009)

Possible Solutions

- Localization
  - Analysis at a grid point is only influenced by observations close-by
  - LETKF of Hunt et al. (2007)
  - Local SIR filter (LSIR)
- Observation Averaging
  - Averaged observations over a region to reduce intermittency
  - Sometimes called „super-obbing“
  - AETKF and ASIR

Toy Model

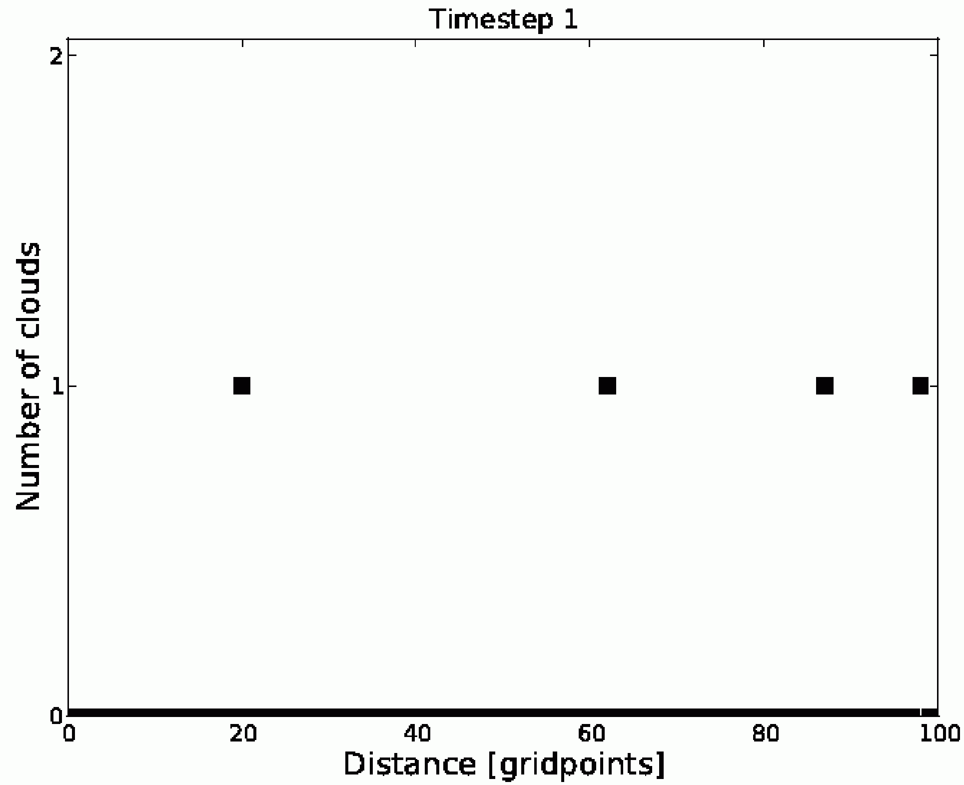
- One spatial and one temporal dimension
- Cloud distribution determined by birth/death probabilities

➡ Leads to a mean cloud half life and an average density of clouds

- Model & observations are perfect

Example run:
- 100 gridpoints
- half life 30 time steps
- average density of 0.1 clouds

- black squares ■ mark the number and position of clouds



Toy Model

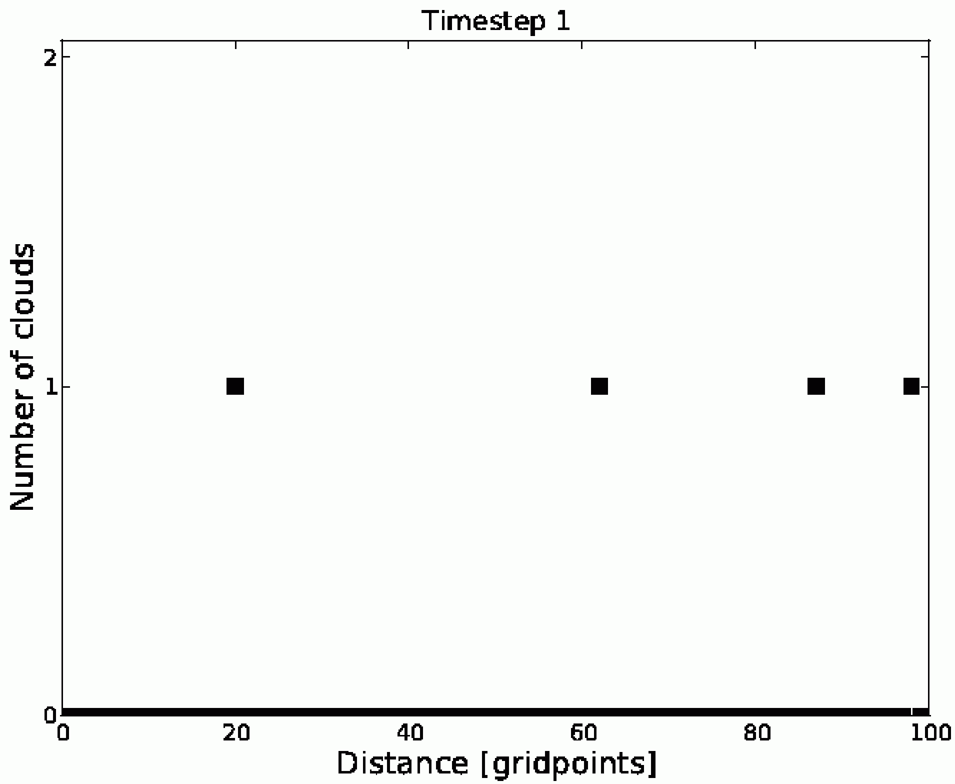
- One spatial and one temporal dimension
- Cloud distribution determined by birth/death probabilities

➡ Leads to a mean cloud half life and an average density of clouds

- Model & observations are perfect

Example run:
- 100 gridpoints
- half life 30 time steps
- average density of 0.1 clouds

- black squares ■ mark the number and position of clouds



Model setting

Basic settings:

- 100 grid points
- Half-life (hl) is 30 (varying problem) or 3000 (stationary problem) time steps
- Average cloud density is 0.1 per grid point
- hl30 corresponds to 2.5 hours

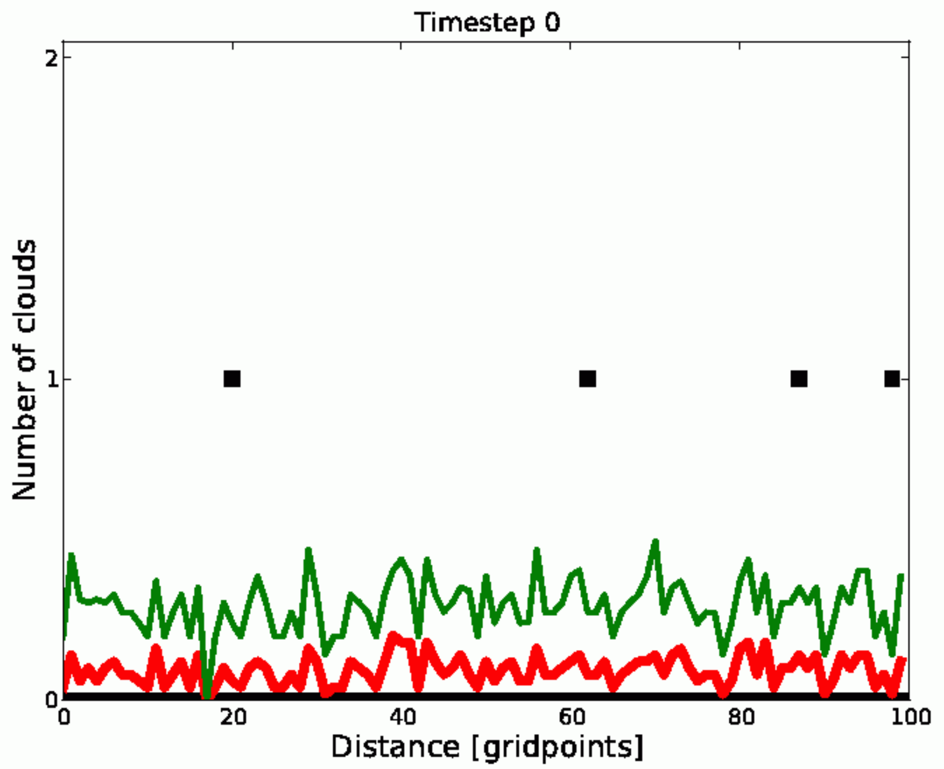
Special settings:

- Averaging is applied for regions of 10 grid points
- Localization is applied to every grid point with a localization radius of 0

Example Runs

- Ensemble size: 50
- Half life: 3000
- Method: ETKF

- **Red line:** ensemble mean
- **Green line:** ensemble spread

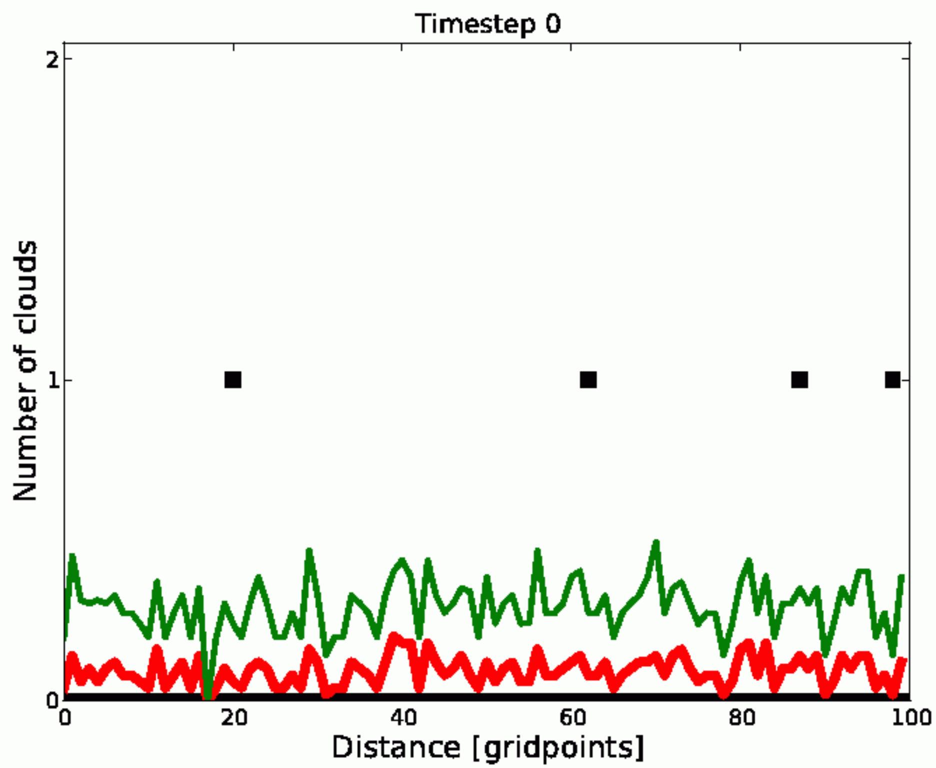


Example Runs

- Ensemble size: 50
- Half life: 3000
- Method: ETKF

- **Red line:** ensemble mean
- **Green line:** ensemble spread

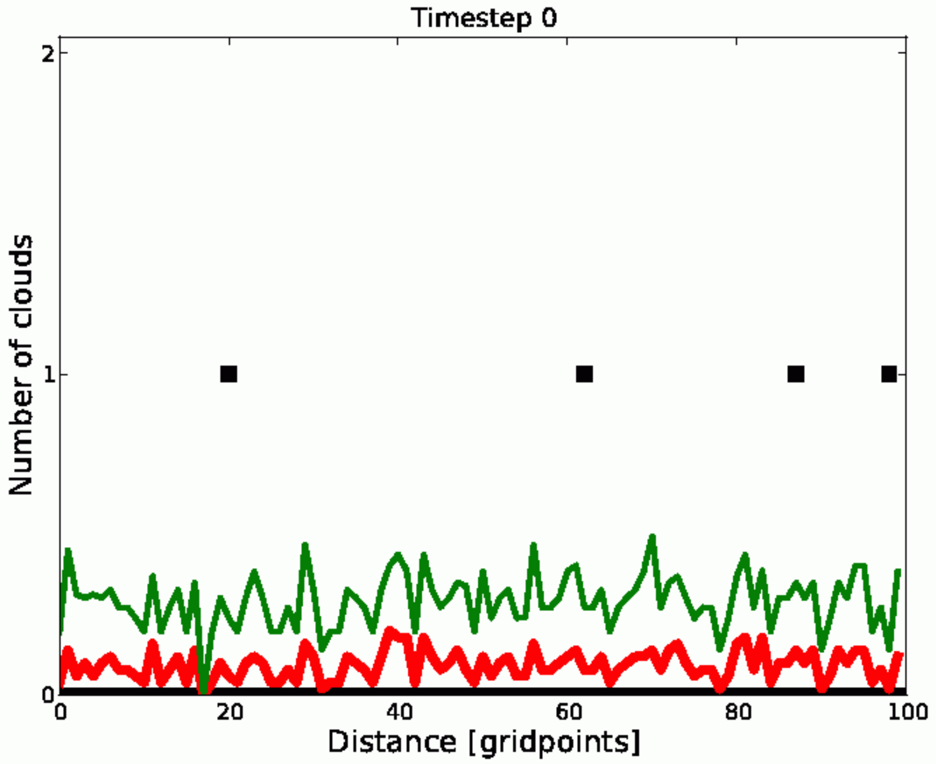
- Clouds are assimilated fast
- Noise stays long
- Some „wrong“ clouds



Example Runs

- Ensemble size: 50
- Half life: 3000
- Method: SIR

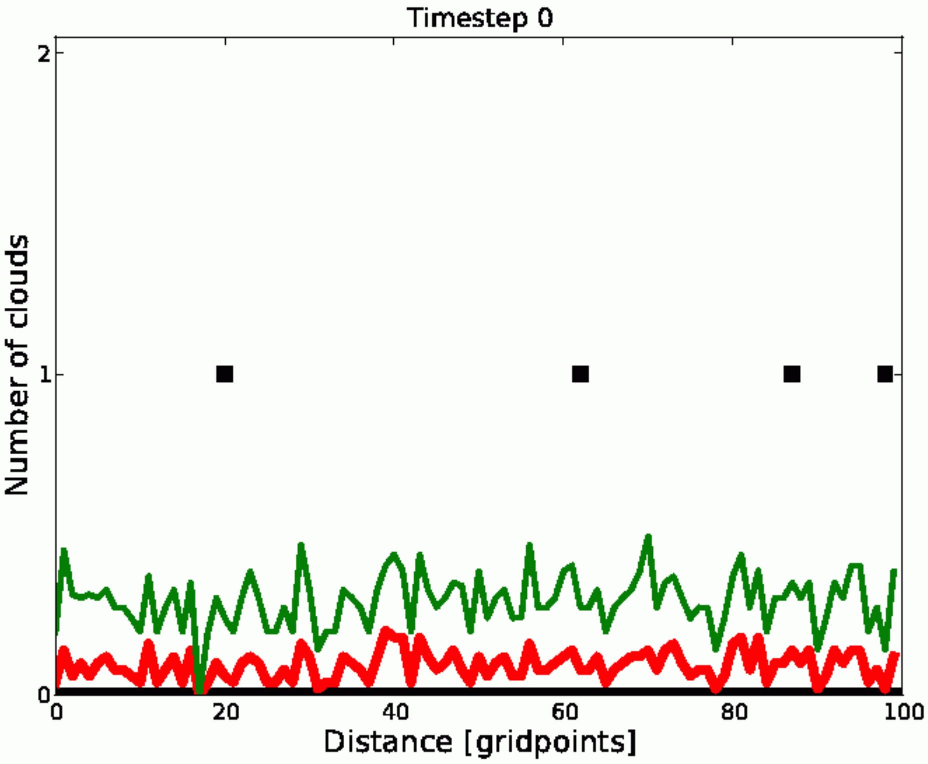
- **Red line:** ensemble mean
- **Green line:** ensemble spread



Example Runs

- Ensemble size: 50
- Half life: 3000
- Method: SIR

- **Red line:** ensemble mean
- **Green line:** ensemble spread



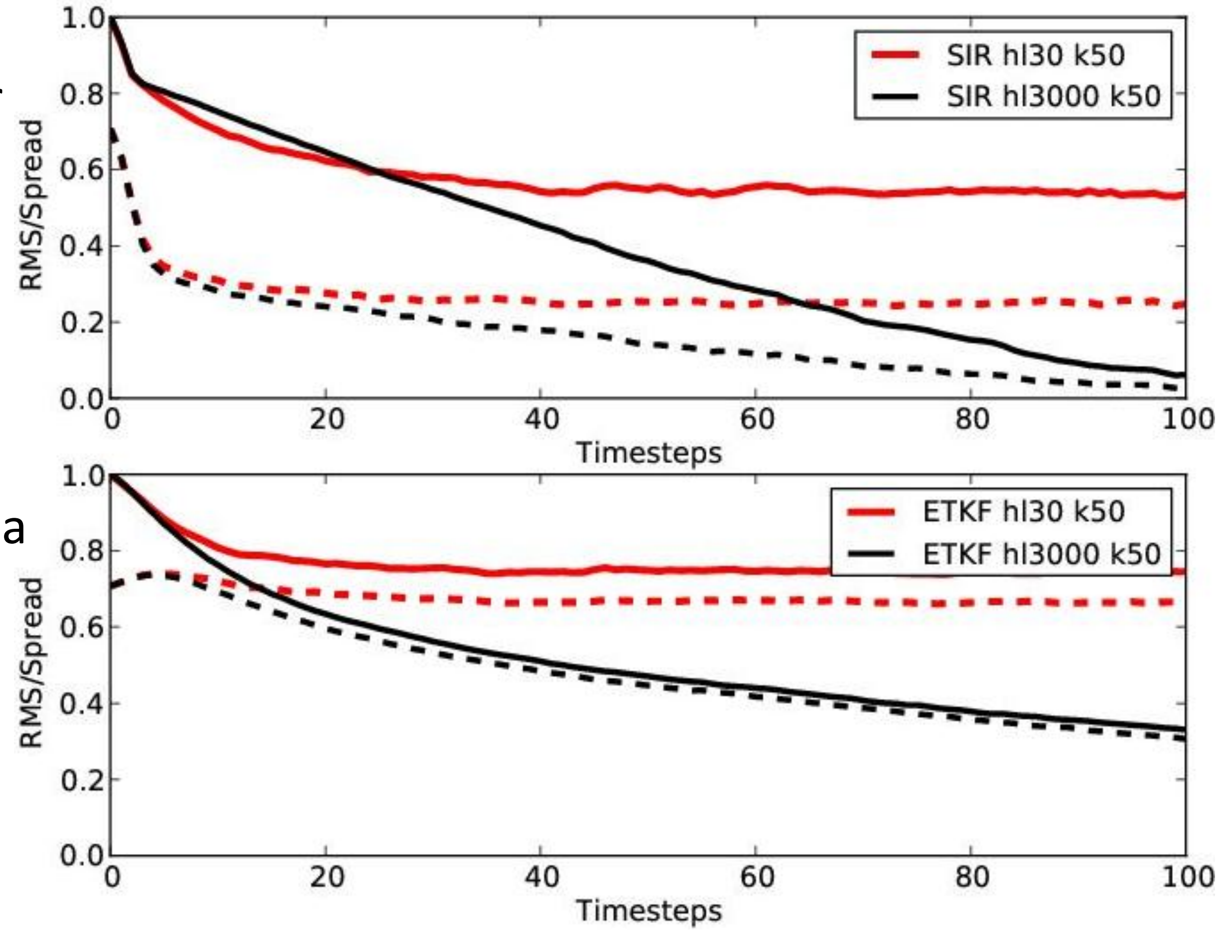
- Clouds are assimilated slowly
- Noise is gone very fast
- Many „wrong“ clouds

Results – Example Runs

- SIR:
  - Hl3000: Almost perfect after 100 time steps
  - Hl30: After 40 time steps more or less constant error
- ETKF:
  - Hl3000: slowly converges to a small error
  - Hl30: Does not get better than 0.8
  - Spread and error are similar

______ Continuous line: RMSE

_ _ _ _ Dashed line: Ensemble spread

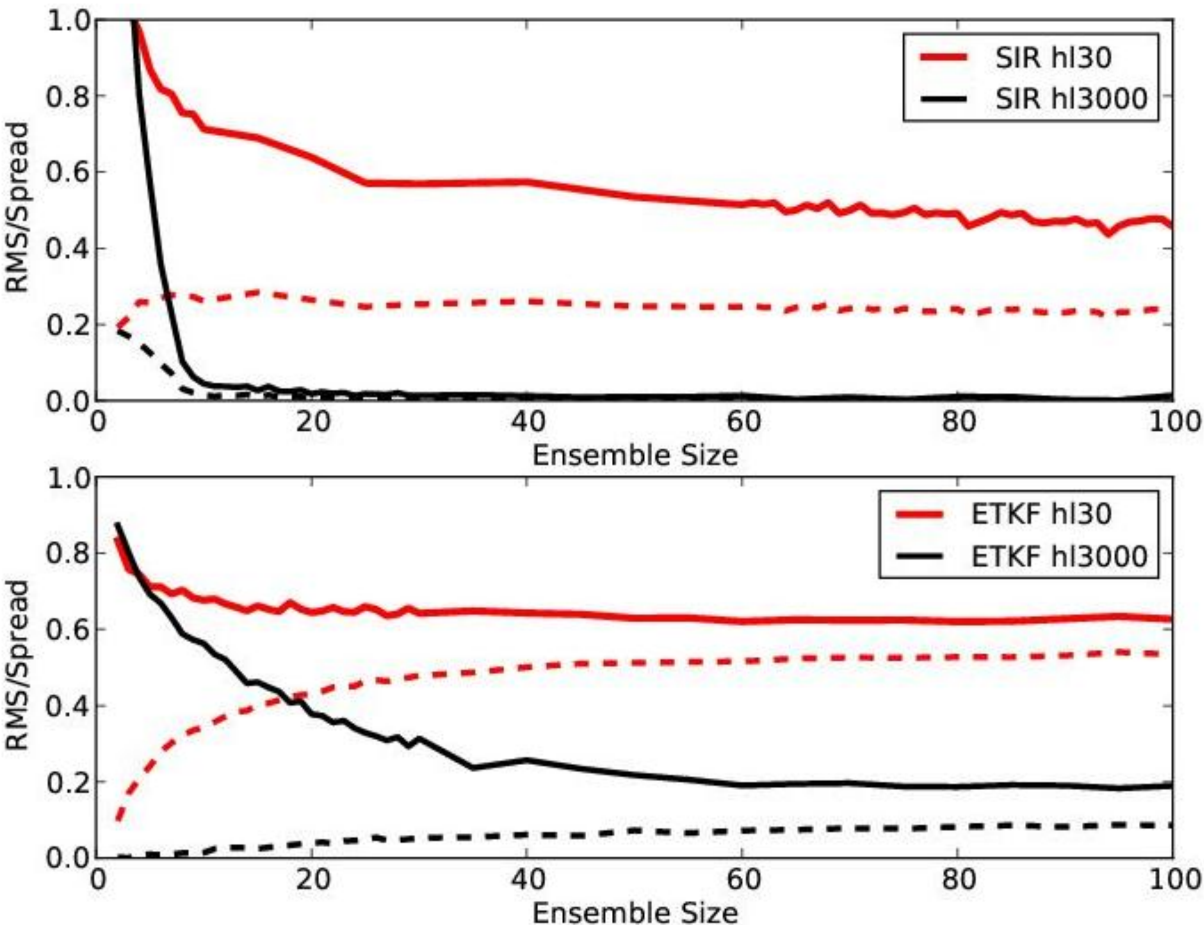


Changing Ensemble size

- SIR:
  - Hl3000: SIR is almost perfect
  - Hl30: Decreases slowly with increasing ensemble size. Still at 0.4 with 200 members
- ETKF:
  - Hl3000: Error only decreases significantly up to 40 members
  - Hl30: Almost no improvement
  - For larger ensembles error and spread are almost constant

______ Continuous line: RMSE

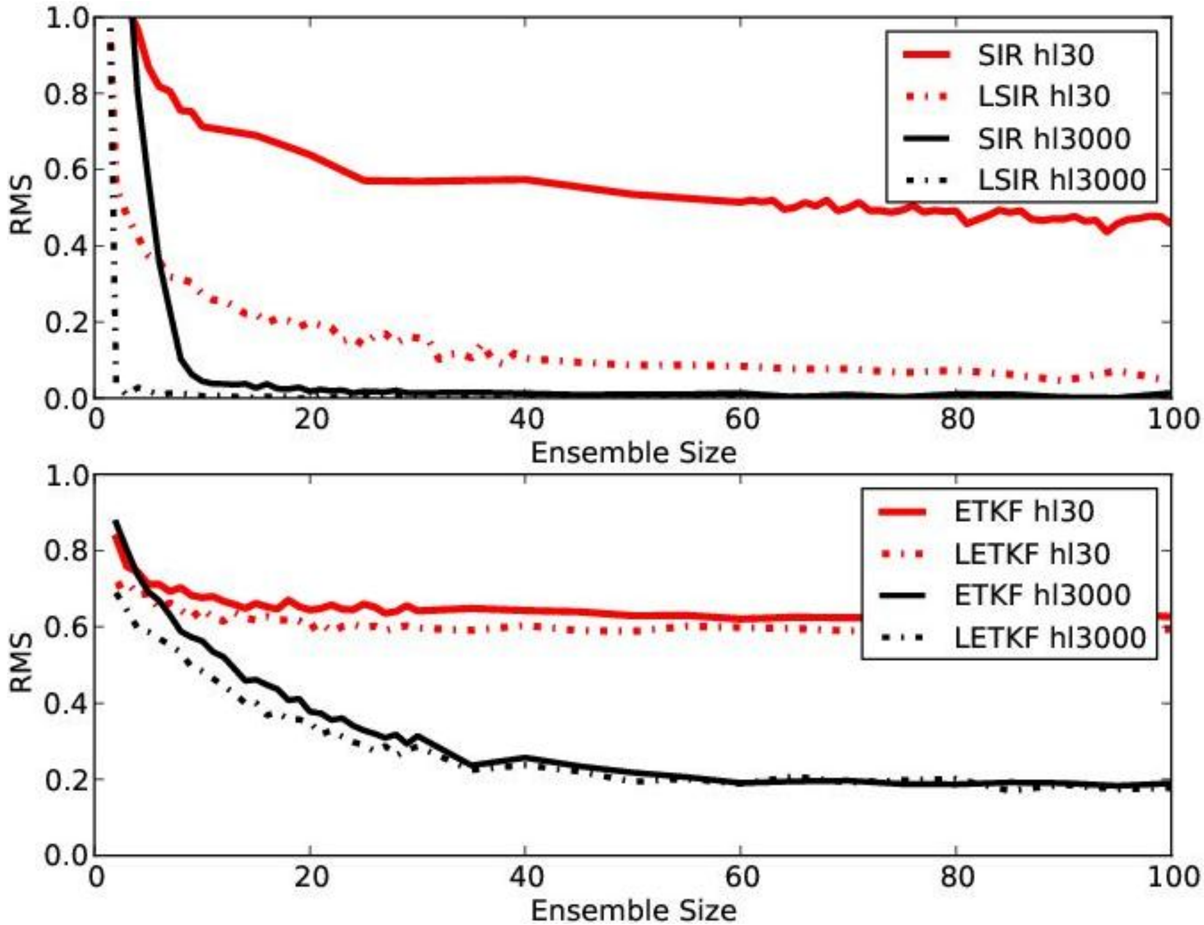
_ _ _ _ Dashed line: Ensemble spread



Localization

- SIR:
  - Hl3000: Already perfect with 3 ensemble members
  - Hl30: Big upgrade compared to standard version

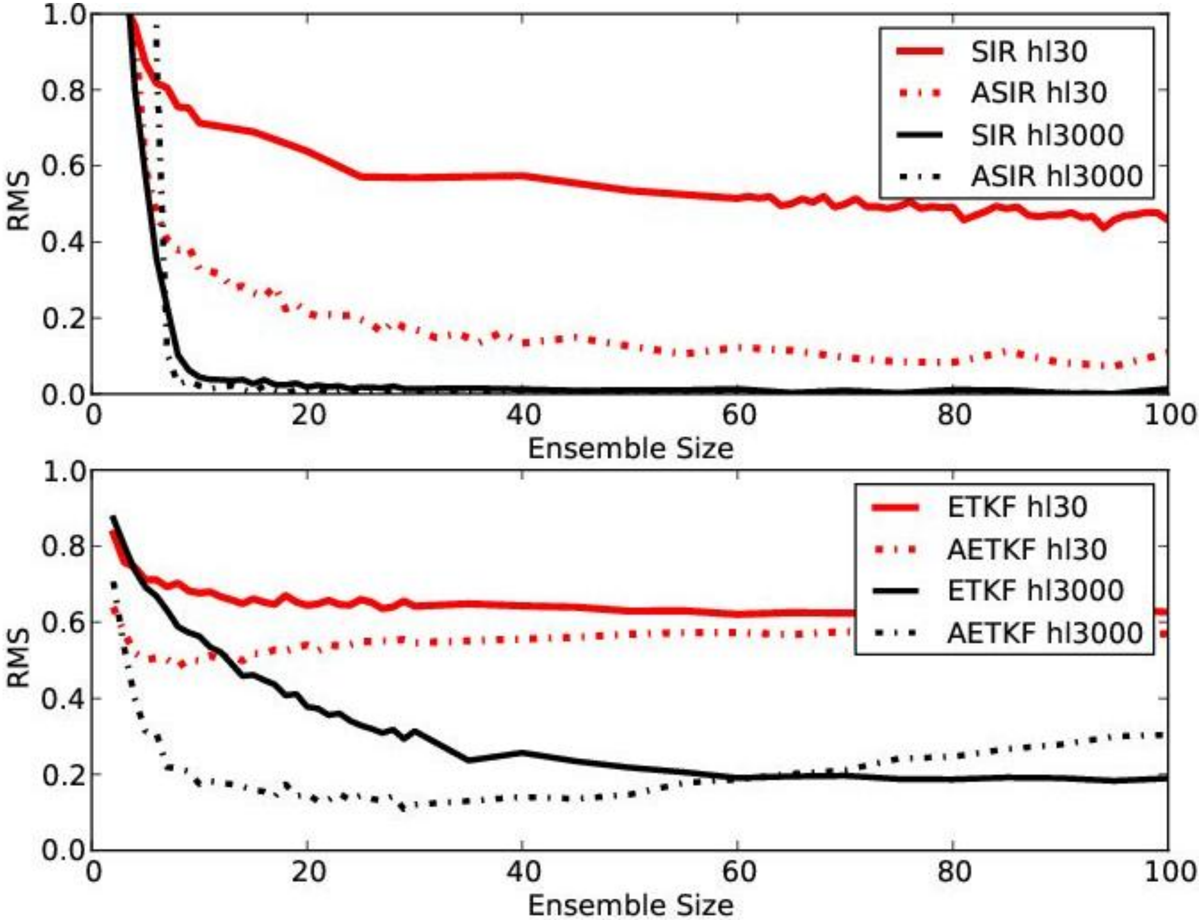
- ETKF:
  - Hl3000+Hl30: Almost no difference to standard version



Averaging

- SIR:
  - Hl3000: No improvement
  - Hl30: Not better than localization

- ETKF:
  - Hl3000: Reduction in ensemble size to get error around 0.2



Conclusions

- Introduced a stochastic model that captures the key features of convective-scale data assimilation
- SIR can give good results, but ensemble size is related to the dimensionality of the problem (can be very large)
- Both standard methods fail when posed with the dynamical situation
- Localization works very well for the SIR and has a smaller effect for the ETKF
- Averaging seems more useful for the ETKF, especially for small ensembles

Outlook

- Use a more complex model (shallow-water) to test interaction with gravity waves
- Do idealised problems with the systems of COSMO/KENDA (Km-Scale Ensemble-Based Data Assimilation)

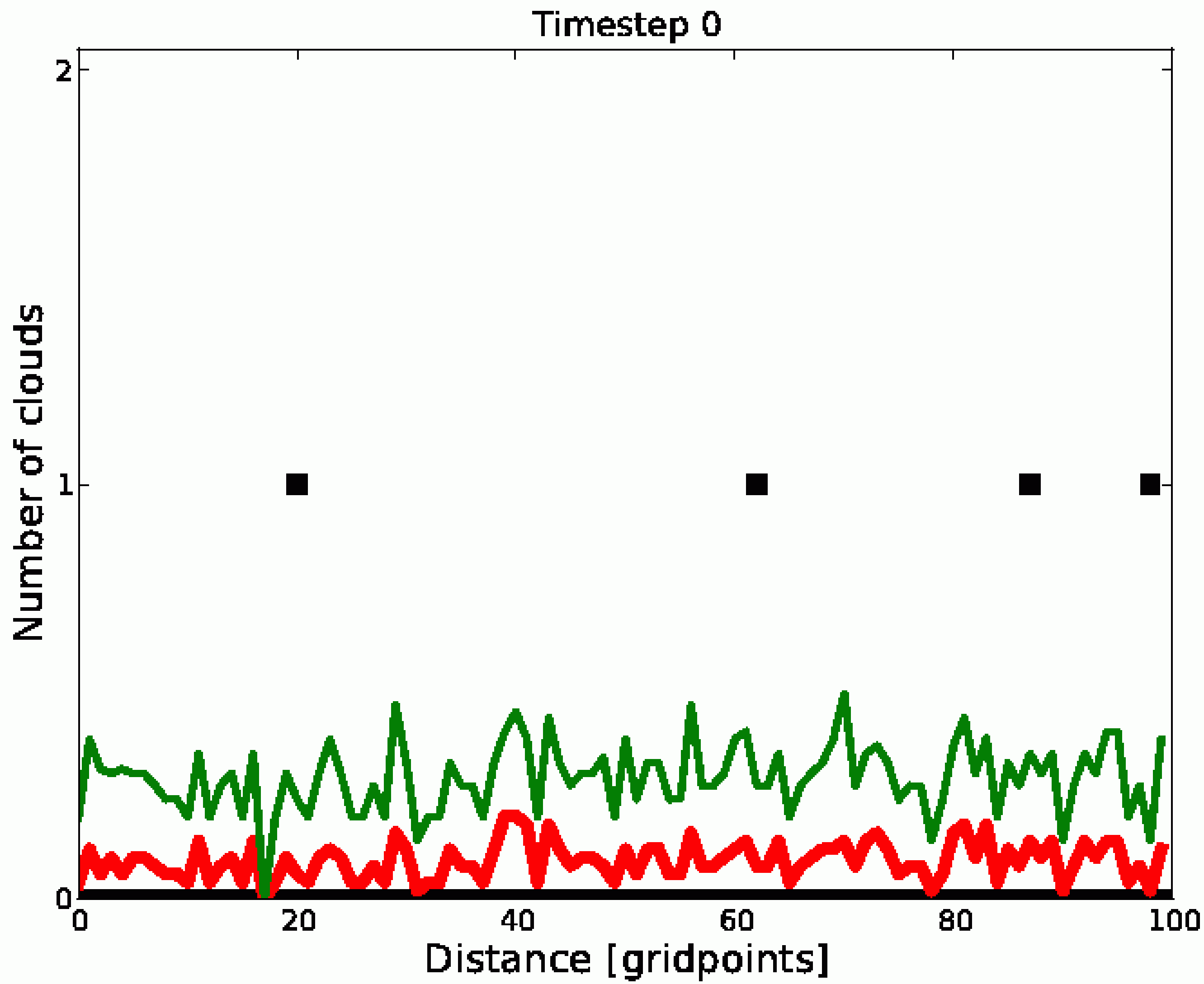
Extra slide - Tuning

- **What happens when tuning SIR:**
- One can change how large the random perturbation is. If I make it larger, there is more noise and more correct clouds and RMS stays almost the same. The question is what one wants. Few correct clouds and little spread, or the opposite.

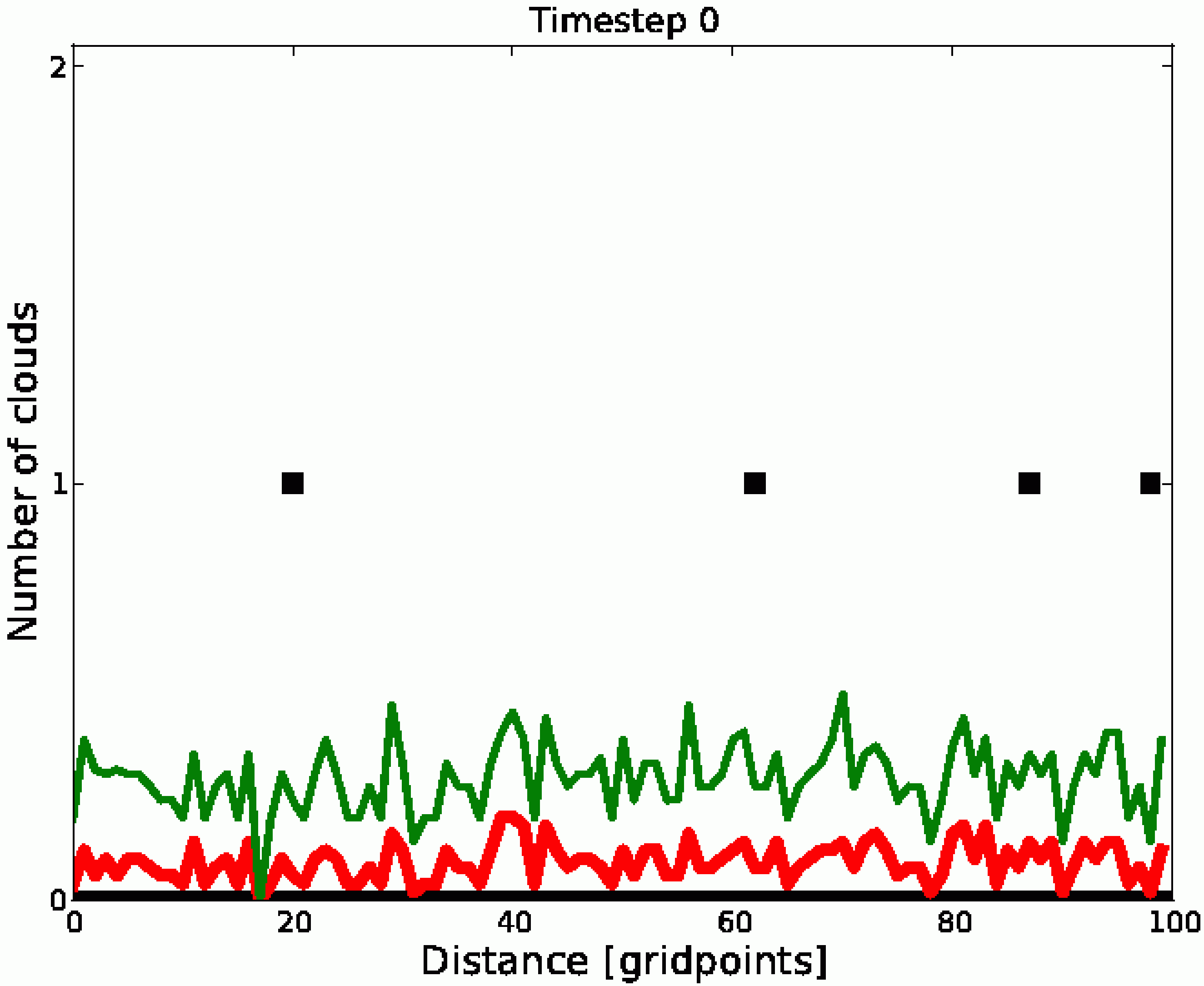
- **What happens when tuning ETKF:**
- Same as for SIR. More spread also increases the error. At least for bigger ensemble sizes, where ensemble and spread are already quite similar. It does not have to be absolutely identical. But with inflation, clouds are faster at the correct position.
- Deflation makes the error go down faster initially (less spread), but final error is the same again.

- **Averaging ETKF:**
- Because the RMS punishes errors of 2 or 3 clouds too much, the spread is too large and covariance deflation is needed. Up to 0.7.
- It would vary too fast if not.

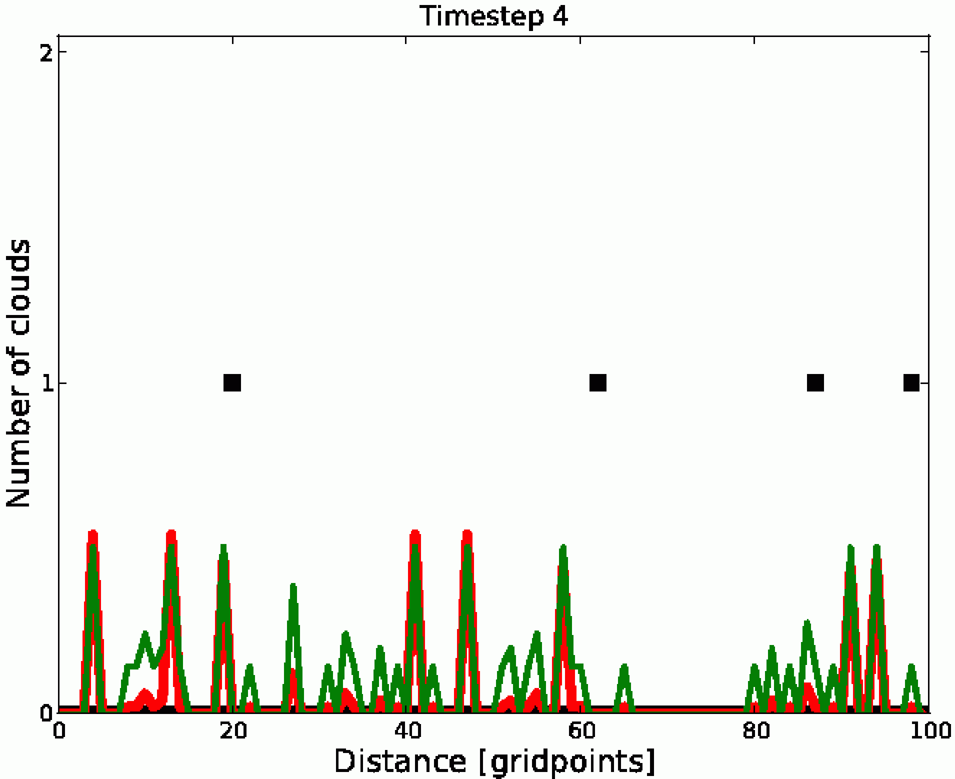
ETKF hl30, 50 ensembles

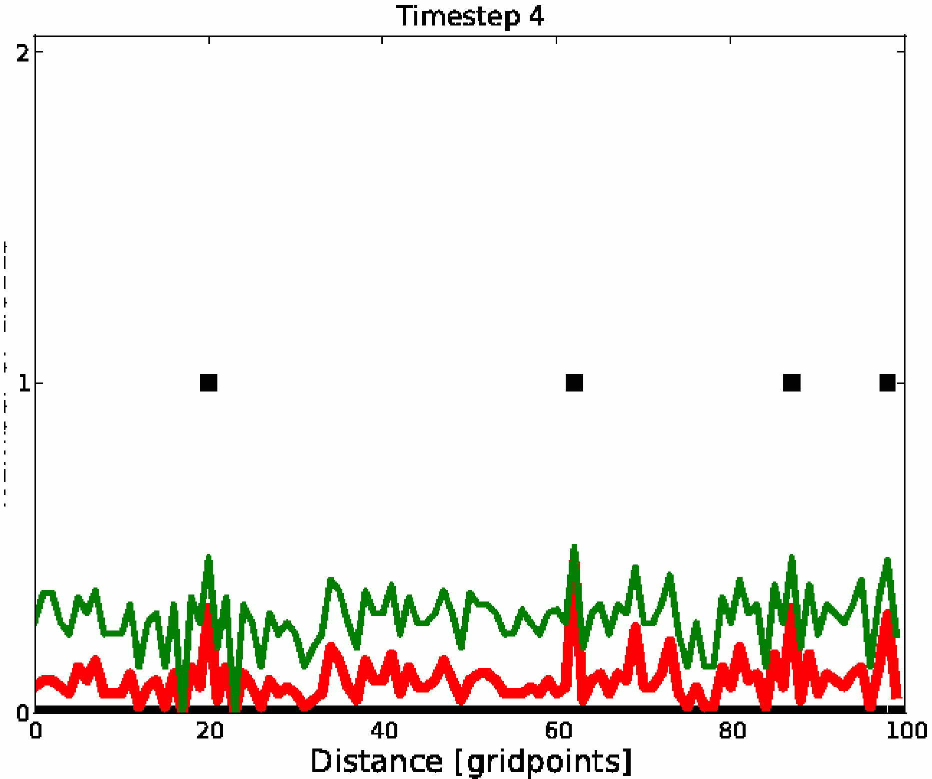


SIR hl30, 50 members



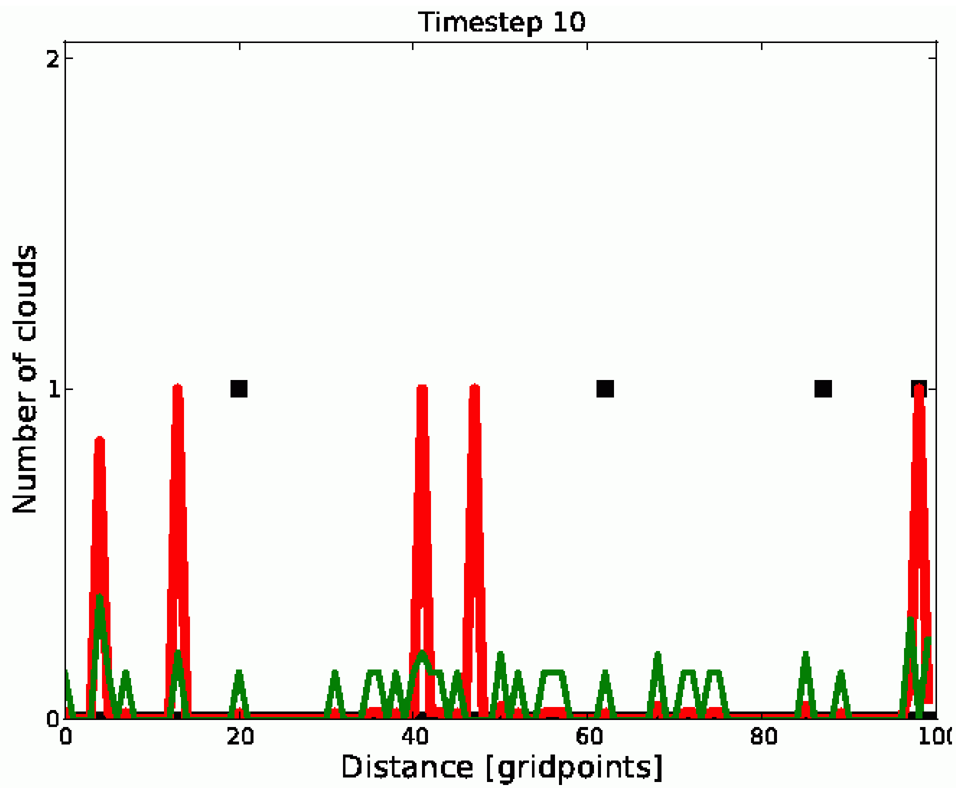
Comparison after 4 steps



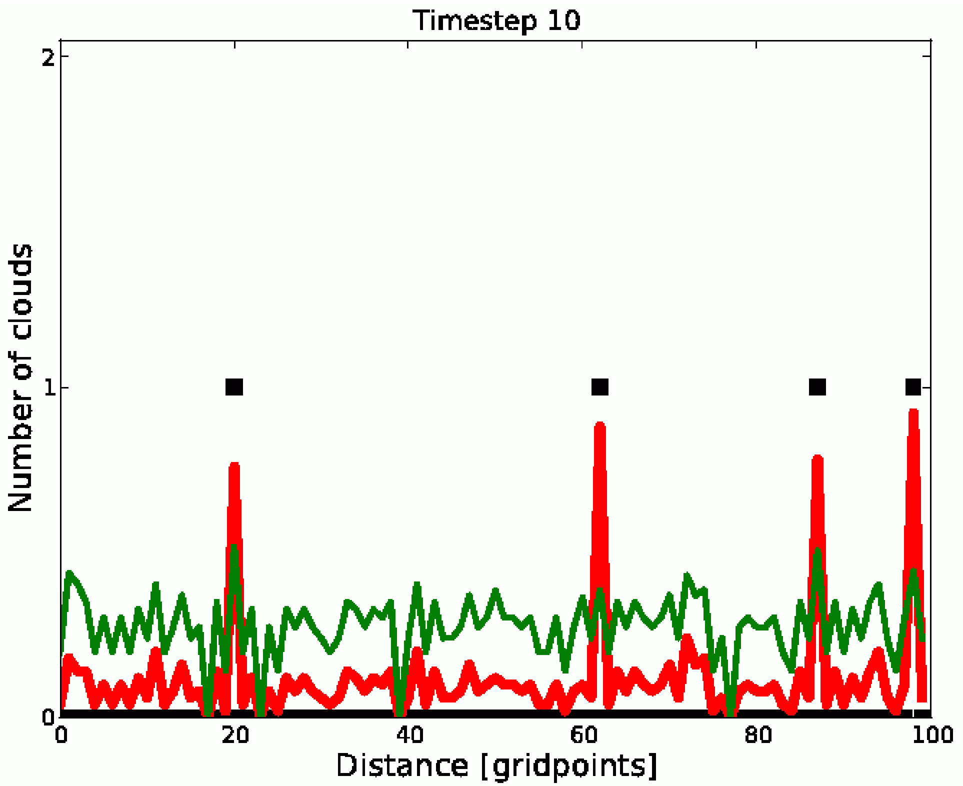


Comparison after 10 steps

SIR

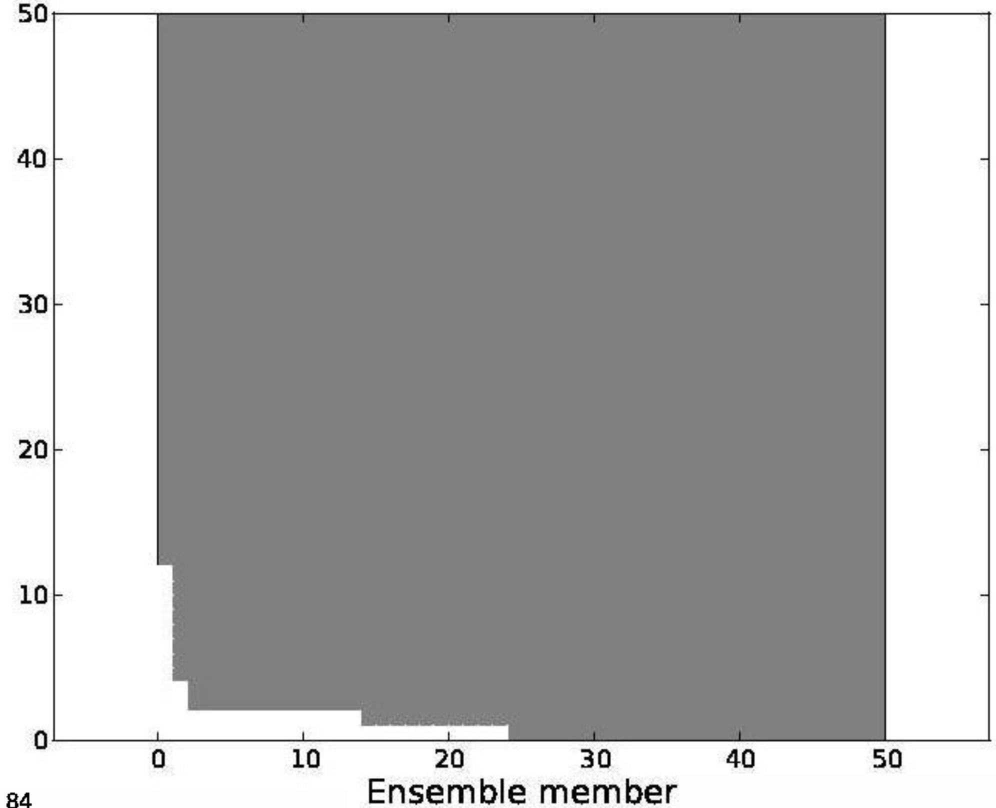


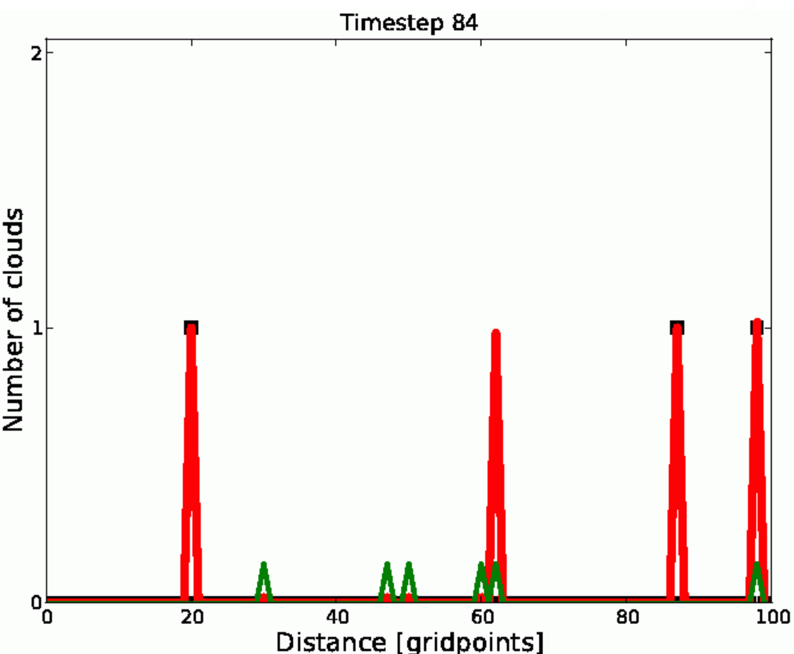
ETKF

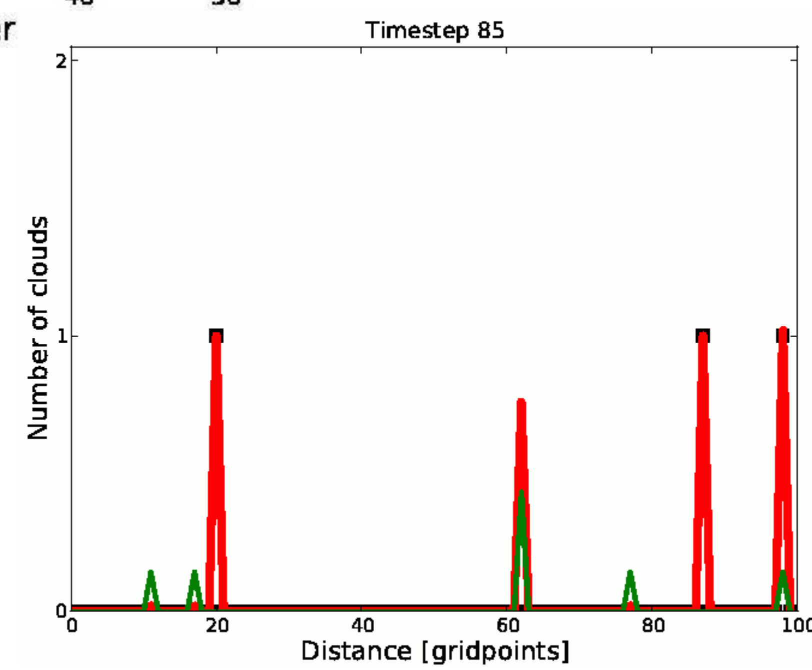


W-Matrix Hinton Diagrams

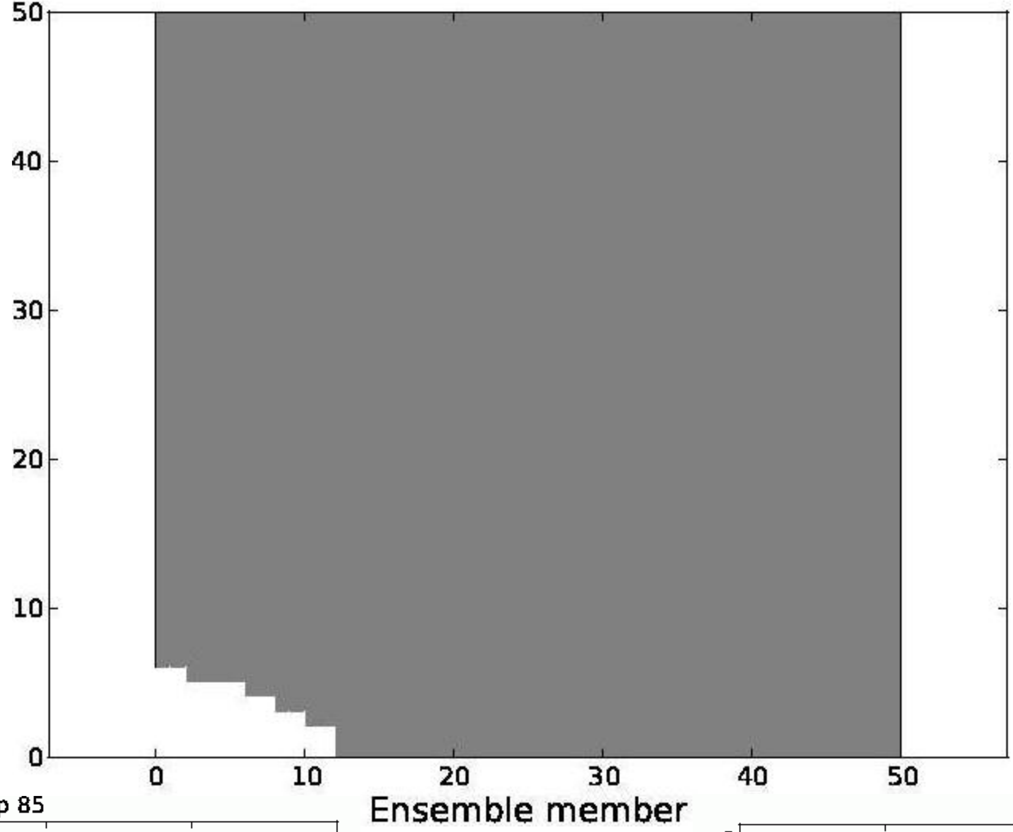
Example 1 SIR, hl 3000,k 50 Step 84-85

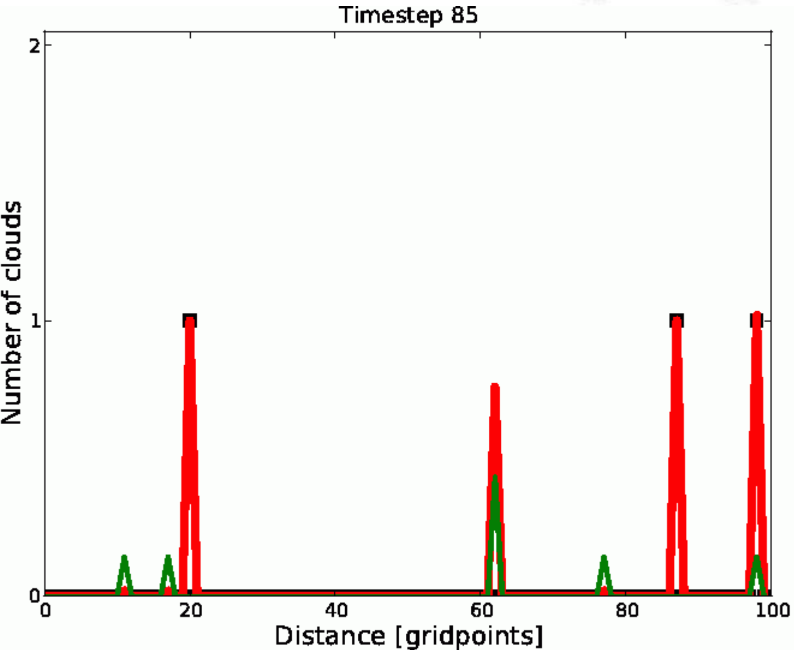


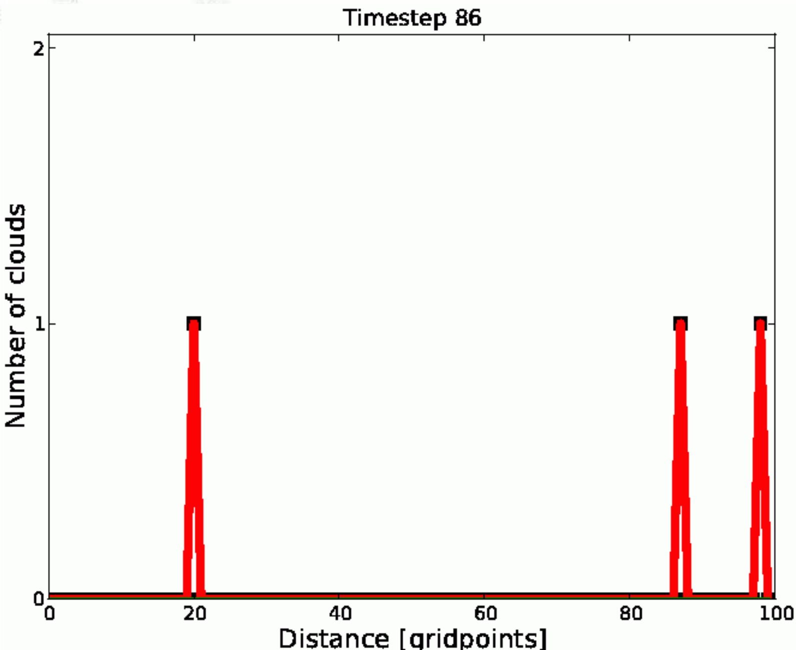




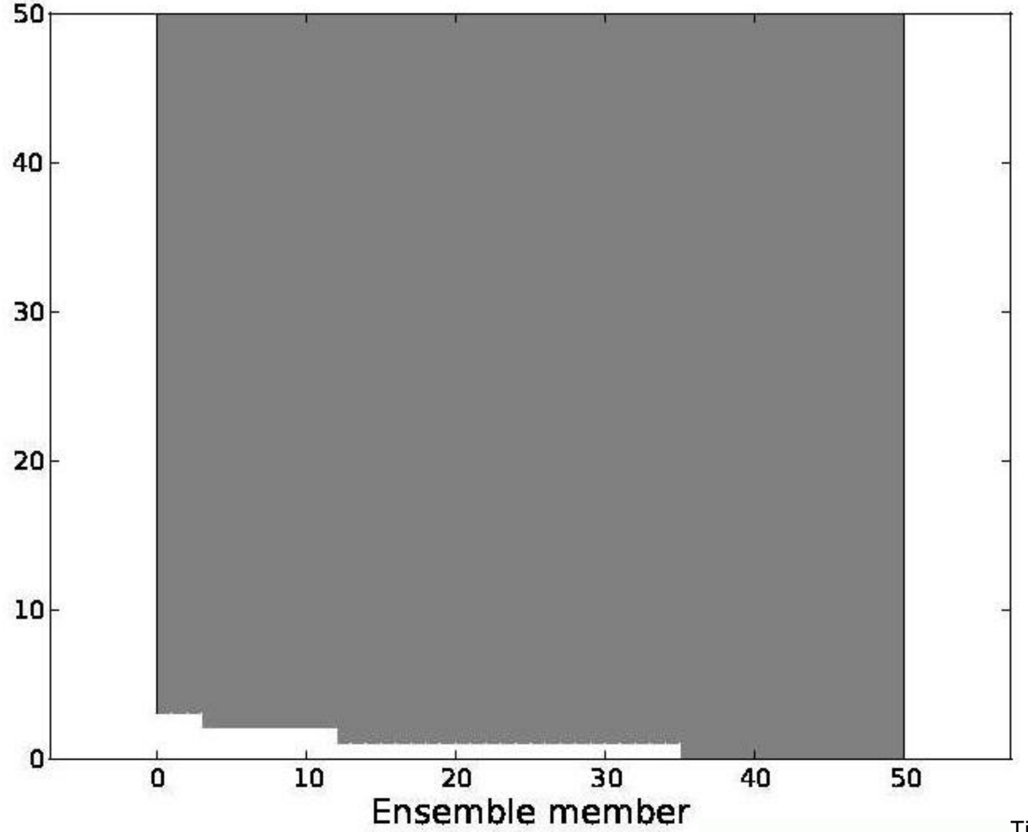
Example 1 SIR, hl 3000,k 50 Step 85-86

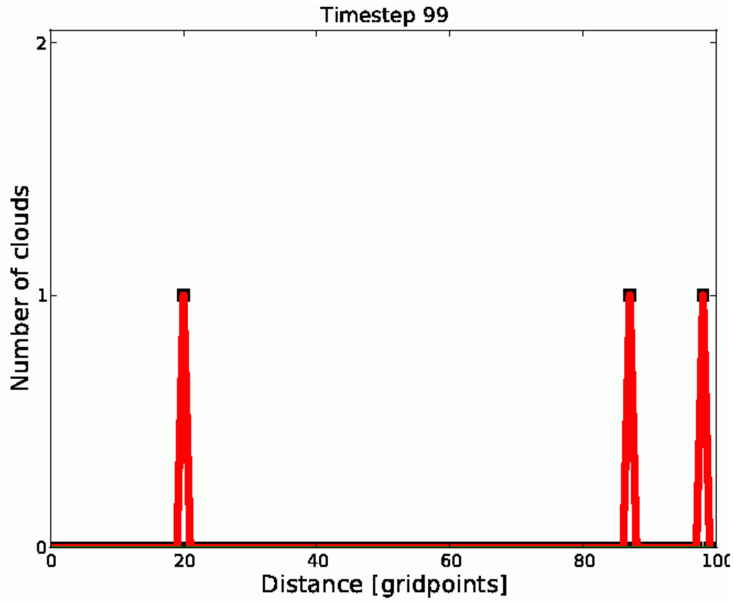


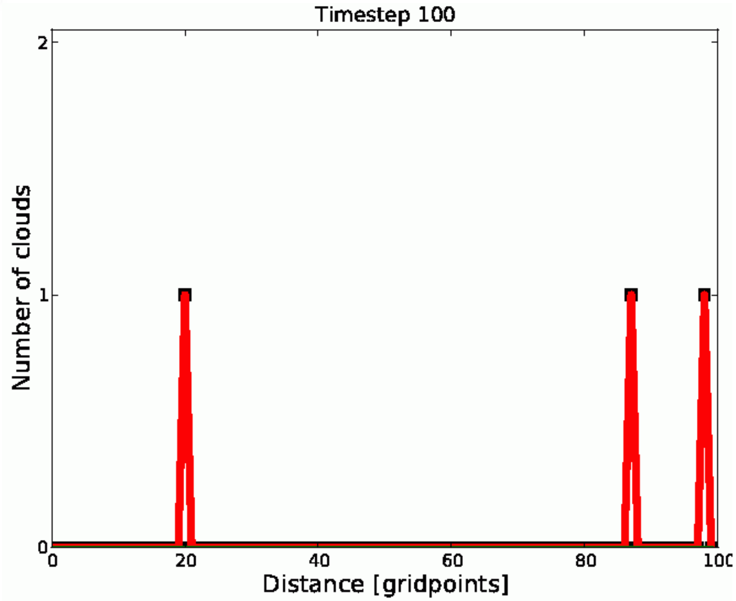




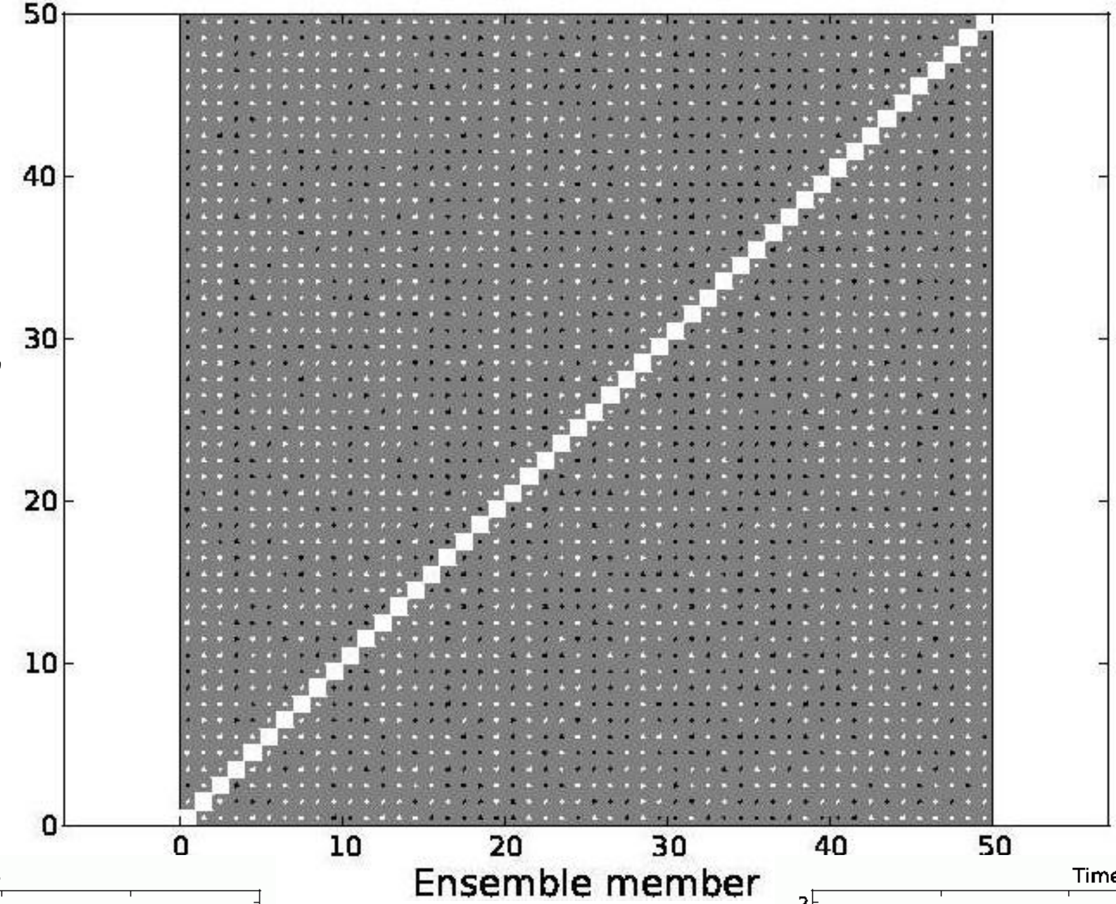
Example 1 SIR, hl 3000,k 50
Step 99 to 100

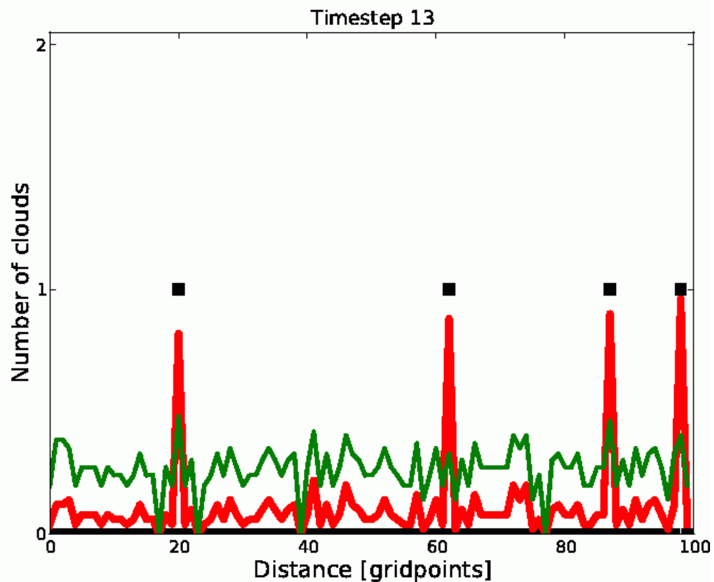


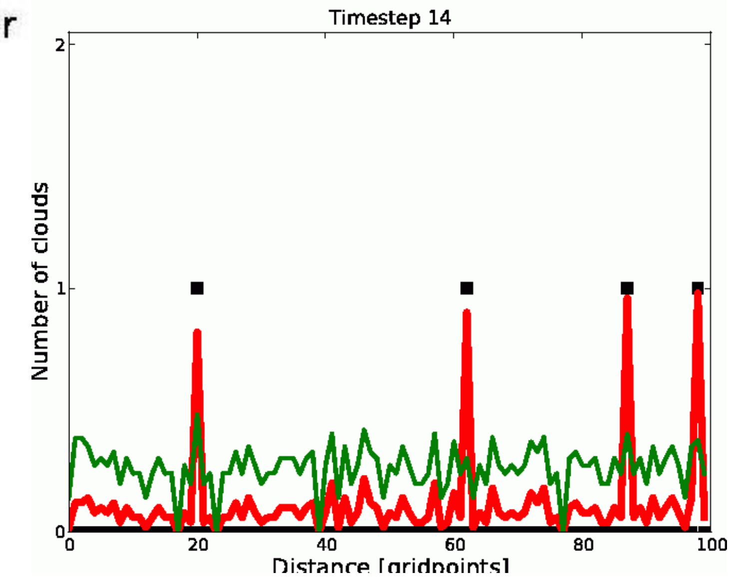




Example 1 ETKF, hl 3000,k 50 Step 13 to 14







Example 1 ETKF, hl 3000,k 50 Step 99 to 100