Advances in the Ensemble Kalman Filter for Historical Reanalysis

Jeff Whitaker and Gil Compo

....or

How to get even more from almost nothing!

Jeff Whitaker and Gil Compo

Data Assimilation for Reanalysis
(as opposed to NWP)

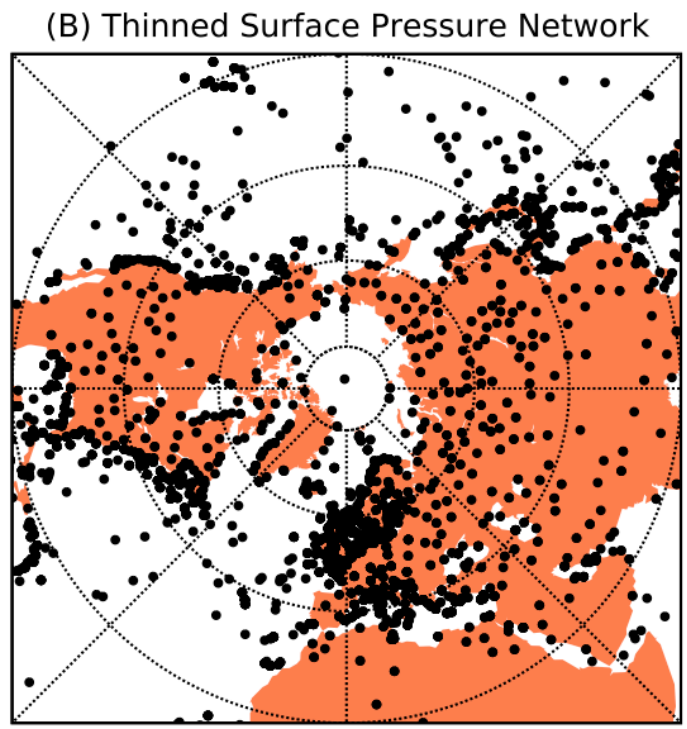
- Need to get the most out of very sparse data (100's or 1000's of obs versus millions).
  - Need flow-dependent background error covariance model.
- Need an estimate of time-varying analysis uncertainty (to estimate error bars).
  - Analysis-error variance for each variable.
- Willing to sacrifice some accuracy for the sake of homogeneity.

Comparison of 3DVar, 4DVar, EnKF

(*Whitaker, Compo and Thepaut, 2009, MWR, p. 1991)*

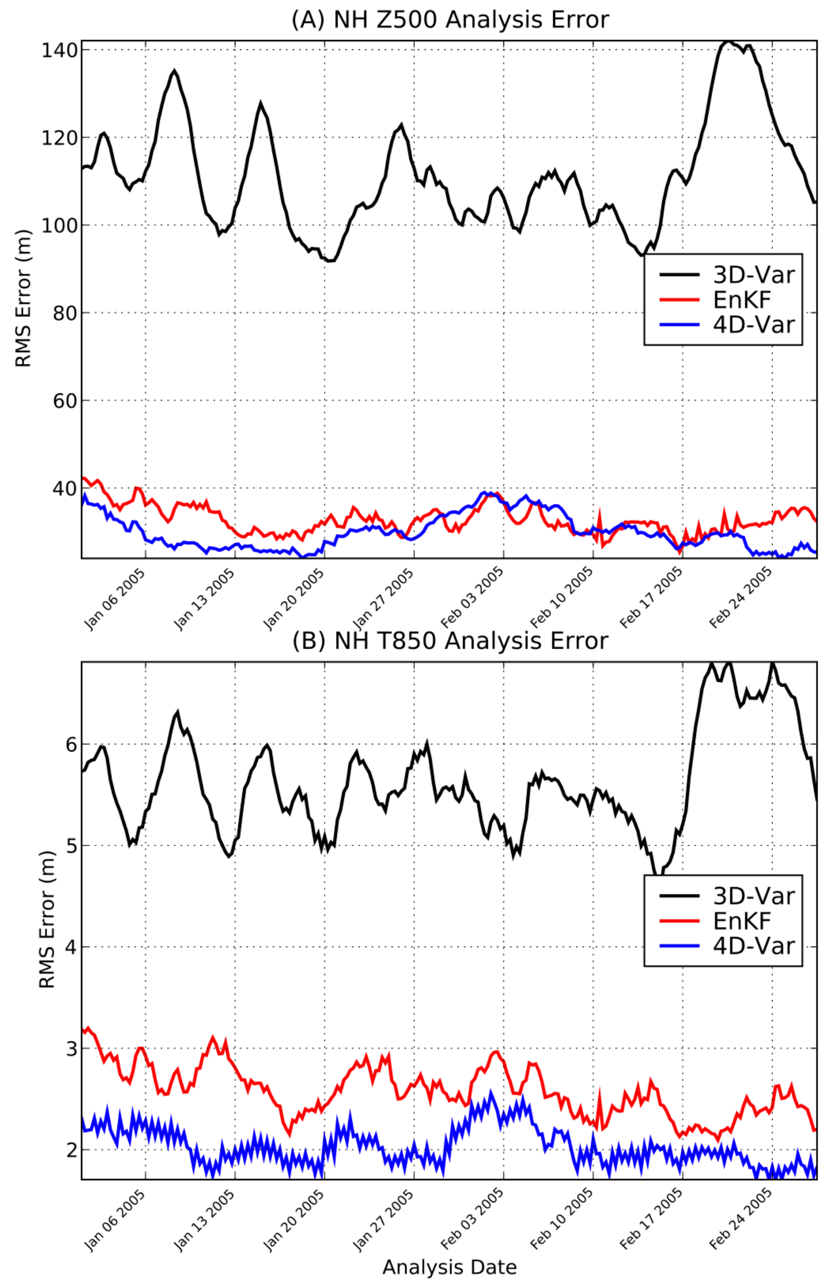
- Surface pressure only (2005412-200502), network thinned to look like 1930's

- Expts with ECWMF IFS (3D & 4DVar with retuned **B**) and NCEP GFS with EnKF.

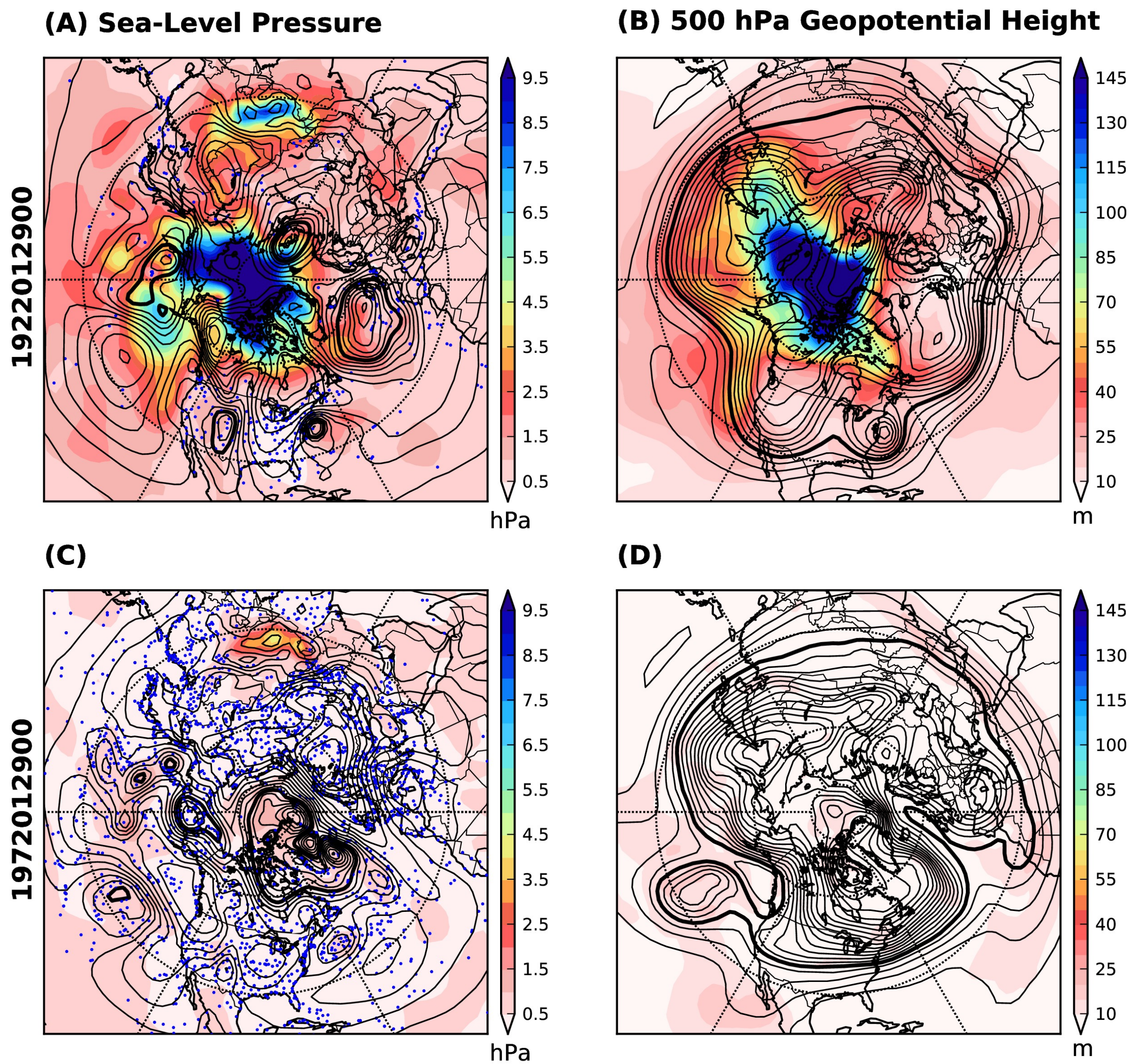


Comparison of 3DVar, 4DVar, EnKF

- 4DVar appears slightly better than EnKF, esp. for 850 T.
- When same model is used in the EnKF, these differences disappear (recent work with M. Hamrud and M. Bonavita at ECMWF)

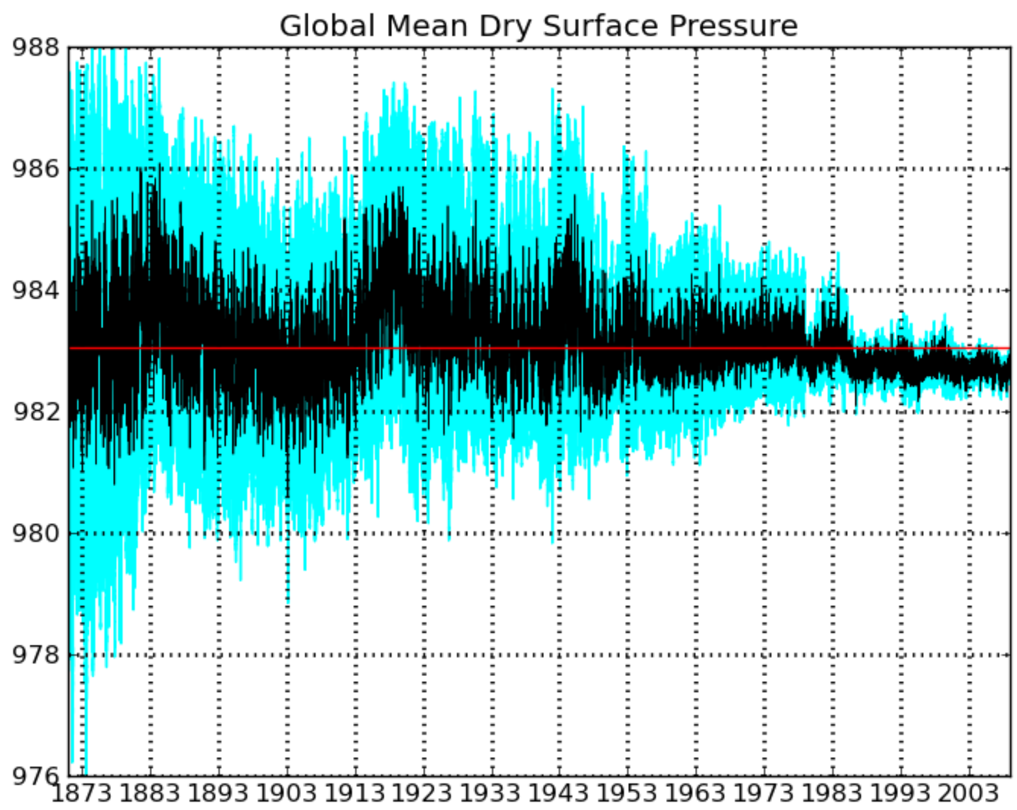


Time-varying analysis uncertainty



Time-varying analysis uncertainty

Ensemble mean (black) global mean dry surface pressure. Ensemble range shaded blue.



EnKF developments

- New multiplicative covariance inflation algorithm.
- Non-gaussian observation errors (account for gross observation errors).
- Ability to assimilate all types of obs (including satellite). Investigating ship winds…
- Fixed-lag smoother (Khare et al, 2008, Tellus, p. 97).

New multiplicative inflation

Applied to posterior – relaxes variance back to prior $\sigma_a^2 = (1-\alpha)\sigma_a^2 + \alpha\sigma_b^2$

$$\mathbf{x}_\mathrm{a}^{'} = \mathbf{x}_\mathrm{a}^{'} \sqrt{\alpha \frac{\sigma_b^2 - \sigma_a^2}{\sigma_a^2} + 1}$$

Motivation:

- sampling error largest where σ_b/σ_a is large (Sacher and Bartello 2008 MWR).
- model error is a larger fraction of background error in regions of dense/accurate obs (where σ_b/σ_a is large, Daley and Menard 1993 MWR).
- adaptively estimated inflation (Anderson 2009) looks like σ_b/σ_a

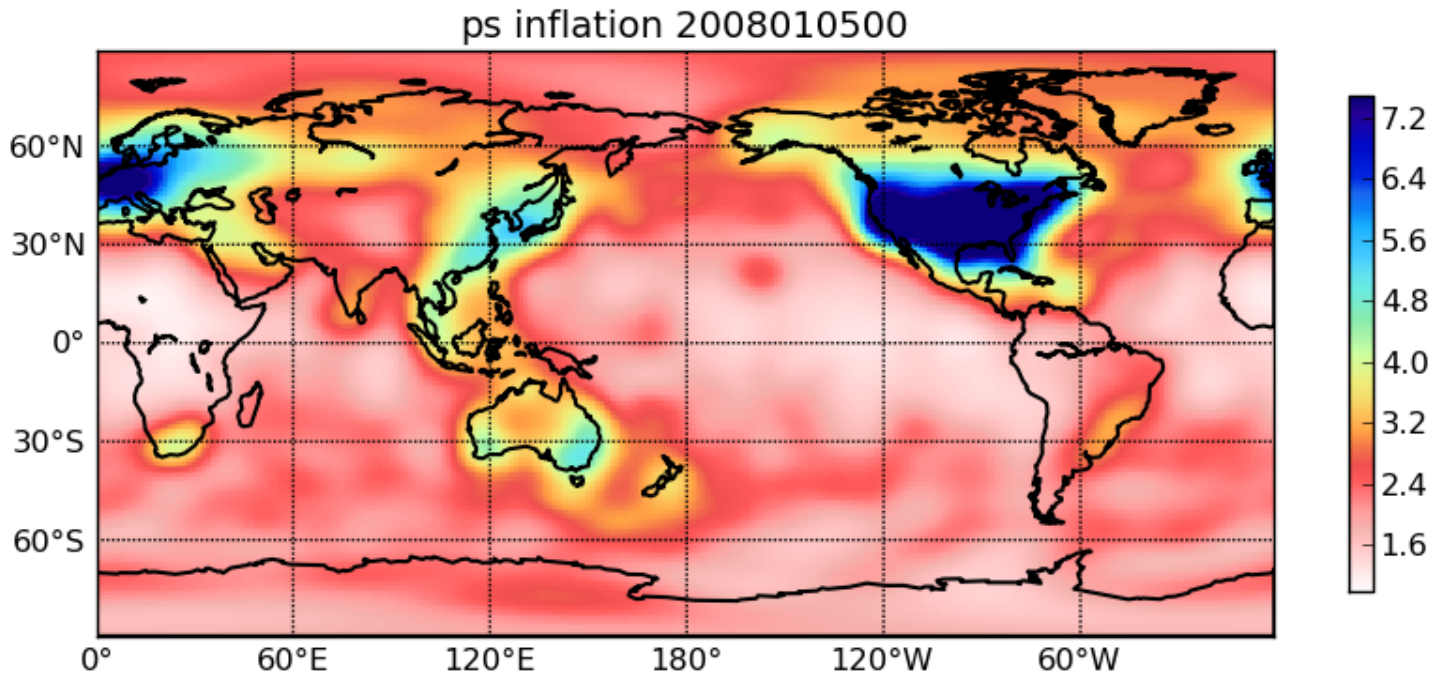
Pros:

- no inflation where there are no increments.
- more inflation where there are dense/accurate obs.

Cons:

- potentially large spatial gradients in inflation may disrupt growing structures.

New multiplicative inflation



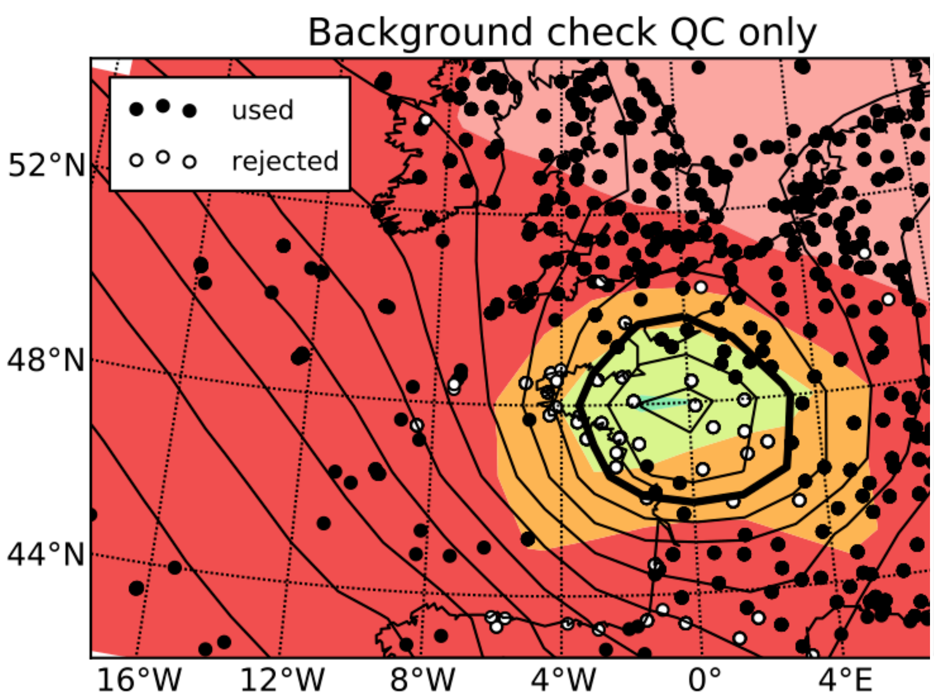
• 20CRv2 had too little inflation (too little spread) in data dense regions, too much in data sparse regions.

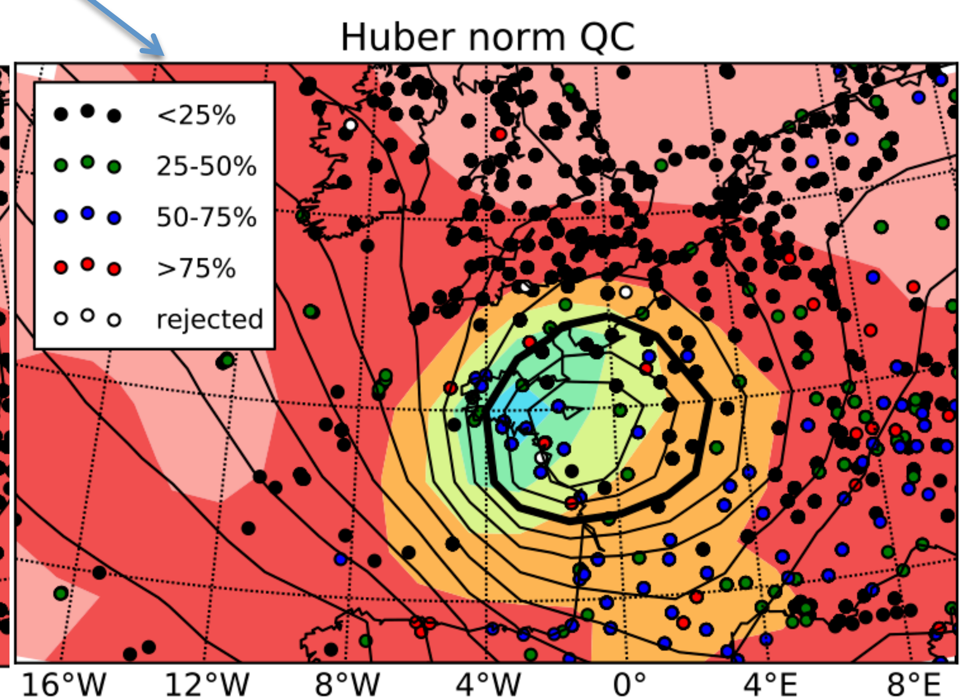
Dealing with non-gaussian ob errors

- Reject erroneous obs, but use good obs that are far from the background.
- Analagous to 'VarQC' in variational systems (Huber norm ob error PDF – gaussian with exponential tails).
- Alternative to 'buddy-check' used in 20CRv2.
- Algorithm:
  - Iteration solution in observation space.
  - Ob error modified at each iteration to account for heavy tails (variance multiplied by inverse probability that observation ***does not*** have a gross error – prob. of gross error is linearly proportional to distance from background).
  - If other obs support an outlier, probability of gross error will be decreased within the iteration.

Test of new QC – 'Lothar' storm Dec. 1999

probability of gross error



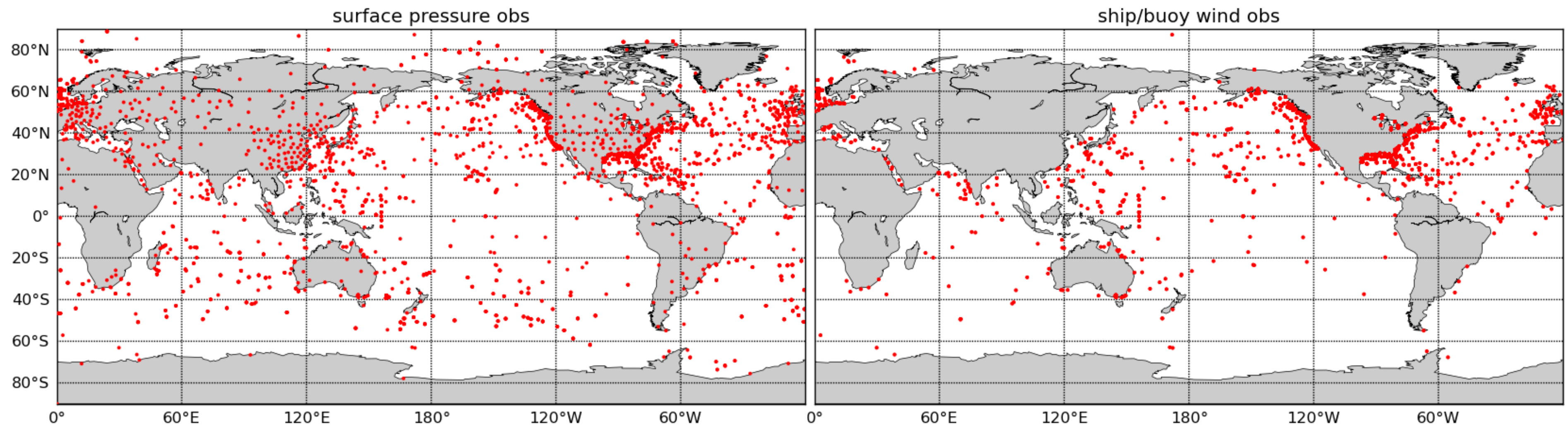


Ob rejected of departure greater than
3.2 times sqrt(ens spread + ob error var)

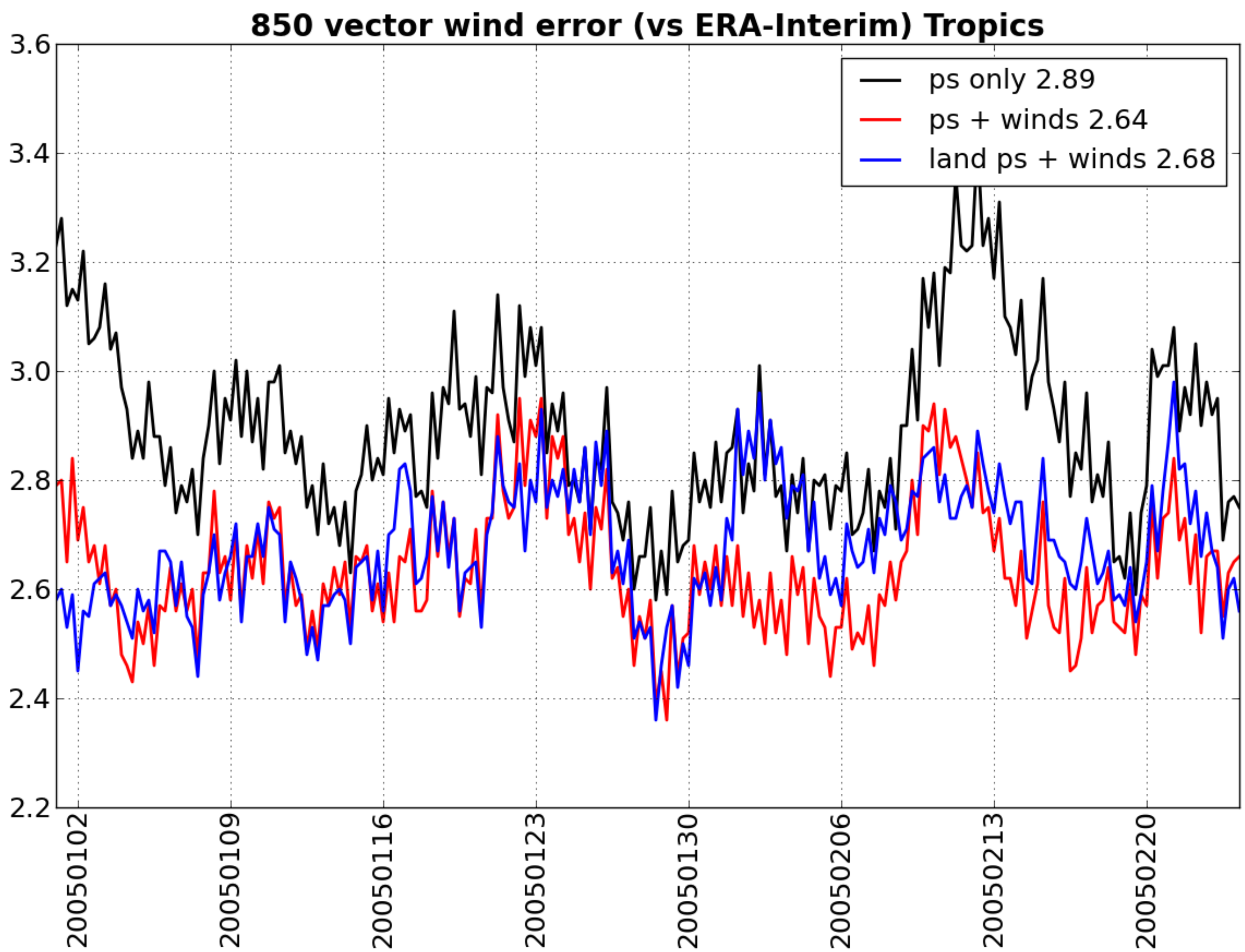
Ob rejected of departure greater than
16 times sqrt(ens spread + ob error var)

Should we try to assimilate ship winds? Test for Jan/Feb 2005

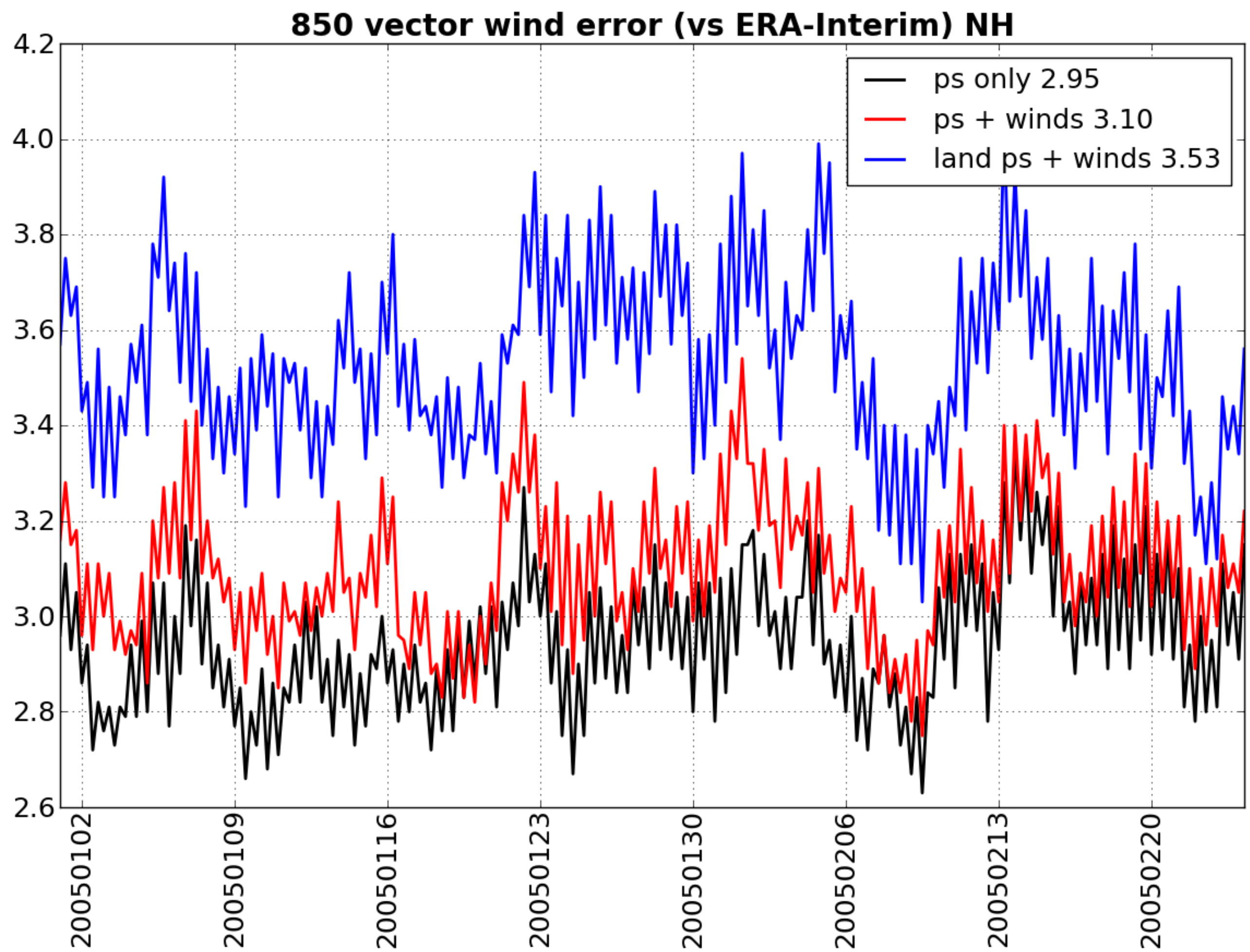
obs in NNR for 2005010100



Should we try to assimilate ship winds?
Test for Jan/Feb 2005



Should we try to assimilate ship winds?
Test for Jan/Feb 2005



Would more smoothing help?

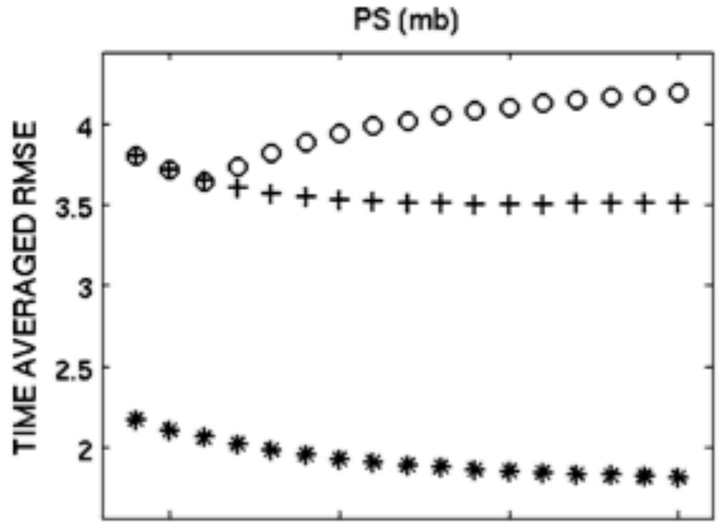
- Khare et al, 2008, Tellus, p. 97 present a fixed-lag smoother algorithm.
- Iterative process, starting with filter solution. Lag-k smoother uses obs k times past analysis time.

Would more smoothing help?

Open circles – no localization

Top line N=20,

Bottom line N=50



Effectiveness depends on:

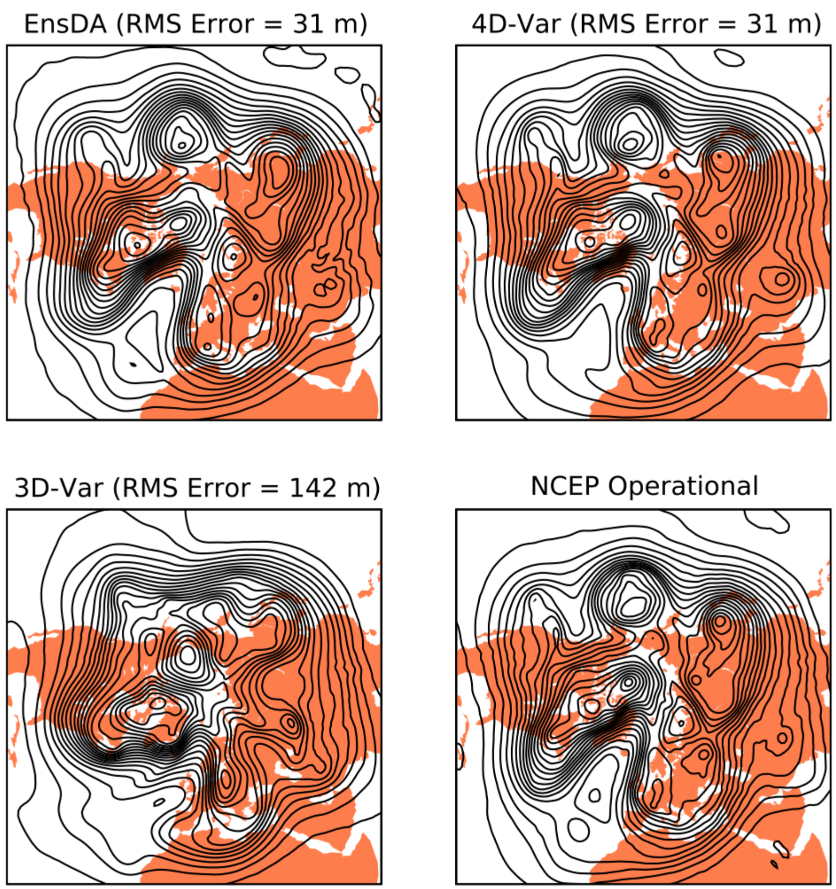
- Observation accuracy.
- Sparseness of network.
- Ensemble size (sampling error increases with lag).
- 10% improvement for N=20, 20% for N=50 in a perfect model.

Summary

- Observation QC and bias correction are **very important**.
- Model error treatment (systematic and random components) critical.
  - How to distinguish model from observation bias?
- Surface marine wind obs should help in tropics, but may be redundant in mid-lats.
- Fixed-lag smoothers should be investigated.

Comparison of 3DVar, 4DVar, EnKF

- 4DVar and EnKF both capture details of synoptic weathe 3DVar does not.



Time-varying analysis uncertainty

Predicted versus actual RMS difference
between ensemble mean first guess and obs.