Localized ensemble-based tangent linear models and their use in propagating hybrid error covariance models

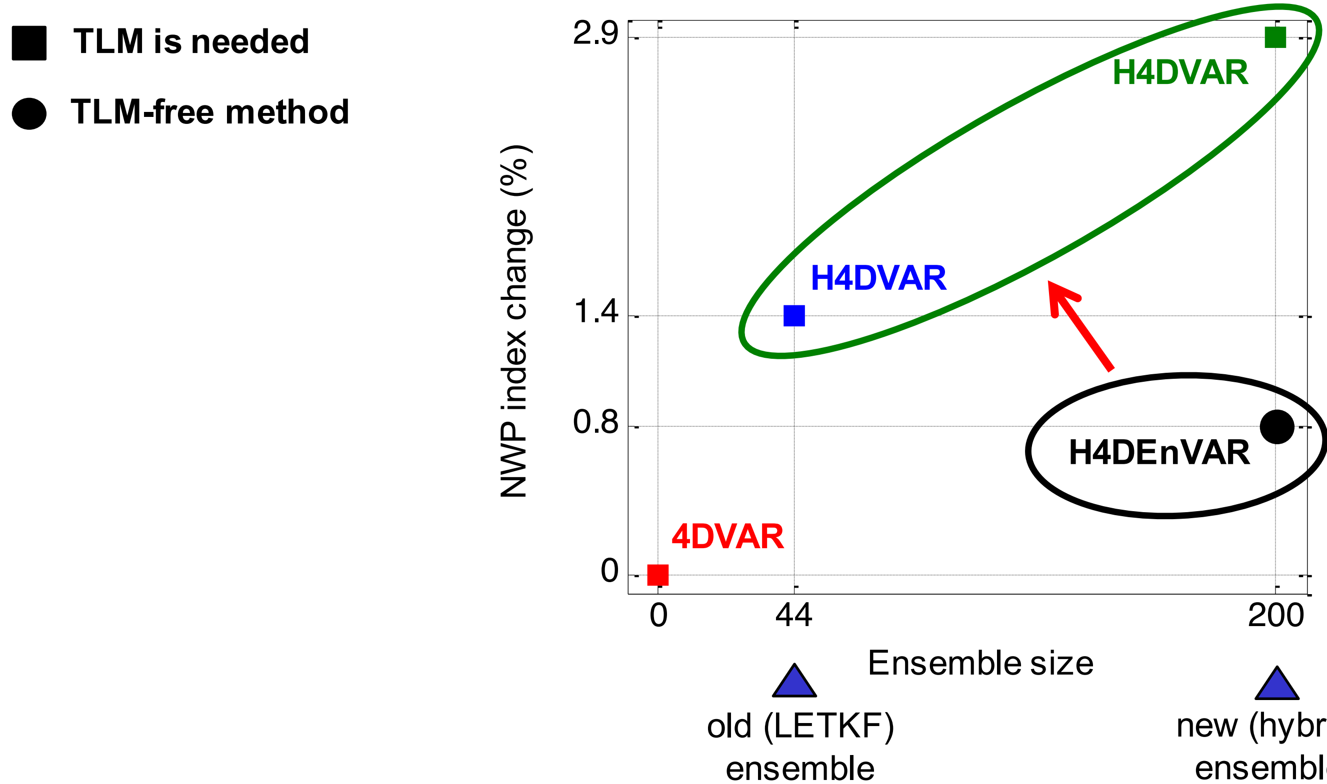
Sergey Frolov (UCAR), Craig H. Bishop (NRL-MRY)

Doug Allen (NRL-DC), David Kuhl (NRL-DC), Karl Hoppel (NRL-DC)

7th EnKF workshop
May 2016

Why we might want an ensemble-based TLM?

NWP index change data from UKMO



TLM-Based Hybrids (H4DVAR) are currently better than TLM-Free Hybrid (H4DEnVAR)

- Traditional TLMs are hard to maintain, are not projected to scale on the next generation computers, and are often not available for coupled models.
- Can we secure benefits of TLM-based estimation while maintaining computational scalability and easy-of maintenance of ensemble-based systems?

Data adapted from Barker (2016)

What is an Ensemble-based TLM?

Traditional TLM

$$\begin{cases} \delta \mathbf{x}_j = M_{j,i}\left[\mathbf{x}_i + \delta \mathbf{x}_i\right] - M_{j,i}\left[\mathbf{x}_i\right] = \mathbf{M}_{j,i}\delta \mathbf{x}_i + \varepsilon_i \\ \lim\limits_{\delta \mathbf{x}_i \to 0}\left[\varepsilon_i\right] = 0 \end{cases}$$

- The Jacobian of a non-linear model
- Optimized for infinitesimal perturbations
- Optimal for finding the <u>MODE</u> of the distribution using outer loops of the 4DVAR

Statistical TLM

$$\mathbf{M}_{j,i}^{stat} = \left\langle \left\{ M_{j,i}\left[\overline{\mathbf{x}_i} + \delta \mathbf{x}_i\right] - \overline{M_{j,i}\left[\mathbf{x}_i\right]} \right\} \delta \mathbf{x}_i^{\,T} \right\rangle \left\langle \delta \mathbf{x}_i \delta \mathbf{x}_i^{\,T} \right\rangle^{-1}$$

- Optimized for finite-size perturbations
- Optimal for finding the <u>MEAN</u> of the distribution without an outer loop in the 4DVAR

Ensemble-based TLM

$$\mathbf{M}_{j,i}^{stat} \approx \mathbf{M}_{j,i}^{\mathrm{ens}} = \mathbf{X}_j \mathbf{X}_i^T \left(\mathbf{X}_i \mathbf{X}_i^T \right)^{-}$$

OR

$$\mathbf{M}_{j,i}^{stat} \approx \mathbf{M}_{j,i}^{\mathrm{ens}} = \mathbf{X}_j \left(\mathbf{X}_i^T \mathbf{X}_i \right)^{-} \mathbf{X}_i^T$$

$$size(\mathbf{X}) = [\mathrm{npoints}\ ;\ \mathrm{nens}]$$

$$(.)^{-} = pinv(.)$$

- An ensemble approximation to the statistical TLM
- When ensemble matrix **X** is full-rank and ensemble perturbations are small, ensemble-based TLM is the same as a Traditional TLM

Need for localization of ETLM

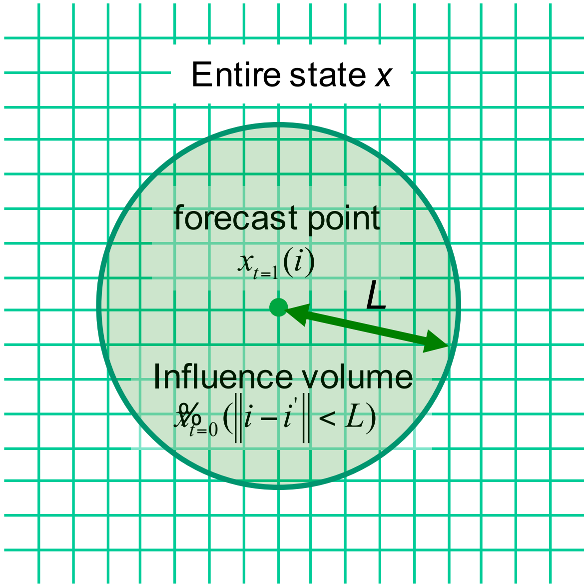
- Previous uses of ETLM:
  - Low-dimensional (1D) systems, ensemble sensitivity, EOF-based modeling
  - Few ensemble members are needed to maintain the full rank of $\mathbf{M}^{\text{ens}}$.

- For hybrid-4DVAR
  - We need to propagate full-rank P^{clim}:

$$\mathbf{P}_{k,k}^{\text{clim}} = \mathbf{M}_{k,0}\mathbf{P}_{0,0}^{\text{clim}}\left(\mathbf{M}_{k,0}\right)^{T}$$

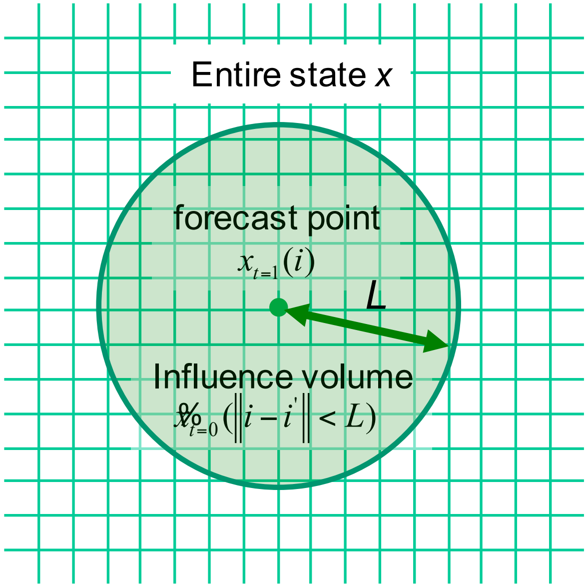
  - Updated P^{clim} will only maintain full-rank if TLM and ADJ are full rank
  - Localization is needed because we have fewer ensemble members than state variables for a typical NWP model (Frolov and Bishop 2016, MWR)

Local ensemble-based TLM



– Similar to finite difference models, ETLM updates each point based on the influence volume (stencil)

Local ensemble-based TLM



for each points "i" build an ensemble regression model

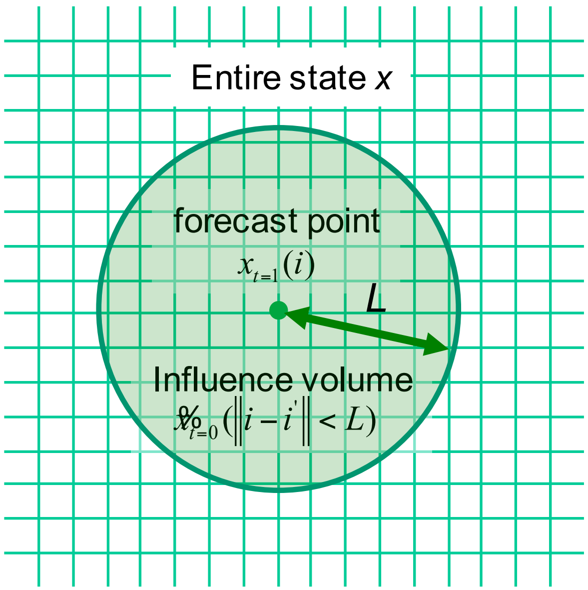
$$\underbrace{x_{t=1}(i) = \tilde{\mathbf{X}}_1}\underbrace{\left[\tilde{\mathbf{X}}_0^T\tilde{\mathbf{X}}_0\right]^{-1}\tilde{\mathbf{X}}_0^T\tilde{x}_0}$$

Defined at the prediction point at future time step

Defined for the influence volume at the previous time step

- Similar to finite difference models, ETLM updates each point based on the influence volume (stencil)
- Update is computed using an ensemble regression model for each point

Local ensemble-based TLM

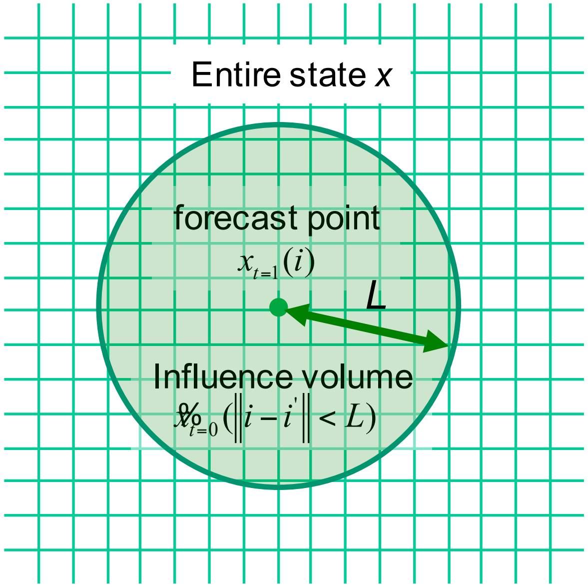


for each points "i" build an ensemble regression model

$$\underbrace{x_{t=1}(i) = \tilde{\mathbf{X}}_1}_{\text{Defined at the prediction point at future time step}} \underbrace{\left[\tilde{\mathbf{X}}_0^T \tilde{\mathbf{X}}_0\right]^{-1} \tilde{\mathbf{X}}_0^T \tilde{x}_0}_{\text{Defined for the influence volume at the previous time step}}$$

- Similar to finite difference models, ETLM updates each point based on the influence volume (stencil)
- Update is computed using an ensemble regression model for each point
- If number of ensemble members is greater than number of grid points in the stencil, the ETLM computation is exact.

Local ensemble-based TLM



for each points "i" build an ensemble regression model

$$\underbrace{x_{t=1}(i) = \tilde{\mathbf{X}}_1}_{\text{Defined at the prediction point at future time step}} \underbrace{\left[\tilde{\mathbf{X}}_0^T\tilde{\mathbf{X}}_0\right]^{-1}\tilde{\mathbf{X}}_0^T\tilde{x}_0}_{\text{Defined for the influence volume at the previous time step}}$$

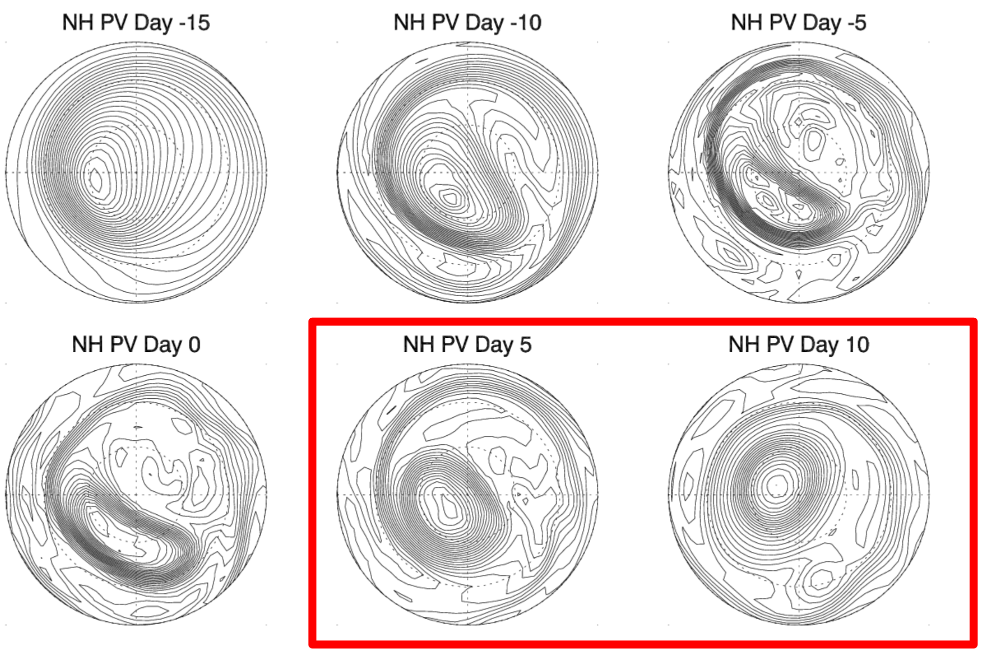
- Similar to finite difference models, ETLM updates each point based on the influence volume (stencil)
- Update is computed using an ensemble regression model for each point
- If number of ensemble members is similar to the DOF that predict future state x_1 the ETLM computation is accurate.

Validation tests

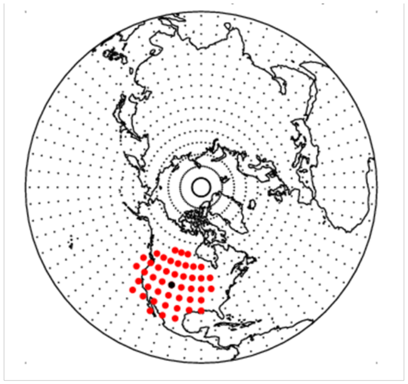
- Test with 1D wave dispersion problems (not shown)
  - [Frolov Bishop 2016 MWR]
- Tests with coupled Lorenc-93 model (not shown)
  - [Bishop et.al. under review QJRMS]
- Tests with 2D shallow-water model on a sphere (shown)
  - [Allen et.al. under review MWR]
- Tests with a coarse resolution (T47) NWP model(shown)
  - Active research

Shallow water model

Snapshots of potential vorticity in NH

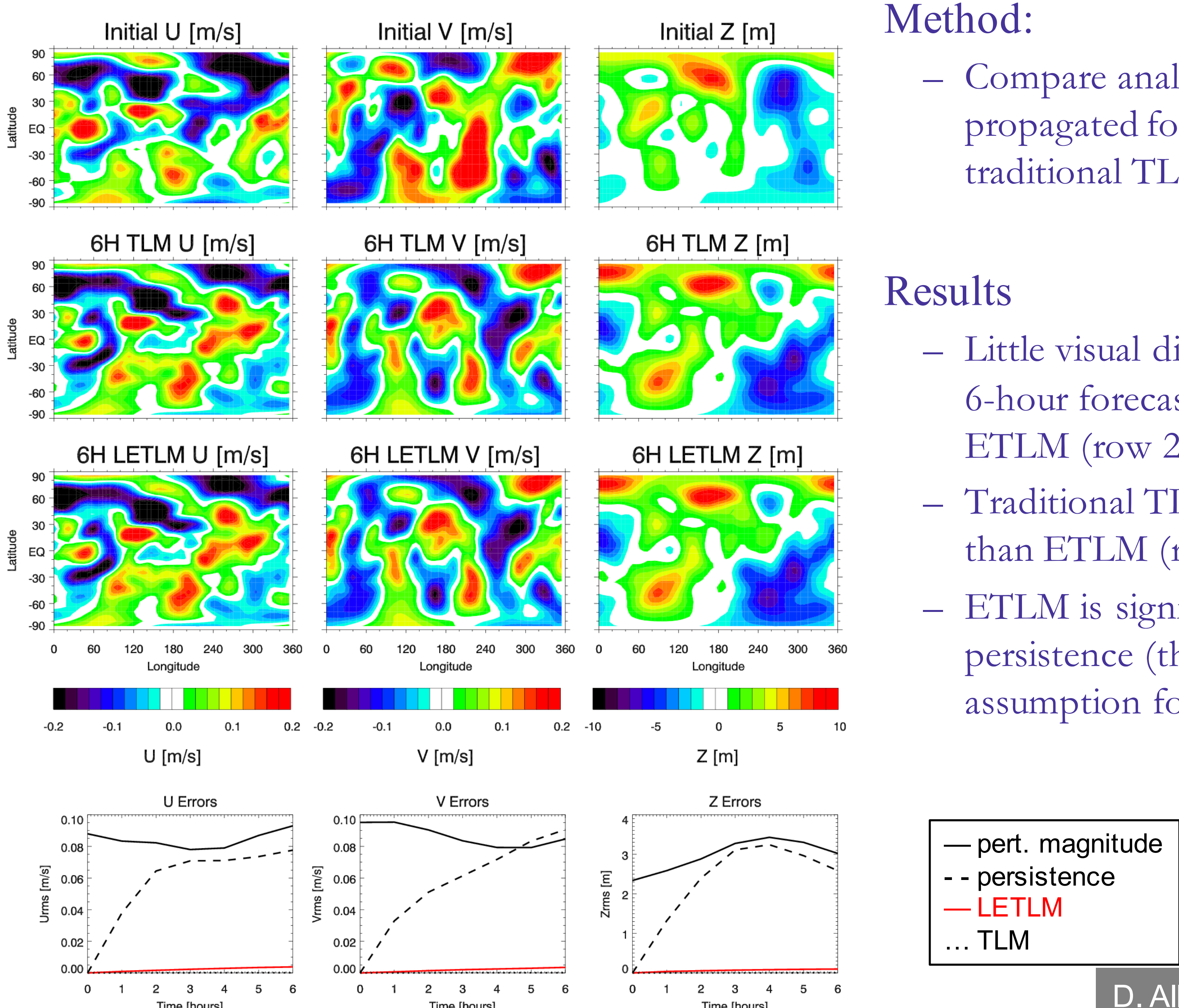


Example of the LETLM stencil



- System:
  - Shallow water equation + ozone as a tracer
  - Spectral T21 resolution (5.8° lon spacing at equator)
  - Truth run with wave 1 topographic forcing (top left figure)
  - Experiment runs during day 5-10 (highlighted in red)
  - Perturbed observations EnKF with 100 members
  - Stencil size 1000km, LETLM timestep 1 hour

Comparisons with a traditional TLM



Method:

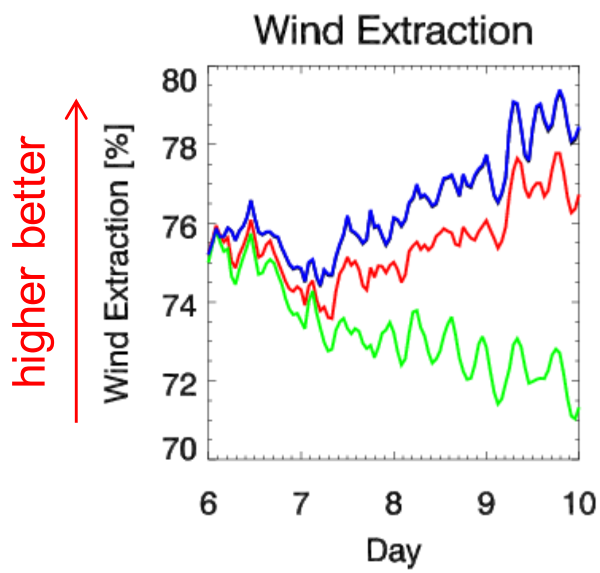
- Compare analysis correction propagated for 6 hours using traditional TLM and LETLM

Results

- Little visual difference between 6-hour forecast with TLM and ETLM (row 2 and 3)
- Traditional TLM is slightly better than ETLM (row 3)
- ETLM is significantly better than persistence (the EnVar assumption for P^{clim})

D. Allan et. al. (submitted MWR)

Testing of ETLM in a Hybrid-4DVAR system



Experiment	WEP
Hybrid-4DVar (with LETLM)	78.22
Hybrid-4DVar	78.17
Hybrid-3EnVar	76.71
NODATA	72.12

Method:

- Compare Hybrid-4DVar executed with traditional TLM and LETLM

Results:

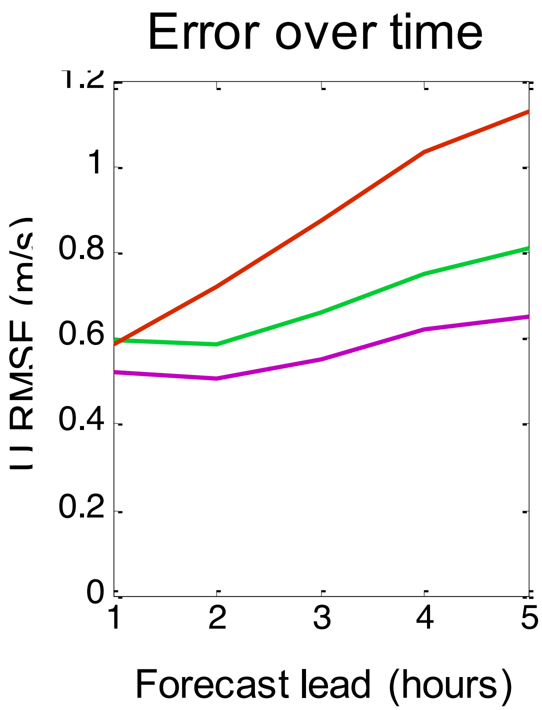
- Little difference between results with TLM and LETLM

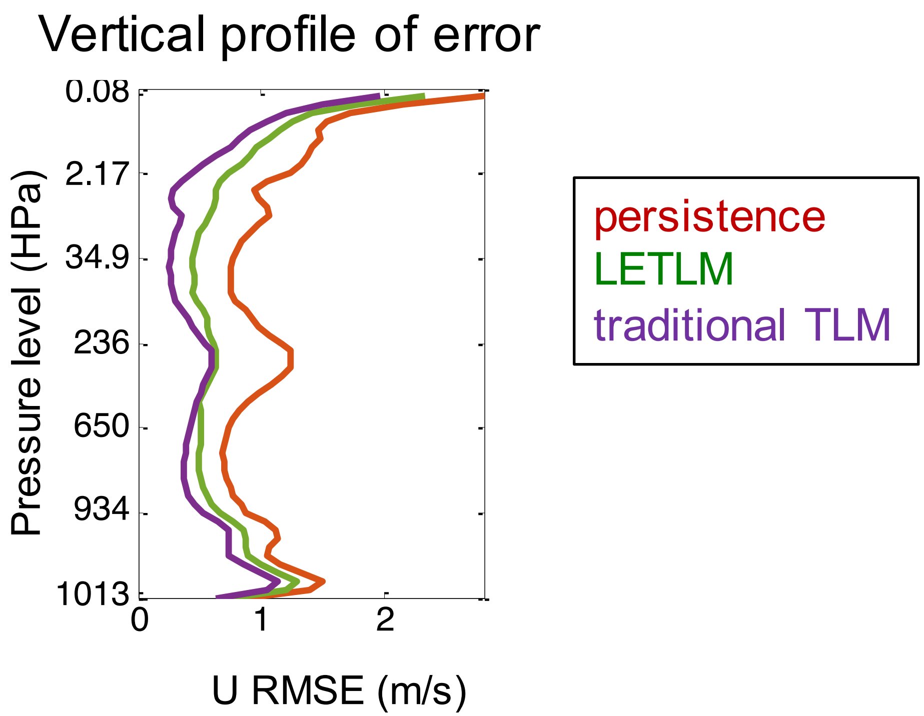
D. Allan et. al. (submitted MWR)

Experiments with an NWP model

- Ensemble generation
  - NAVGEM global model (T47L60)
  - NAVDAS-AR (4DVAR) was cycled for 5 days
  - 200 ET ensembles members were cycled for 5 days
- ETLM
  - Optimal (global) stencil size: L=500 km and +-2 levels in the vertical
    - Equivalent to 2 grid points at the equator or speed of 136 m/s
  - State variables: U,V,T, and geopotential
- Perturbations
  - 4DVAR analysis increment
- Validation
  - Difference between two nonlinear runs
  - NOGAPS TLM

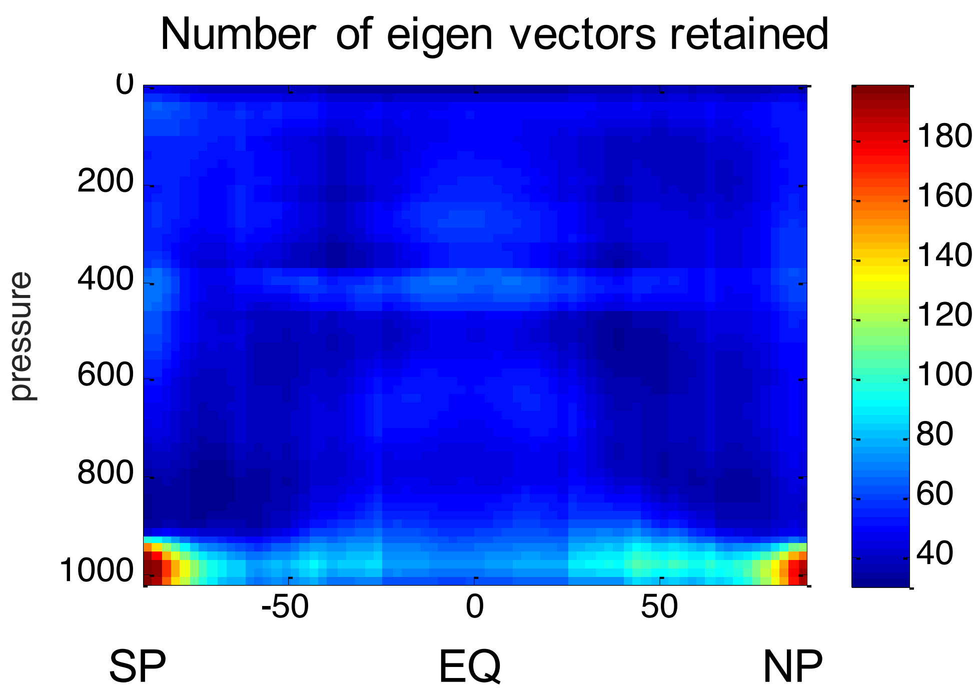
EARLY Results for 3D NWP (illustrated for U velocity)

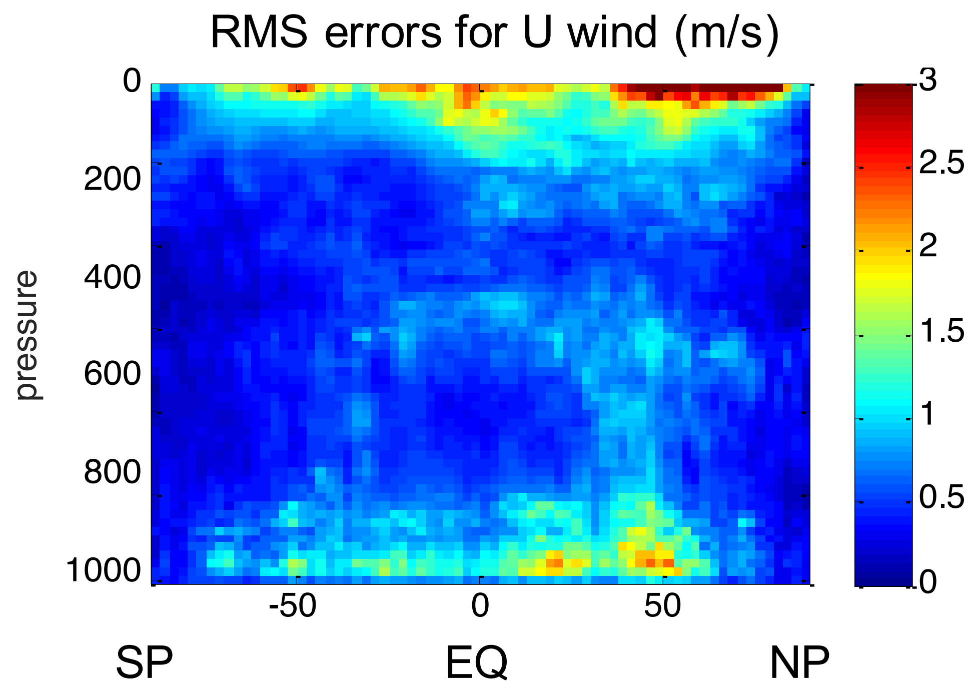




- LETLM has skill over persistence and is slightly worse than the traditional TLM
- LETLM is very young (~2 month of work) vs. traditional TLM (~6 years)

Next steps for NWP LETLM





- Improve ensemble (switch to perturbed obs. from ET?):
  - on average, 45 eign-modes are retained out of 200 ens. members;
  - ET pert. grow differently than forecast error;
- Tune LETLM differently with latitude and height.

Summary

- LETLM is a promising novel way to construct TLM and ADJ operators for data assimilation and sensitivity studies
- ETLM will converge to the true TLM given enough ensemble members
- Localization method is proposed to deal with low-rank ensemble approximation

- Results in the 2D shallow water model show that LETLM is as skillful as a traditional TLM in a Hybrid-4DVAR study

- Active research is underway to develop a skillful LETLM for a 3D global weather model
  - ~100 members might be enough?