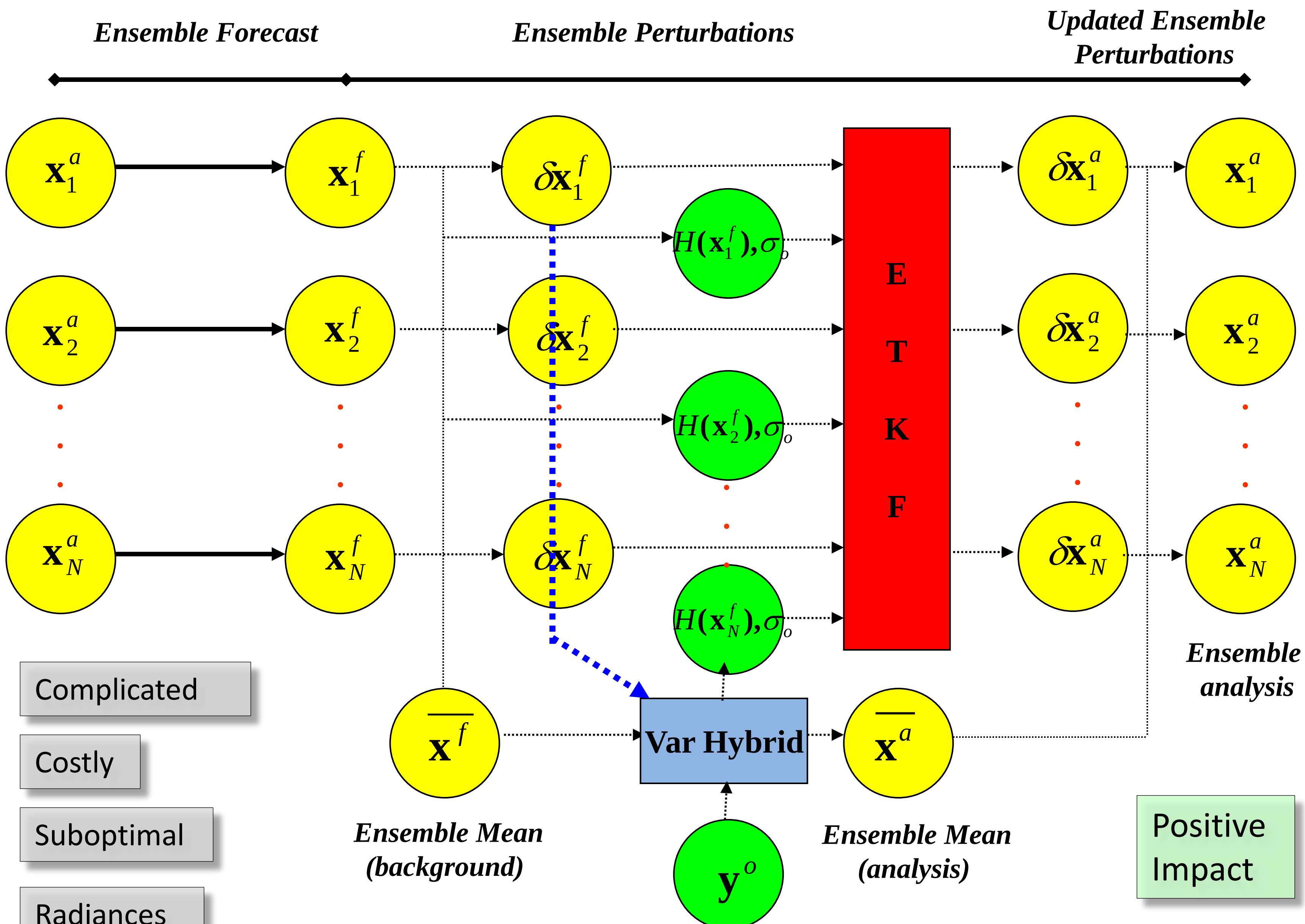
An Integrated Ensemble/Variational Hybrid Data Assimilation System

DAOS Working Group Meeting
September 2012

Tom Auligné, *National Center for Atmospheric Research*

Acknowledgements:
Contributors: Luke Peffers (AFTAC), Chris Snyder (NCAR)
Funding: AFWA

Variational/Ensemble Hybrid DA



Complicated

Costly

Suboptimal

Radiances

Positive Impact

Variational/Ensemble Hybrid DA

Ensemble Covariance included in 3D/4DVAR cost function through state augmentation
(Lorenc 2003, Wang et al. 2008)

$$J(\mathbf{x}_1^{'},\boldsymbol{\alpha}) = \beta_1 J_1 + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2}\mathbf{x}_1^{'T}\mathbf{B}^{-1}\mathbf{x}_1^{'} + \beta_2 \underline{\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{C}^{-1}\boldsymbol{\alpha}} + \frac{1}{2}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}^{'}\right)^T\mathbf{R}^{-1}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}^{'}\right)$$

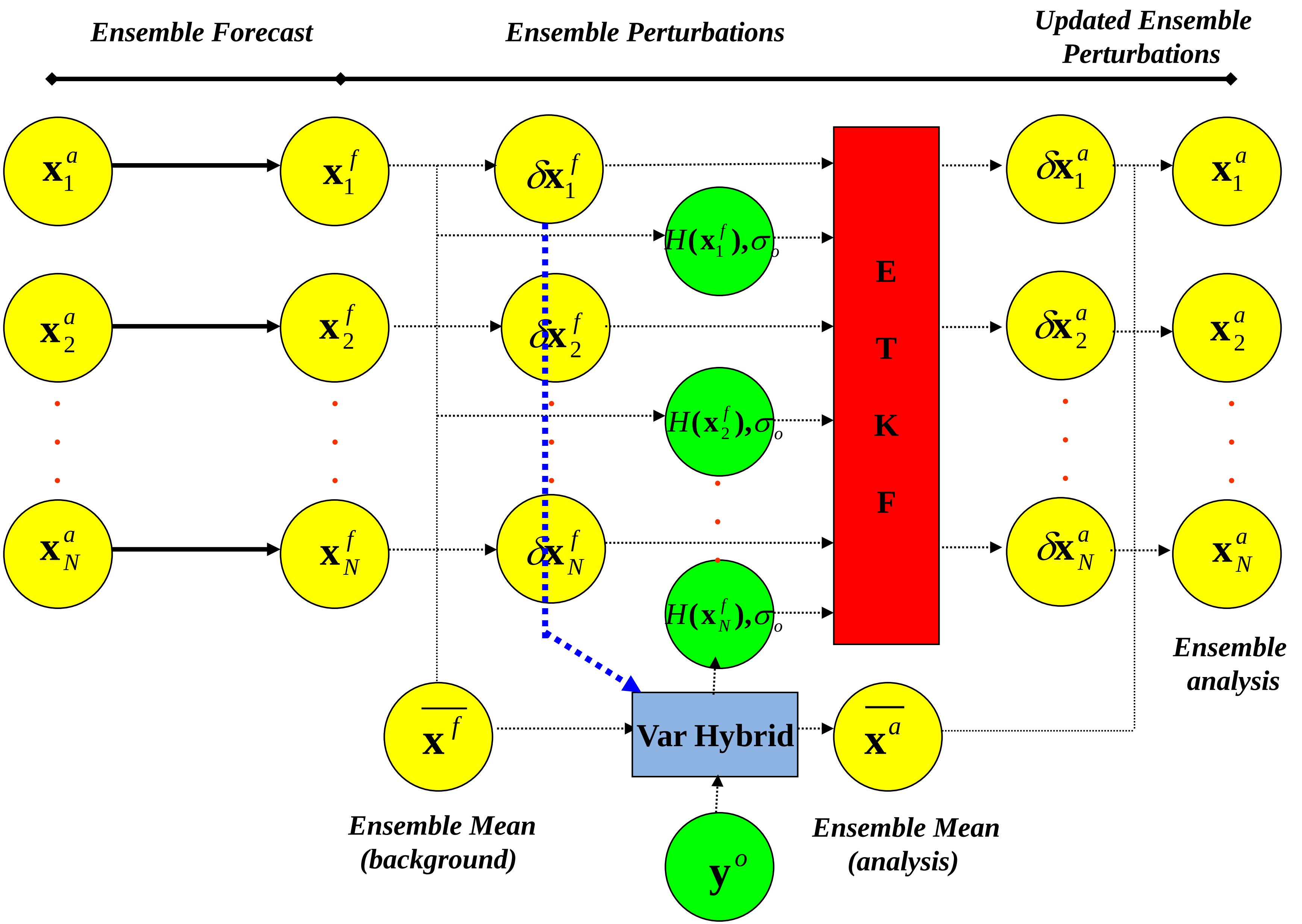
Extra term associated with extended control variable

$$\mathbf{x}^{'} = \mathbf{x}_1^{'} + \underline{\sum_{k=1}^{K}\left(\boldsymbol{\alpha}_k \circ \mathbf{x}_k^e\right)}$$

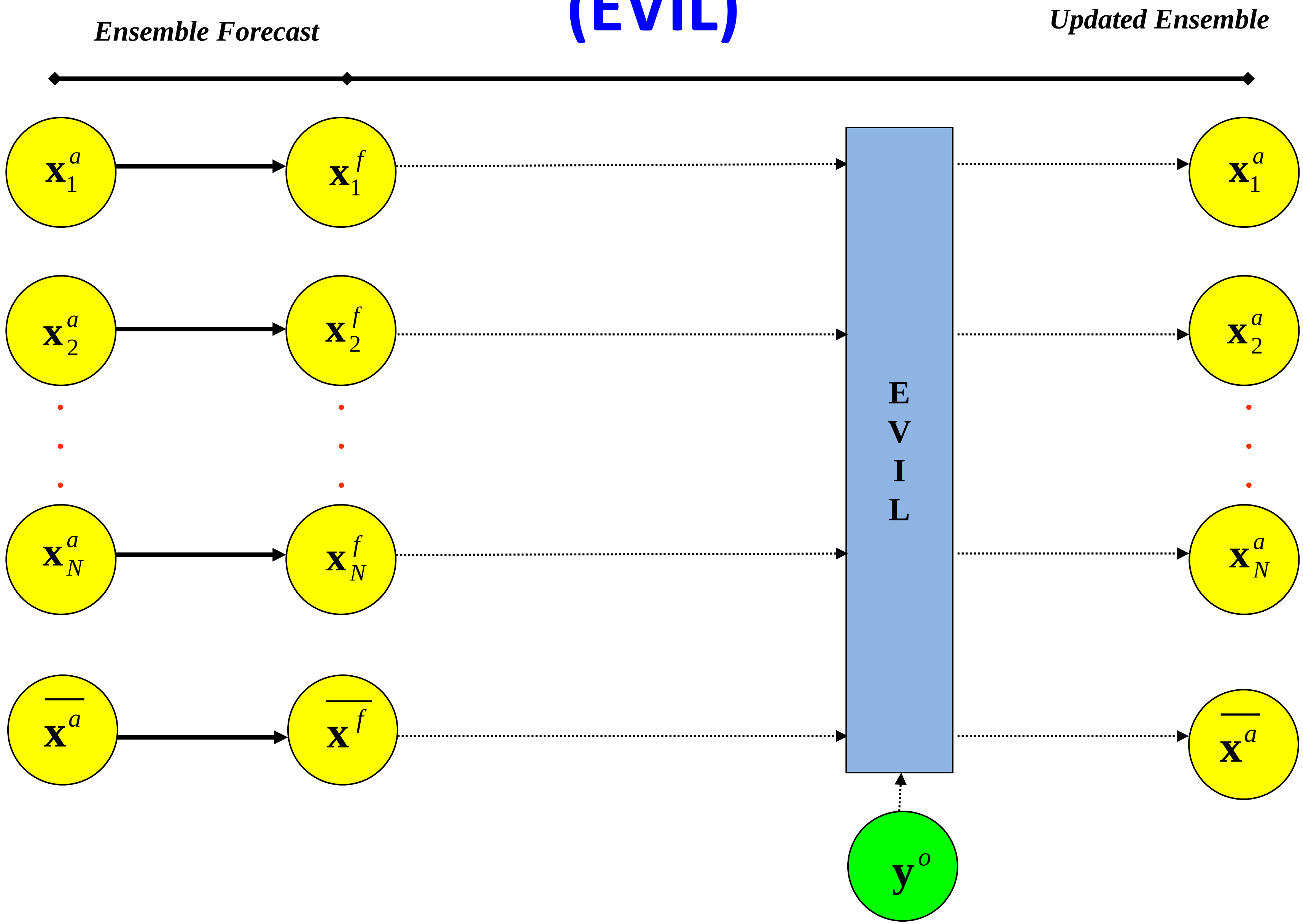
Extra increment associated with ensemble

<u>Missing piece</u>: update of the ensemble perturbations in VAR

Variational/Ensemble Hybrid DA



Ensemble Variational Integrated Lanczos (EVIL)



EVIL: Methodology

1. "Classic" Hybrid Ensemble Variational Data Assimilation
2. Update posterior ensemble perturbations

$$P_a = [J''(x_a)]^{-1}$$

$$P_a^{1/2} = [J''(x_a)]^{-1/2}$$

related to MLEF approach (Zupanski, MWR 2005)

$$P_a^{1/2} = V\Lambda^{-1/2}V^T P_f^{1/2}$$

with [λ,v] eigenpairs of J" (rank N)

$$P_a^{1/2} \approx [I + \tilde{V}(\tilde{\Lambda}^{-1/2} - I)\tilde{V}^T]P_f^{1/2}$$

(rank K << N)
related to Var Preconditioning (Fisher and Courtier, 1995)

<u>**Lanczos minimization:**</u>

$$[\Theta, Z] \approx [\tilde{\Lambda}, \tilde{V}]$$

with [θ,z] Ritz pairs

In "full ensemble" mode, the posterior ensemble perturbations:

$$\delta x_a^k \approx [I + Z(\Theta^{-1/2} - I)Z^T]\delta x_f^k$$

related to ETKF

EVIL: Methodology

Localization:

$$\tilde{\alpha} = C_\alpha^{-1/2}\alpha$$ through Recursive Filters (Purser et al., MWR 2003)

In "full ensemble" mode:

$$\delta x_a^k \approx C_\alpha^{1/2}[I + Z(\Theta_i^{-1/2} - I)Z^T]C_\alpha^{-1/2} \circ \delta x_f^k$$

Hybrid mode (*i.e.* combination of P_f and B)**:**

$$\tilde{\delta x}_f^k = \begin{pmatrix} \delta x_f^k \\ Rand(B) \end{pmatrix} \qquad \tilde{C} = \begin{pmatrix} C_\alpha & 0 \\ 0 & I \end{pmatrix}$$

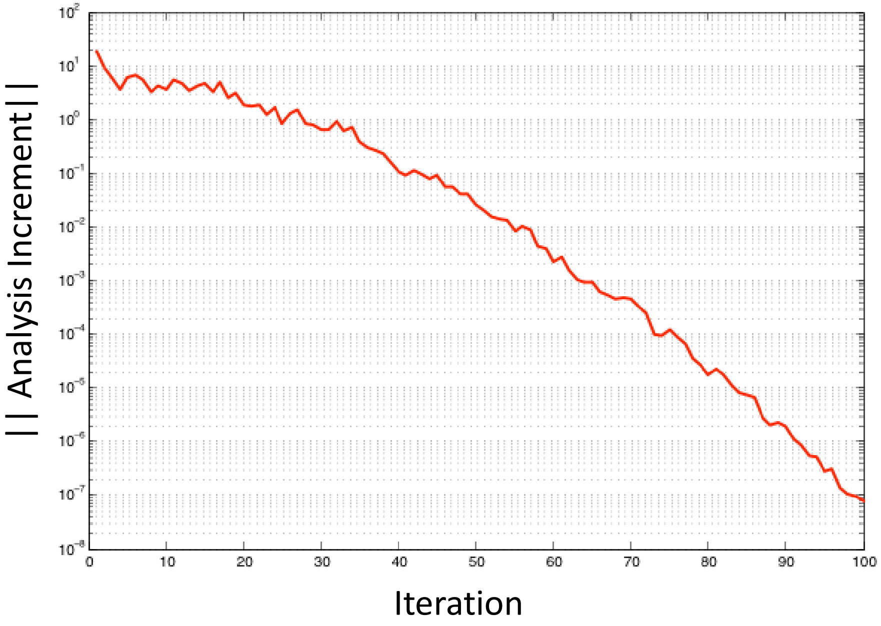
$$\tilde{\delta x}_a^k \approx \tilde{C}^{1/2}[I + Z(\Theta_i^{-1/2} - I)Z^T]\tilde{C}^{-1/2} \circ \tilde{\delta x}_f^k$$

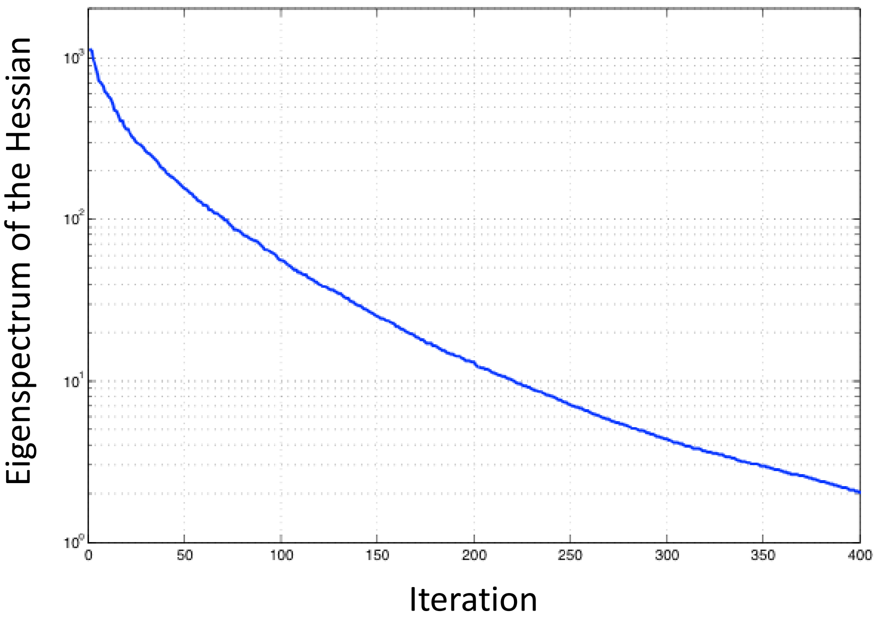
EVIL approach is directly and "fully" *hybridizable*:
- indifferent to the amount of hybridization (0% → 100%)
- directly includes radiances, (Var)QC, VarBC, etc
- "enriching" posterior ensemble with new directions from climatological B (cf. model error)

EVIL: Experimental Design

- EVIL + inflation calculation implemented within WRF variational data assimilation
- WRF over CONUS domain, 100km horizontal resolution.
- Assimilate all conventional observations + Satellite winds + GPS-RO
- 20 ensemble members, no vertical localization
- Prior = WRF forecasts initialized 12h earlier from a randomization of B
- Test EVIL in “full ensemble” mode
- Single Assimilation Cycle (2009060112)

EVIL: Single Assimilation

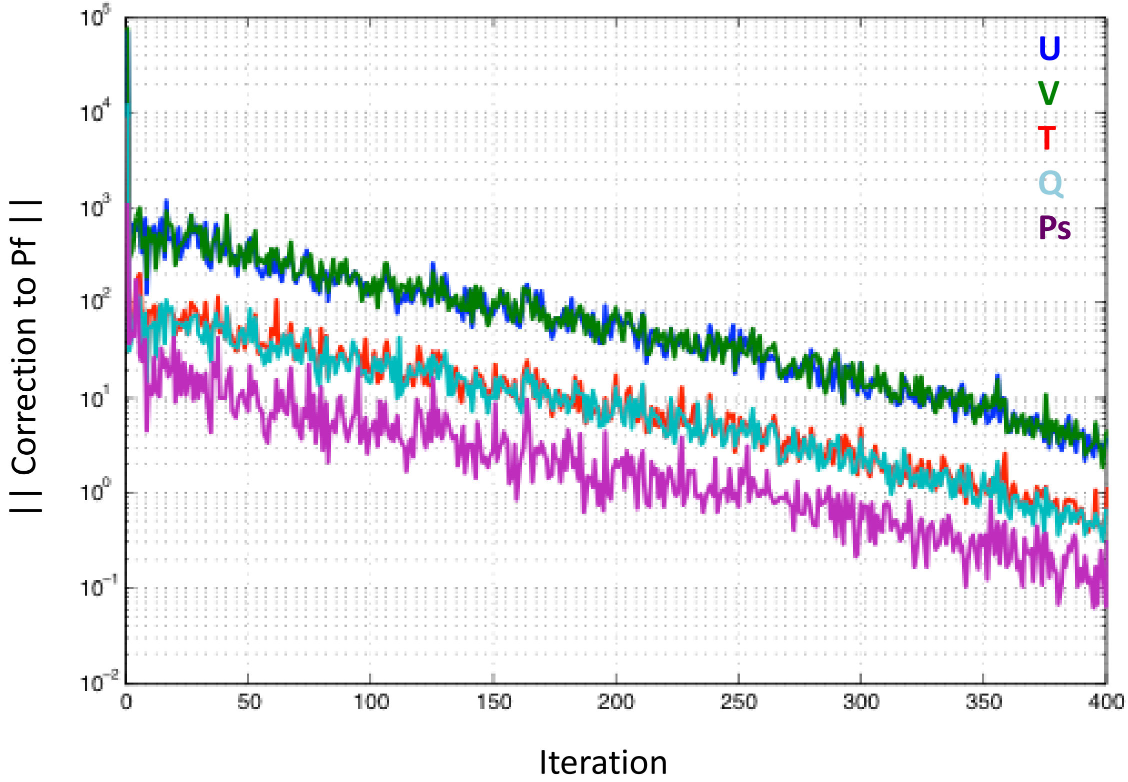




EVIL: Single Assimilation

$$\delta x_a^k \approx C_\alpha^{1/2}[I + Z(\Theta_i^{-1/2} - I)Z^T]C_\alpha^{-1/2} \circ \delta x_f^k$$

$$\left\| C_\alpha^{1/2}[Z_i(\Theta_i^{-1/2} - I)Z_i^T]C_\alpha^{-1/2} \circ \delta x_f^k \right\|$$



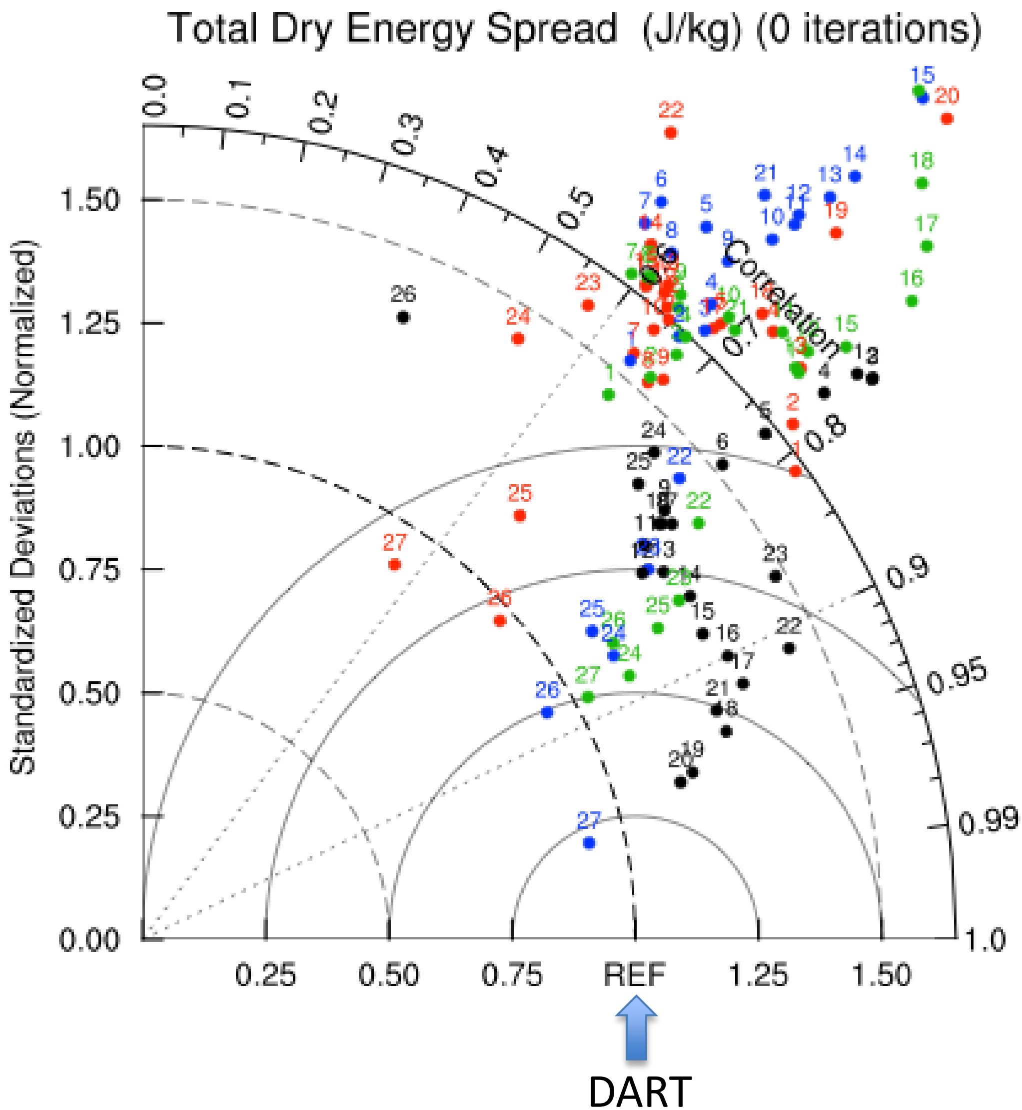
EVIL: Experimental Design

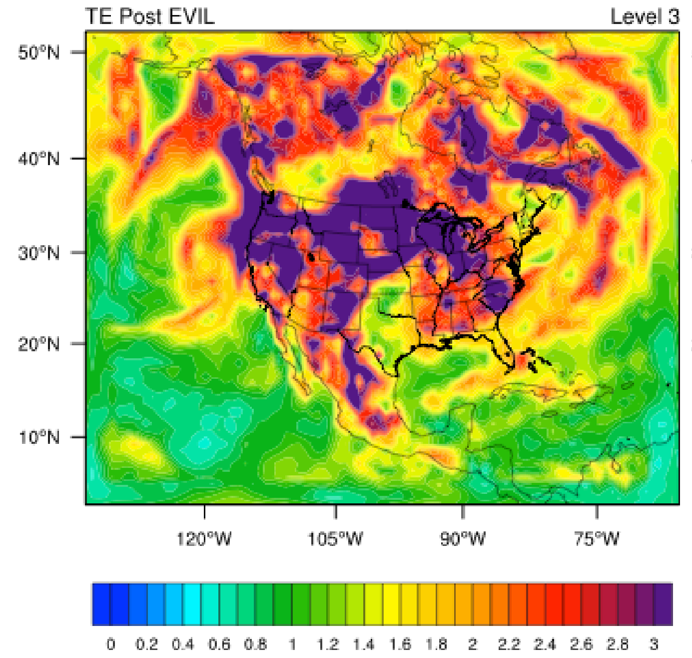
- Test EVIL in “full ensemble” mode

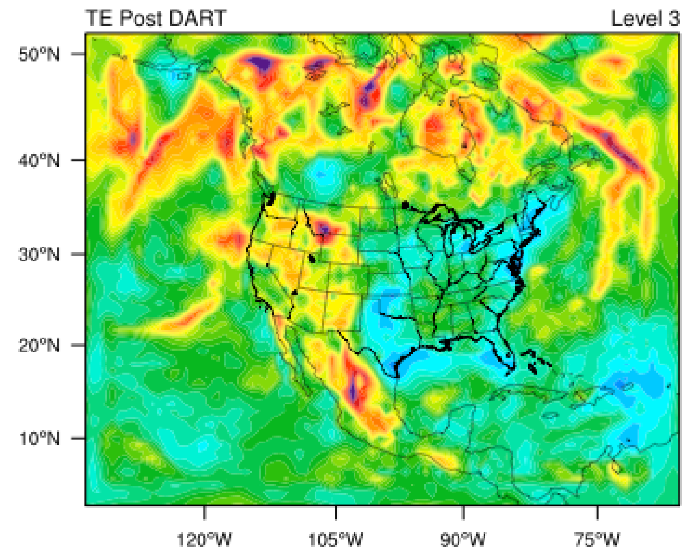
- Comparison to DART (EAKF, Anderson 2001) with same observations and QC

- DART versus EVIL, why should they differ?
  - EVIL is a maximum likelihood method (DART = minimum variance)
  - EVIL is closer to the LETKF (DART is a sequential scheme in obs space)
  - Localization functions are slightly different (Caspari and Cohn vs. RF)
  - EVIL uses H_TL around the deterministic bckg (DART uses an ensemble of H)
  - …

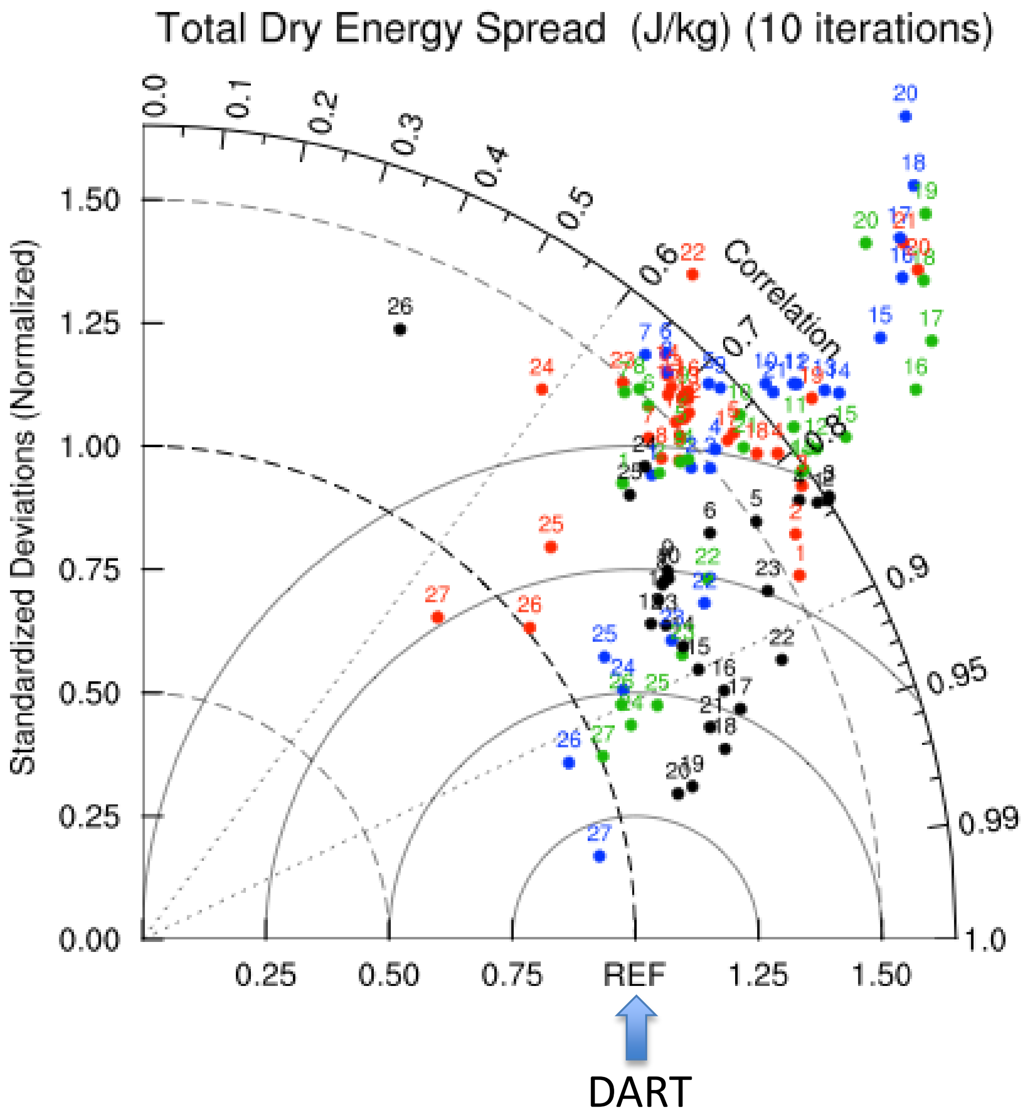
EVIL: Single Assimilation

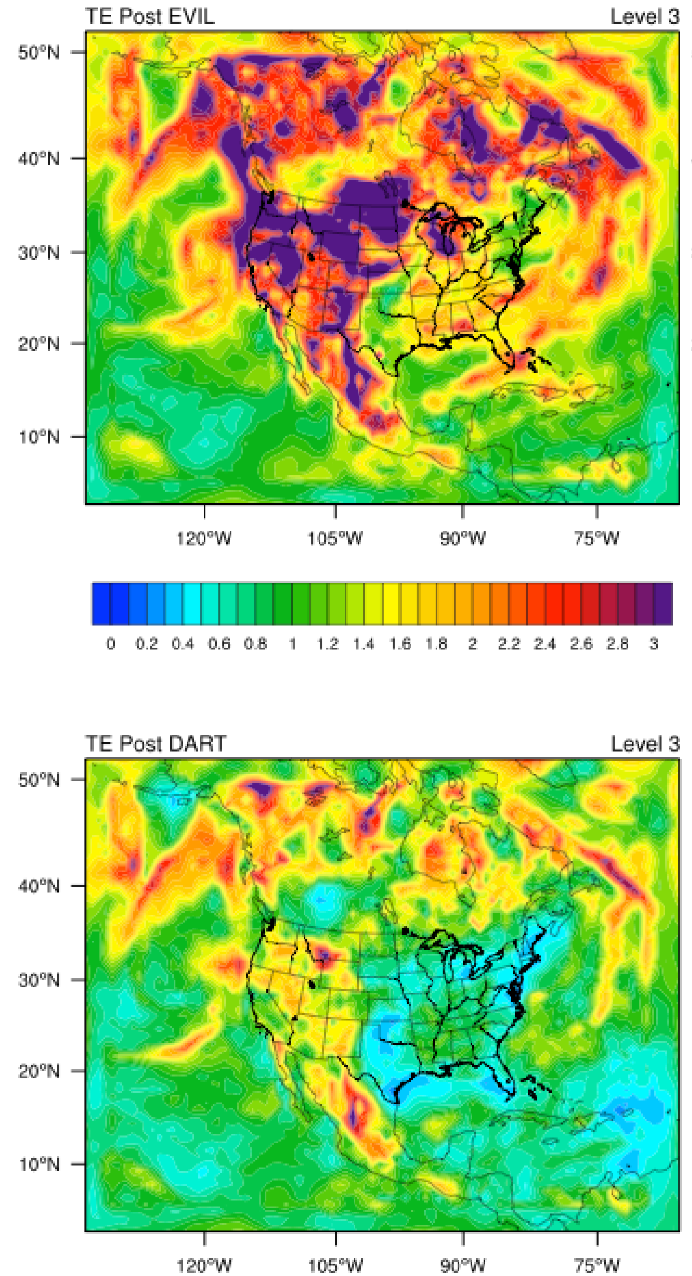




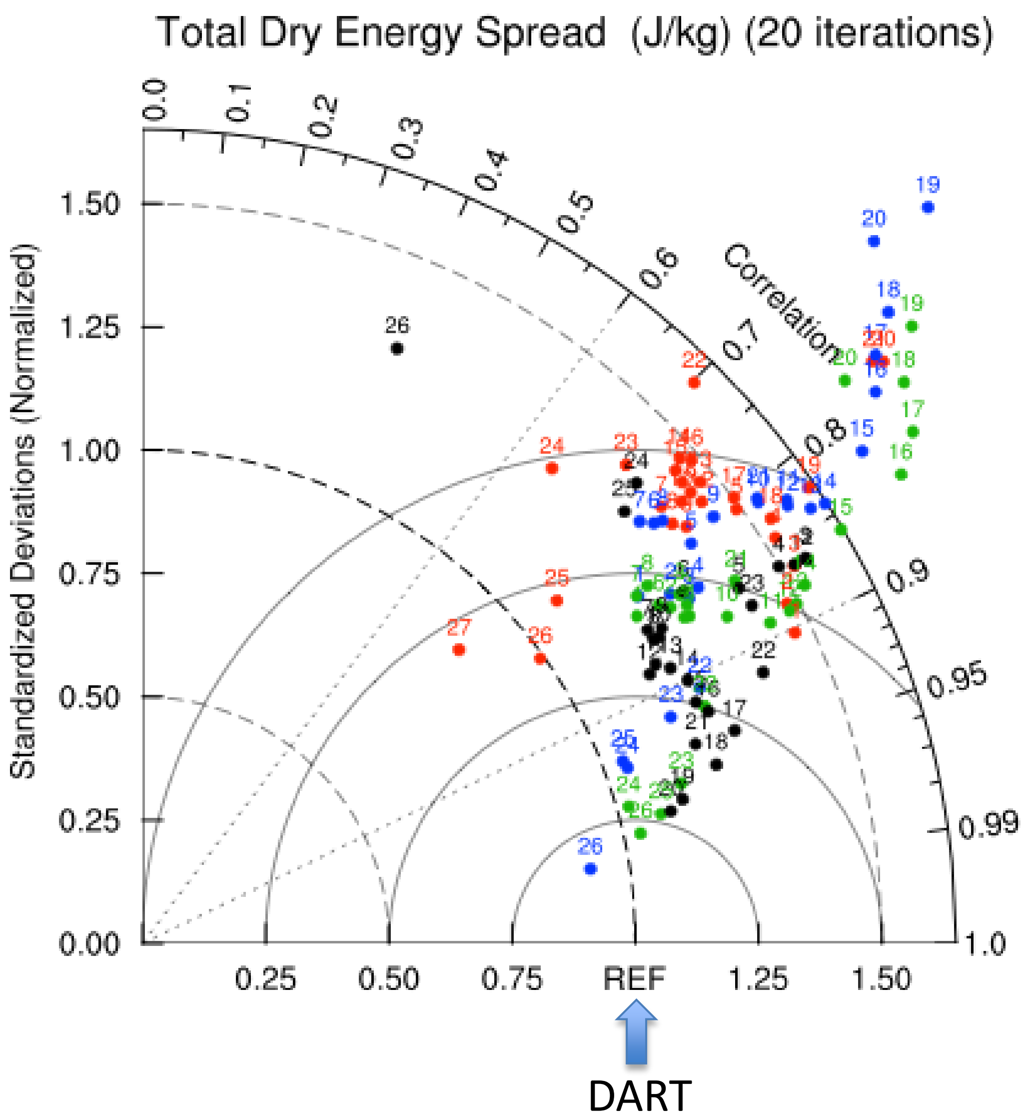


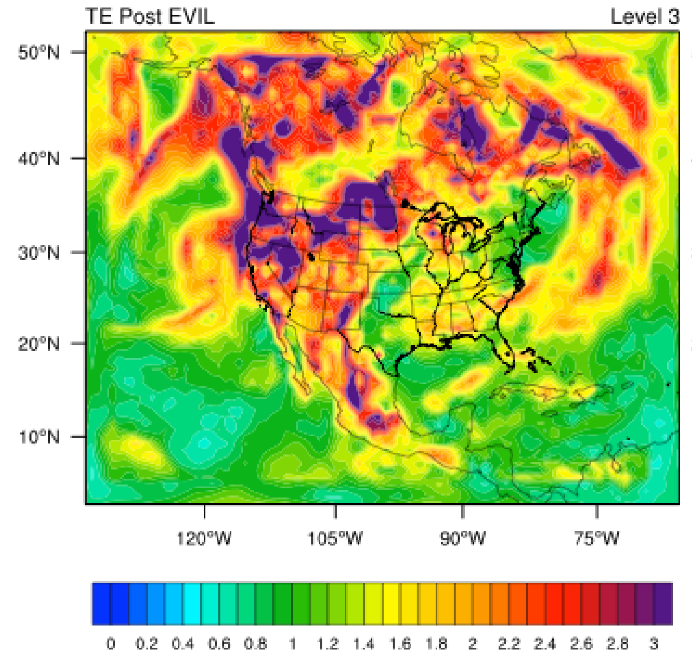
EVIL: Single Assimilation

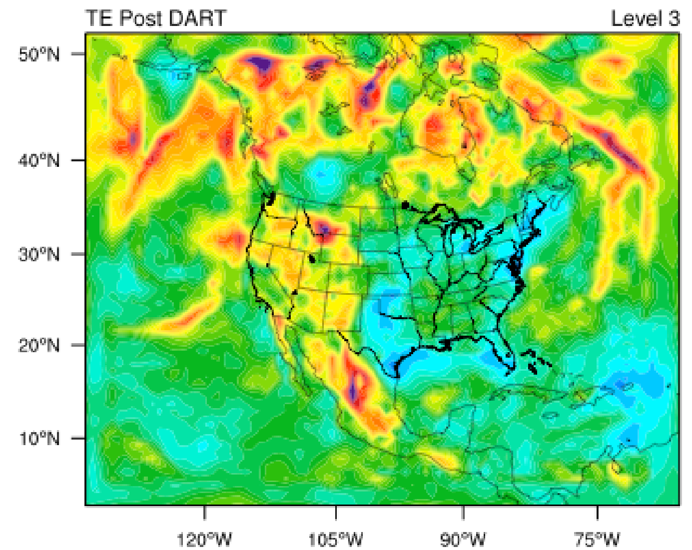




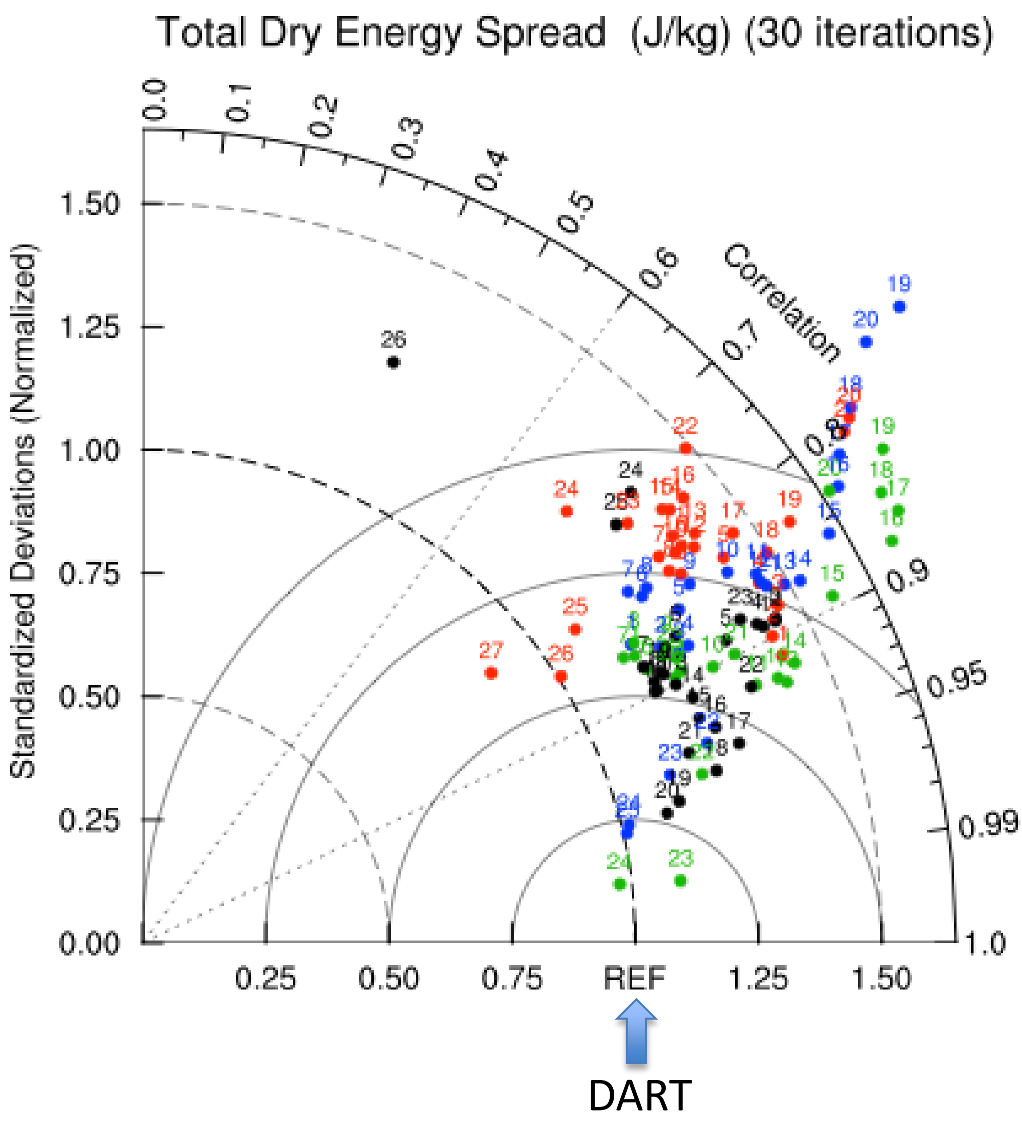
EVIL: Single Assimilation

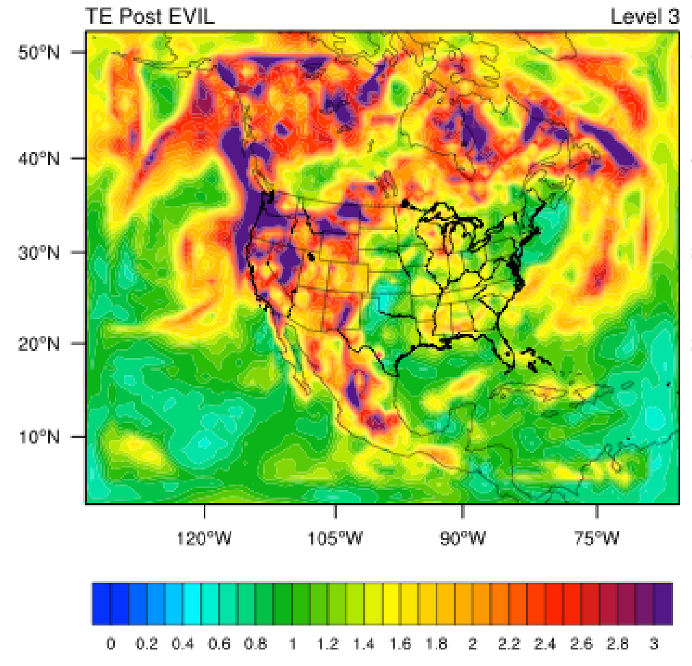


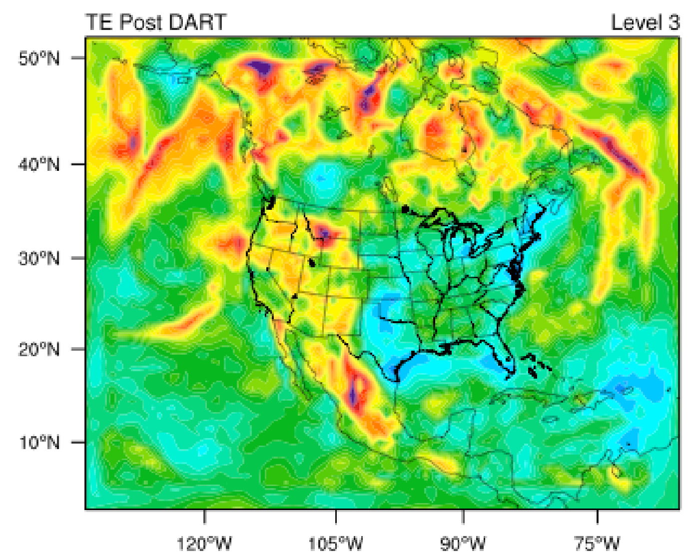




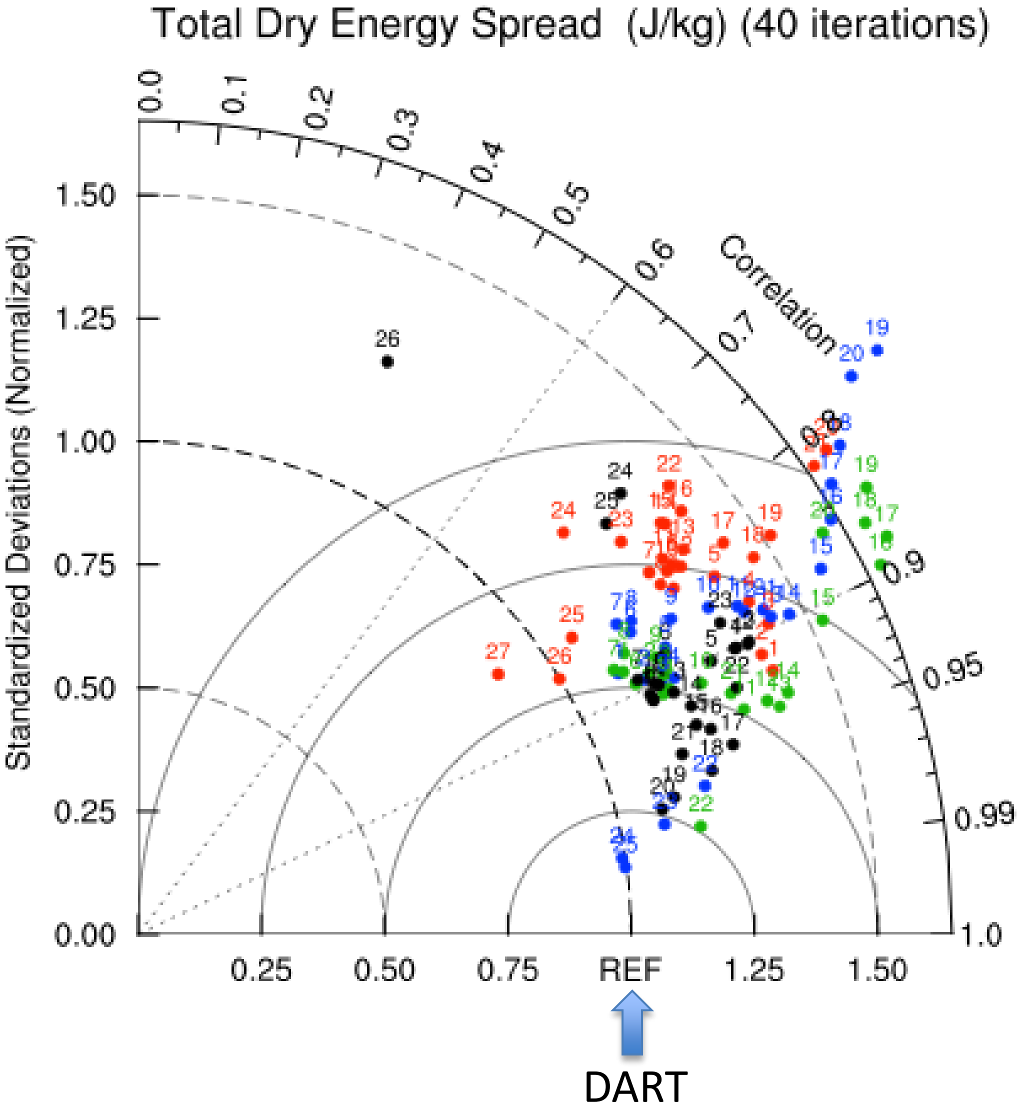
EVIL: Single Assimilation

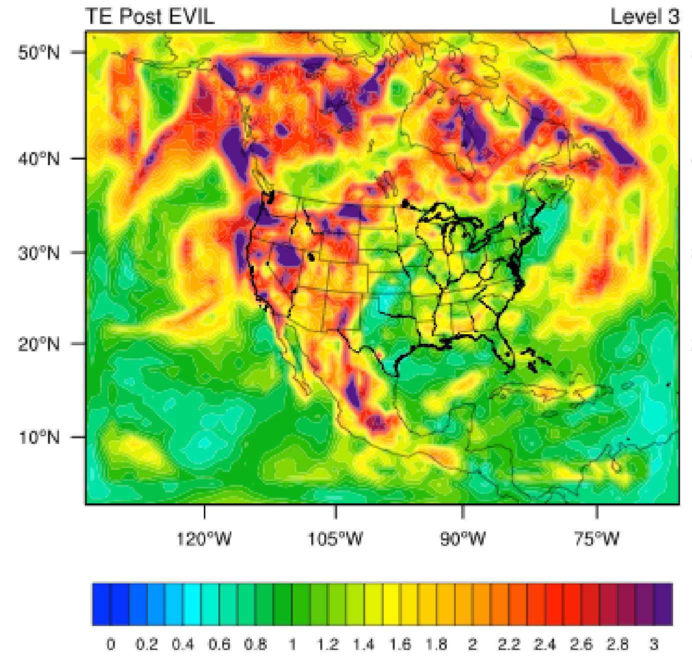


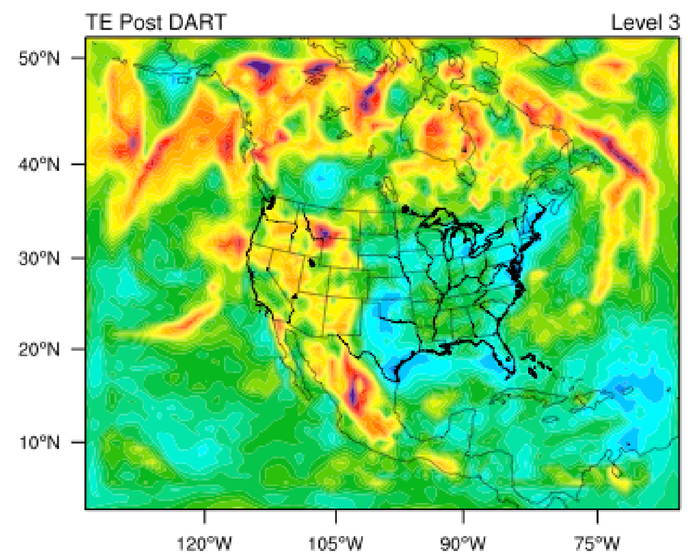




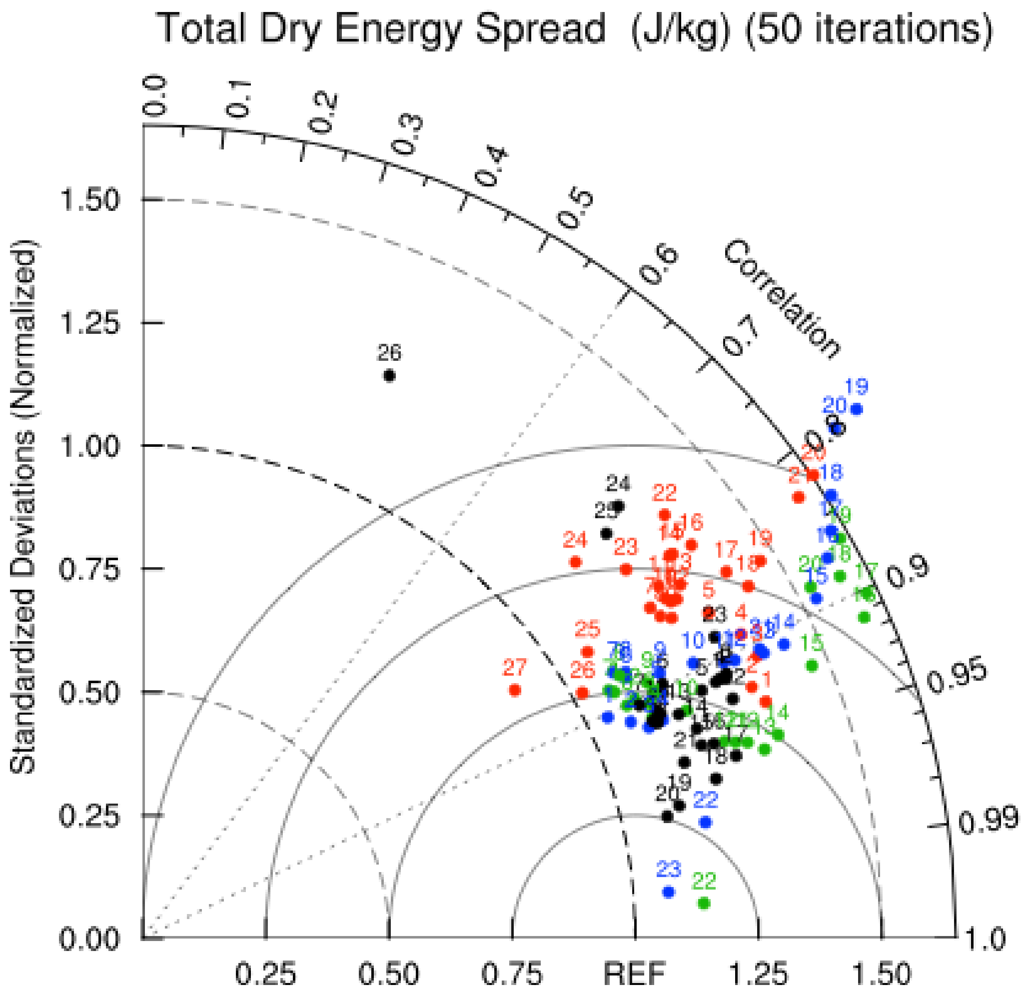
EVIL: Single Assimilation



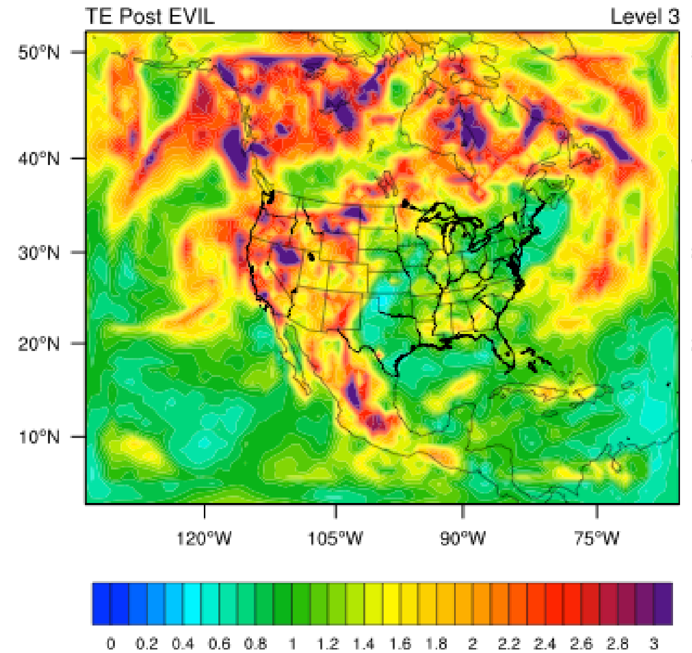


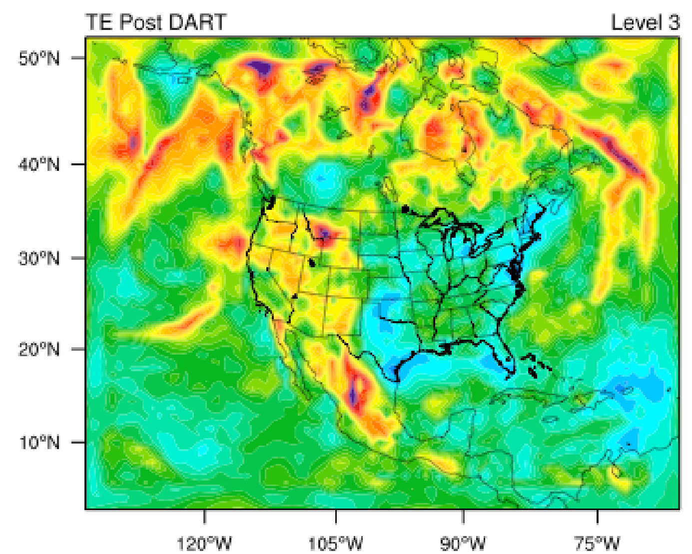


EVIL: Single Assimilation

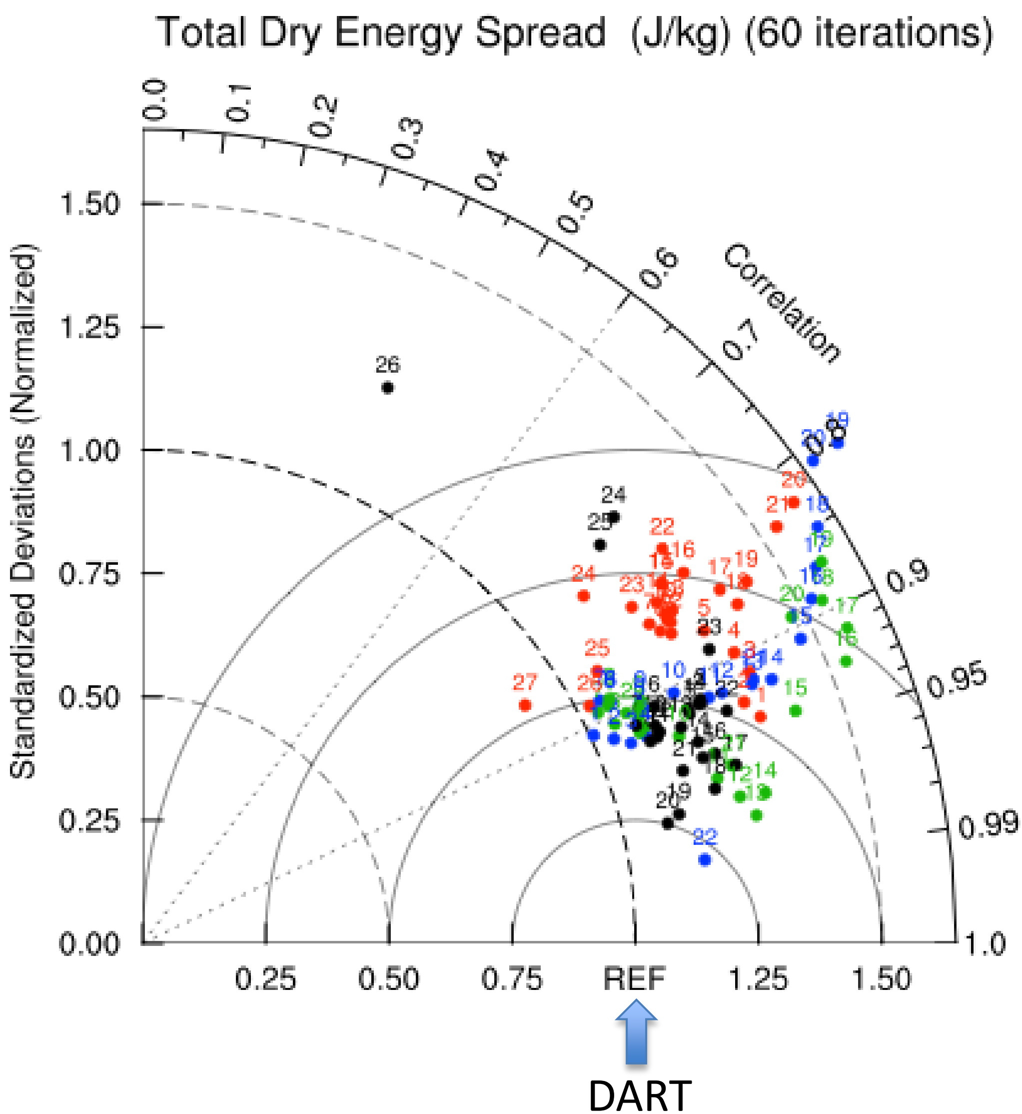


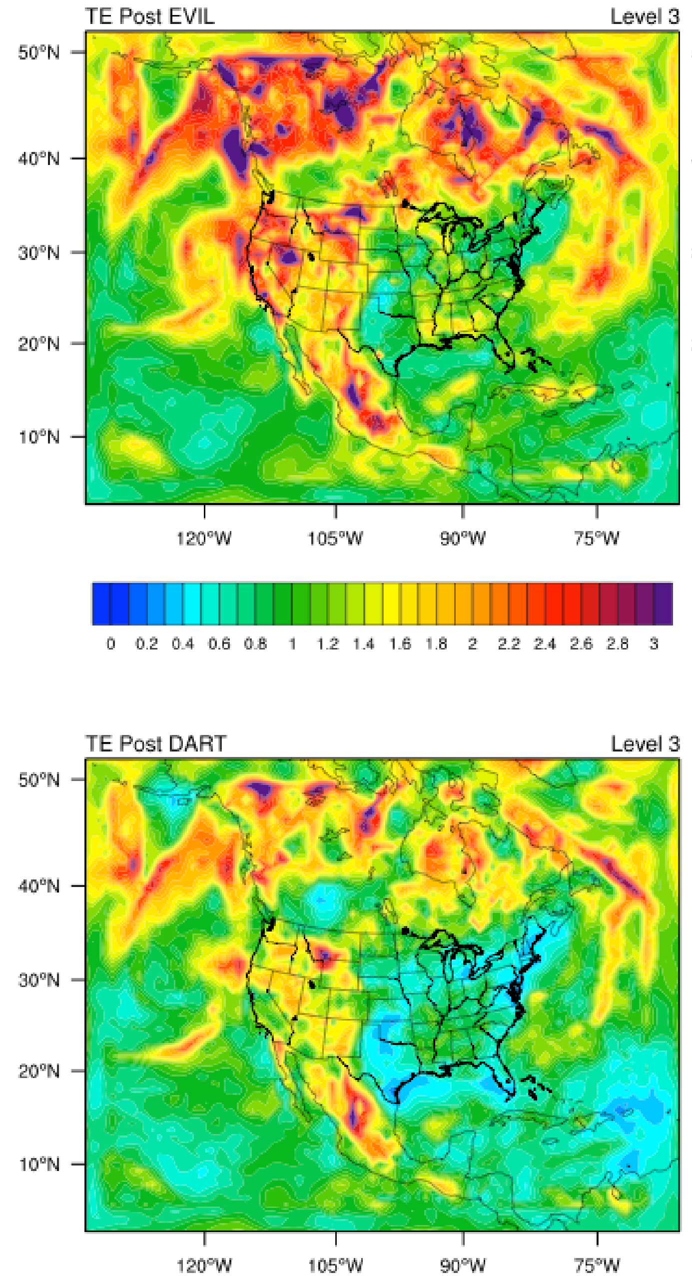
DART



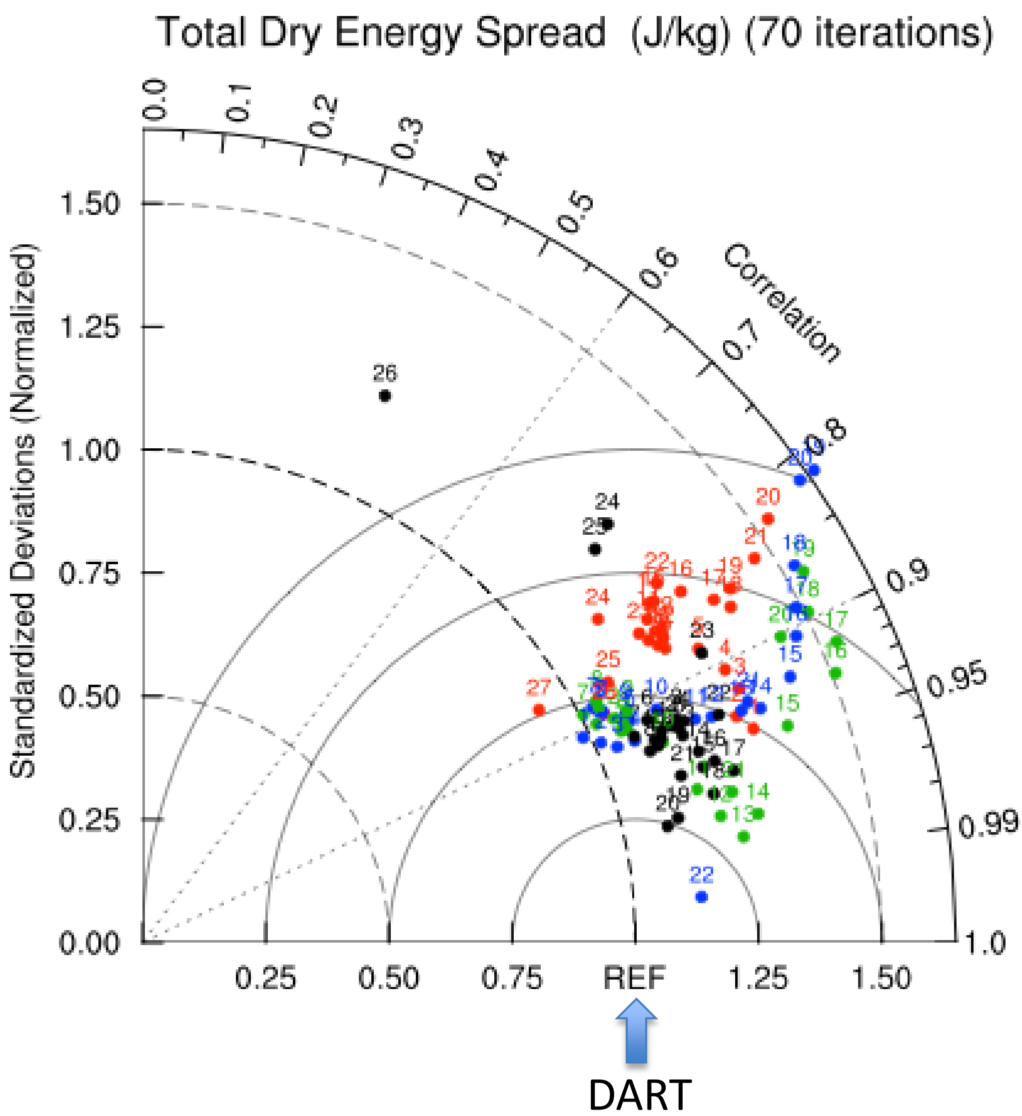


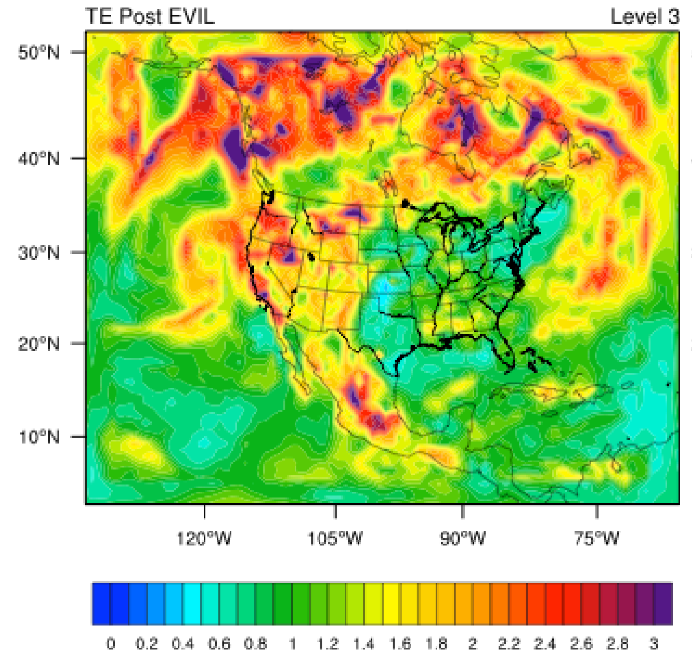
EVIL: Single Assimilation

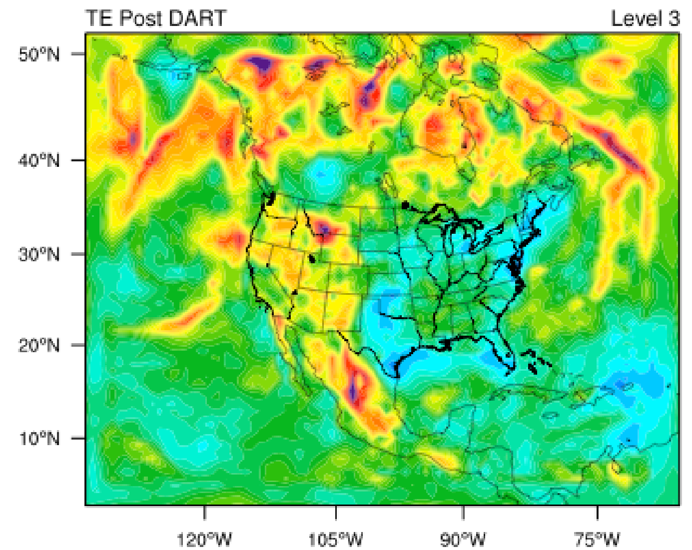




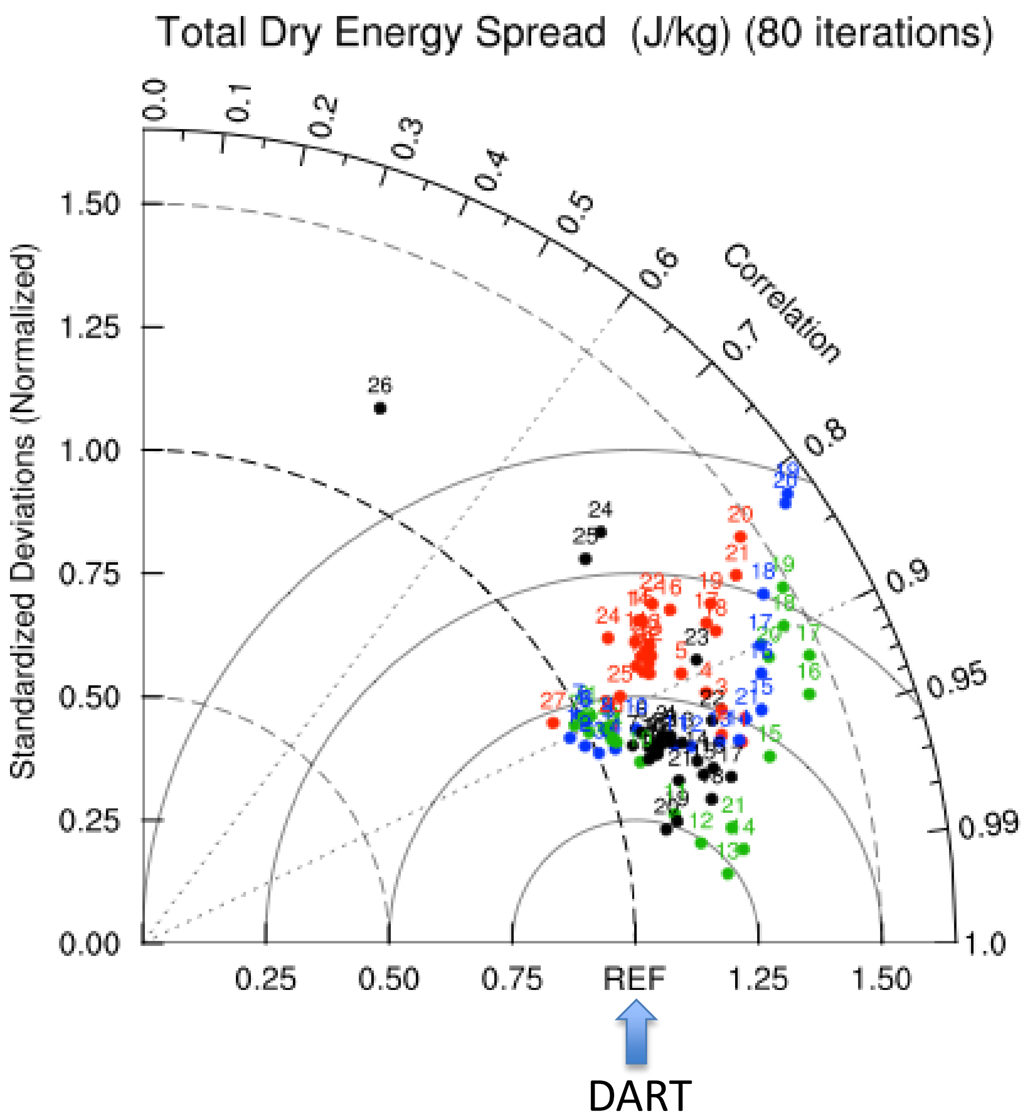
EVIL: Single Assimilation

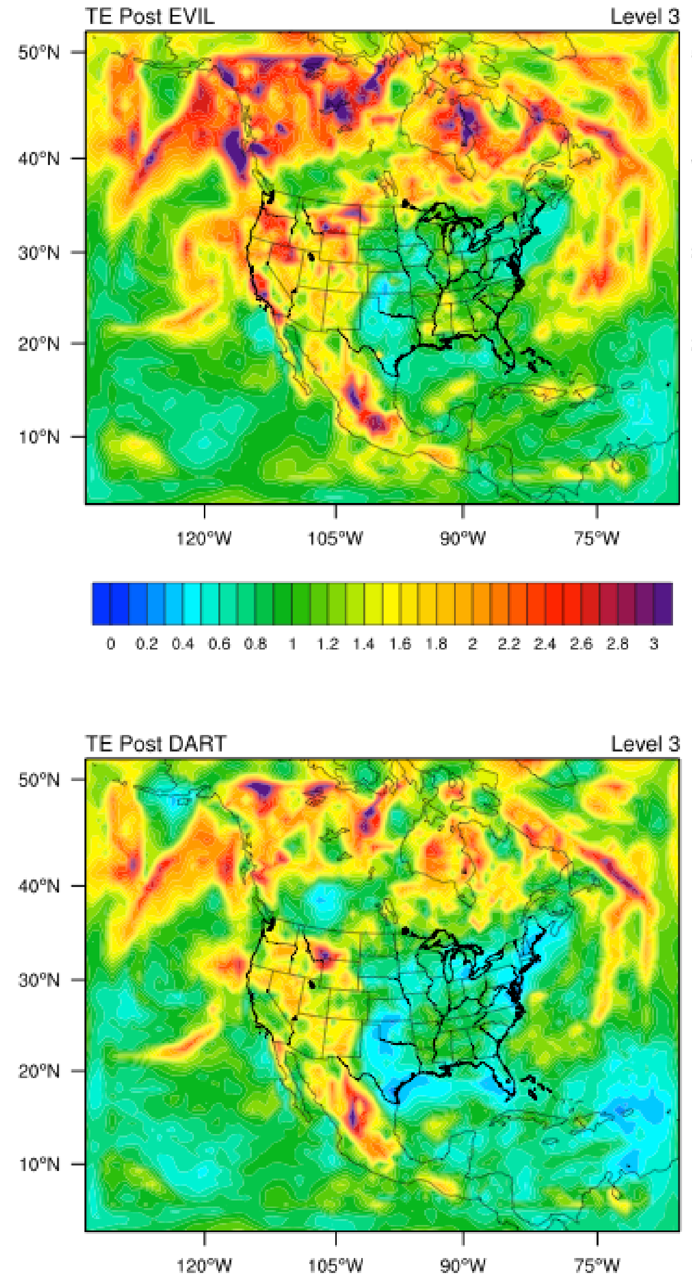




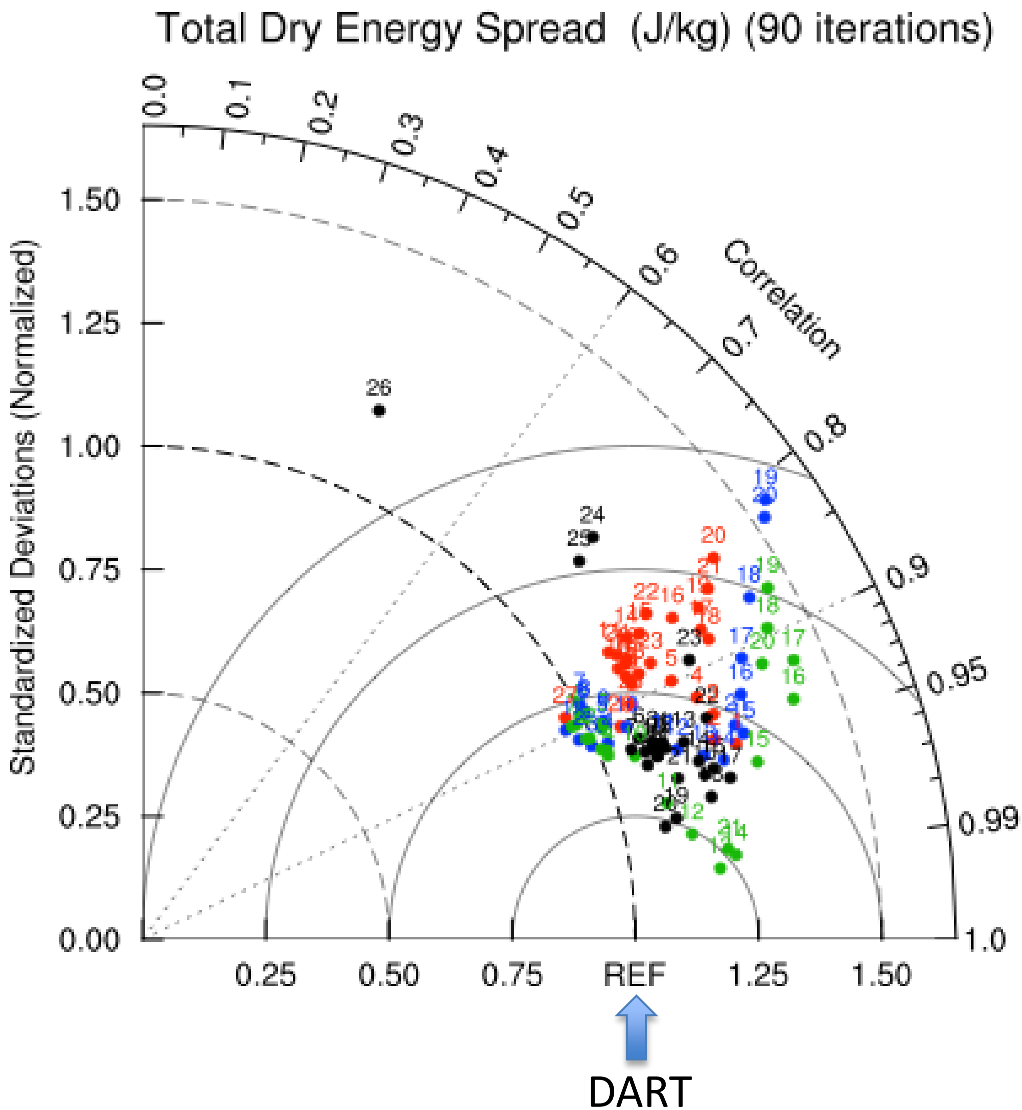


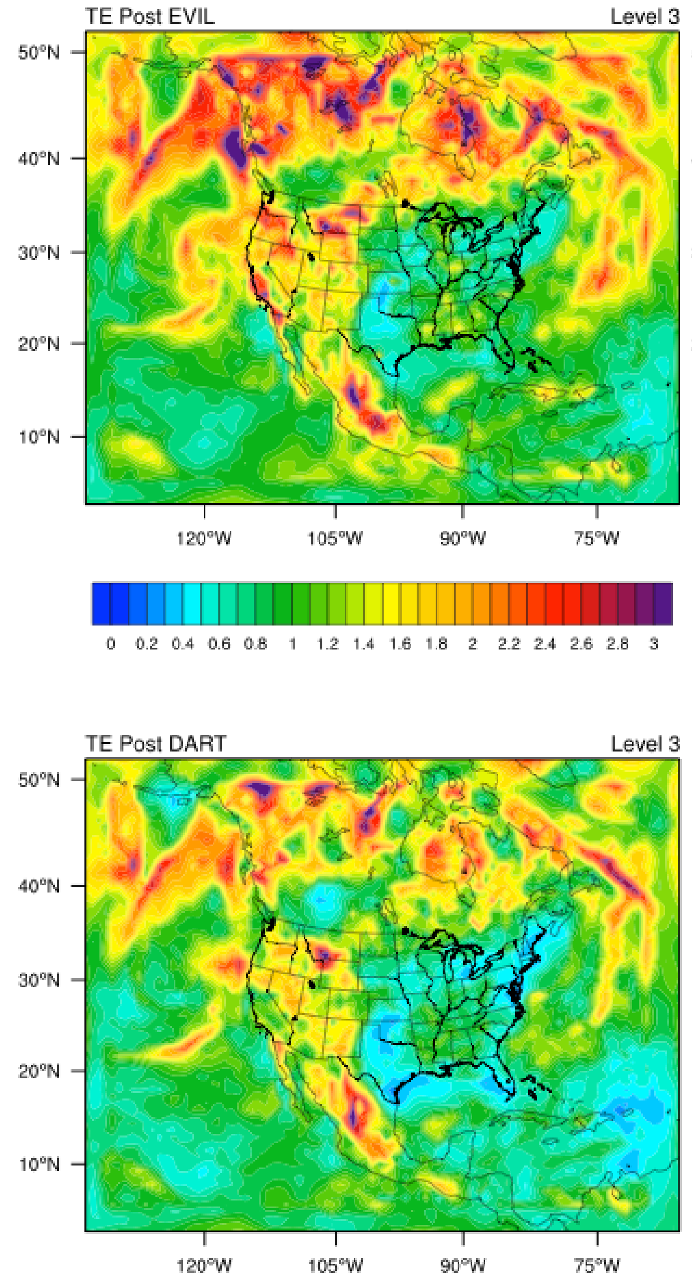
EVIL: Single Assimilation



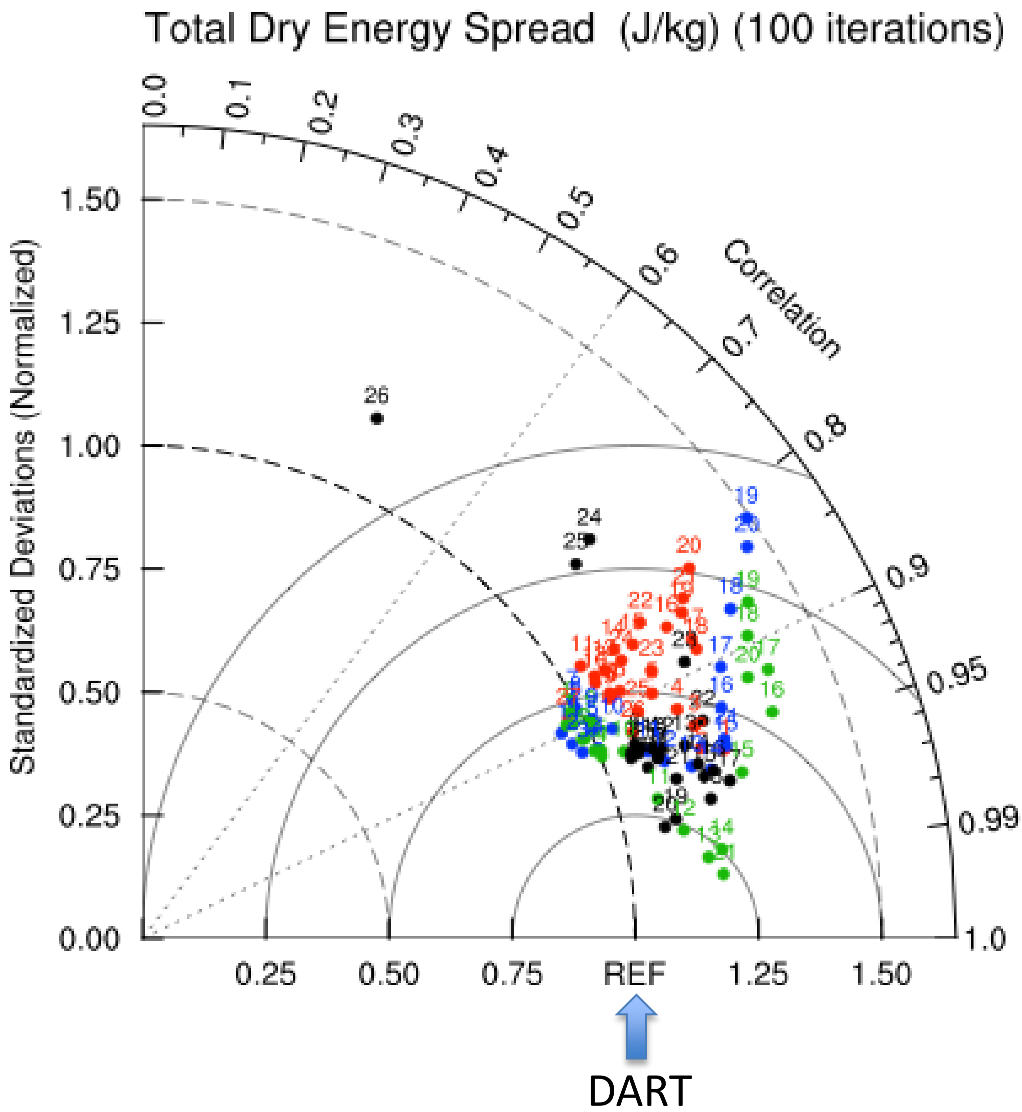


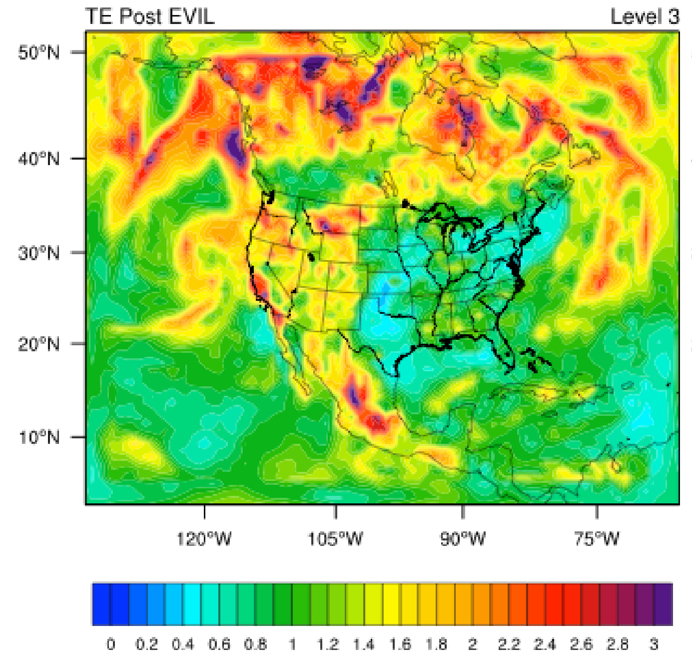
EVIL: Single Assimilation

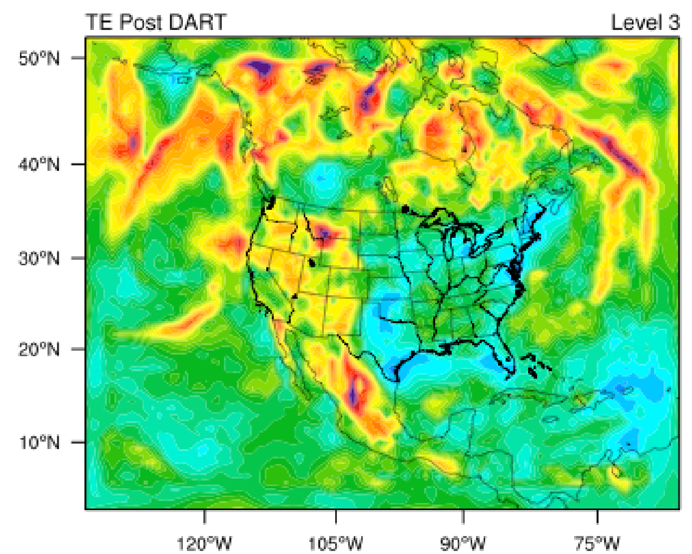




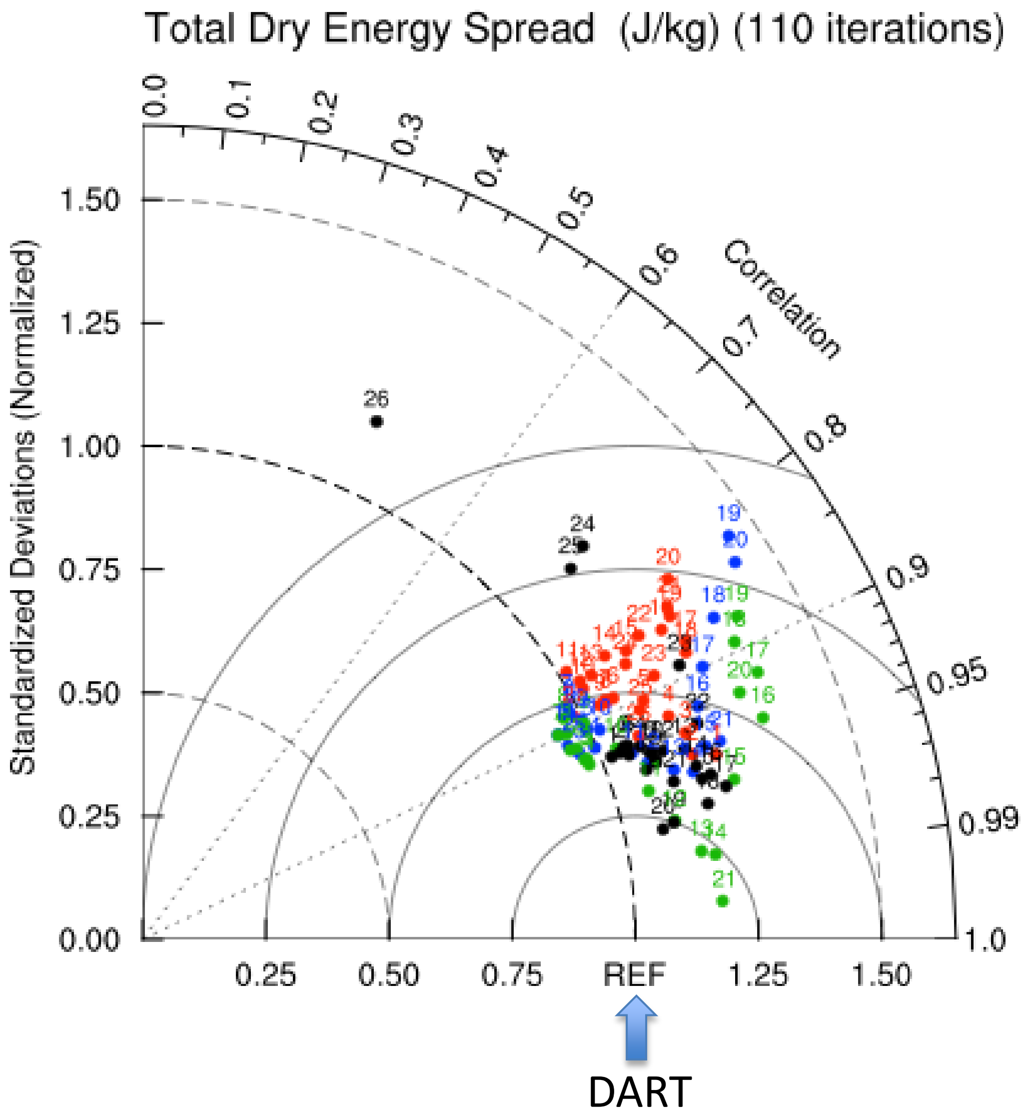
EVIL: Single Assimilation

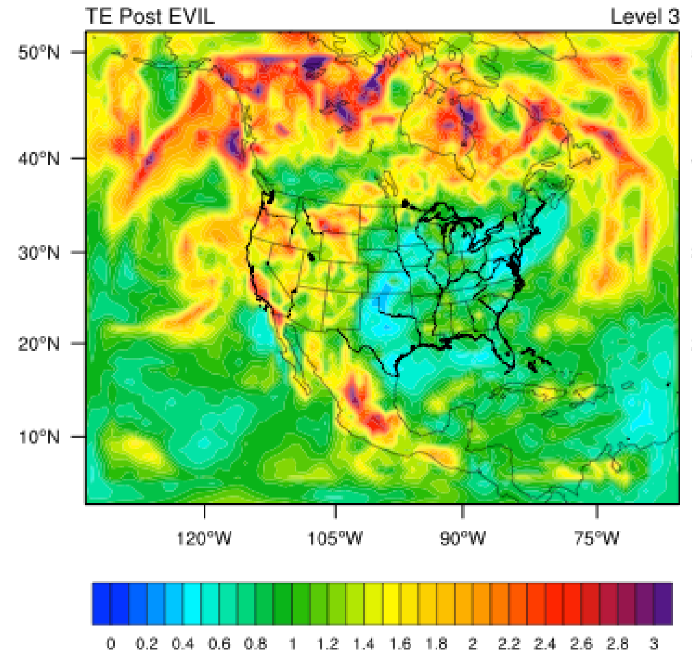


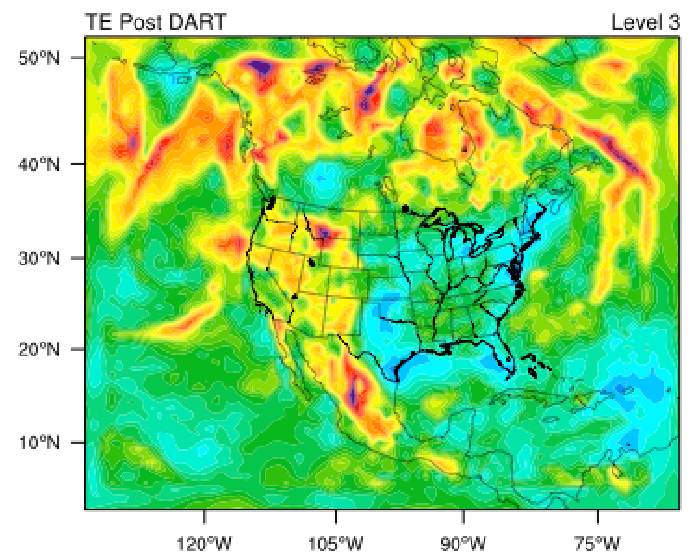




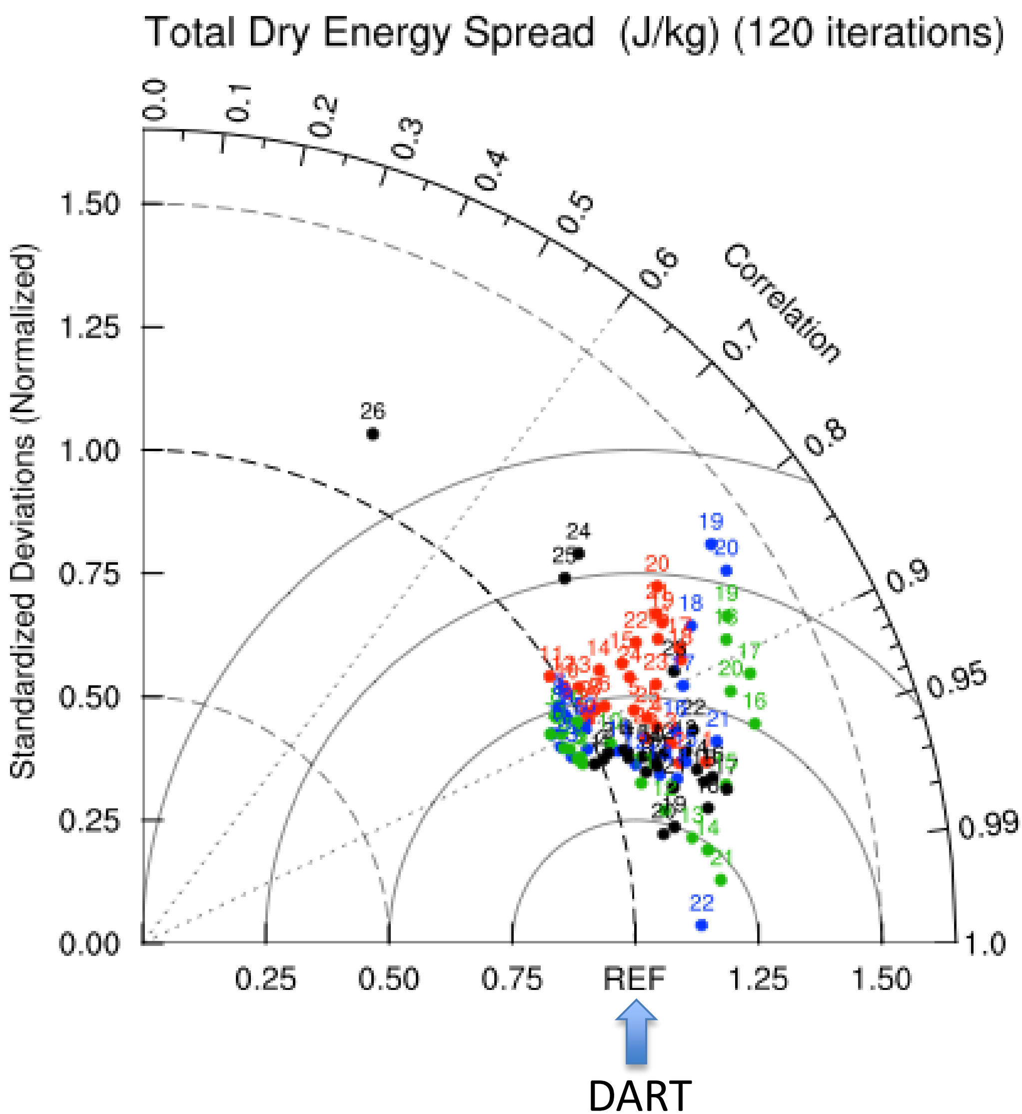
EVIL: Single Assimilation

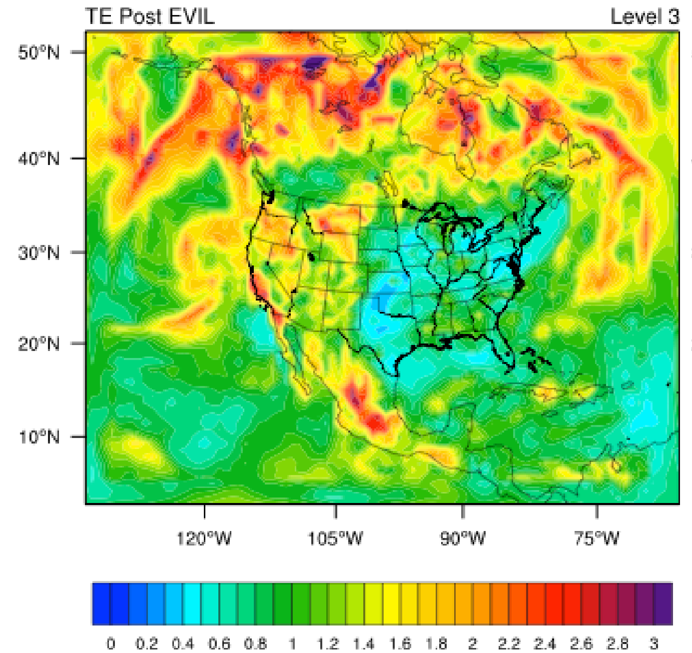


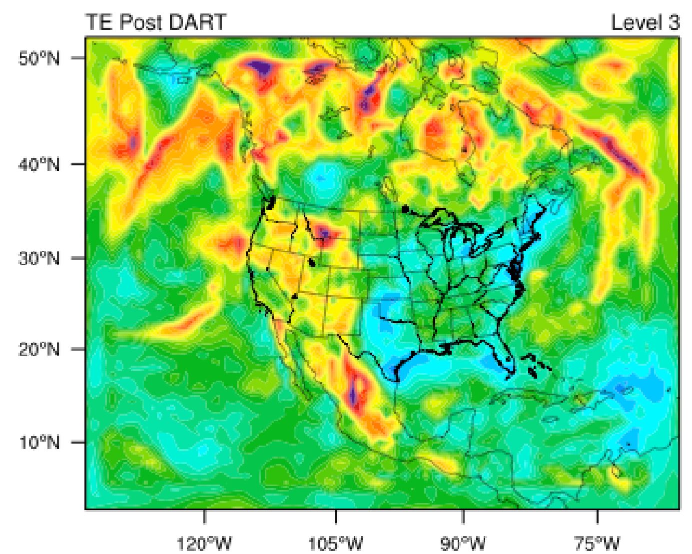




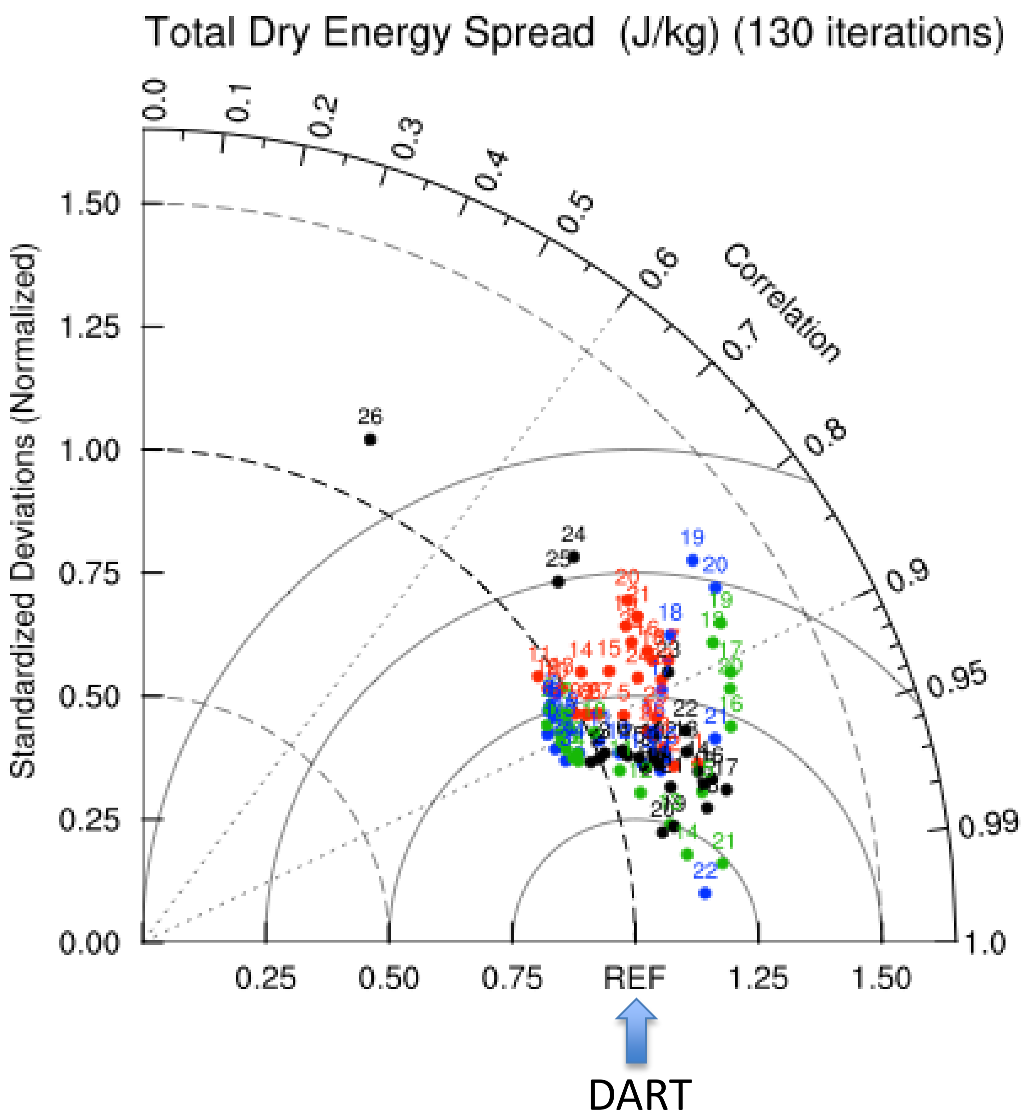
EVIL: Single Assimilation

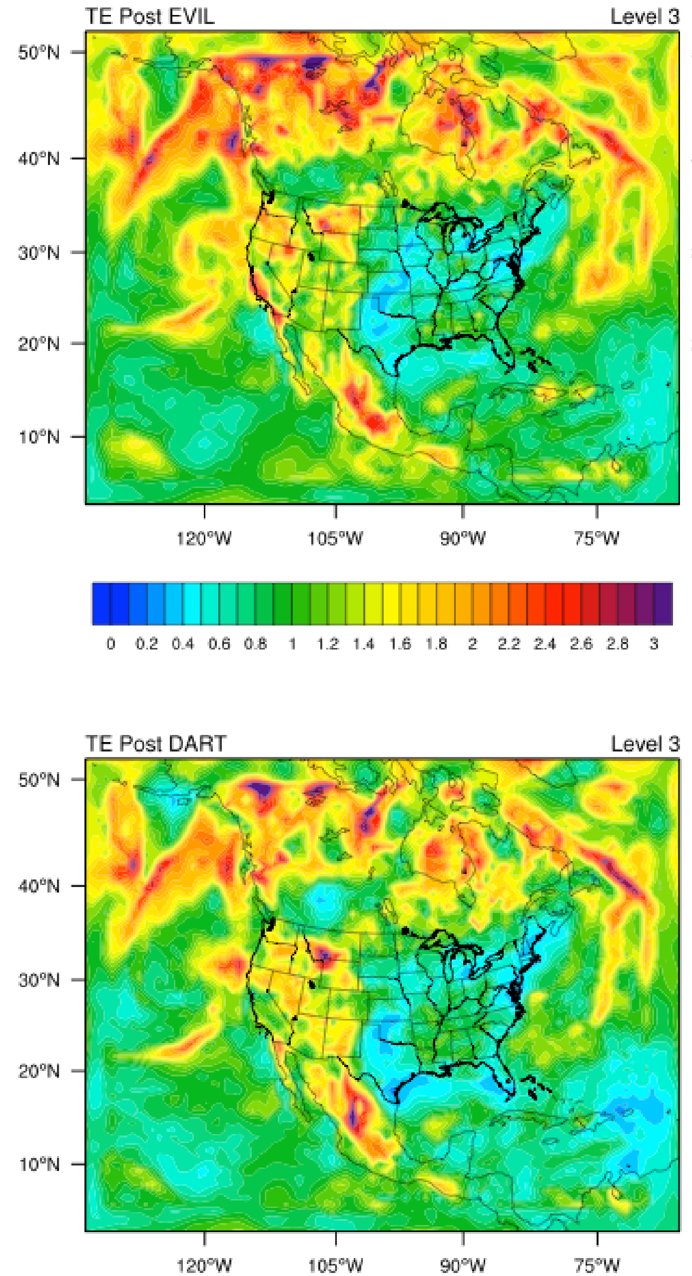




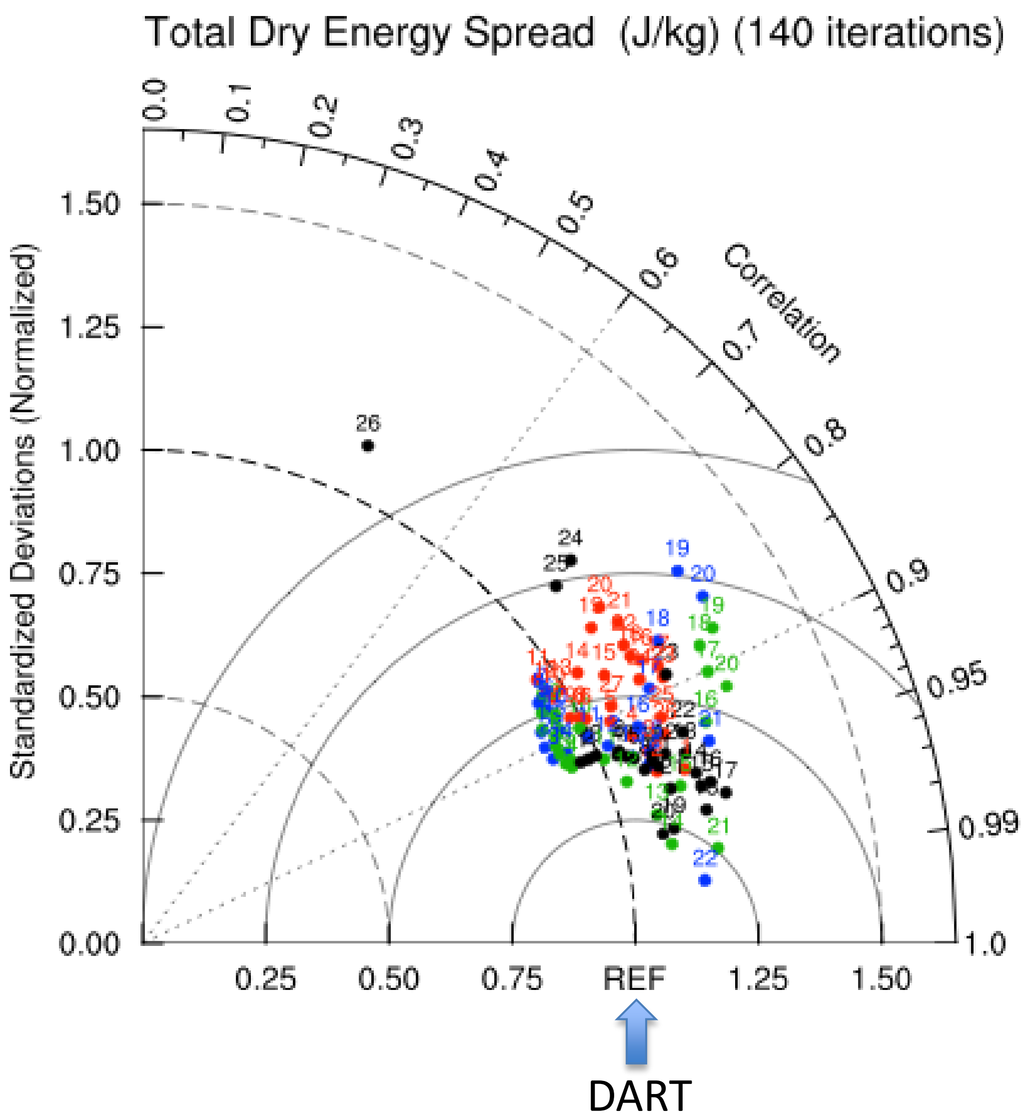


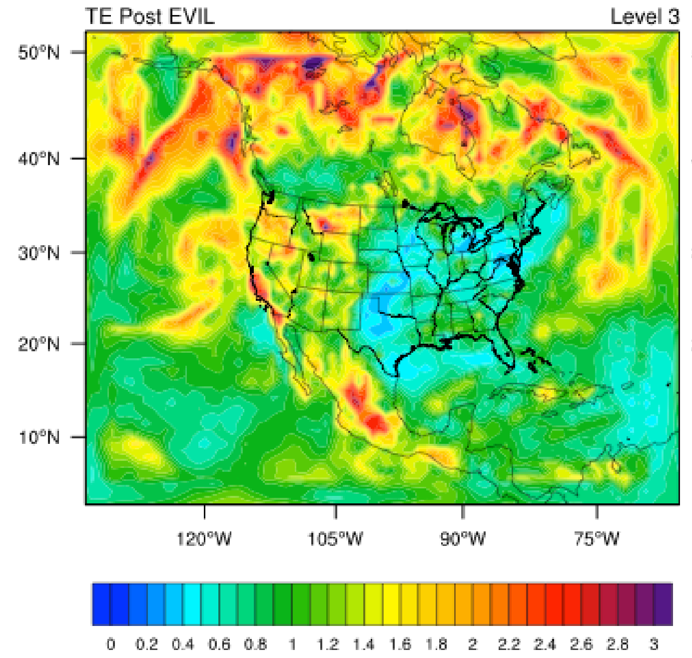
EVIL: Single Assimilation

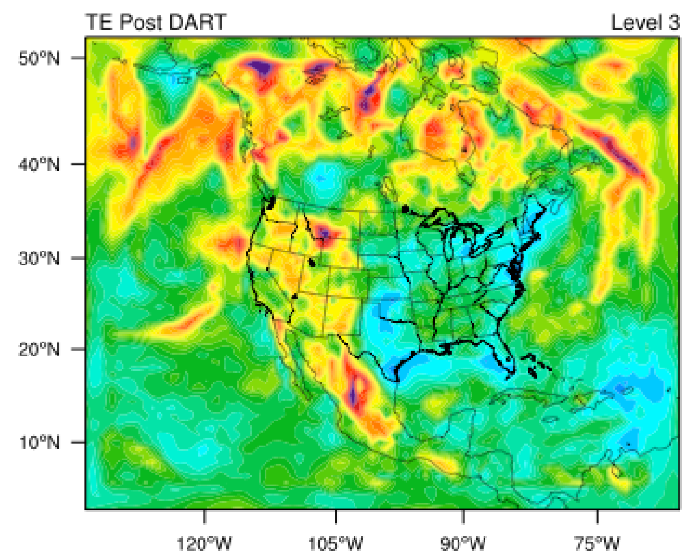




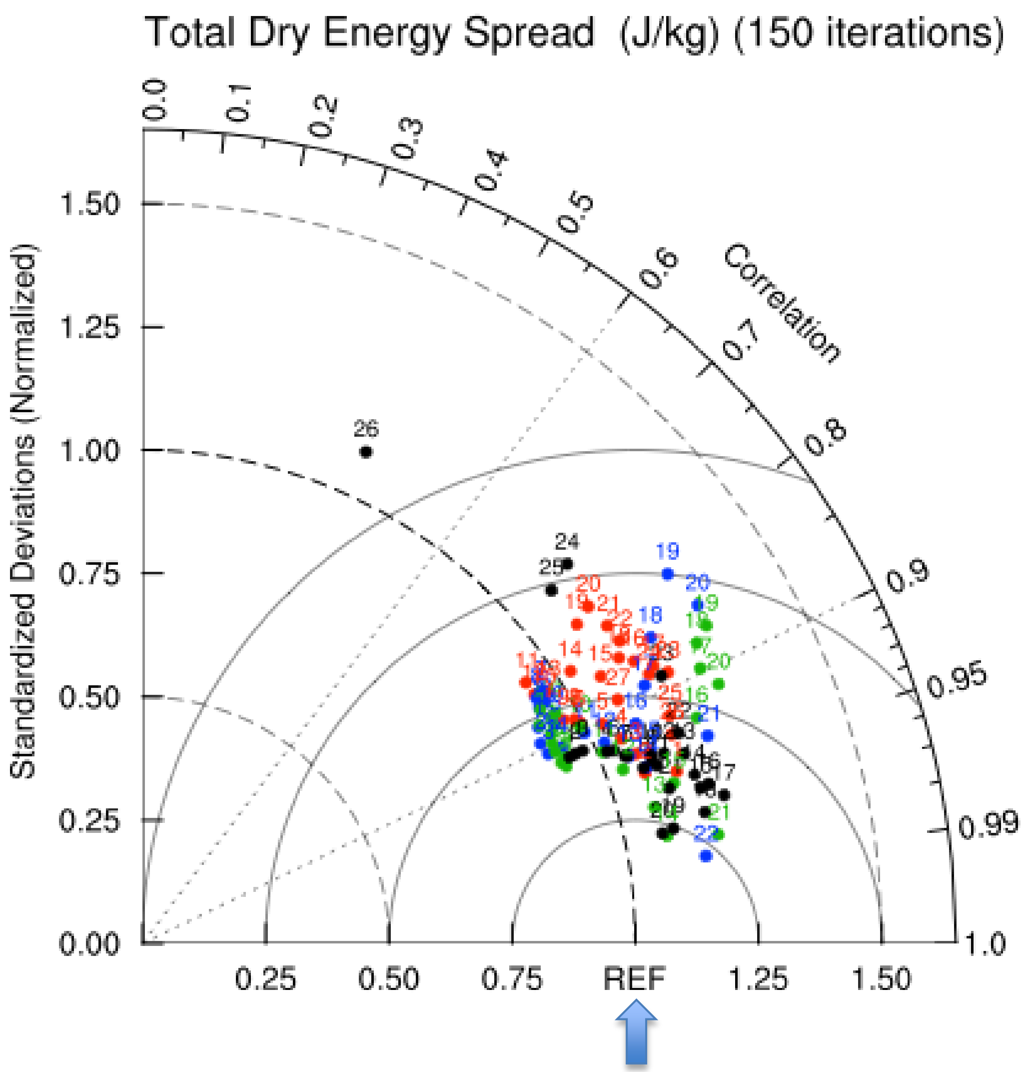
EVIL: Single Assimilation



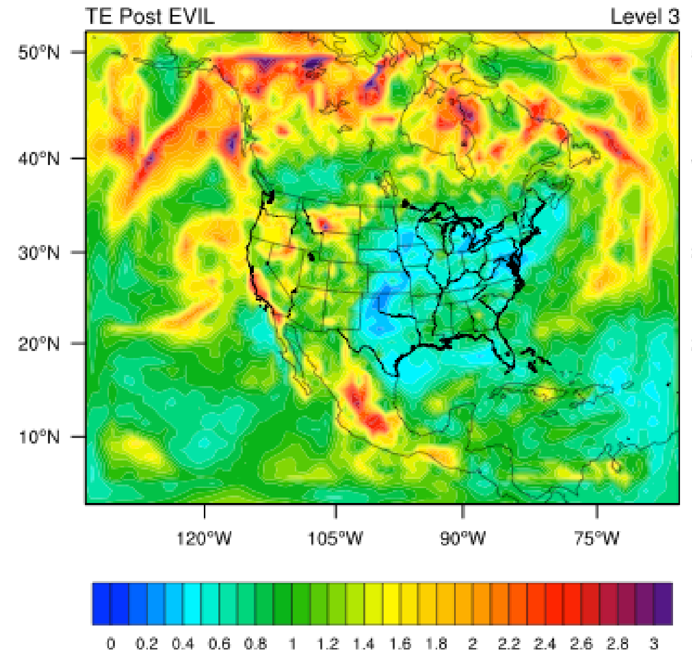


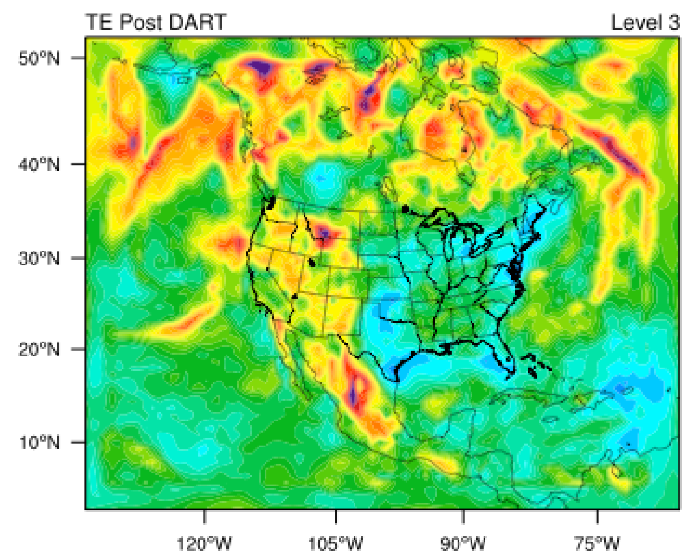


EVIL: Single Assimilation

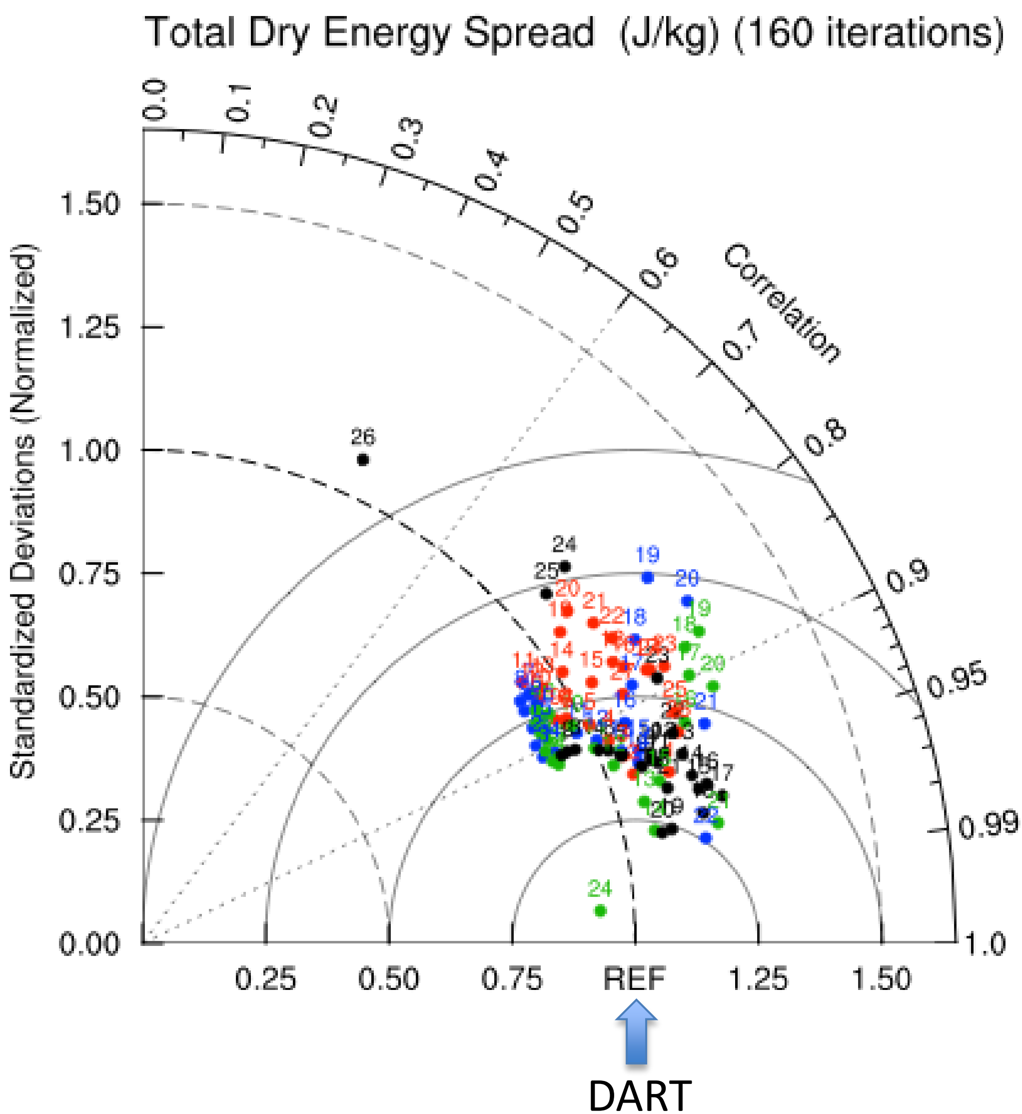


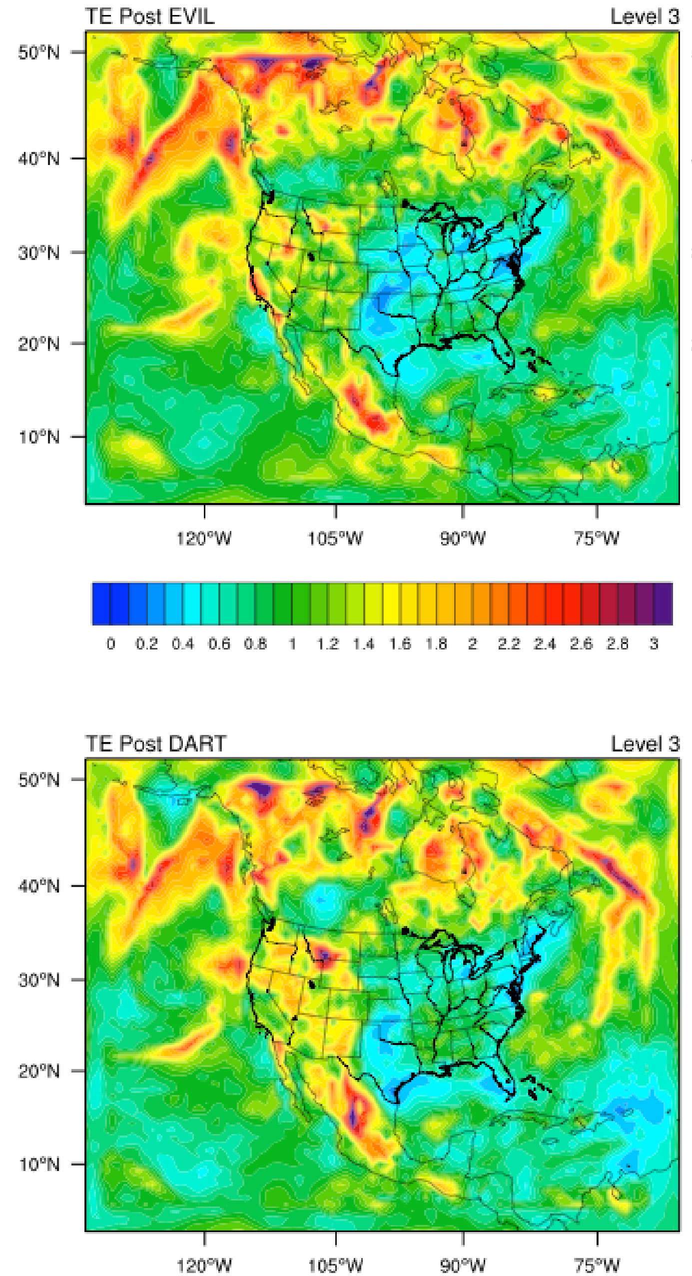
DART



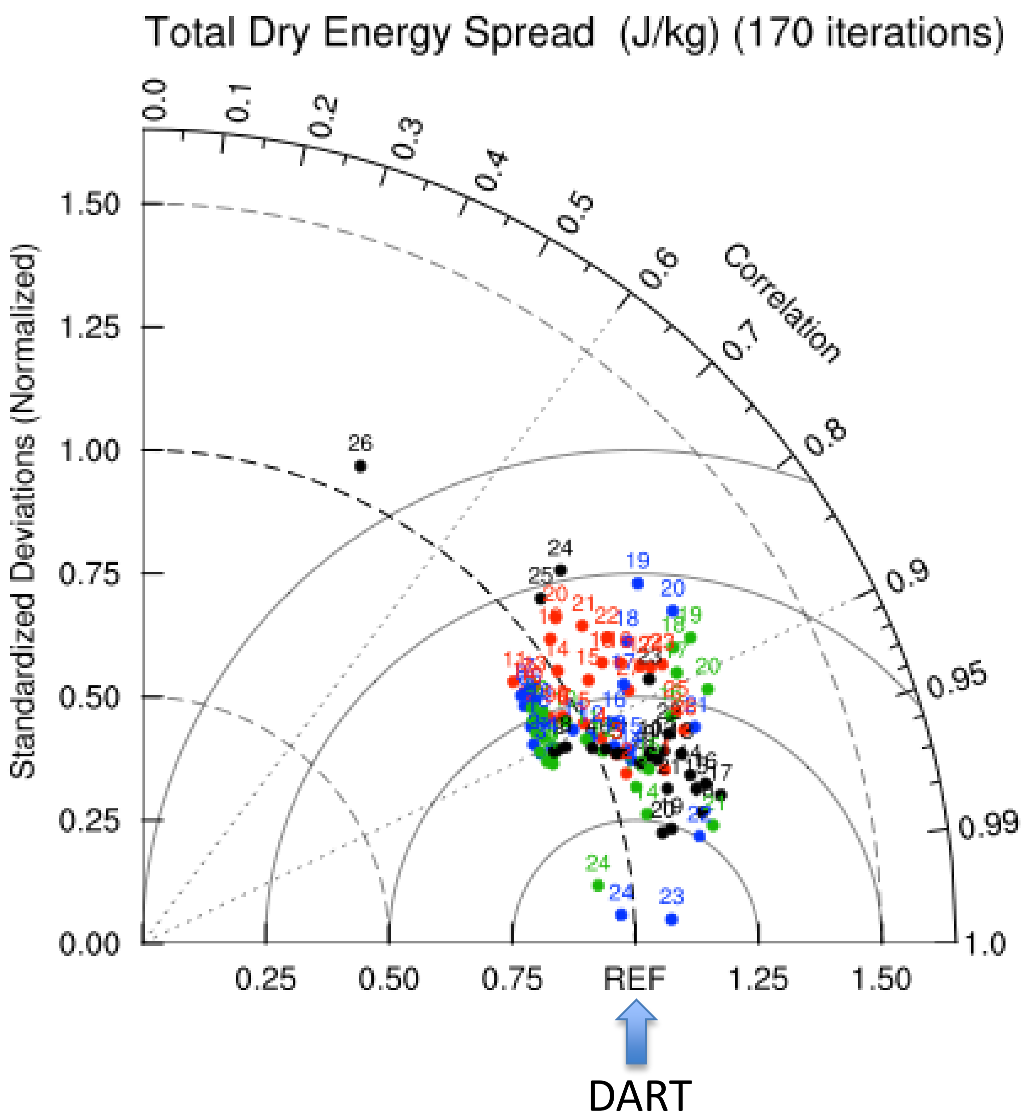


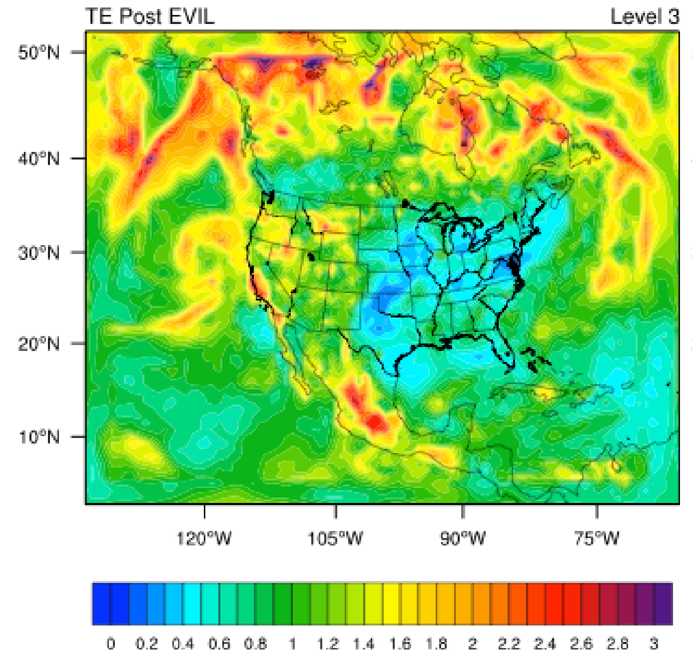
EVIL: Single Assimilation

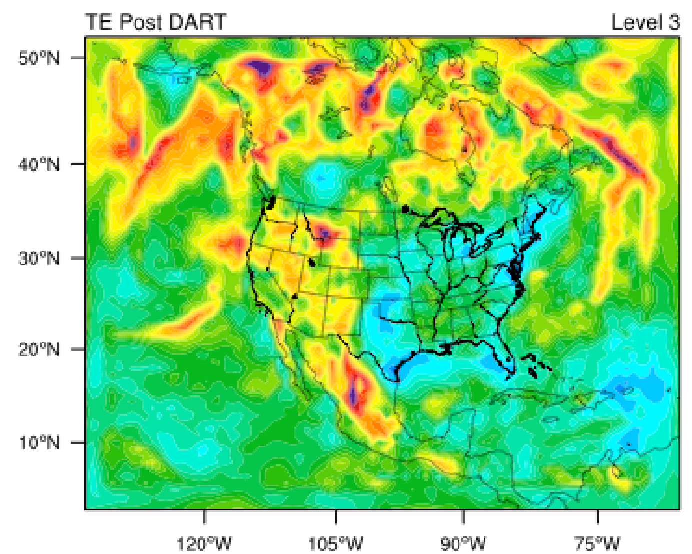




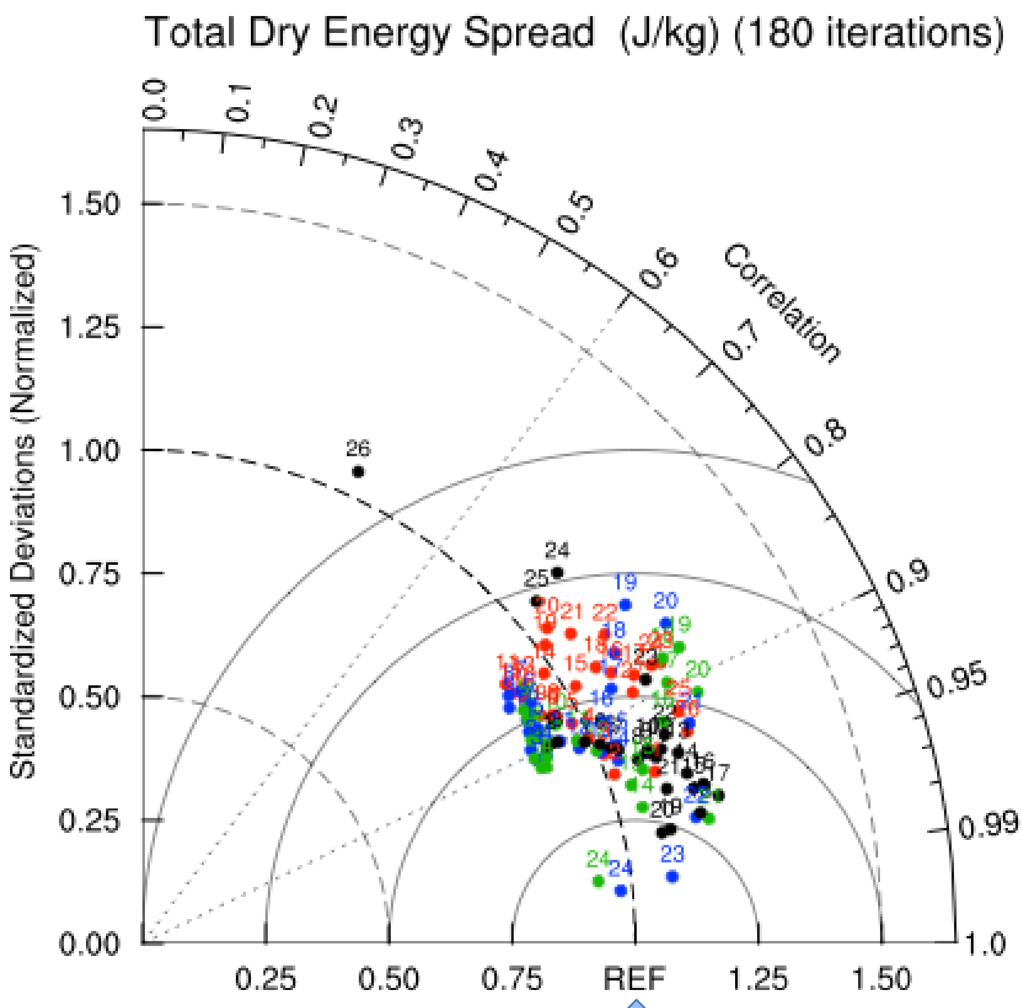
EVIL: Single Assimilation



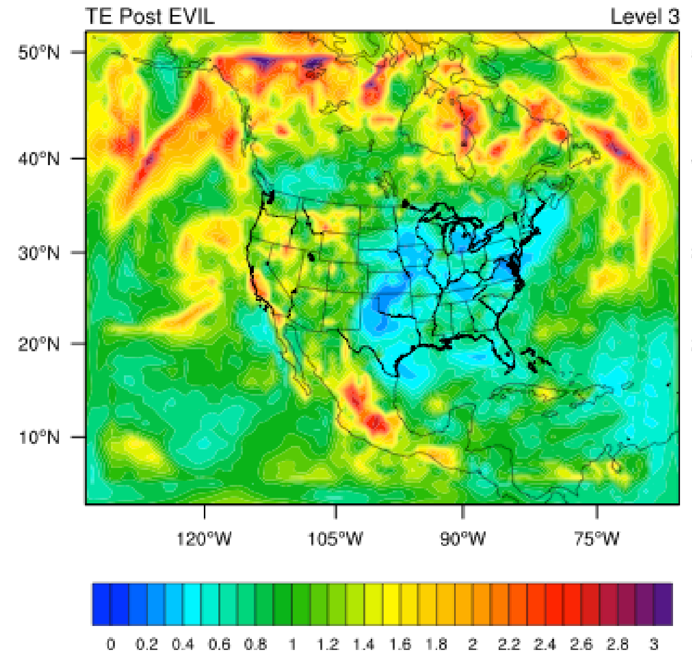


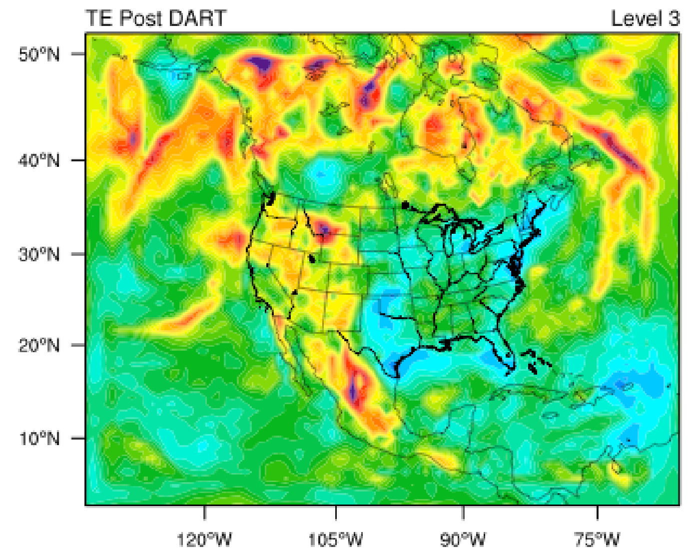


EVIL: Single Assimilation

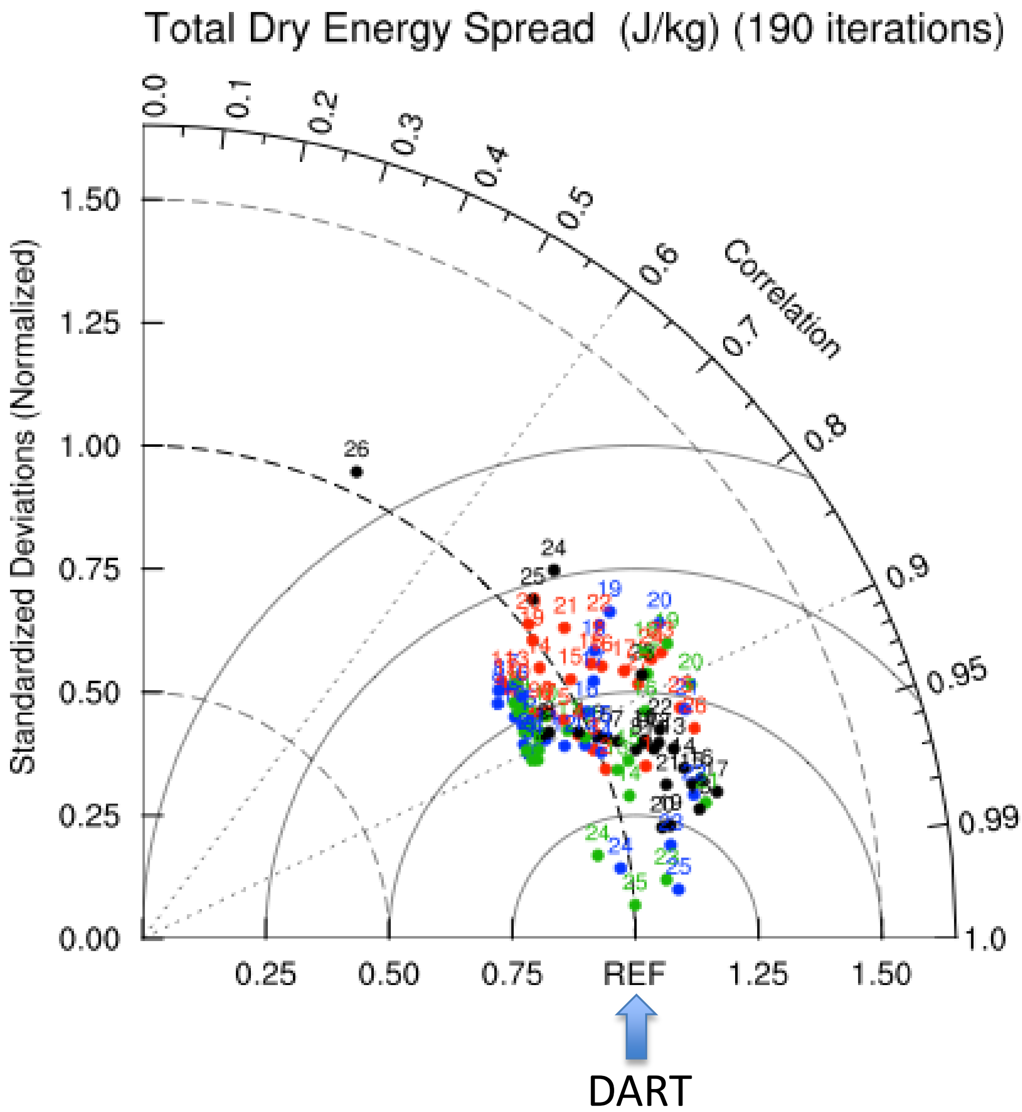


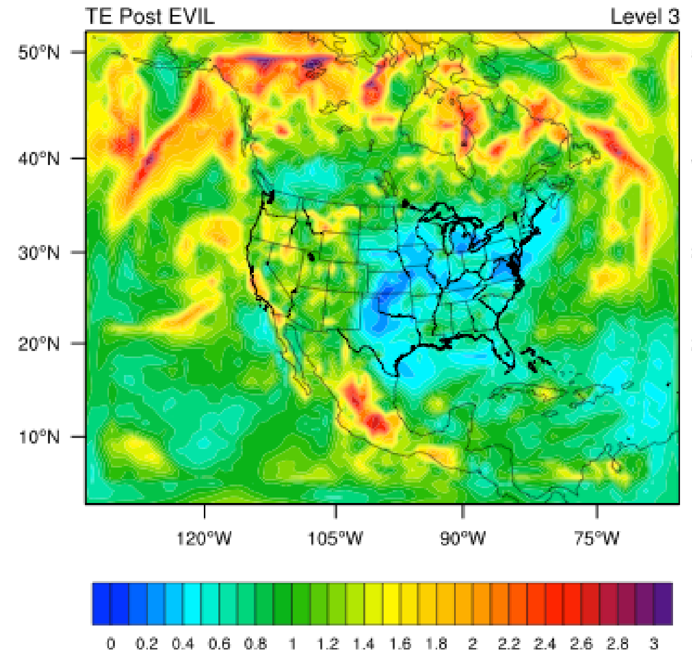
DART

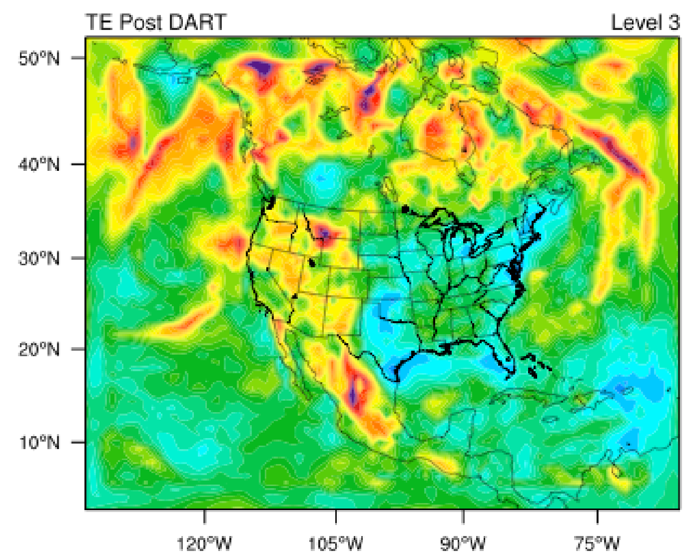




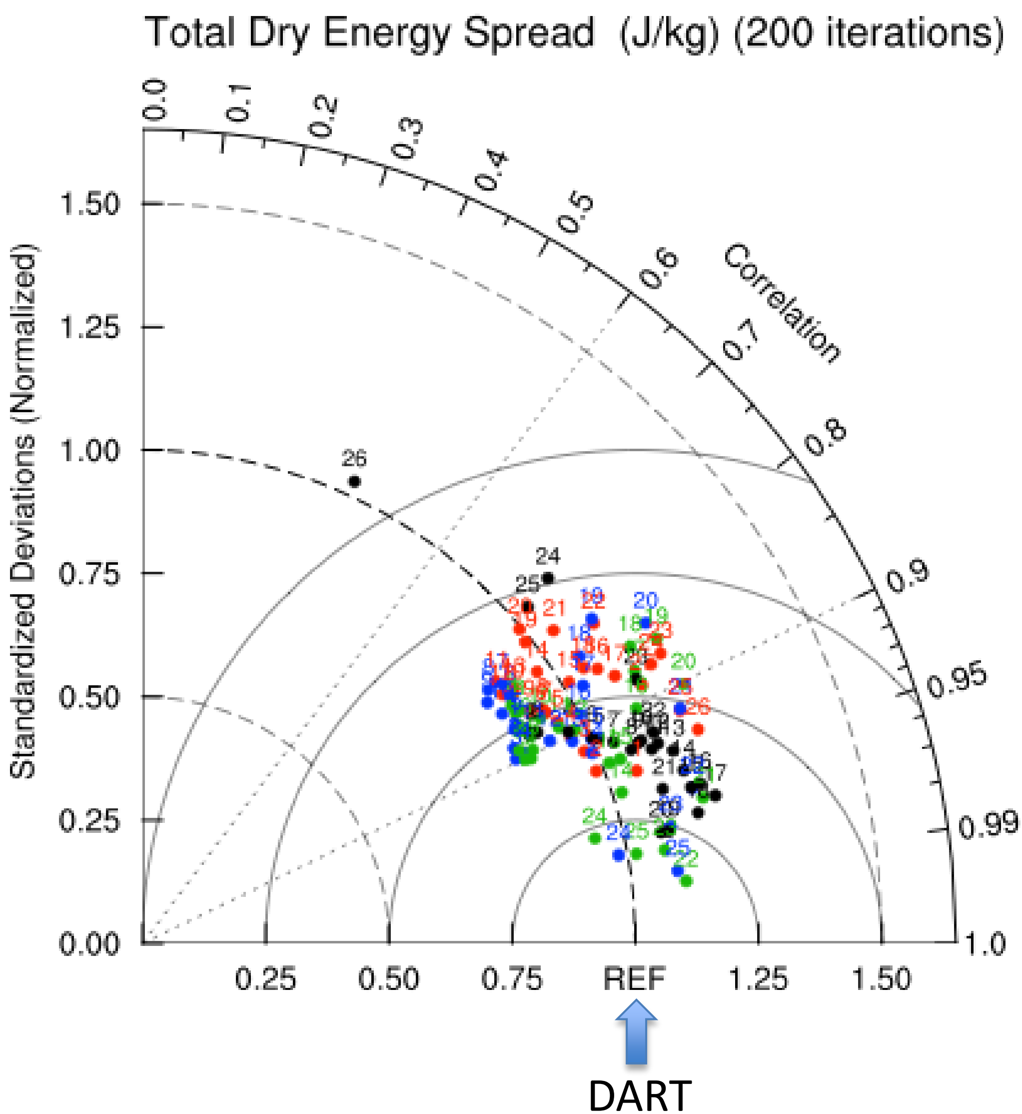
EVIL: Single Assimilation

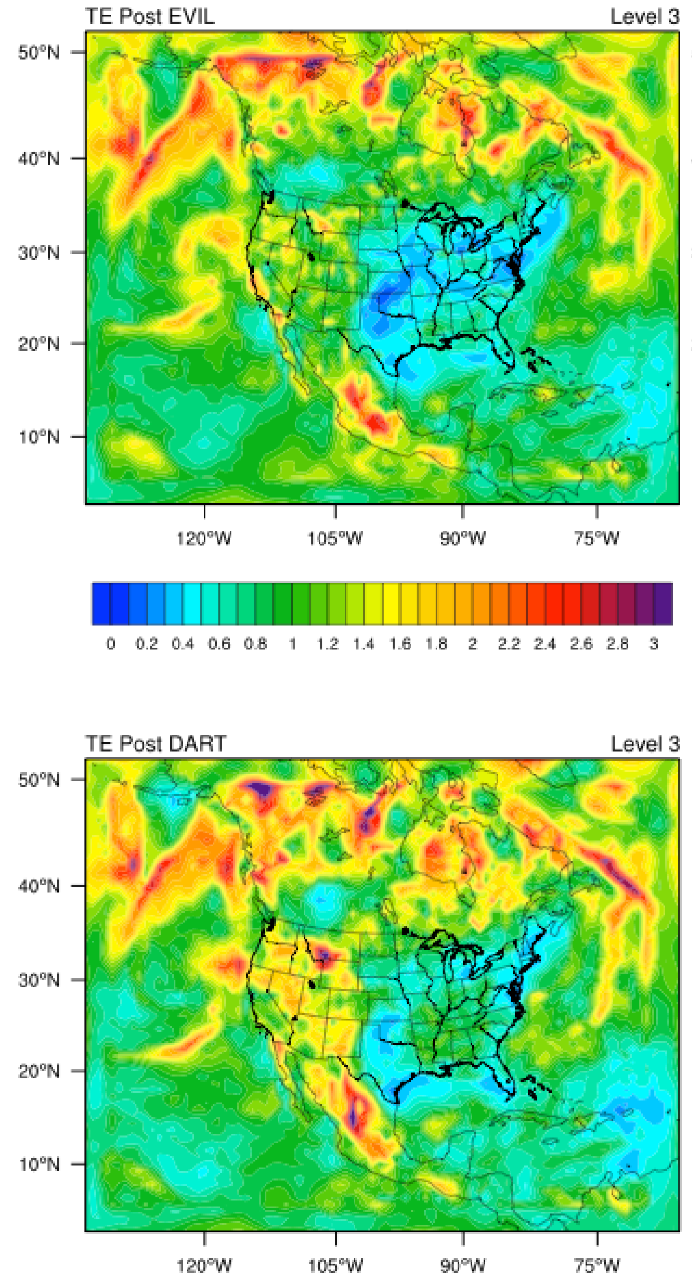




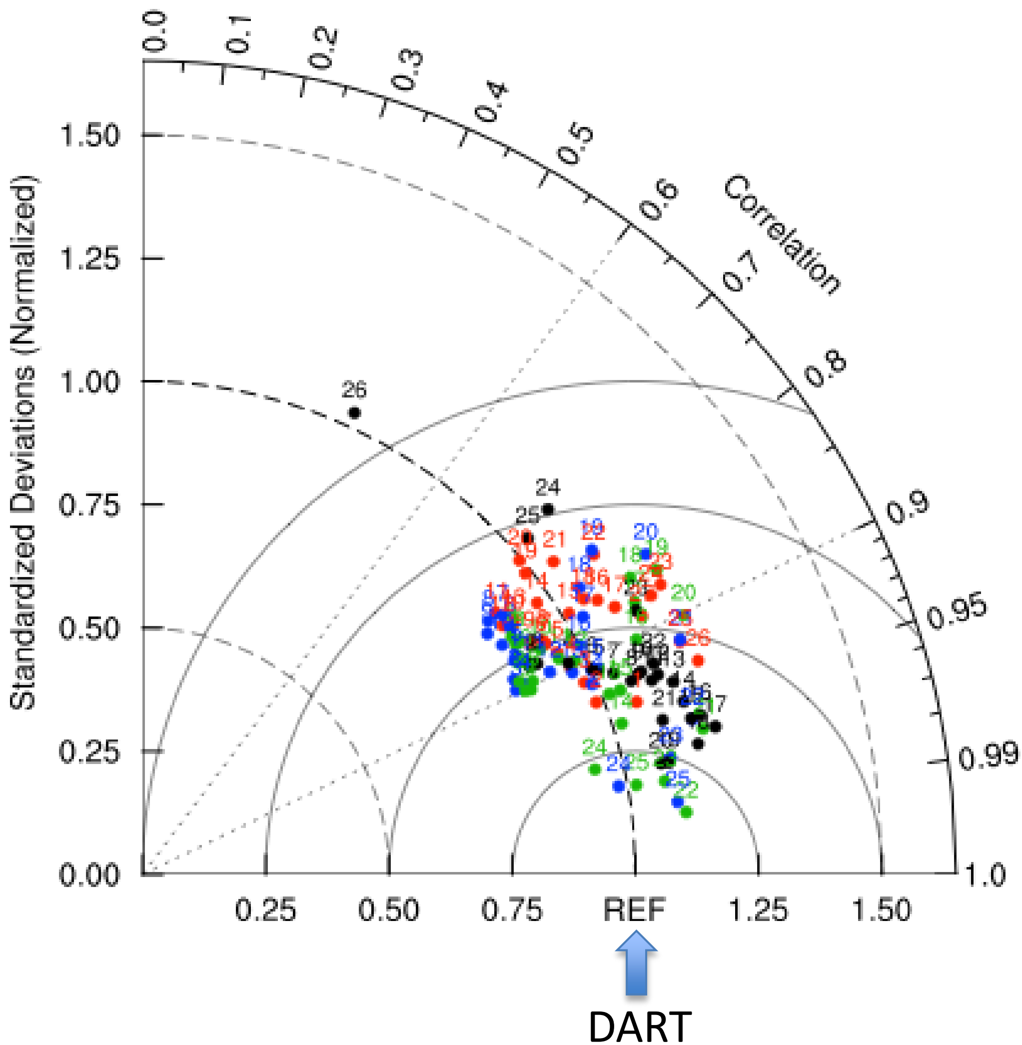


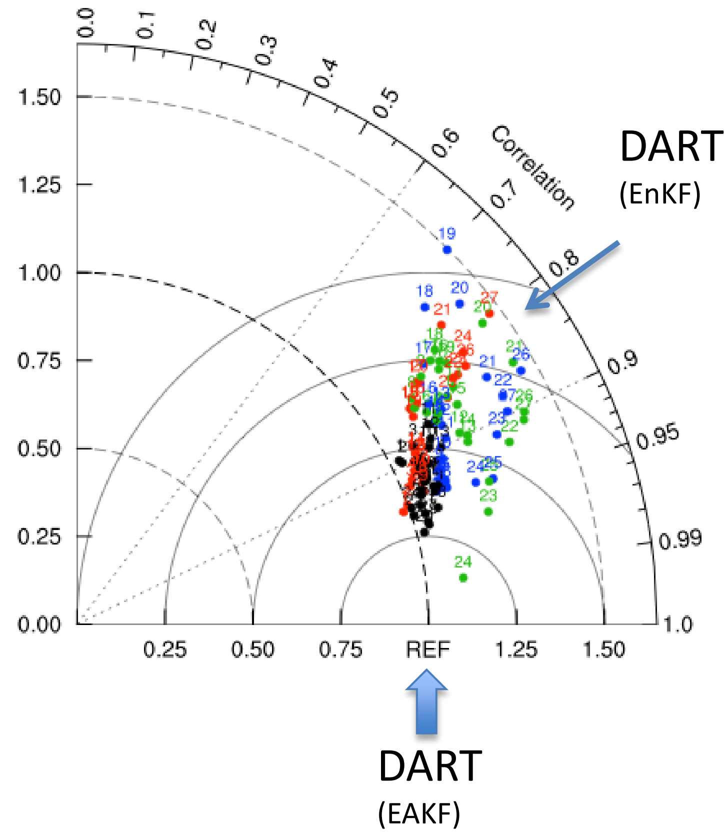
EVIL: Single Assimilation

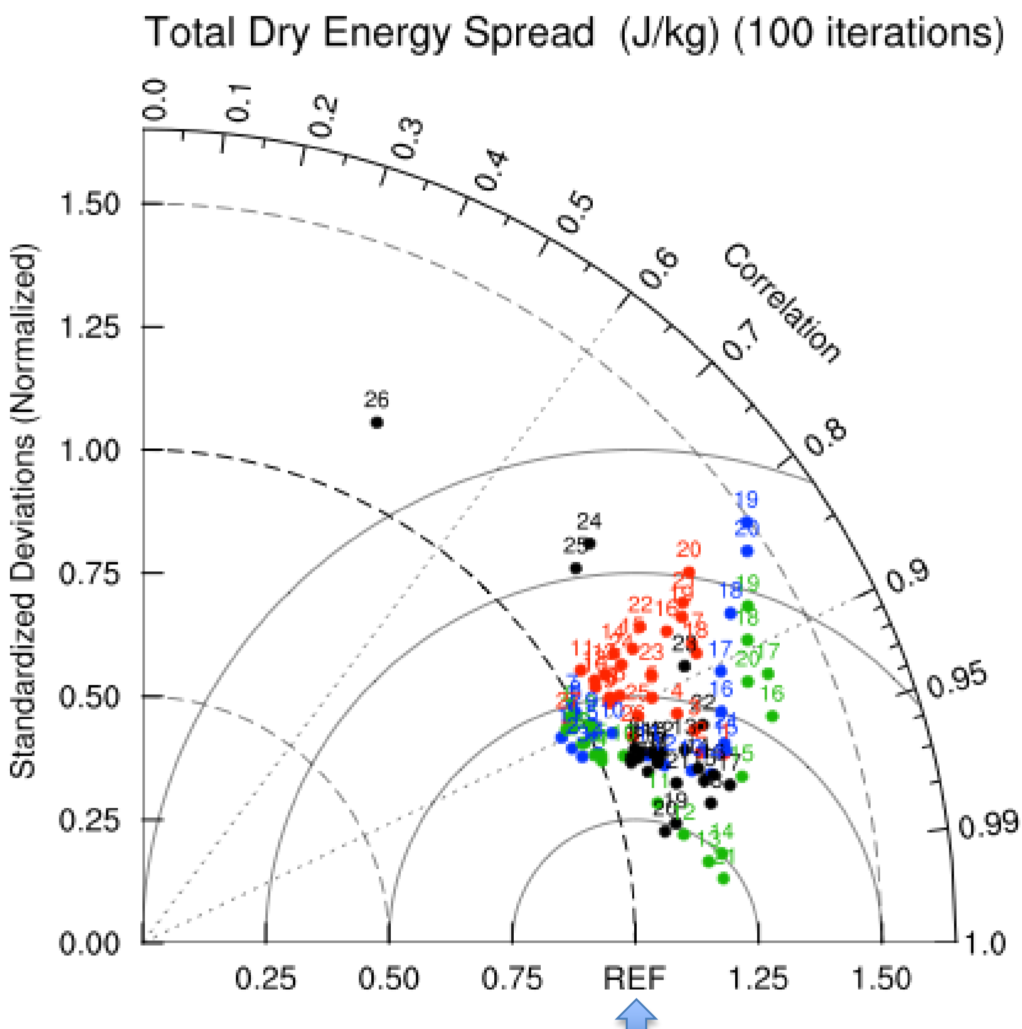


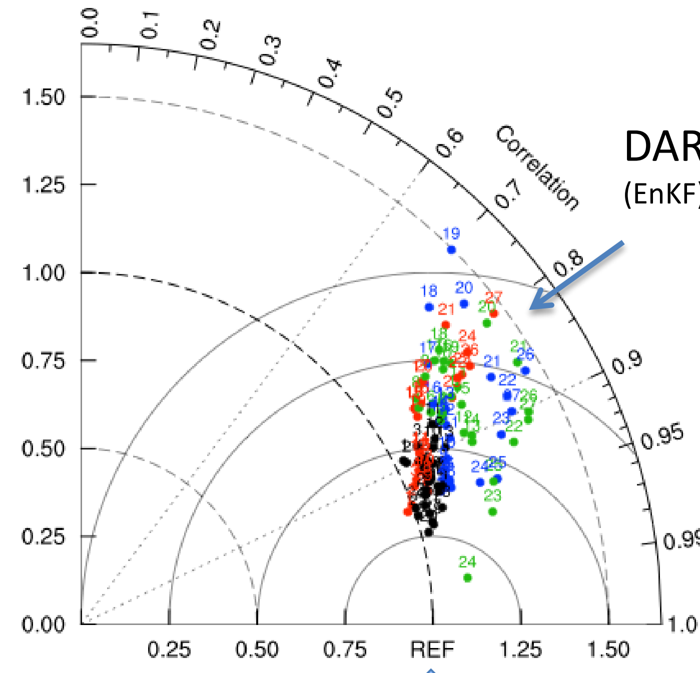


Total Dry Energy Spread (J/kg) (200 iterations)





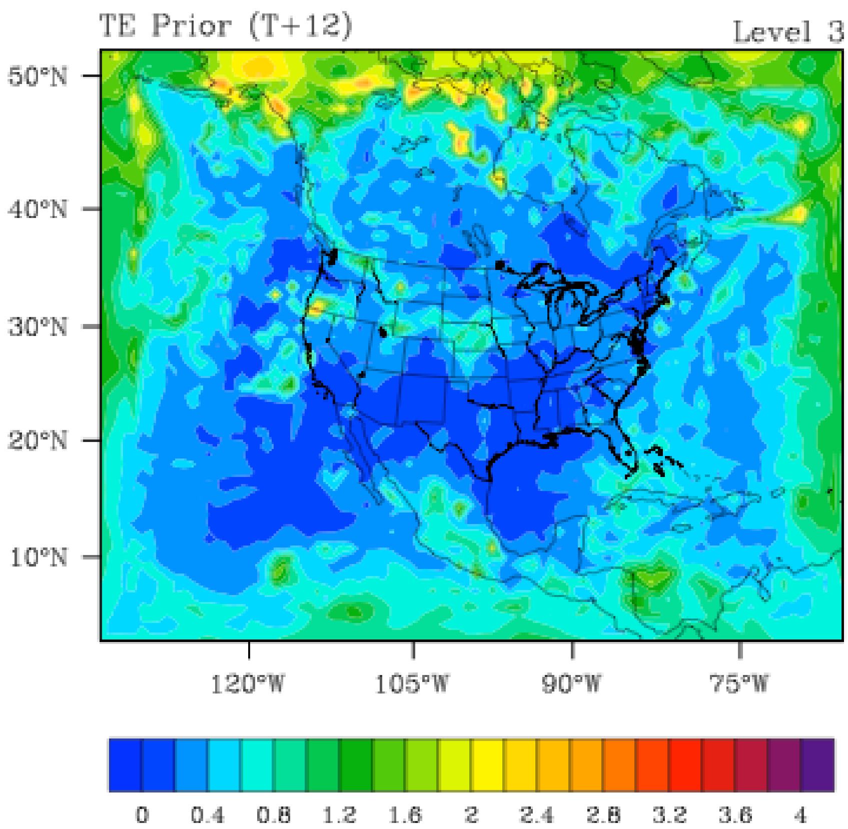

Total Dry Energy Spread (J/kg) (100 iterations)
Standardized Deviations (Normalized)
Correlation
REF
DART

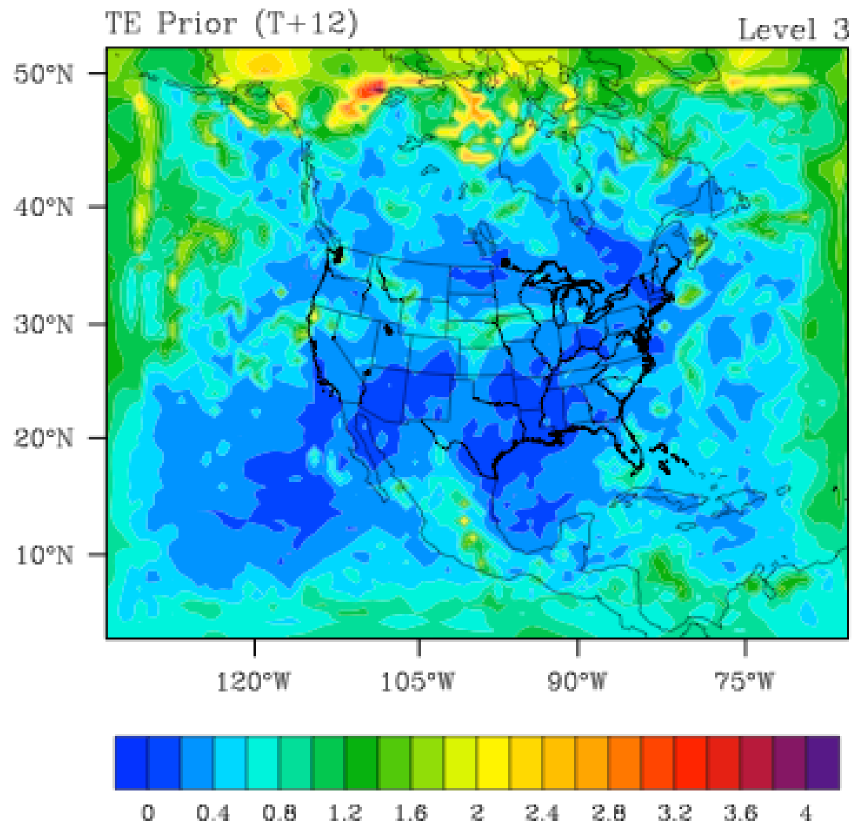

Correlation
REF
DART
(EnKF)
DART
(EAKF)

EVIL: Experimental Design

- One-week full cycling experiment (2009060100-2009060712)
- Analyses at 00 and 12UTC
- Constant inflation factor (1.02)
- 100 iterations for EVIL

Prior Spread after cycling independently for 7 days





EVIL **DART**

EVIL: “early stopping”

- Inflation is a necessary evil for Ensemble Kalman Filters
    - Un-represented forecast model error
    - Under-estimation of posterior spread

- Each EVIL analysis is an iterative process (like 3DVar)

- In machine learning, stopping the minimization is a common technique to avoid over-fitting noisy targets.

- Similarly, “early stopping” can be used in EVIL to avoid “over-confidence” in the posterior spread

- **When do we stop the minimization?**

EVIL: "early stopping"

- **<u>Option 1</u>**: stop when the posterior inflation factor = 1.0

$$d = y^o - H(x_b) - \overline{H}\delta x$$

$$E(dd^T) = R + \mu\overline{H}P_a\overline{H}^T$$

$$\mu \approx Tr[E(dd^T)] - \frac{Tr(R)}{Tr(\overline{H}P_a\overline{H}^T)}$$

$$\mu = Tr[E(R^{-1}dd^T)] - \frac{Tr(I_p)}{Tr(\overline{H}P_a\overline{H}^T R^{-1})}$$

$$\mu \approx \frac{2Jo(x_a) - p}{DFS}$$

$$DFS \approx \sum[Rand(\overline{H}P_a\overline{H}^T R^{-1})]$$

$$P_a = C_\alpha^{1/2}[I + Z(\Theta^{-1/2} - I)Z^T]C_\alpha^{-1/2}P_f$$

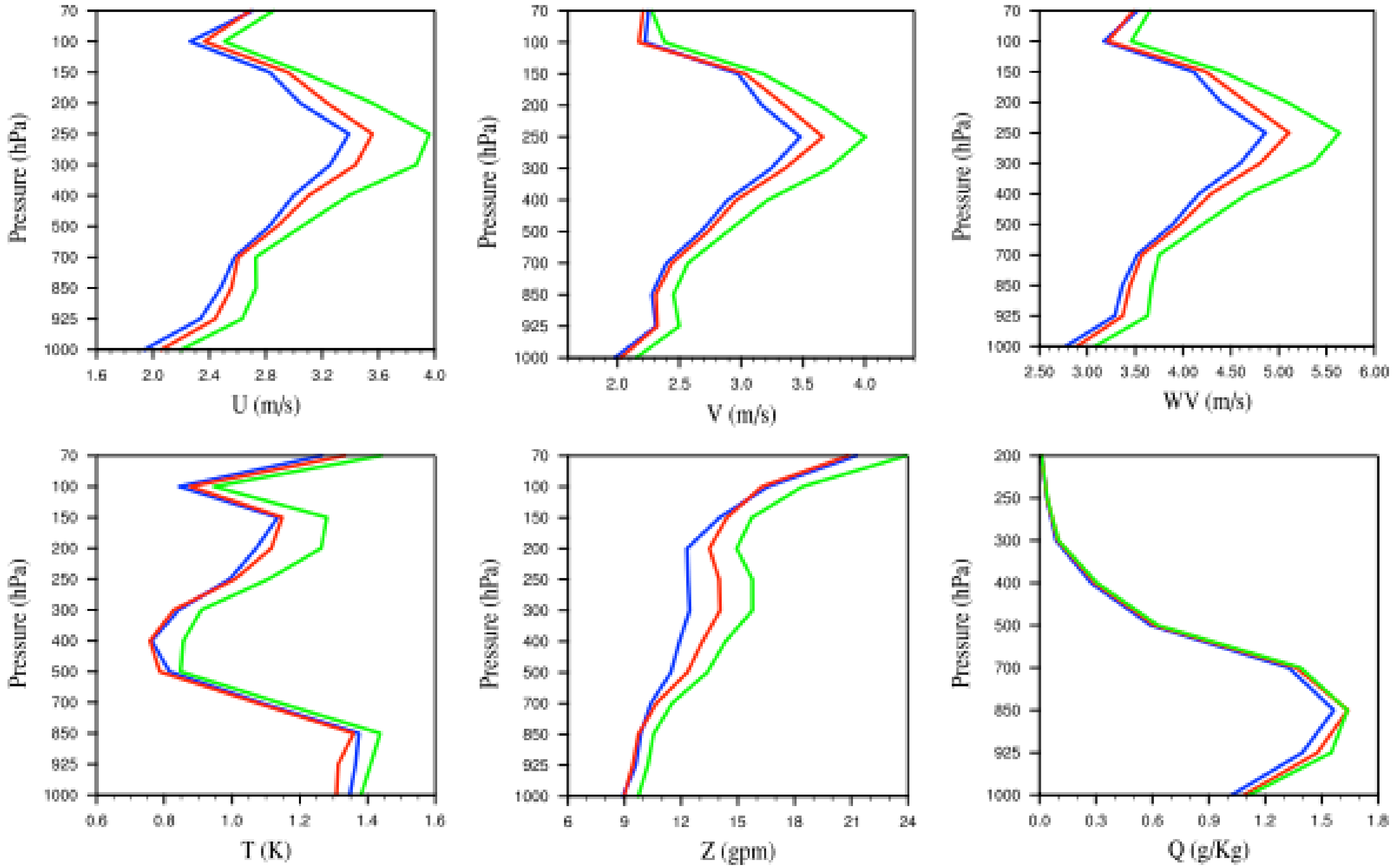
EVIL: "early stopping"

- **<u>Option 2</u>**: stop when the correction to Pf spread gets within sampling noise

$$\delta x_a^k \approx C_\alpha^{1/2}[I + Z(\Theta_i^{-1/2} - I)Z^T]C_\alpha^{-1/2} \circ \delta x_f^k$$

$$\frac{\sum_{k=1}^{N_e}\left\| C_\alpha^{1/2}[Z_i(\Theta_i^{-1/2} - I)Z_i^T]C_\alpha^{-1/2} \circ \delta x_f^k \right\|}{\sum_{k=1}^{N_e}\left\| \delta x_f^k \right\|} \approx \left\| Z_i(\Theta_i^{-1/2} - I)Z_i^T \right\| \le \frac{1}{\sqrt{N_e - 1}}$$

RMSE 03 - 07 Jun 2009 (12-Hourly Cycle) vs. FNL analysis



DART 1.02 — Constant Inflation (1.02)

DART — Adaptive Inflation

EVIL — Early Stopping at 50 iterations (**no inflation**)

Conclusion

- Introduced a new Ensemble/Variational "integrated" approach: EVIL
(based on the Hybrid, updating ensemble perturbations inside VAR)

- Initial testing in "full-ensemble" : EVIL behaves similarly to DART.

- "Early stopping" seems to reduce the need for inflation.

- Future plans:
  - Extended-period testing with more members + vertical localization
  - Investigate optimal inflation vs. early stopping
  - Test in Hybrid mode
  - Suggestions???