# Predicting forecast uncertainty with singular vectors and ensemble-derived analysis error covariances

Mark Buehner

Meteorological Service of Canada
Dorval, Quebec

Thanks to Ayrton Zadra and Peter Houtekamer

Internal Seminar, February 20, 2004.

# Introduction

- several applications require accurate prediction of forecast uncertainty or its origins: ensemble prediction, observation targeting, data assimilation, data QC
- singular vectors (SVs) identify the most unstable subspace of the (linearized) forecast model according to specified norms at initial and optimization times
- current applications (ECMWF, NRL) use SVs with total energy (TE) norms for ensemble prediction and observation targeting
- if initial-time norm is inverse analysis error covariances, SVs optimally describe forecast uncertainty at optimization-time (if model error negligible, TL error growth)
- EnKF provides estimate of flow-dependent analysis error covariances via Monte Carlo simulation to model effect of all error sources in forecast-analysis cycle
- goal: evaluate use of these flow-dependent or (more easily obtained) stationary analysis error cov from ensemble approaches to define initial-time norm for SVs

# Outline

- SV theory and calculation
- approaches for estimating analysis error cov with:
  1. EnKF (flow-dependent)
  2. perturbed 3d-var (stationary)
- impact of covariance norms on structure of SVs (vs TE)
- impact of norms on distribution of sensitive regions (targeting application)
- in context of ensemble prediction (fraction of explained NH fcterr), compare:
  1. SVs
  2. EnKF members
  3. random perturbations
- conclusions and future directions

# SV Theory and Calculation

- given an analysis, find perturbation $\mathbf{x}_0$ that maximizes ratio:

$$\frac{\mathbf{x}_t^T \mathbf{W}_t \mathbf{x}_t}{\mathbf{x}_0^T \mathbf{W}_0 \mathbf{x}_0}, \quad \text{where } \mathbf{x}_t = \mathbf{M}\mathbf{x}_0$$

- reformulate in terms of control variable $\gamma$, where $\mathbf{x}_0 = \mathbf{W}_0^{-1/2}\gamma$

- results in eigenvalue problem: $\boxed{\mathbf{W}_0^{-T/2}\mathbf{M}^T\mathbf{W}_t\mathbf{M}\mathbf{W}_0^{-1/2}\gamma_k = \sigma_k^2\gamma_k}$

- $\mathbf{W}_t$: final-time norm typically TE, restricted to geographical region of interest

- $\mathbf{W}_0$: initial-time norm commonly TE (TESV), but...

- when $\mathbf{W}_0 = \mathbf{P}_a^{-1}$, the final-time SVs are the leading eigenvectors of $\mathbf{M}\mathbf{P}_a\mathbf{M}^T$ (predictability error $\equiv$ forecast error for perfect model)

# Experimental Setup

- like 4d-var, transfer data between 3d-var software (Lanczos solver and norms) and GEM (TLM and ADJ):

| 3d-var | GEM |
|---|---|
| $\times \mathbf{W}_0^{-1/2}$ | |
| $\longrightarrow$ | run tlm |
| $\times \mathbf{W}_t$ | $\longleftarrow$ |
| $\longrightarrow$ | run adj |
| $\times \mathbf{W}_0^{-T/2}$ | $\longleftarrow$ |

- TLM and ADJ at 120x60x28L with simplified linear physics package

- final time norm is TE north of 30°N, 48 hr optimization time interval

- SVs computed each day, 1-14 Dec, 2003

- different initial time (global) norms used :
  - total energy (as commonly used at several NWP centres)
  - inverse analysis error covariances (stationary and flow-dependent)

# Approaches for Sampling Analysis Error

Kalman Filter equations: $\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}_f, \quad \mathbf{P}_f = \mathbf{M}\mathbf{P}_a\mathbf{M}^T + \mathbf{Q}$

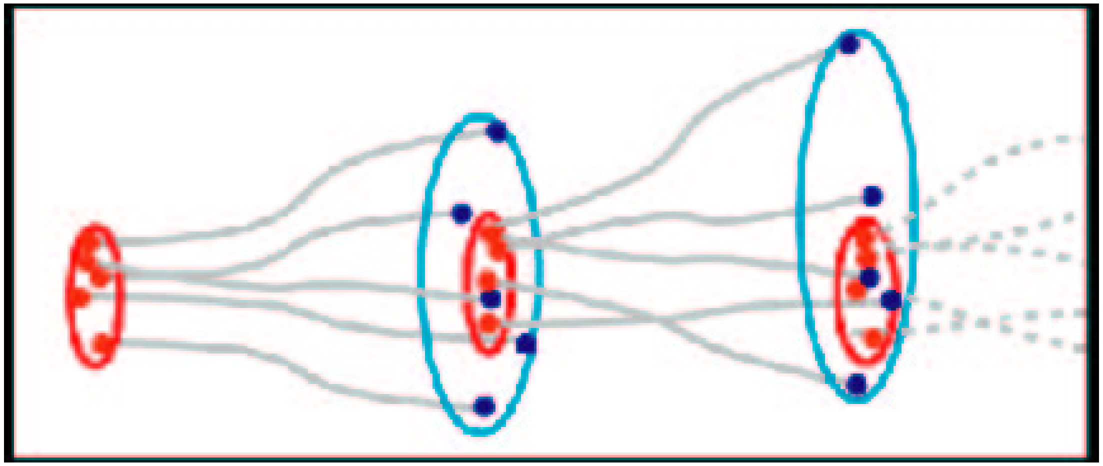
- not feasible for realistic model, only valid for linear model, little knowledge of $\mathbf{Q}$
- approx with Monte Carlo simulation approaches (perturb uncertain quantities):

EnKF: different ensemble of 128 error samples every 6h

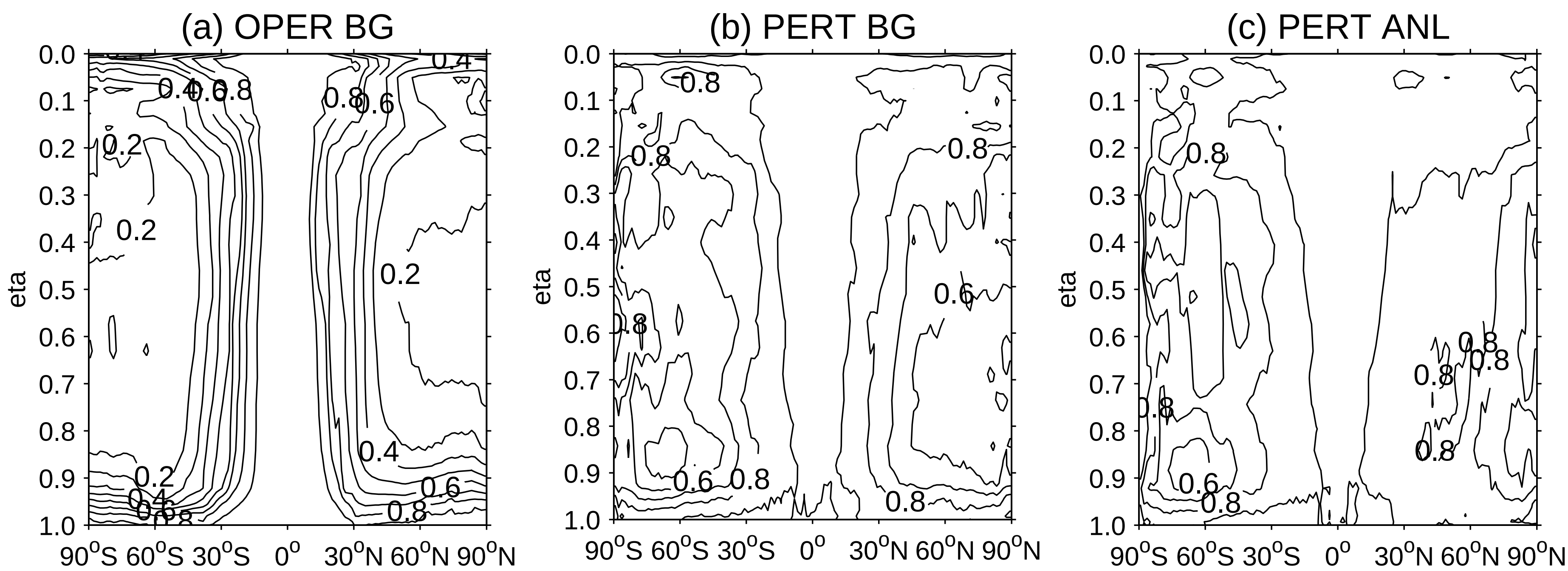
- analysis uses perturbed observations and flow-dependent ensemble covariances
- model error: random perturbations with similar cov as $\mathbf{B}$ in oper. 3d-var
- for SV norm: deviations of analyses from ensemble mean: flow-dependent $\mathbf{P}_a$

Perturbed 3d-var: single ensemble of 138 error samples from 4 week period

- analysis uses perturbed observations and 3d-var (ECMWF; EPS at CMC)
- model error: random perturbations from adaptively scaled $\mathbf{B}$
- for SV norm: differences between analyses from perturbed and unperturbed exp'ts: stationary $\mathbf{P}_a$



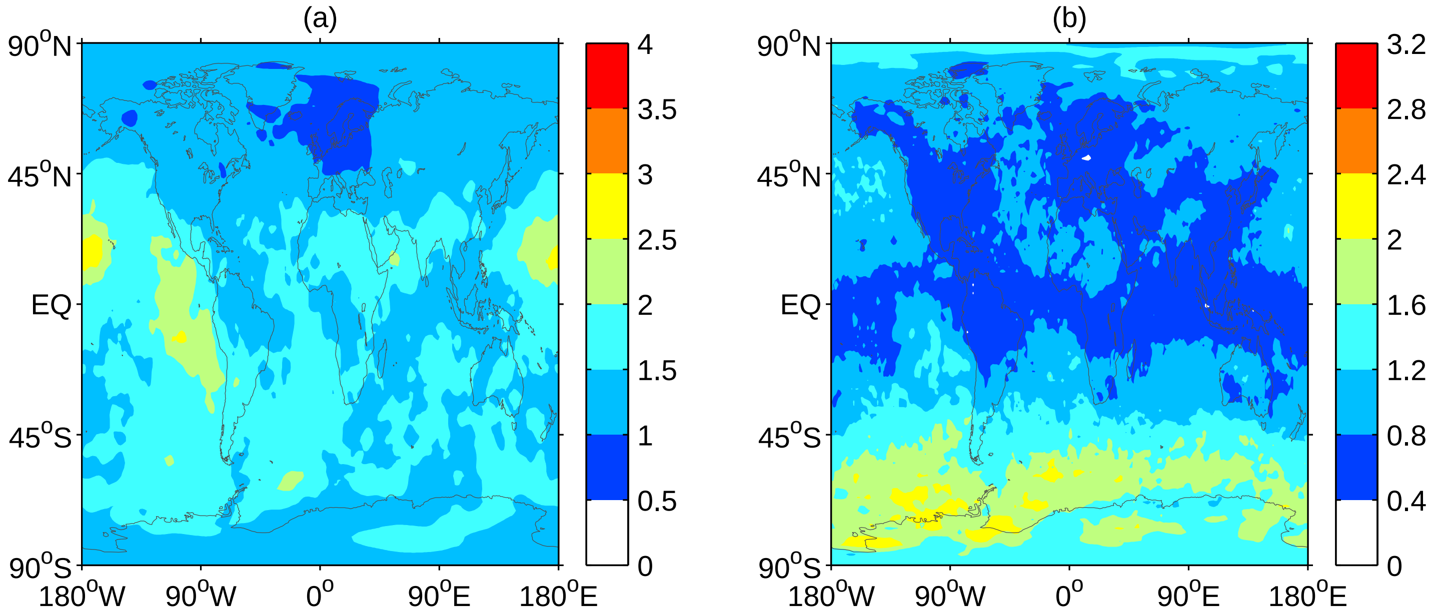
# Geostrophic Balance in Perturbed 3d-var Covariances



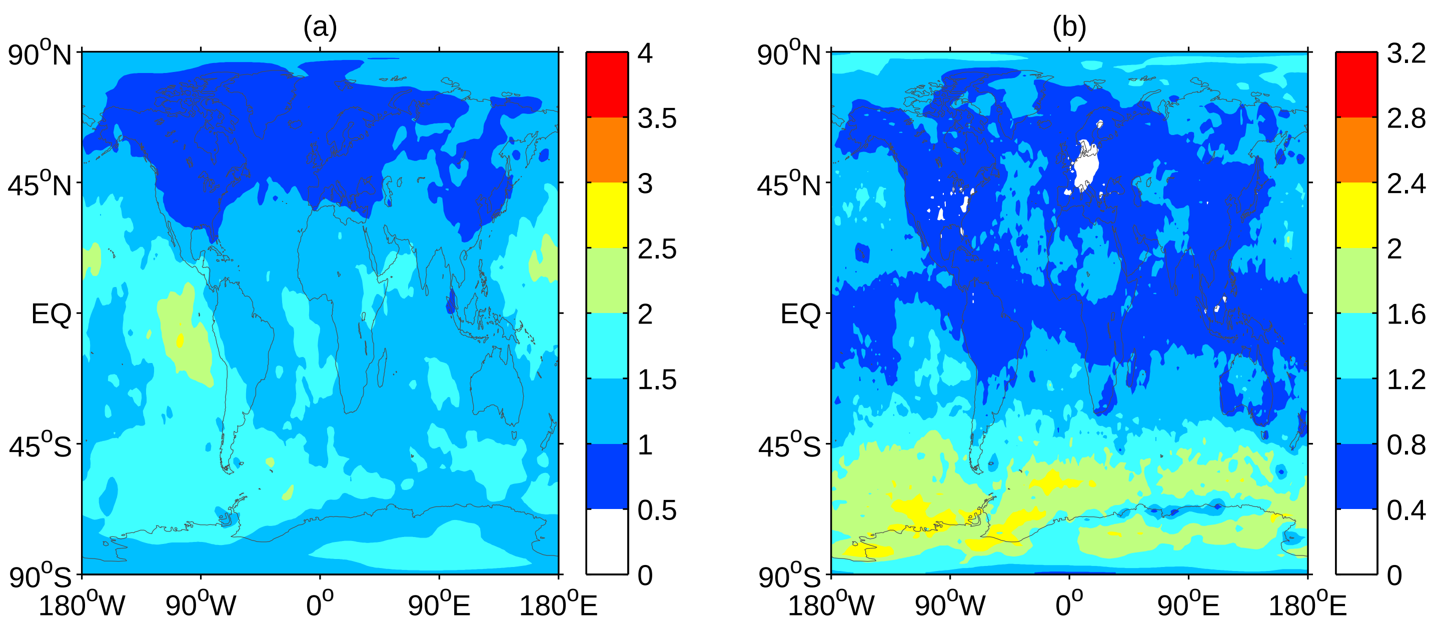

- ratio of unbalanced T to total T variance
- in operational BG covariances T almost completely balanced with winds
- less mass/wind balance in perturbed 3d-var (and EnKF) BG covariances
- analysis error even less balanced than background error

# Std Dev of (a) Ψ at 250hPa and (b) T at 500hPa

BG error cov from perturbed 3d-var (138 error samples from 4 week period)

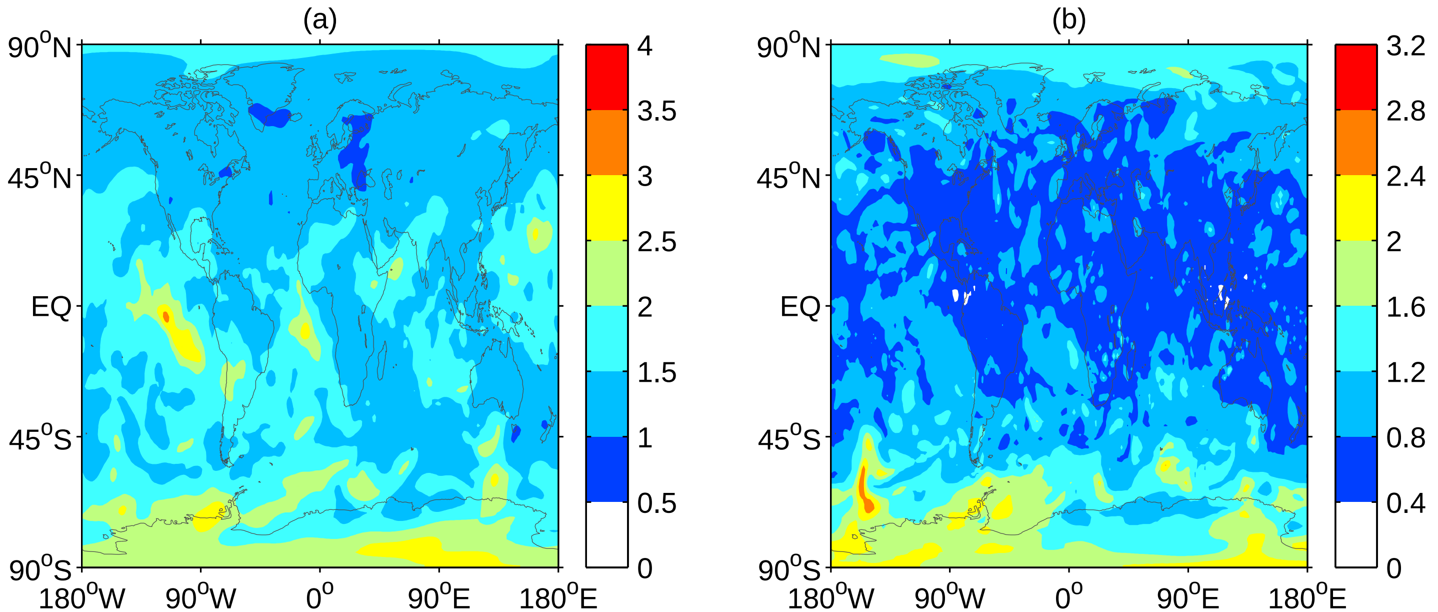



Analysis error cov from perturbed 3d-var (138 error samples from 4 week period)

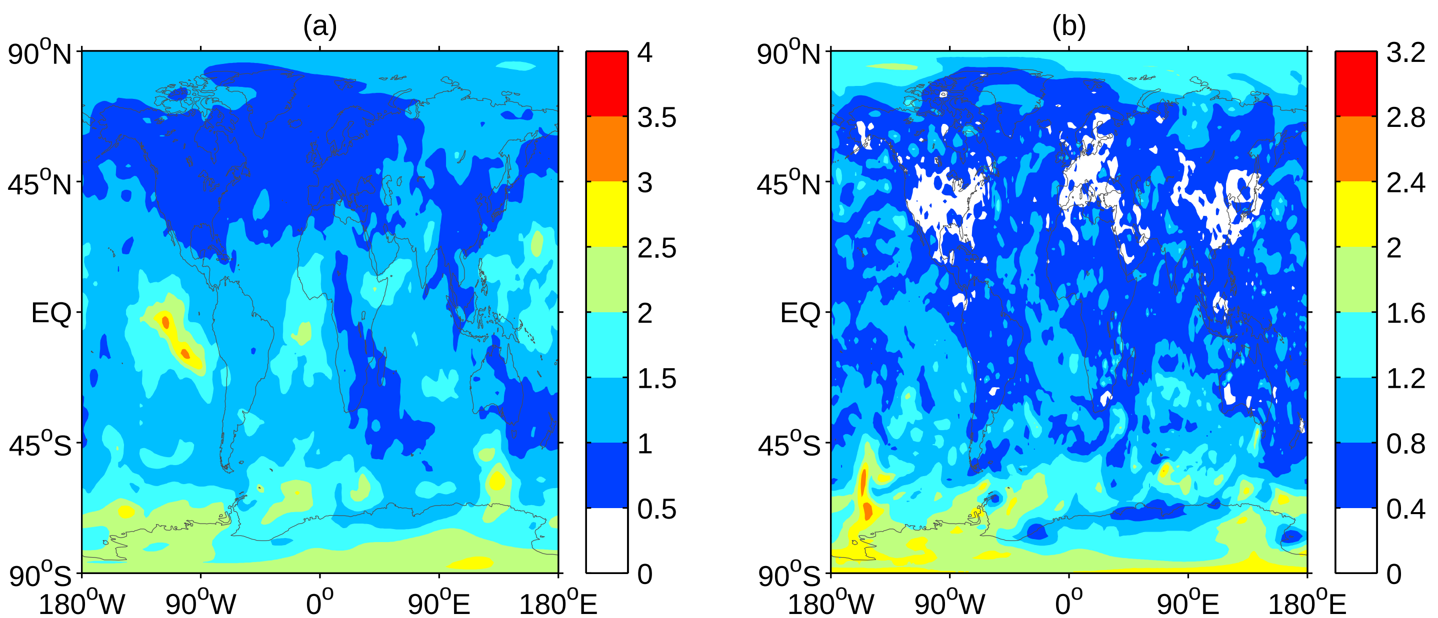


# Std Dev of (a) $\Psi$ at 250hPa and (b) T at 500hPa (2)

BG error cov from EnKF (128 error samples valid 2003121212)




Analysis error cov from EnKF (128 error samples valid 2003121212)



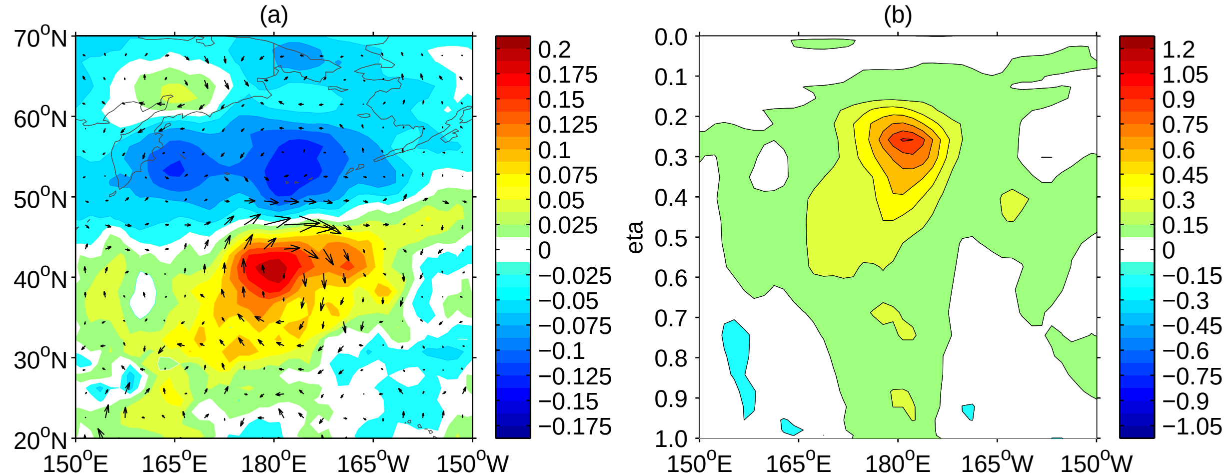
# Estimation of Error Correlations

- cannot accurately estimate full correlation matrix (rank $O(10^6)$) from small number of samples ($O(10^2)$)
- therefore must impose constraints on structure of correlations (extra information)
- operational BG error correlations assume homogeneity/isotropy
- alternatively, we can spatially localize horizontal and vertical correlations (currently done in EnKF analysis and in experimental 3d-var)

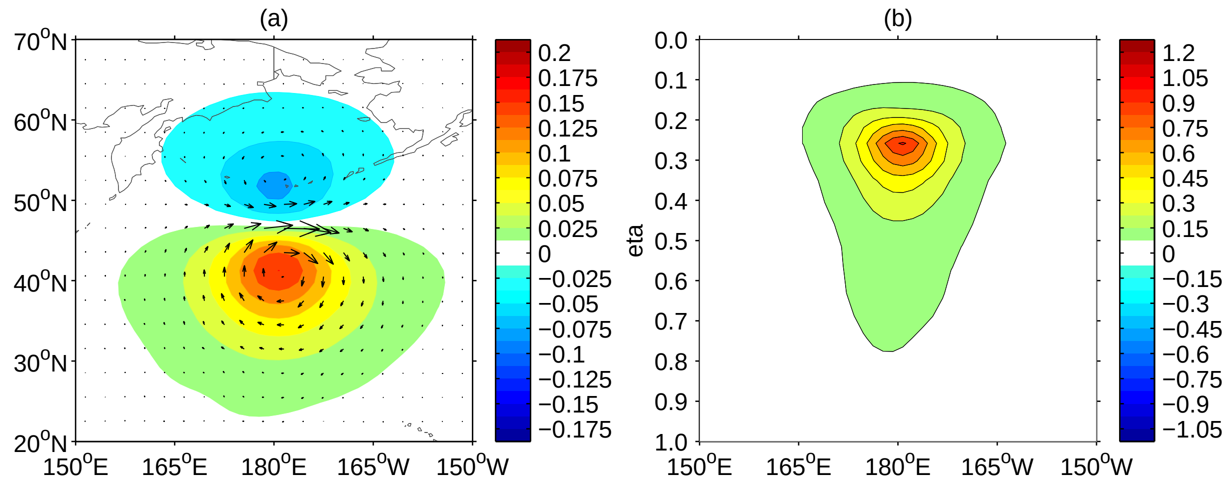
- demonstrate by plotting BG error covariances of U @ 250 hPa with wind and GZ (next slide)

# Estimation of Error Correlations (2)

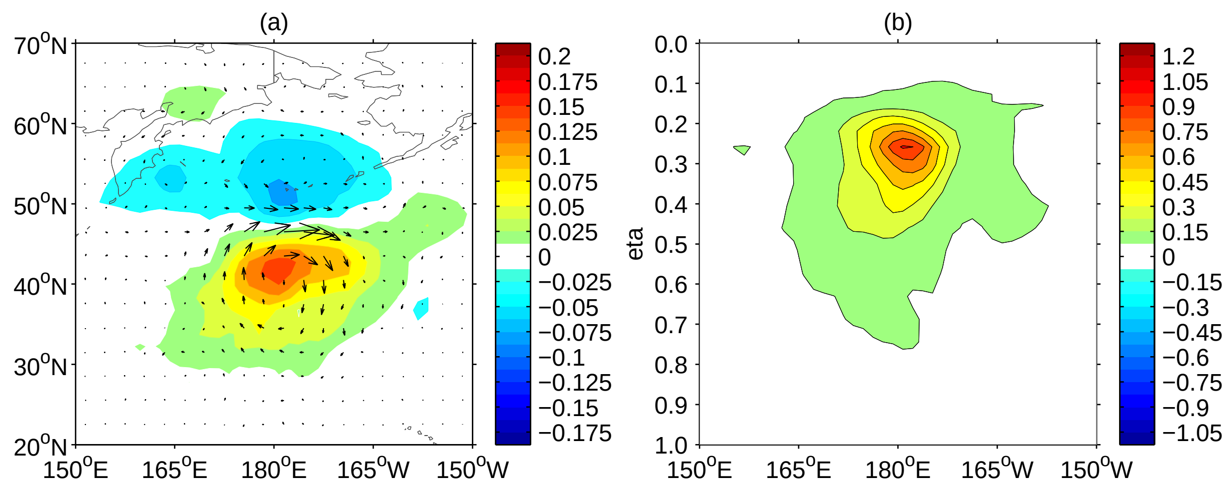
Original sample estimate of correlations:




Homogeneous/isotropic correlations:




Spatially localized correlations:

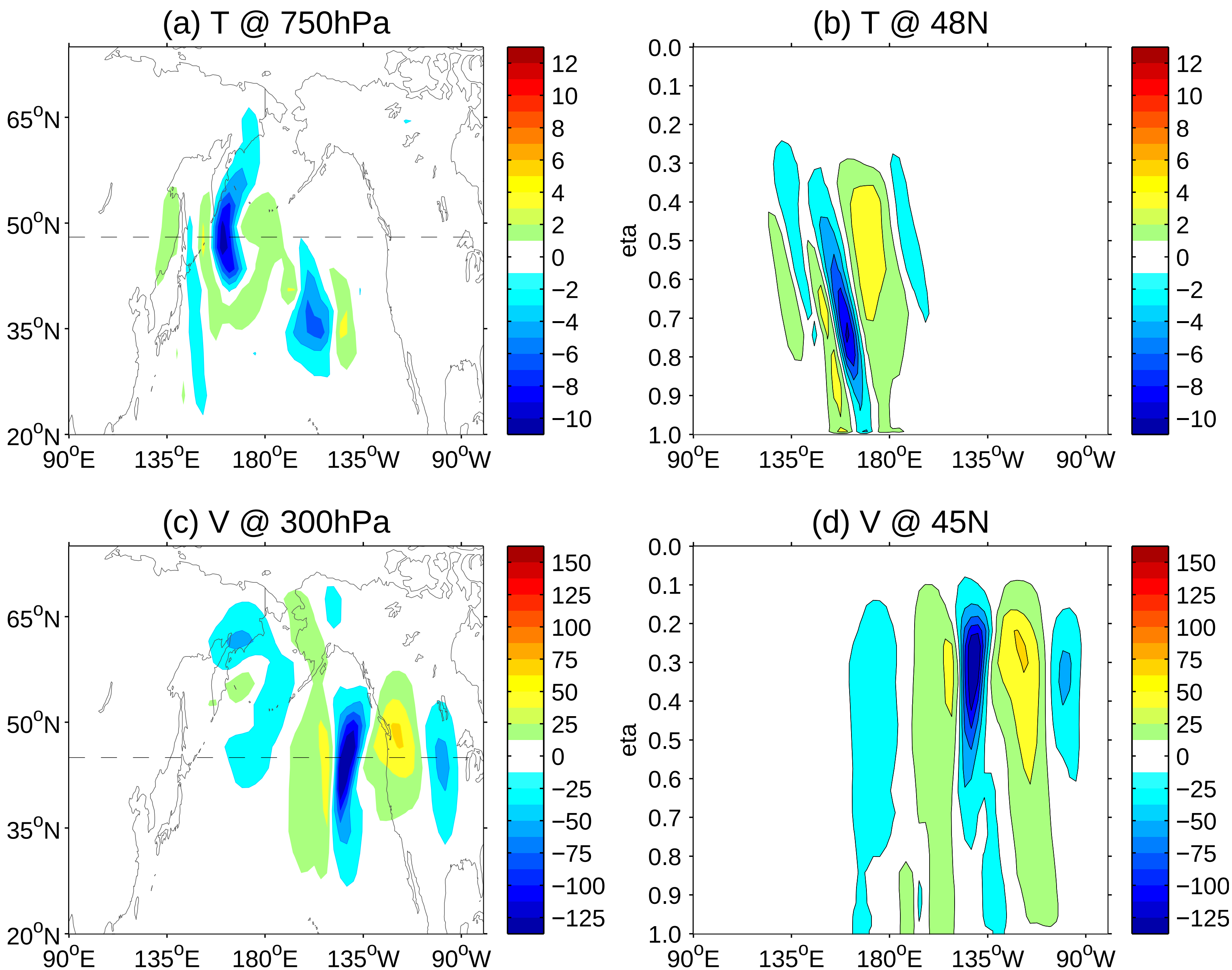


# Impact of Initial-time Norm on SV structure

## Leading total energy SV on 2003121600

initial time:

final time:

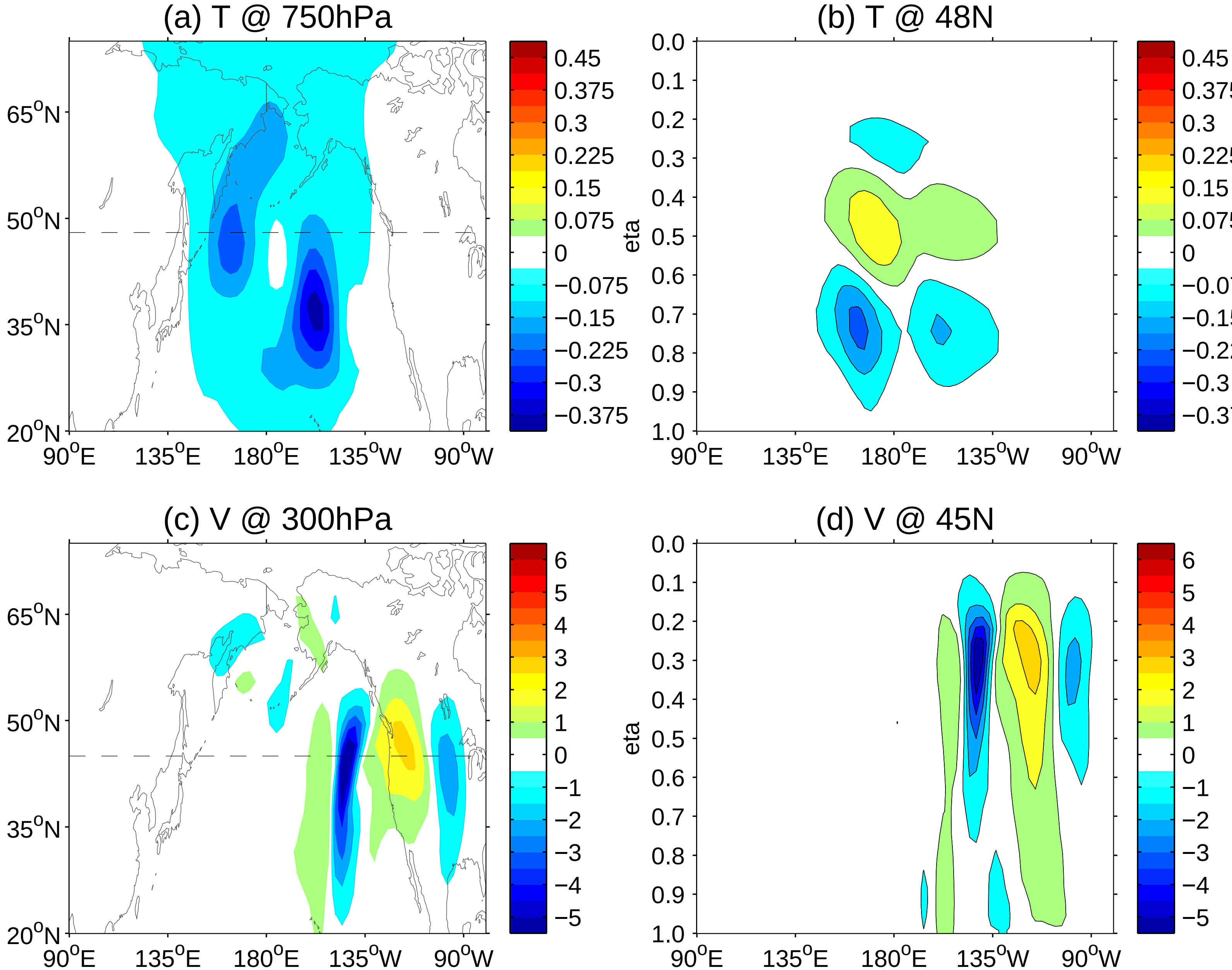


# Impact of Initial-time Norm on SV structure (2)

Leading analysis error covariance SV on 2003121600
(perturbed 3d-var, homogeneous/isotropic correlations)

initial time:

final time:

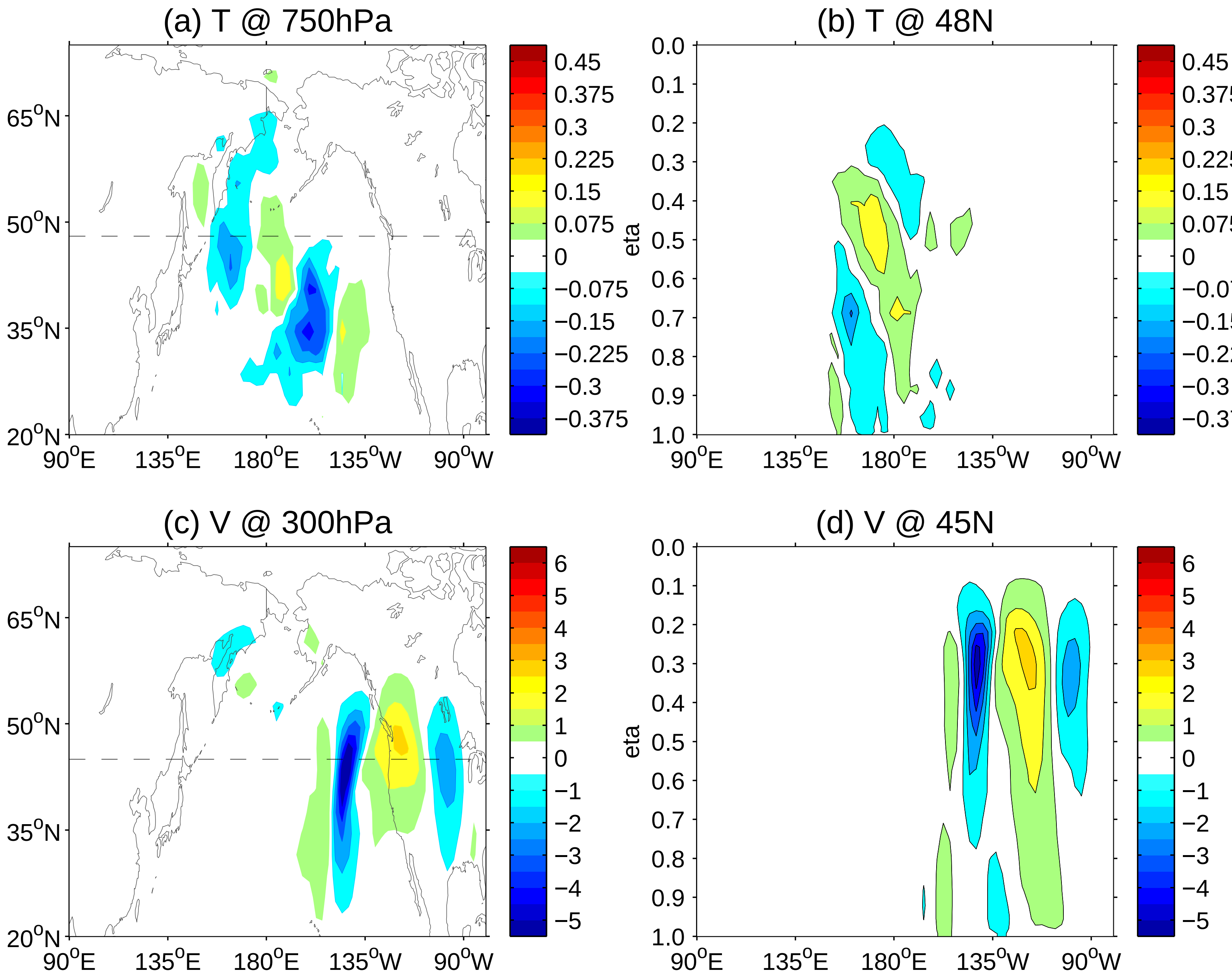


# Impact of Initial-time Norm on SV structure (3)

## Leading analysis error covariance SV on 2003121600 (perturbed 3d-var, localized correlations)

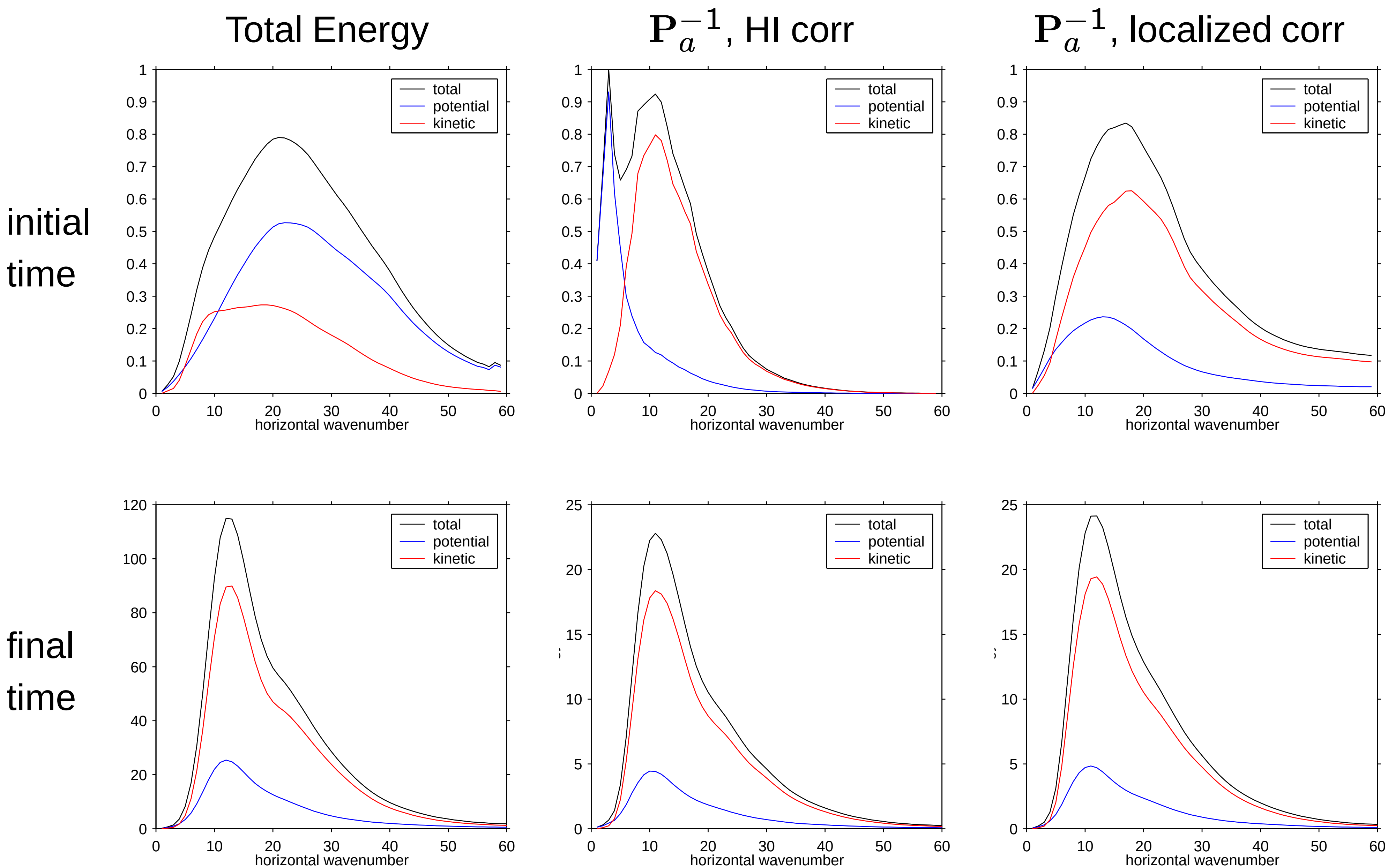
initial time:

final time:



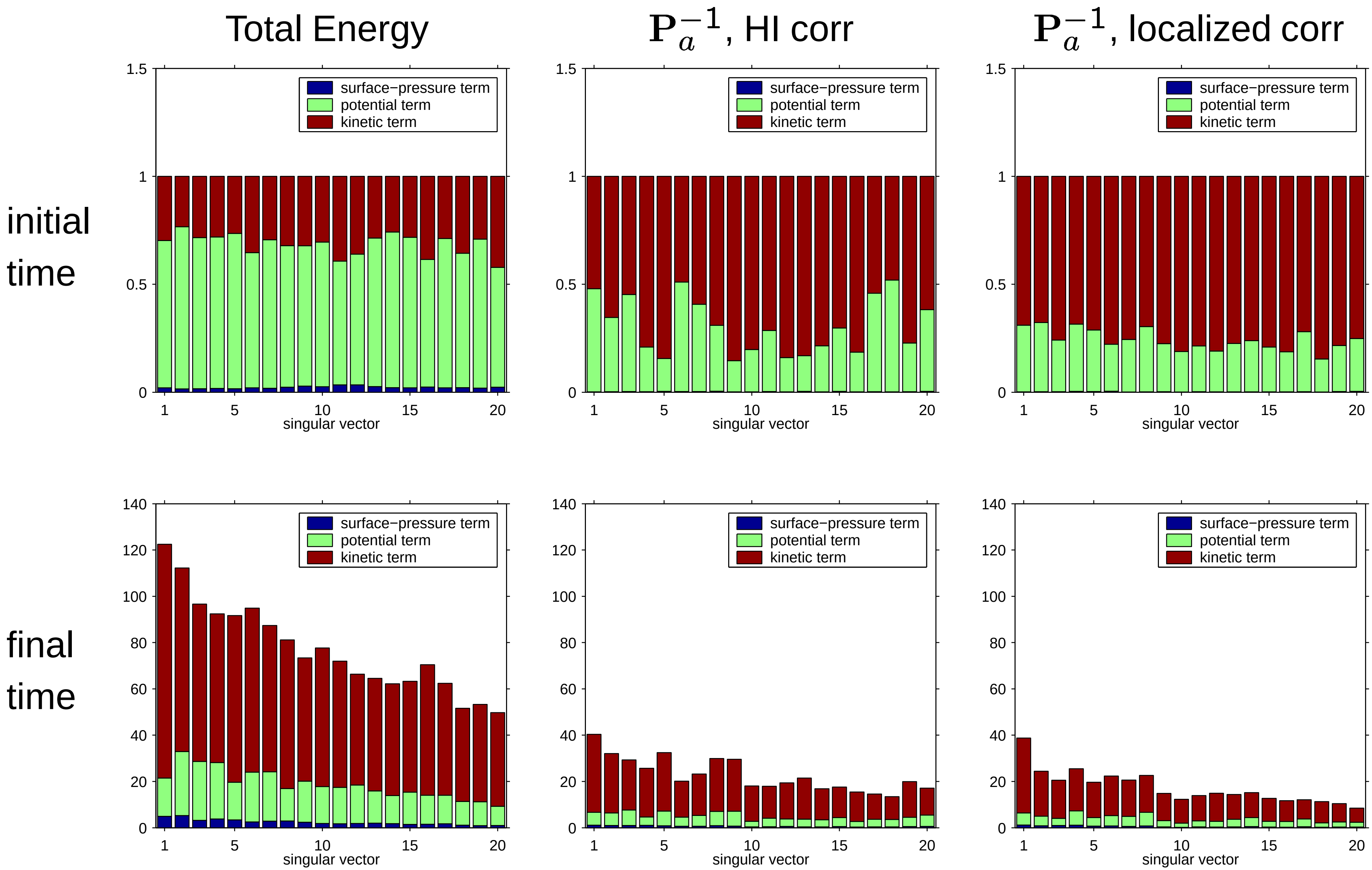
# Impact of Initial-time Norm on SV structure (4)

## Vertically integrated power spectra for 20 SVs on 2003121600 with initial-time norm:



# Impact of Initial-time Norm on SV structure (5)

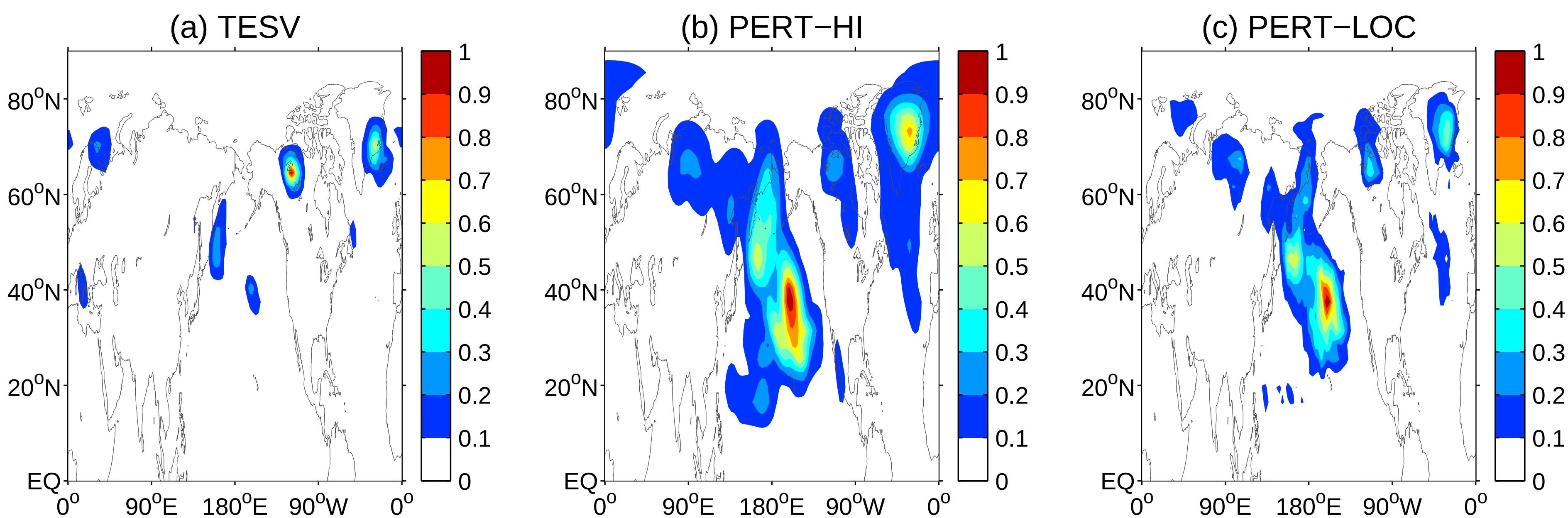
## Partition of energy for 20 SVs on 2003121600 with initial-time norm:



# Impact of Initial-time Norm on Observation Targeting

Initial-time vertically integrated total energy, growth factor weighted average of 20 SVs with initial-time norms on 2003121600:

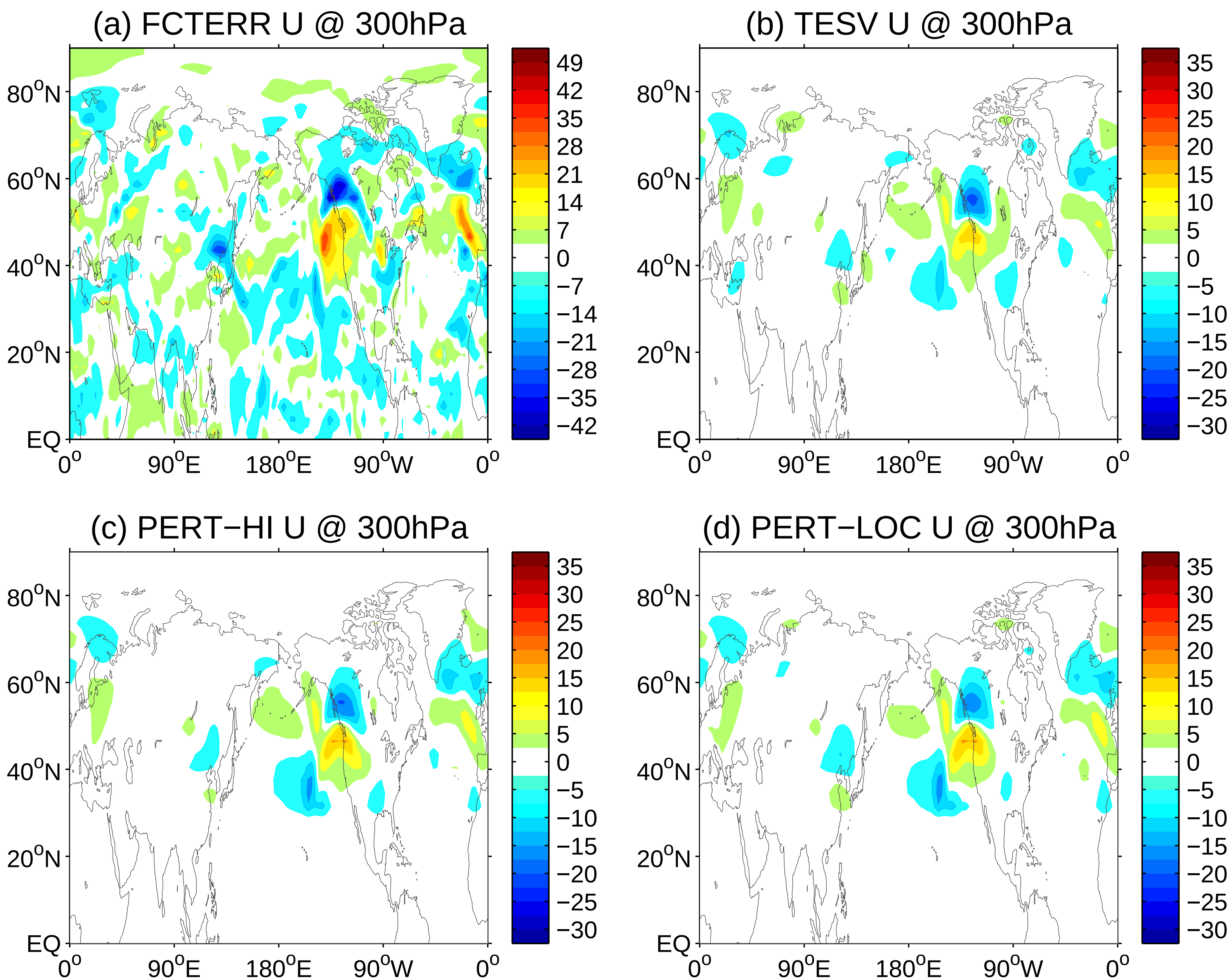
- (a) total energy
- (b) analysis error cov (perturbed 3d-var, homogeneous/isotropic correlations)
- (c) analysis error cov (perturbed 3d-var, localized correlations)




Type of map used to select locations for adaptive observation deployments (e.g. dropsondes)

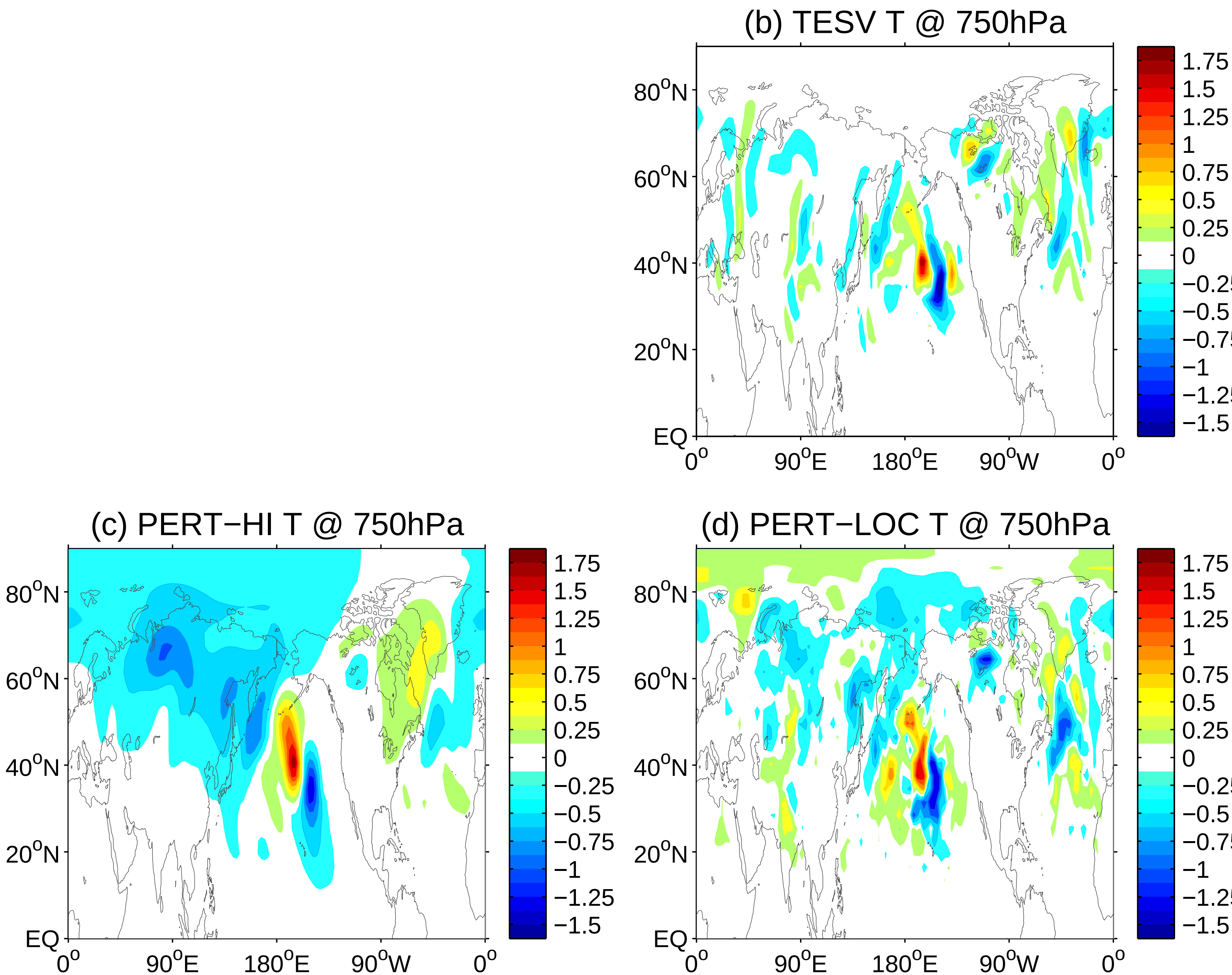
# Impact of Initial-time Norm on Observation Targeting (2)

NH FctErr and its projection onto 20 SVs at optimization time (2003121800)



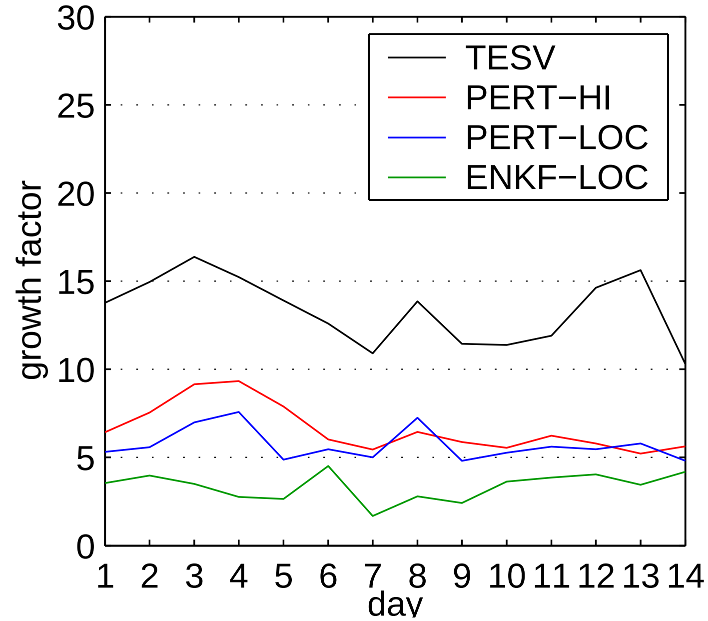
# Impact of Initial-time Norm on Observation Targeting (3)

Pseudo-Inverse of FctErr projection onto 20 SVs - like KAE
(linear combination of initial-time SVs on 2003121600)

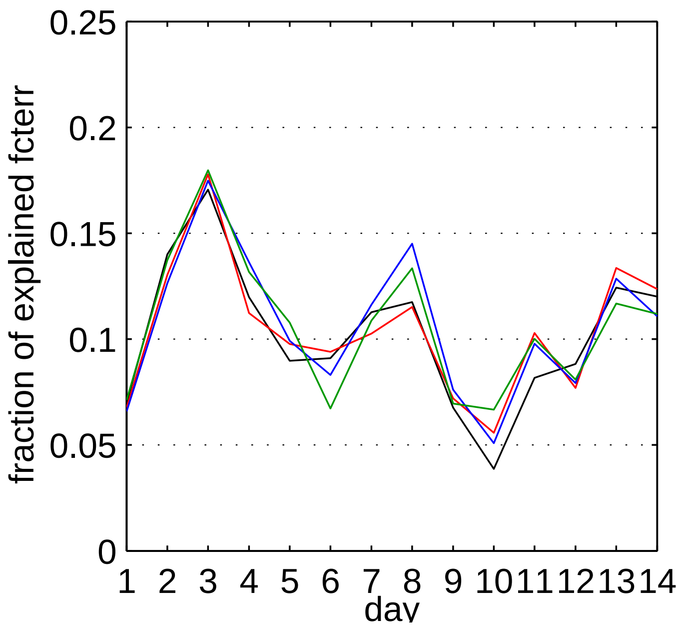


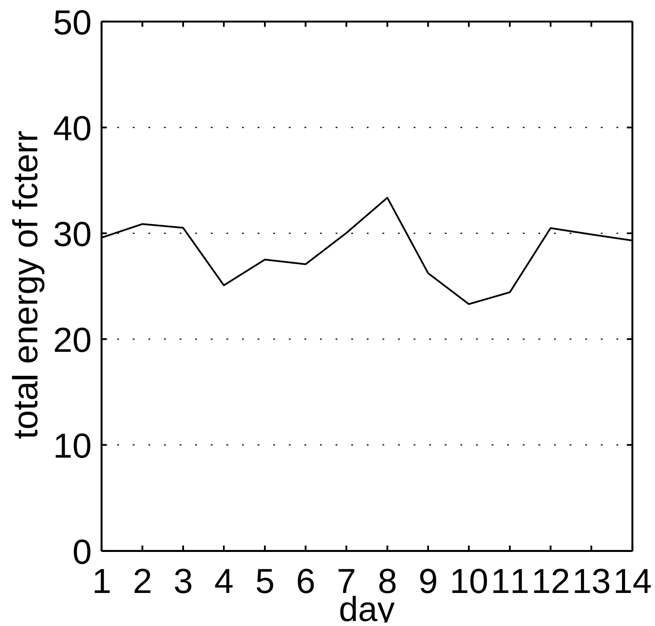
# Impact of Initial-time Norm on Predicting FctErr

- total energy growth factor of leading SV ($\sqrt{TE_{48}/TE_0}$)
  - TESV: initial SV norm is total energy
  - PERT-HI: initial SV norm from perturbed 3d-var analysis error cov with homogeneous/isotropic corr
  - PERT-LOC: initial SV norm from perturbed 3d-var analysis error cov with localized corr
  - ENKF-LOC: initial SV norm from flow-dependent EnKF analysis error cov with localized cor



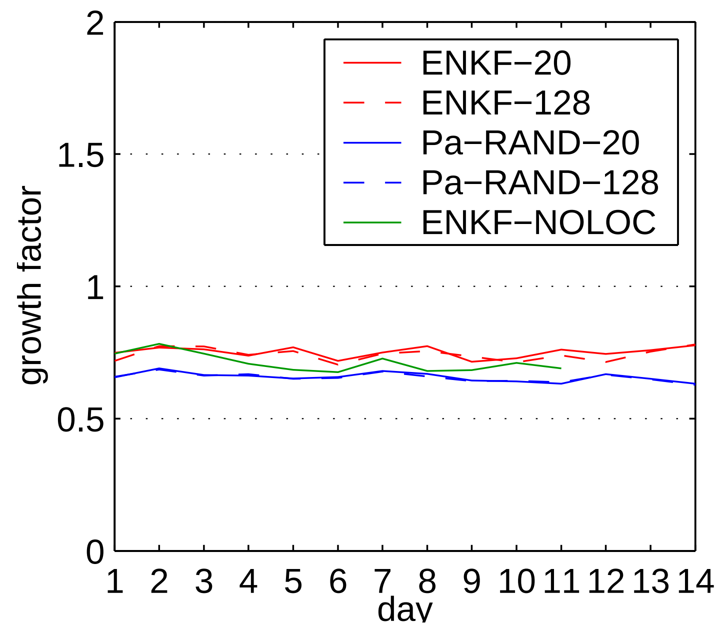

- fraction of 48-hour NH forecast error energy explained by 20 SVs and total fcterr



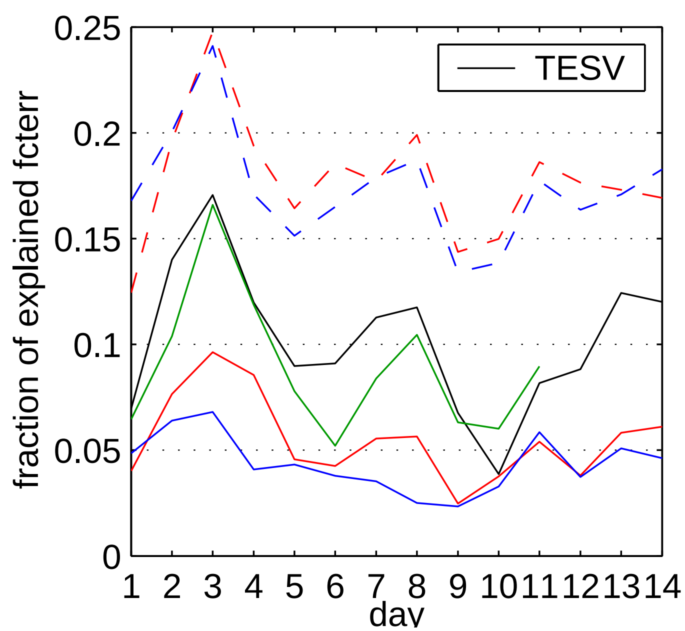



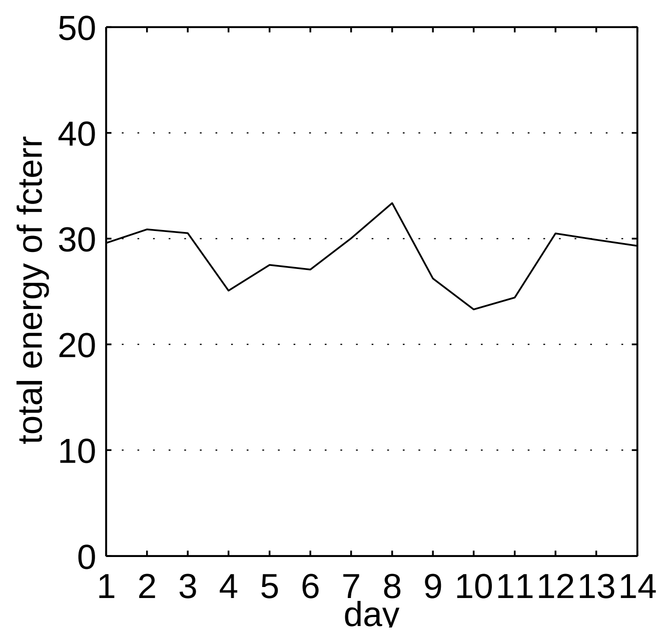
# EnKF Members vs Random Pert. for Predicting FctErr

- total energy growth factor of one ensemble member ($\sqrt{TE_{48}/TE_0}$)
  - ENKF-20,-128: 48h TLM forecasts initialized with 20 or 128 members from analysis ensembles
  - Pa-RAND-20,-128: 48h TLM forecasts initialized with 20 or 128 random perturbations drawn from Gaussian distribution with perturbed 3d-var analysis error cov (homogeneous/isotropic corr)
  - ENKF-NOLOC: 20 SVs from non-localized EnKF $\mathbf{P}_a$ (orthogonal linear combinations of EnKF members)




- fraction of 48-hour NH forecast error energy explained and total fcterr



# Conclusions

1. large impact on spatial and multi-variate structure of initial-time SVs and location of most sensitive regions from using analysis error covariance norms, and
2. for the case examined, SVs using stationary $\mathbf{P}_a^{-1}$ norm better predict sensitive region than TESV (i.e. more consistent with pseudo-inverse of fcterr)
3. using different SV norms results in small impact on final-time SVs and on ability to explain 48-hour forecast error
4. EnKF analysis ensemble members and random pertubations from stationary analysis error distribution explain similar fraction of 48-hour forecast error
5. 20 ensemble members (or random perturbations) explain about half the amount of forecast error as 20 SVs, but 128 members explain about 50%-100% more

# Future Directions

- apply standard ensemble prediction diagnostics to evaluate SVs, EnKF and random perturbations (diagnostics that account for spread)
- perform longer ensemble predictions with nonlinear model
- explore sensitivity to ensemble size for SVs, EnKF and random perturbations (determine relative cost of each approach for equivalent quality)
- explore application of SVs to observation targeting (A. Zadra) and application of analysis error covariance norms to Key Error Analysis (S. Laroche)

The End