1. Preconditioning with Hessian Eigenvectors

2. Singular Vectors: Diagnostic Tool and Covariance Propagation

Mark Buehner

(with help from M. Tanguay, P. Gauthier, S. Pellerin, S. Laroche)

Meteorological Service of Canada
Dorval, Quebec

Outline of Talk

Background:

- Lanczos-type eigenvalue solvers compute leading eigenvalues/vectors of a large matrix $\mathbf{L}$ requiring only: $\mathbf{w} = \mathbf{L}\mathbf{v}$, for arbitrary $\mathbf{v}$
- ARPACK is a freely available package of this type

1. Preconditioning with HEVs:

- 3d/4d-var cost function
- estimation of the Hessian
- results from preconditioning 3d/4d-var

2. Singular Vectors:

- formulation
- identifying numerical instabilities
- structure of TESVs, BGSVs, HSVs
- covariance propagation with HSVs
- RRKF and time-averaged SVs

Cost Function for 3d/4d-var

- 3d/4d-var produces analysis increment by minimizing:

$$
\begin{aligned}
J &= J_{bg} + J_o \\
&= \Delta\mathbf{x}^T \mathbf{B}^{-1} \Delta\mathbf{x} + \left(\mathbf{H}\Delta\mathbf{x} - \mathbf{y}'\right)^T \mathbf{R}^{-1} \left(\mathbf{H}\Delta\mathbf{x} - \mathbf{y}'\right)
\end{aligned}
$$

where $\mathbf{y}' = \mathbf{y} - H\left(\mathbf{x}^{bg}\right)$, and

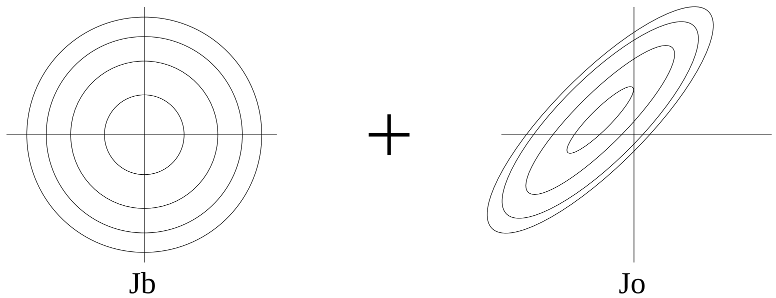
$$
\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^{bg} = [\Delta\psi, \Delta\chi, \Delta T, \Delta P_s, \Delta \ln(q)]
$$

- Note: $\mathbf{H}$ also involves integration of the TLM of GEM for 4d-var
- $\mathbf{B}$ spreads information from observations spatially according to covariance structure
- current $\mathbf{B}$: stationary, correlations horizontally homog./isotropic, variances zonally invariant
- between-variable covariances governed by balance operators (geostrophic and Ekman balances)
- variances, univariate correlations, and balance operators estimated from ensemble of lagged forecast differences (NMC method)

Preconditioning with HEVs

- to improve efficiency J is rewritten in terms of the control vector ξ, where $\Delta \mathbf{x} = \mathbf{B}^{1/2}\xi$:

$$J = \xi^T \xi + \left(\mathbf{H}\mathbf{B}^{1/2}\xi - \mathbf{y}'\right)^T \mathbf{R}^{-1} \left(\mathbf{H}\mathbf{B}^{1/2}\xi - \mathbf{y}'\right)$$



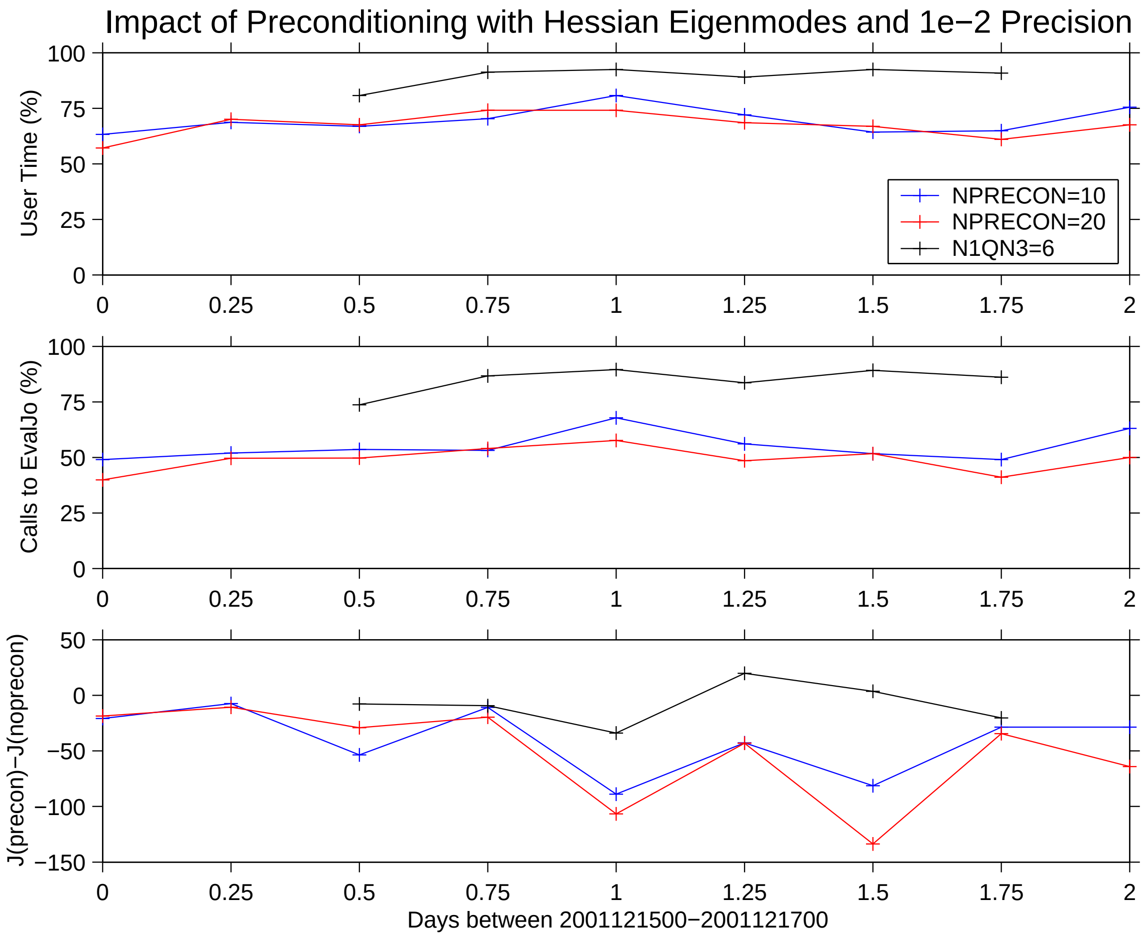
- J possibly still poorly conditioned due to J_o
- solution: use (approximate) Hessian to define norm for minimizer
- Hessian (and inverse analysis error cov) given by:

$$\mathbf{J}_{\xi\xi} = \mathbf{A}_{\xi}^{-1} = \mathbf{I} + \mathbf{B}^{T/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$$

- obtain truncated eigen-decomp of second term (Lanczos algorithm): extracts dominant effect of observations on Hessian (and $\mathbf{A}$)
- for 3d-var: $\mathbf{B}$, $\mathbf{H}$ and $\mathbf{R}$ nearly constant, therefore eigenvectors can be re-used (differences in $\mathbf{H}$ between 0/12UTC and 6/18UTC)
- for 4d-var: dependence of TLM/ADJ on background trajectory imposes more time-dependence

Preconditioning 3d/4d-var

- 3d-var: eigenvectors computed from outside period at 6:00UTC and held fixed

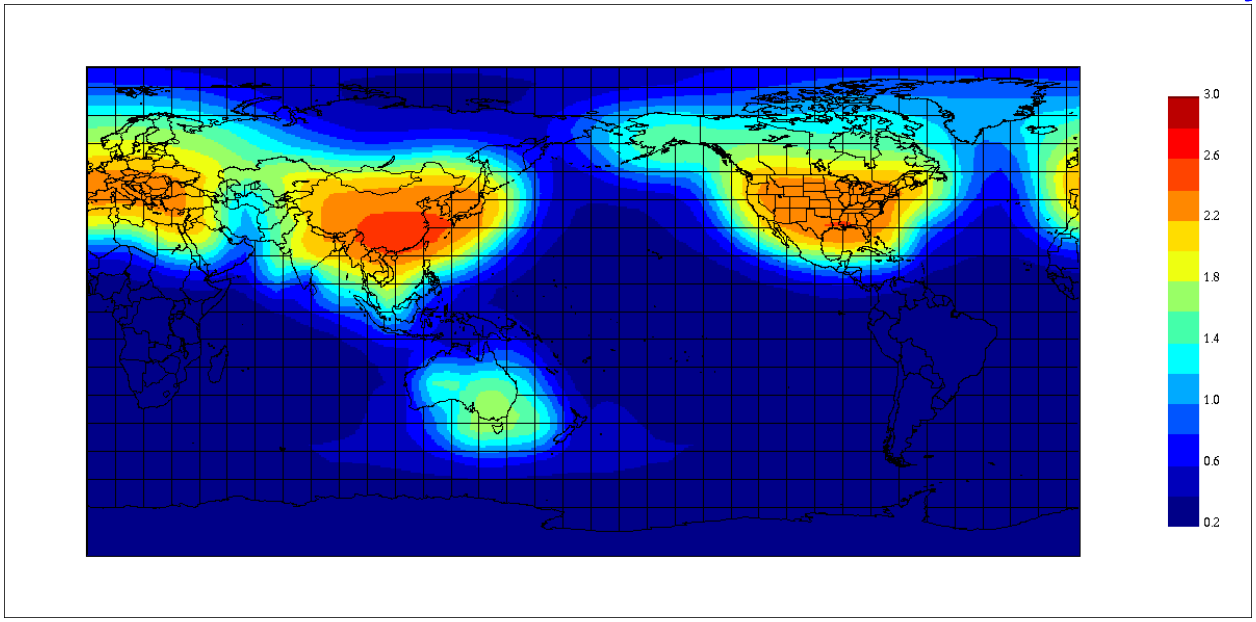


- preliminary result with 4d-var, eigenvectors calculated from first time only:

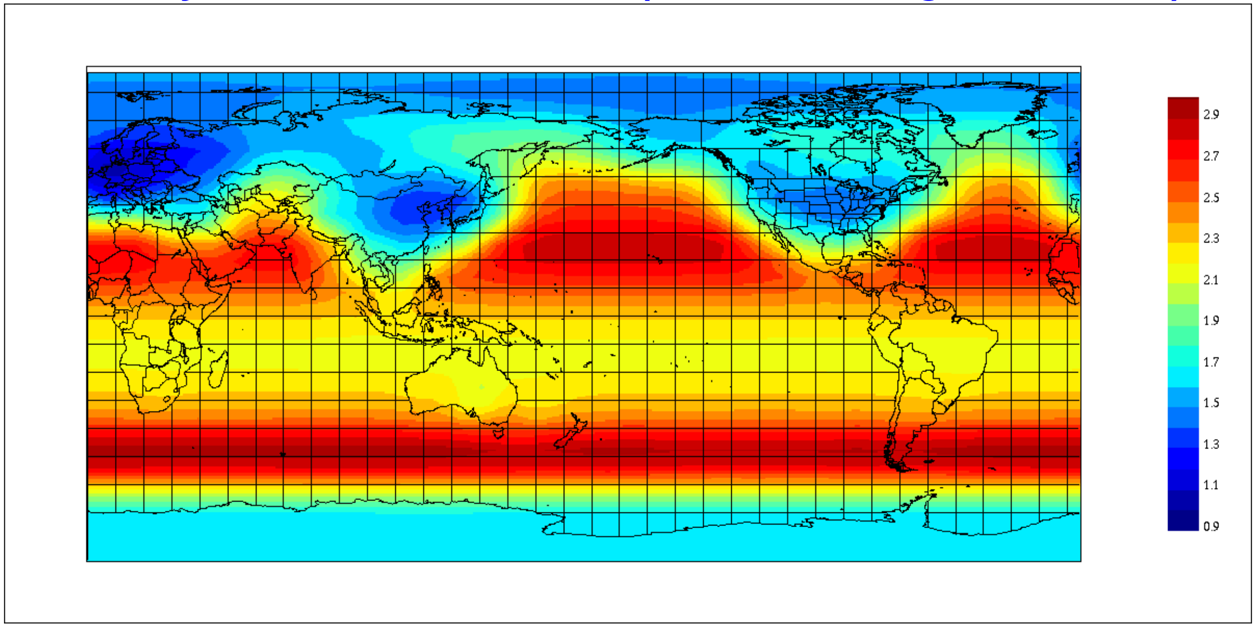
date	*NHEV*	*#iterations*	*User Time*
2001020712	0	131	5943
2001020800	0	139	6565
2001020712	10	92	4096
2001020800	10	94	4400

Estimation of $\mathbf{A}$: ψ, 250 hPa

SQRT of Correction to BG Variances from Radiosonde data only:



Approximate Analysis Error Std.Dev. (from 300 eigenvectors):



Singular Vectors

Background:

- compute perturbations that will optimally grow (linear dynamics) over specified period (t) w.r.t. specified norms
- currently used at ECMWF to define subspace for RRKF and to generate initial perturbations for EPS
- useful for identifying numerical instabilities in the model

Formulation:

- maximize ratio: $\dfrac{\mathbf{x}_t^T \mathbf{W}_t \mathbf{x}_t}{\mathbf{x}_0^T \mathbf{W}_0 \mathbf{x}_0}$, where $\mathbf{x}_t = \mathbf{M}\mathbf{x}_0$
- reformulate in terms of control variable γ, where $\mathbf{x}_0 = \mathbf{W}_0^{-1/2}\gamma$
- results in eigenvector problem: $\boxed{\mathbf{W}_0^{-T/2}\mathbf{M}^T\mathbf{W}_t\mathbf{M}\mathbf{W}_0^{-1/2}\gamma_k = \lambda_k^2\gamma_k}$
- final time norm typically TE, at initial time: TE (TESV), $\mathbf{A}^{-1}$ (HSV), or $\mathbf{B}^{-1}$ (BGSV)
- HSV: only by using approximation to $\mathbf{A}^{-1}$ can $\mathbf{W}_0^{-1/2}$ be calculated (at ECMWF full Hessian used, resulting in a more expensive generalized eigenvalue problem)

Experimental Setup

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

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

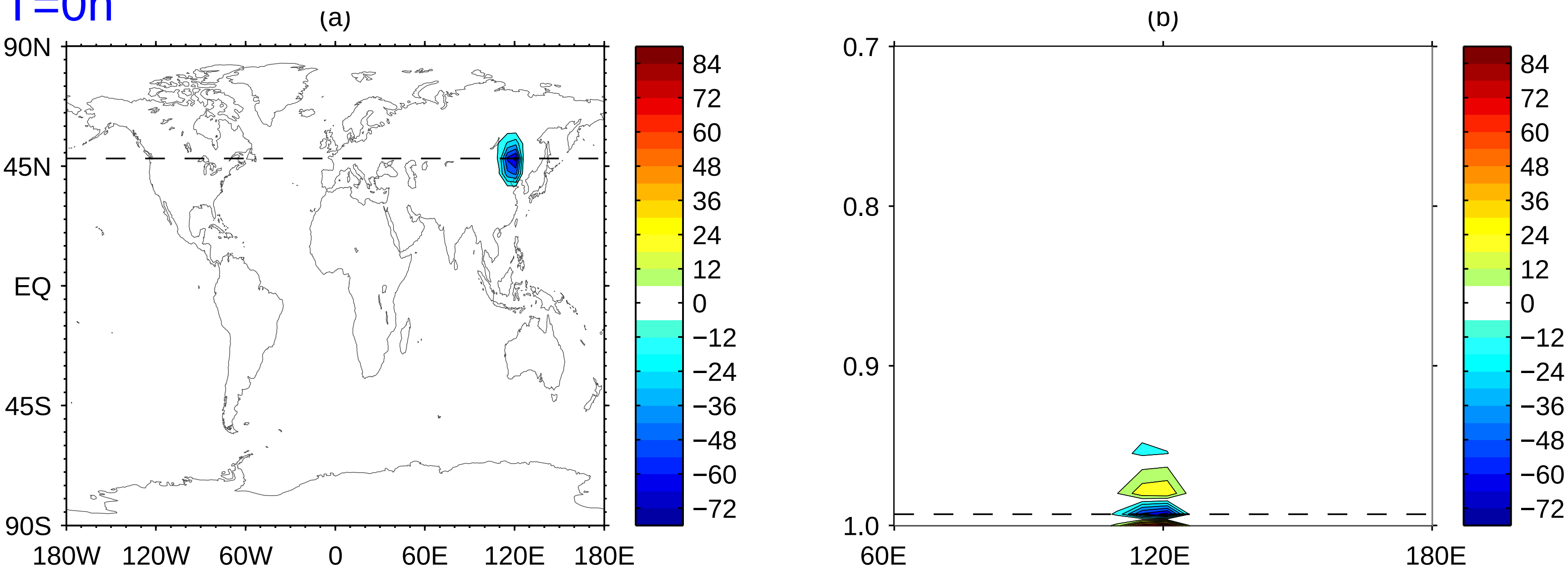
- experiments at low resolution (64x32x28L) with simplified vert diffusion (S. Laroche) and no humidity
- final time norm is TE north of 30°N
- different initial time (global) norms used: TESV, HSV, BGSV
- $\mathbf{A}^{-1}$ estimated using 300 HEVs calculated from 3d-var cost function and full set of obs
- computational cost per SV: approximately 3 times cost of TLM+ADJ integration

Results: Computational Modes in TESVs

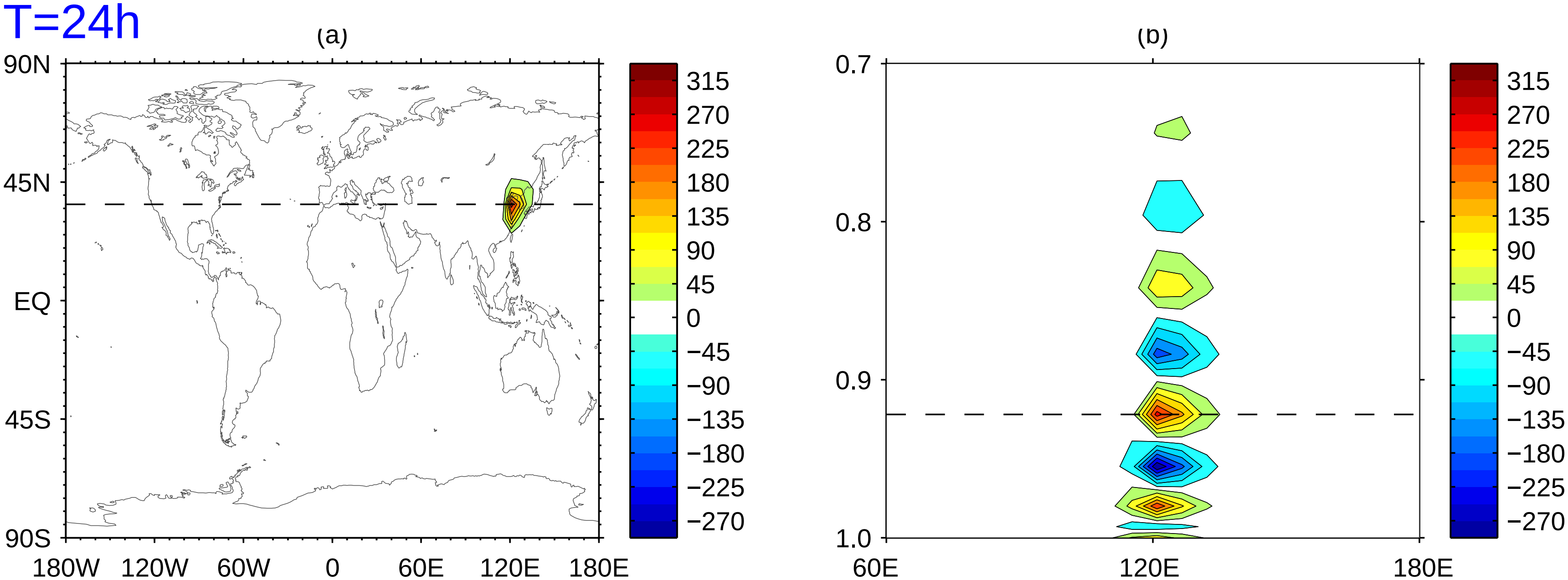
- unlike HSV and BGSV, no balance or smoothness constraints imposed by $\mathbf{W}_0$
- quickly highlights numerical instabilities related to model numerics:
  - with no vertical diffusion: spurious growth at sfc
  - then, modes at pole dominate
  - after applying smoothness constraint or sponge at pole, modes at the top appear
  - after applying sponge at the top or zeroing a few levels at t=0h, localized modes near the jet appear (similar to noise seen by M. Roch?)
- also grow in nonlinear model (lo-res with simplified physics)
- more realistic modes with sponge near pole and zeroing U,V,T down to 200 hPa at the initial time
- further experiments required to determine:
  - if modes grow in model at full-res with full physics and if so, what is their importance generally and for 4d-var
  - best way to eliminate computational modes (at ECMWF also zero values above level 6 at initial time and use enhanced horizontal diffusion in TL/AD and possibly smaller timestep)

No Vertical Diffusion: TESV structure (T)

T=0h

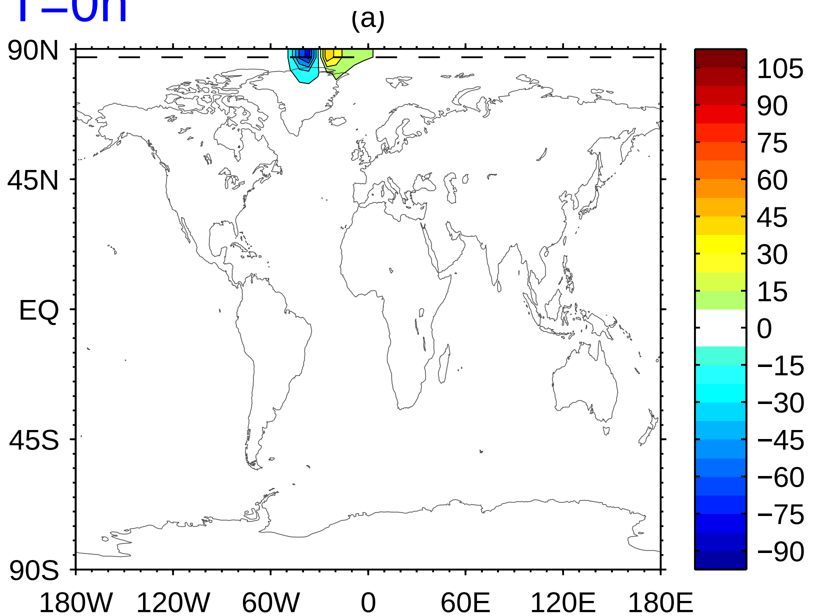


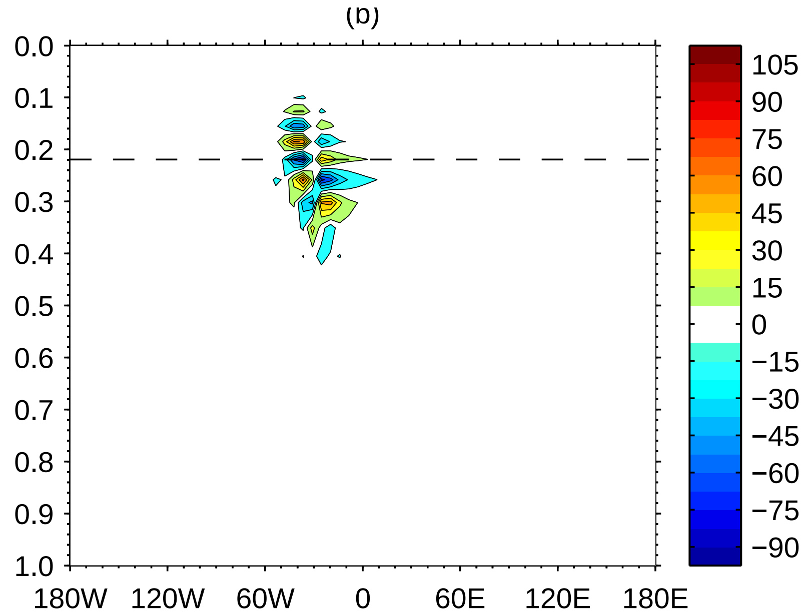
T=24h



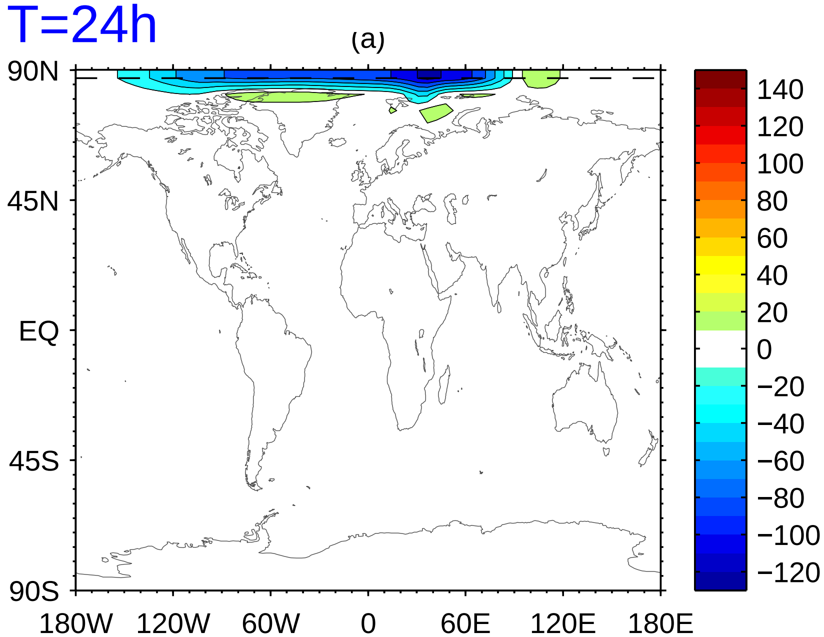
Pole Problem: TESV structure (T)

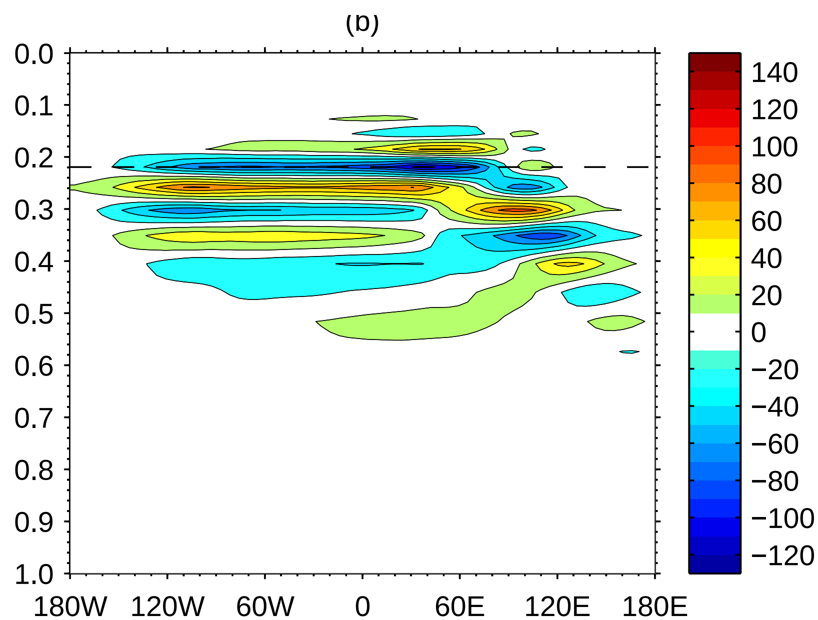
T=0h





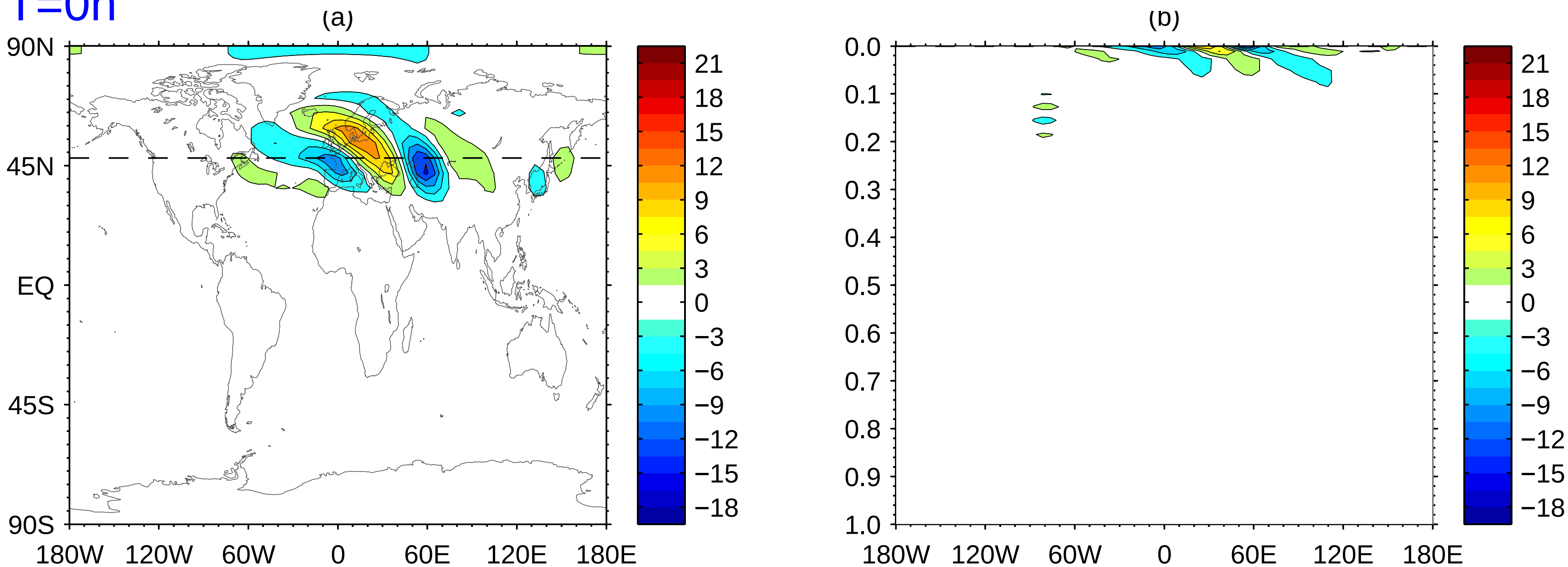
T=24h



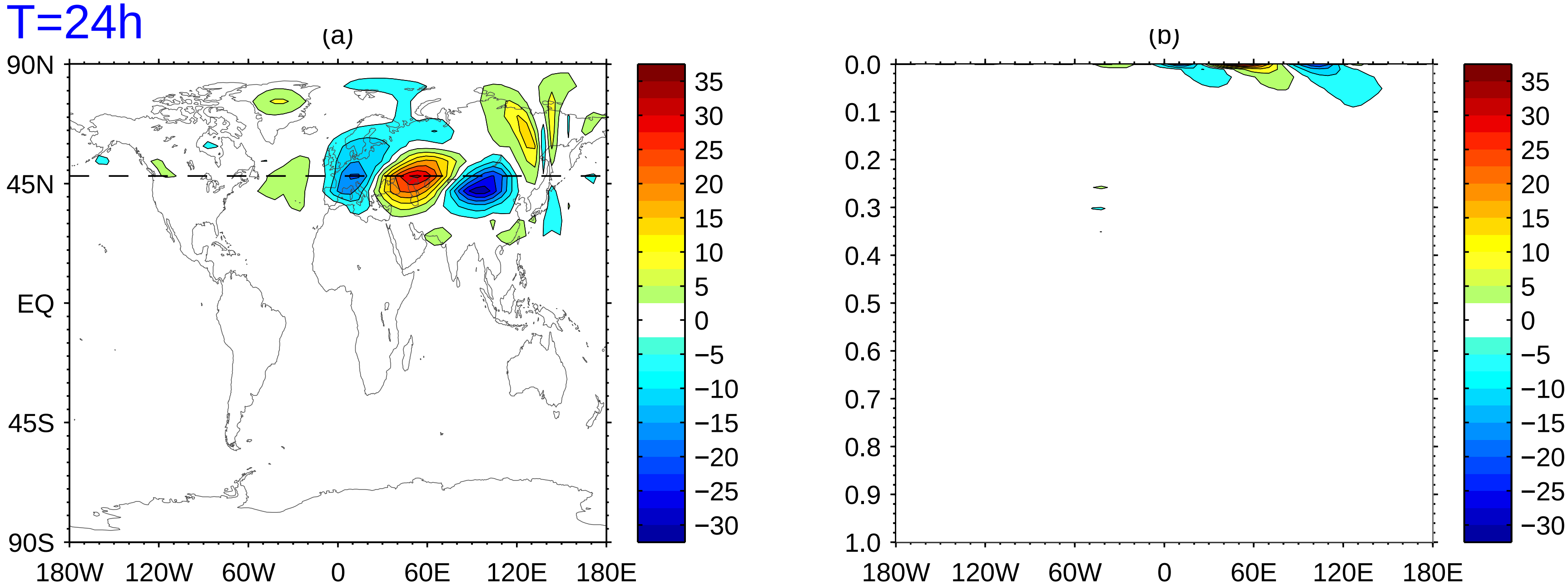


Top Problem (mode 23): TESV structure (T)

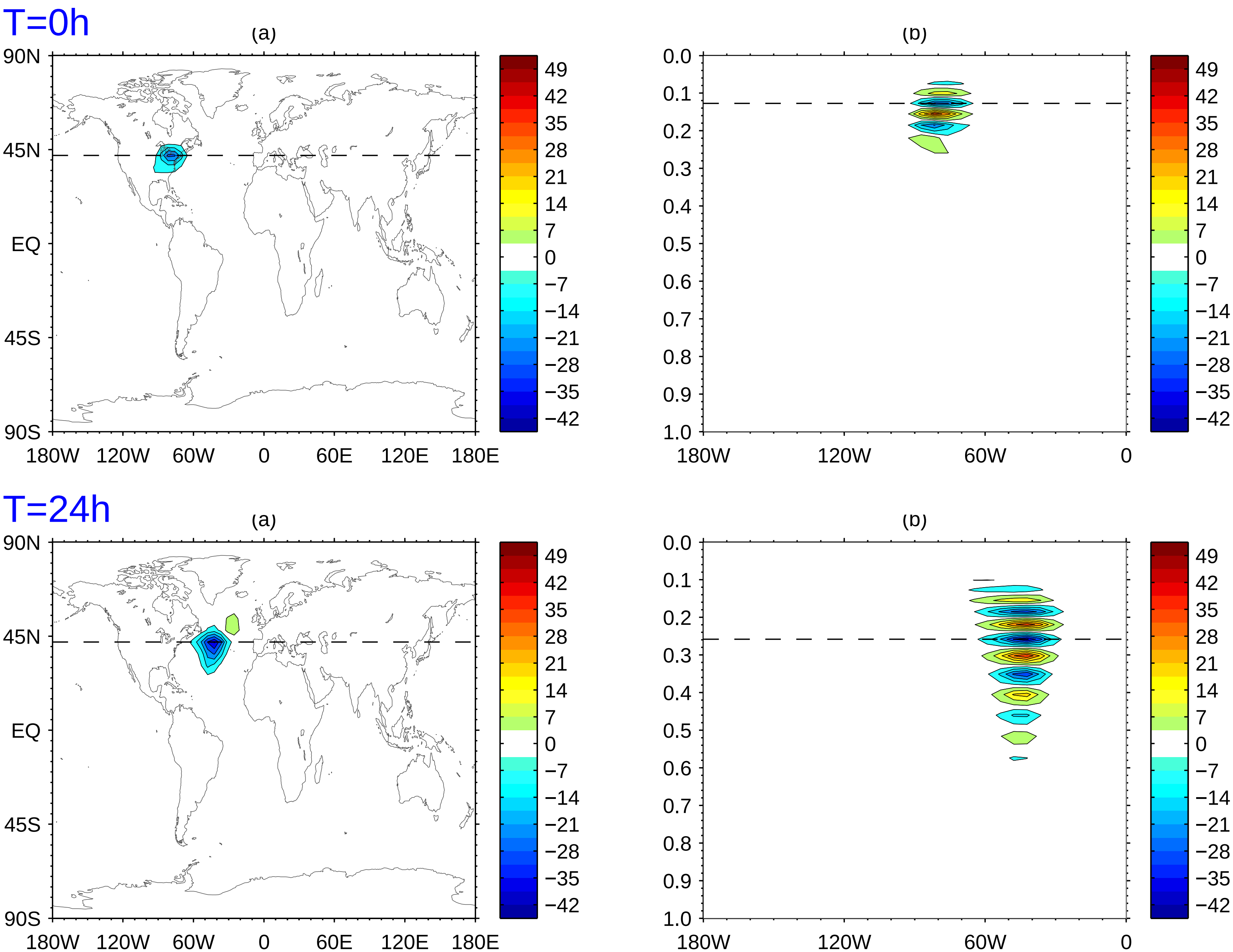
T=0h



T=24h



Near the jet (mode 21): TESV structure (T)



Meteorological Modes Using 3 Norms

At initial time:

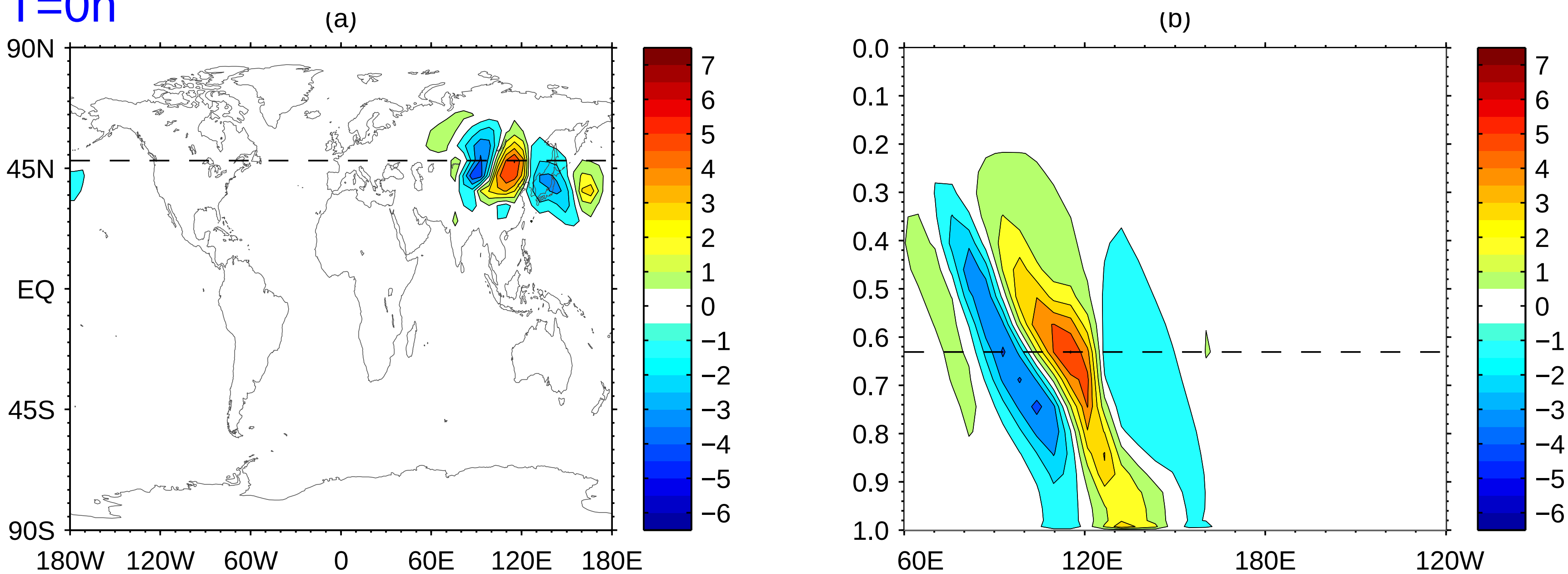
- structures tilted vertically against the shear
- TESV generally sharper hor and vert structure and higher growth rates than BGSV or HSV
- TESV: KE concentrated around 800 hPa, BGSV and HSV: around 650 hPa
- TESV: temperature relatively more dominant at initial time than for BGSV and HSV
- leading TESV and BGSV in similar location, HSV over oceans (high anl error)

At optimization time (48h):

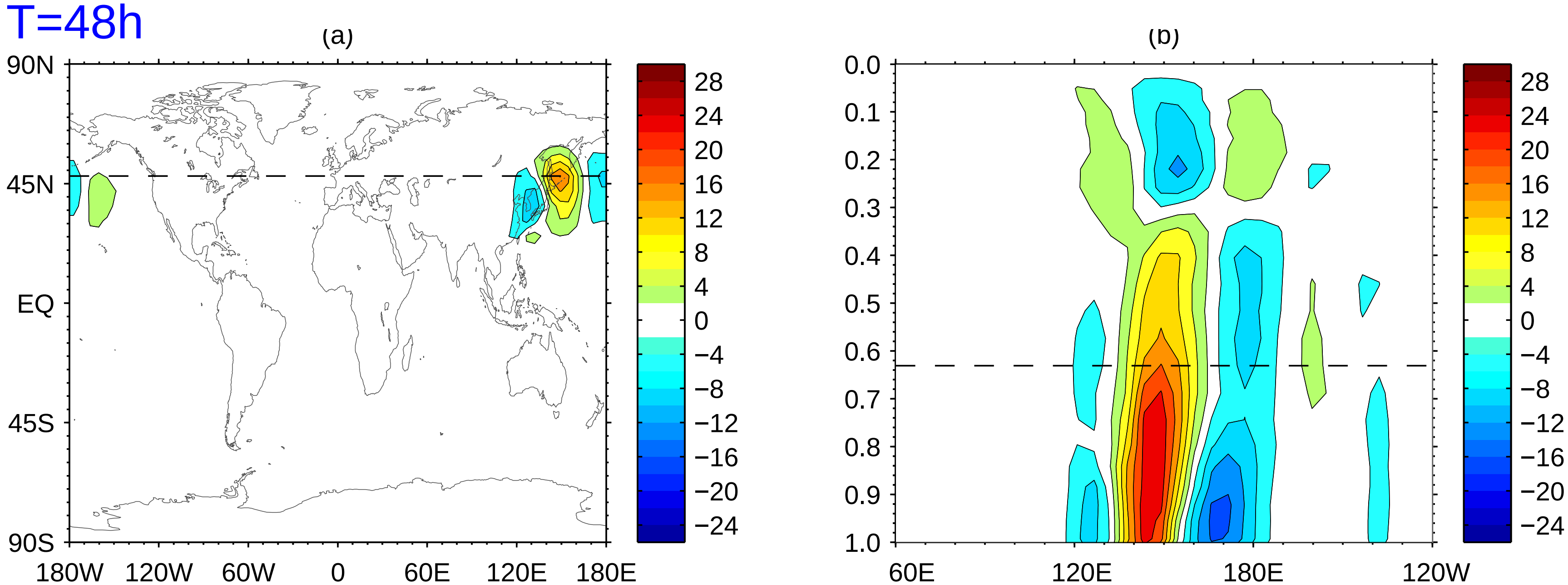
- tilt removed (especially for T)
- leading TESV and BGSV very similar
- all norms give similar structure, KE concentrated around 300 hPa
- nonlinearly evolved SVs very similar structure (no scaling)
- amplification factors lower than other studies (case dependent)

Results: TESV structure (T)

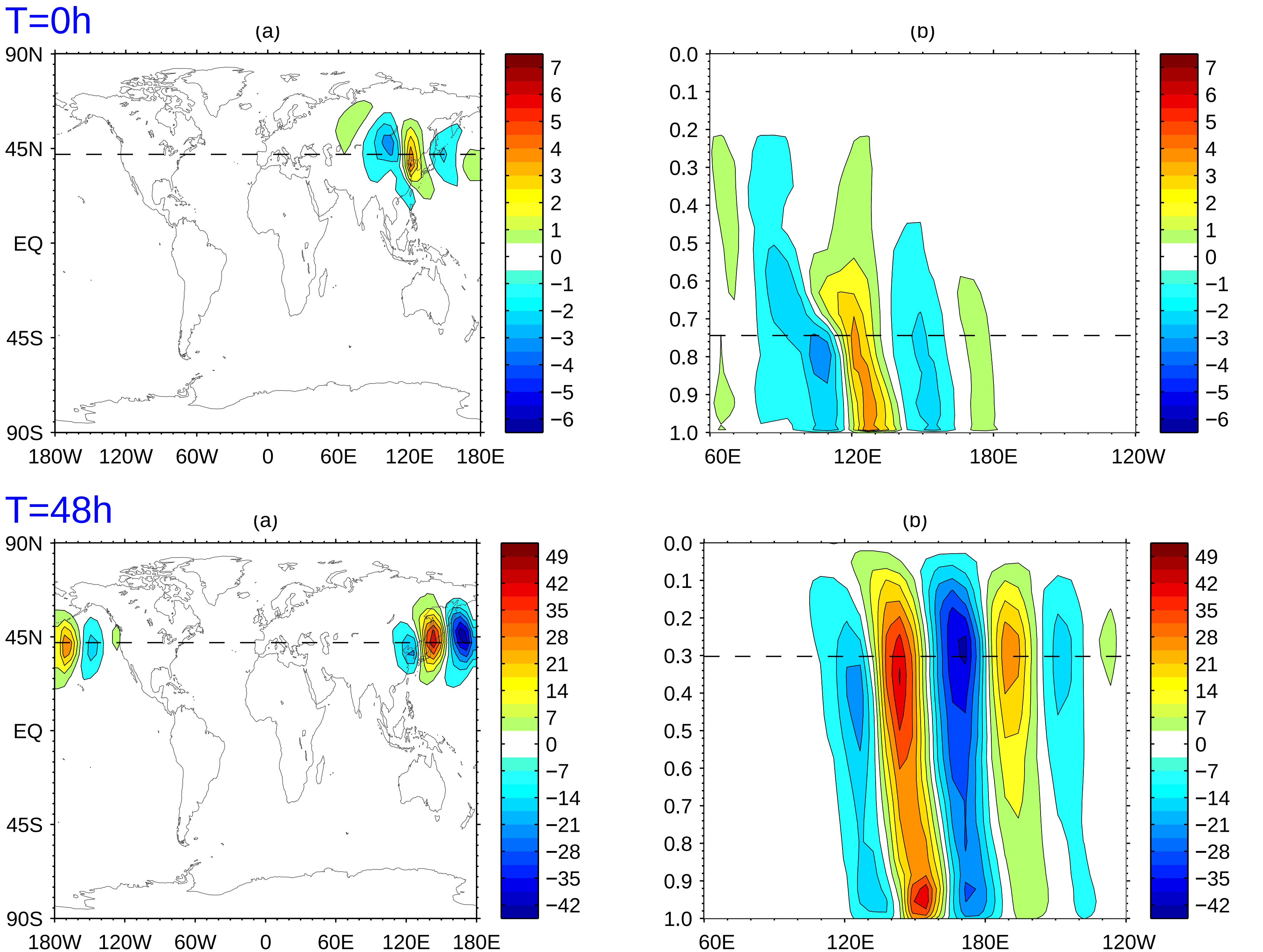
T=0h



T=48h

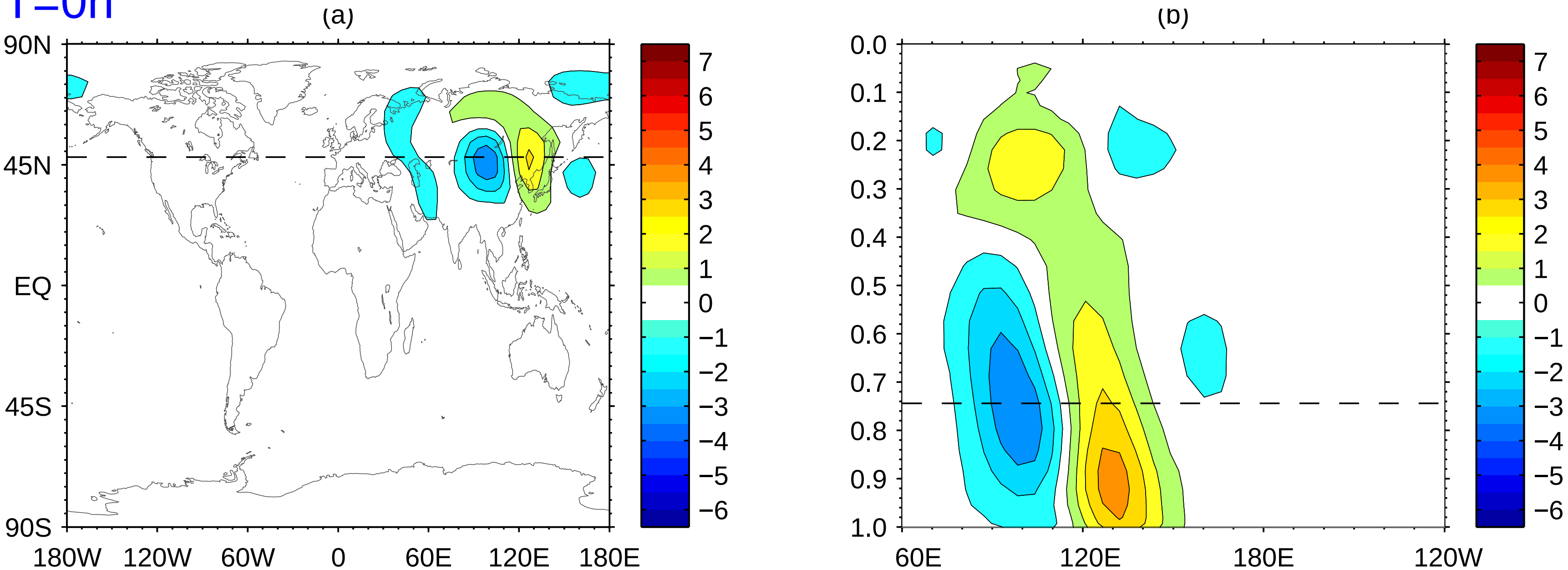


Results: TESV structure (V)

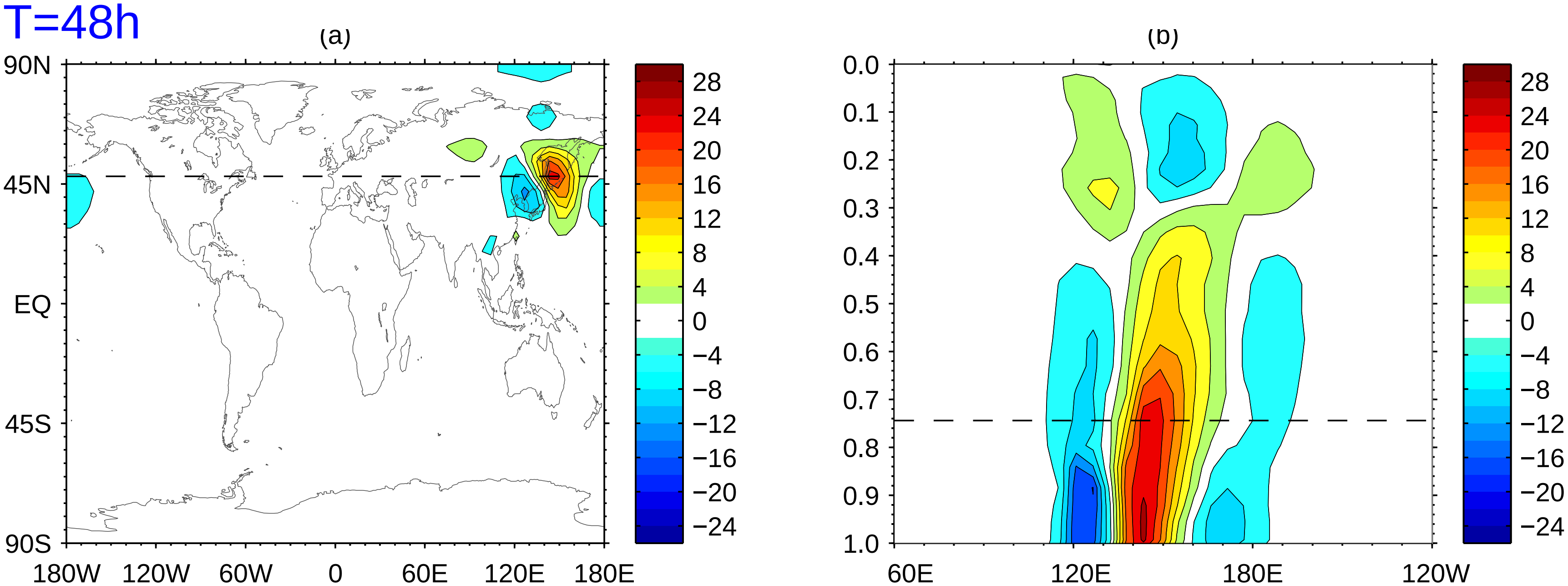


Results: BGSV structure (T)

T=0h

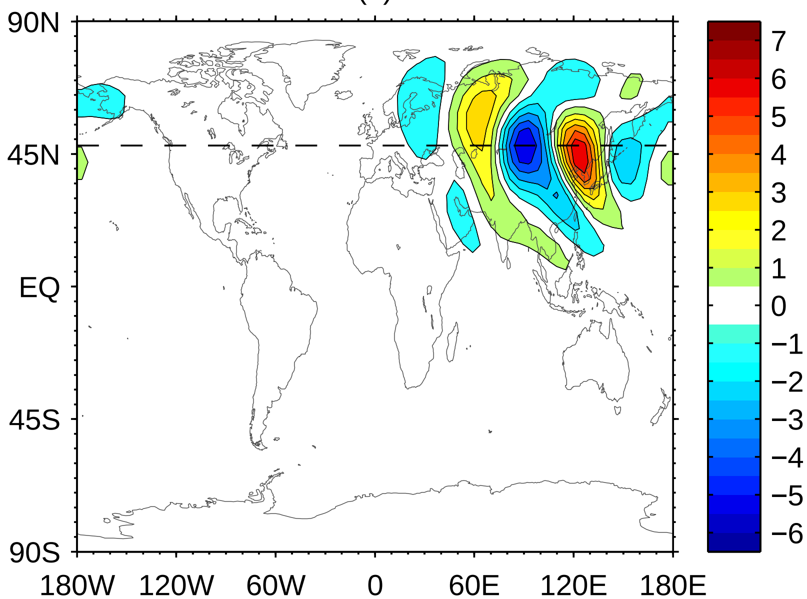


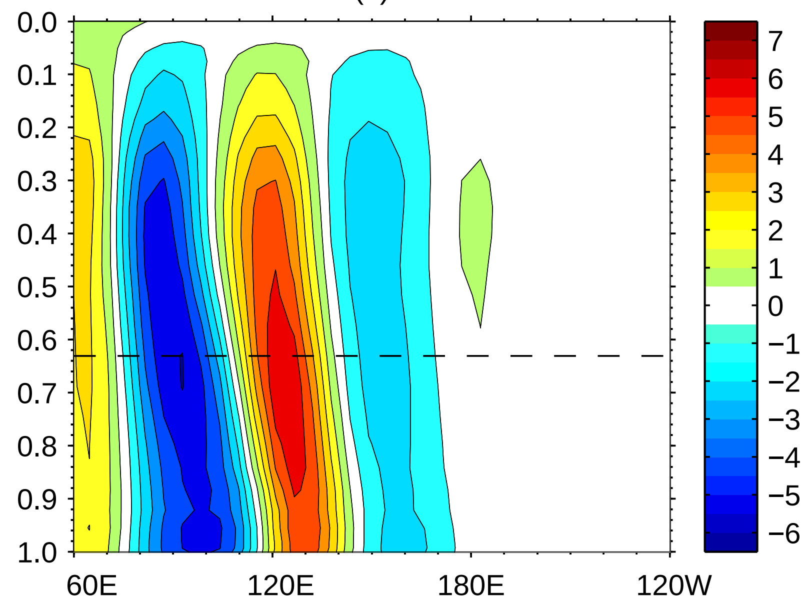
T=48h



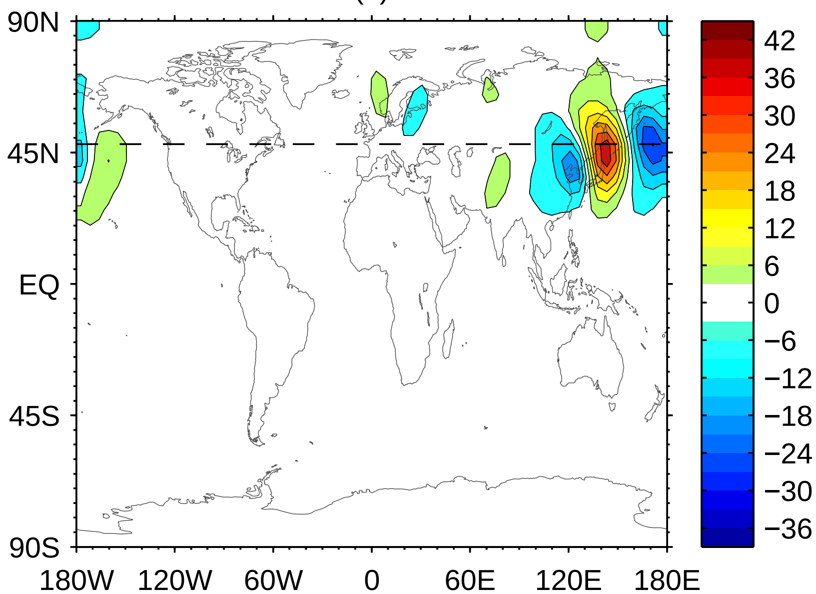
Results: BGSV structure (V)

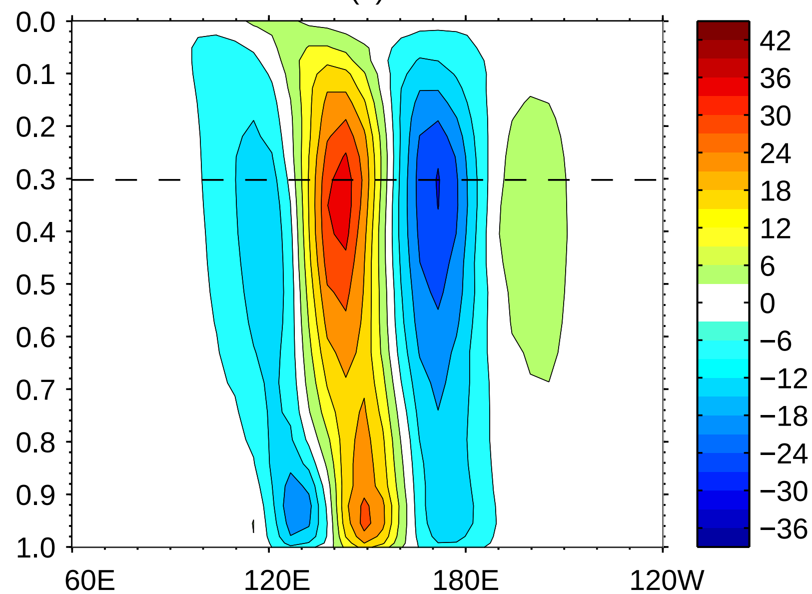
T=0h





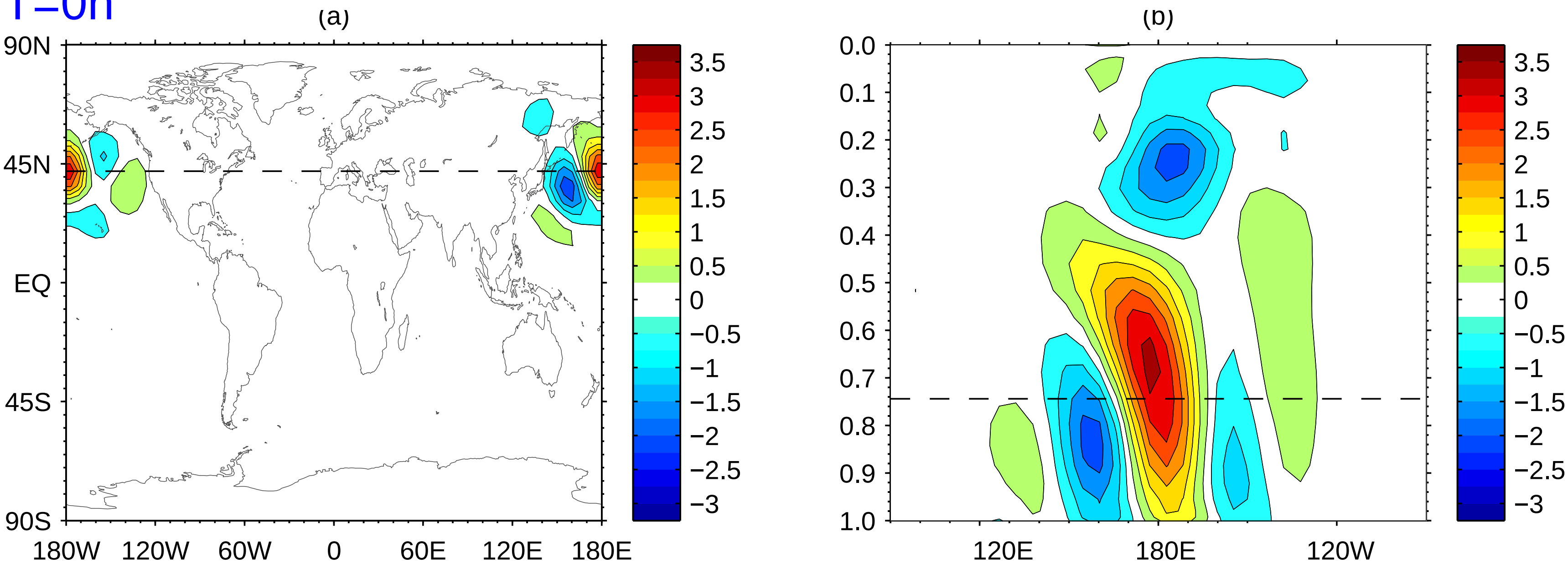
T=48h



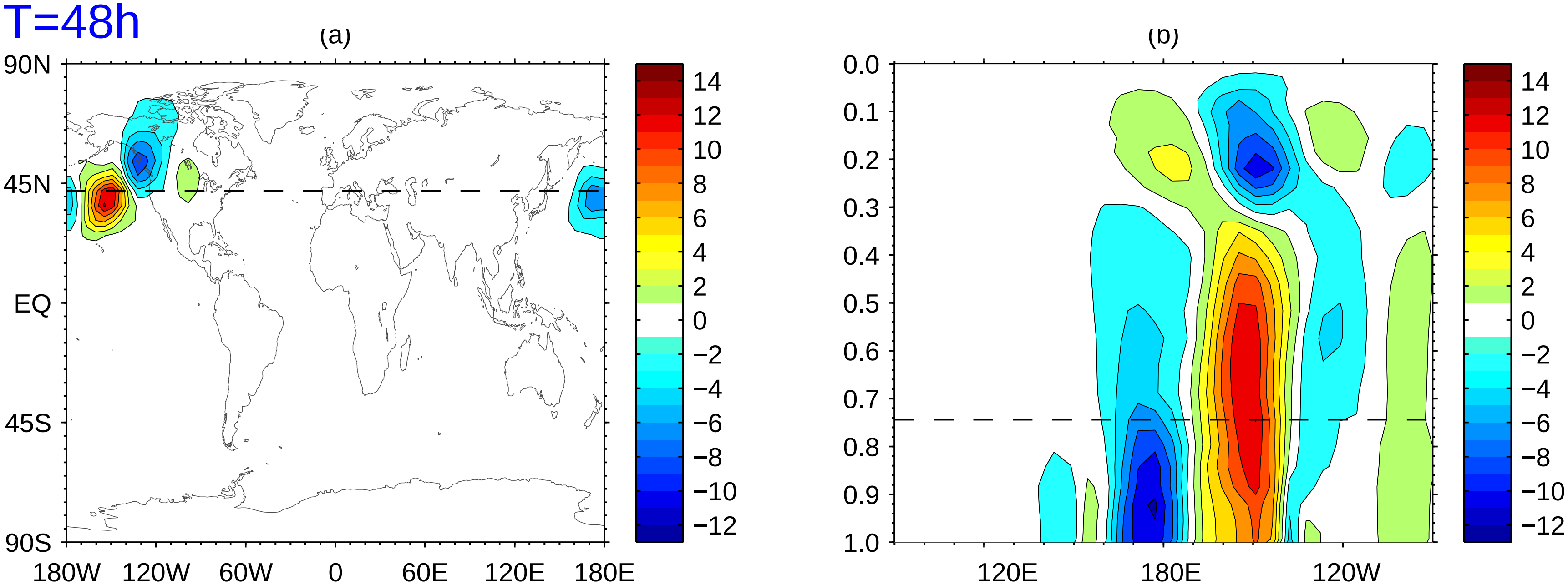


Results: HSV structure (T)

T=0h

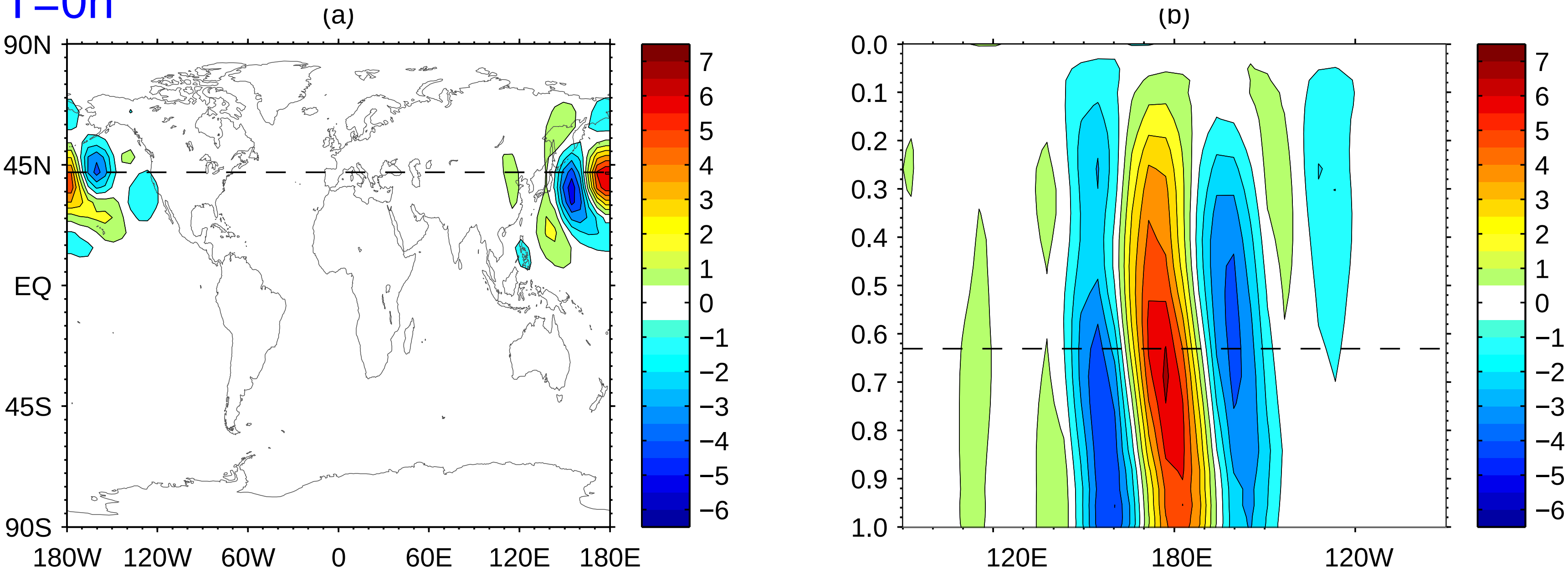


T=48h

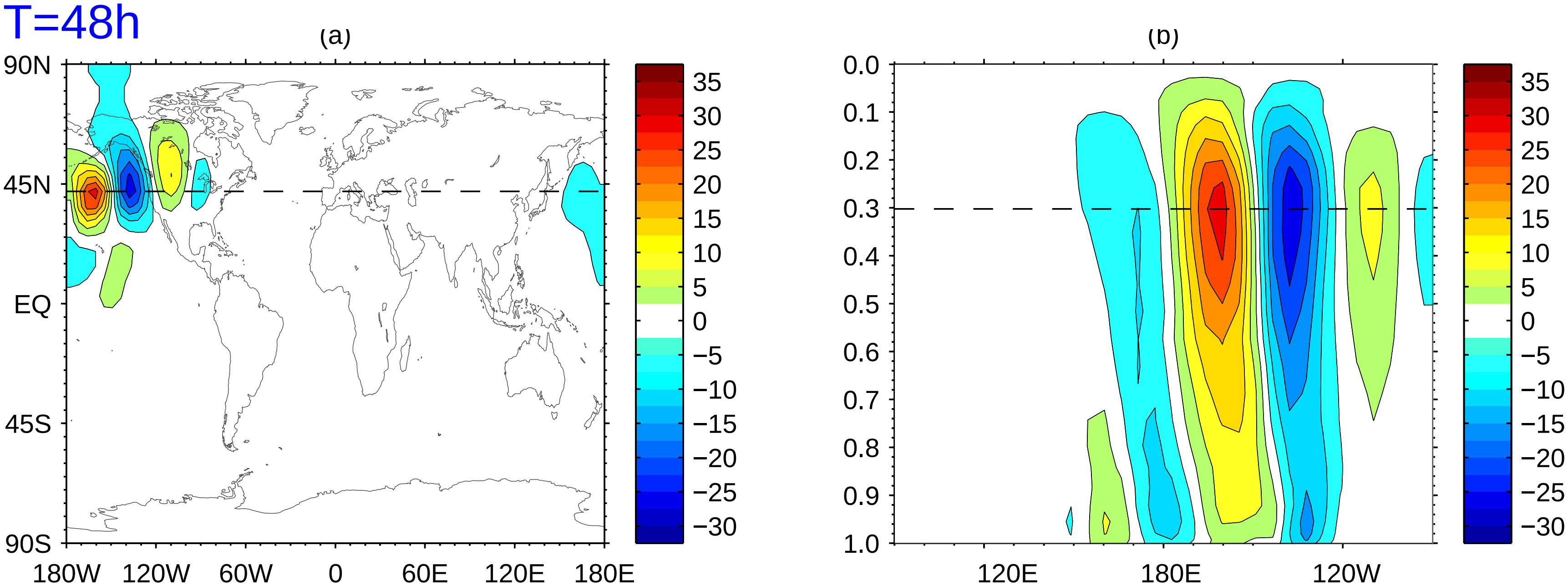


Results: HSV structure (V)

T=0h

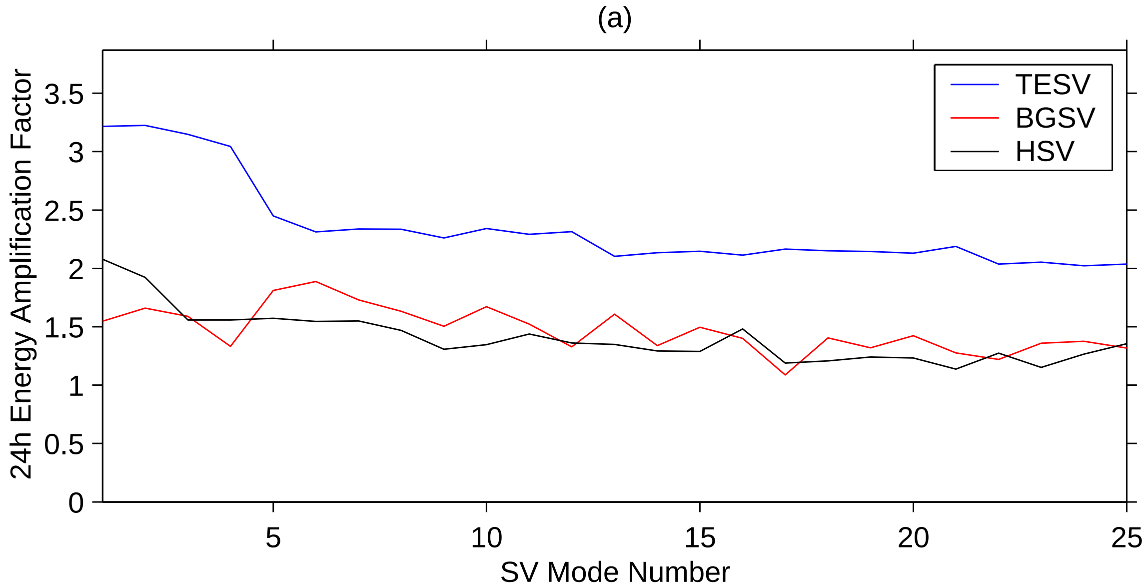


T=48h

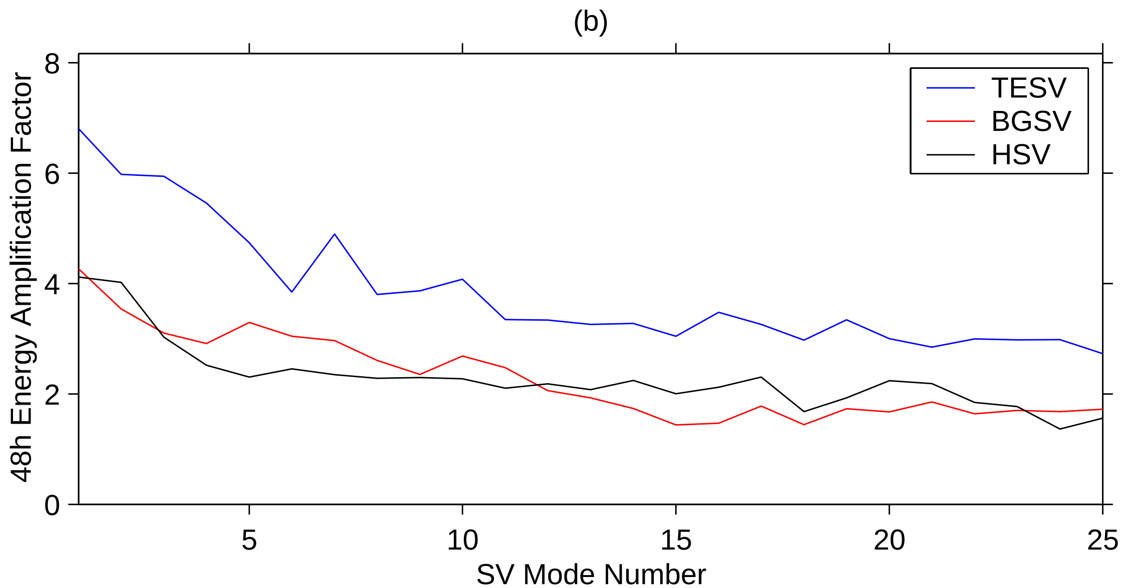


Results: Amplification Factors for Different Norms/Optimization Times

t=24h



t=48h



Time-Averaged SVs

- limitation of SVs: dependent on chosen optimization time, may be significantly sub-optimal at other times (nonmodal growth)

- instead, maximize the average growth over multiple optimization times, e.g.:

$$\frac{\mathbf{x}_1^T \mathbf{W}_1 \mathbf{x}_1 + \mathbf{x}_2^T \mathbf{W}_2 \mathbf{x}_2}{\mathbf{x}_0^T \mathbf{W}_0 \mathbf{x}_0}$$, where $\mathbf{x}_1 = \mathbf{M}_1 \mathbf{x}_0$, and $\mathbf{x}_2 = \mathbf{M}_2 \mathbf{x}_1$

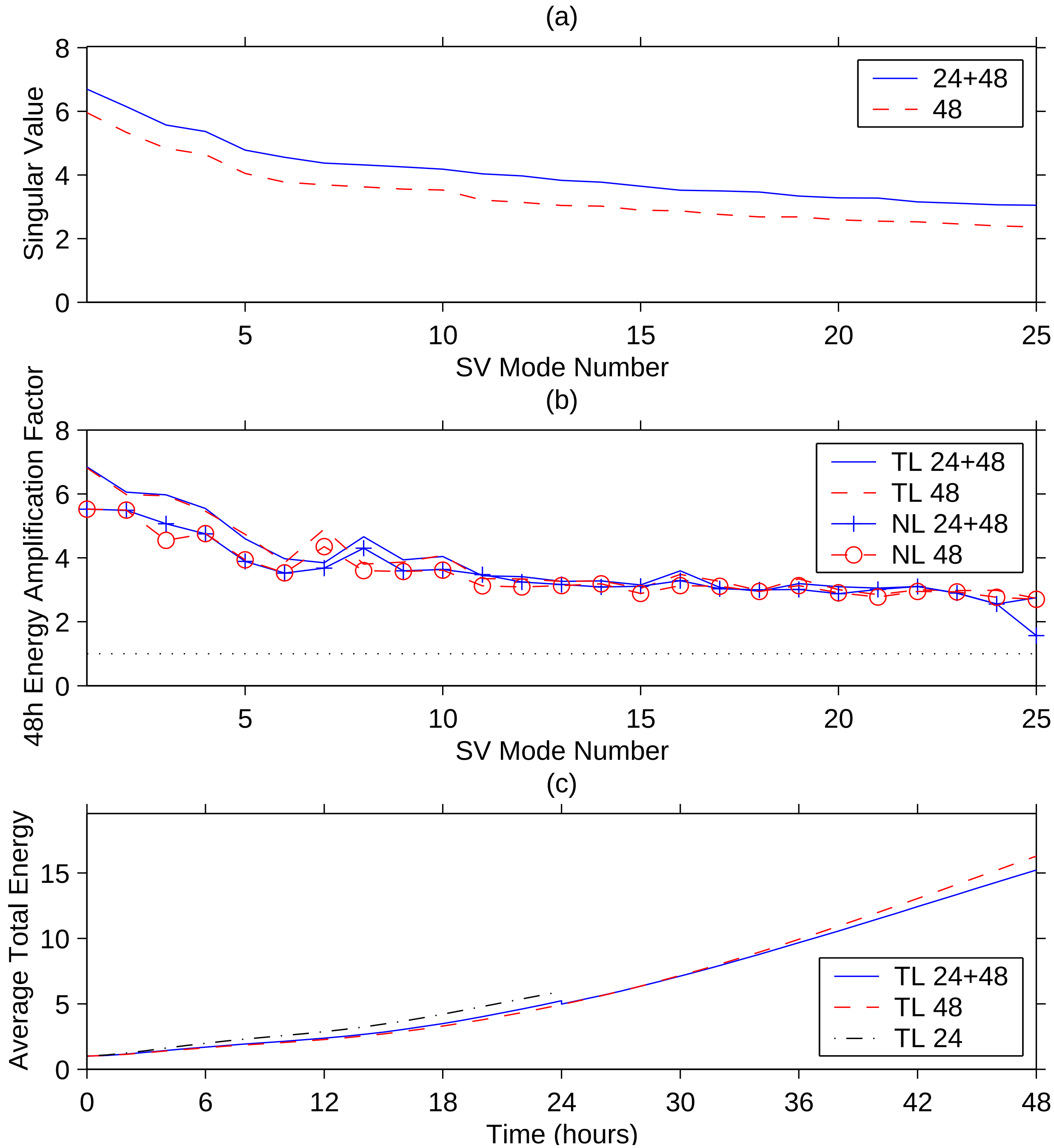
- leads to eigenvalue problem:

$$\boxed{\mathbf{W}_0^{-T/2} \mathbf{M}_1^T \left(\mathbf{W}_1 + \mathbf{M}_2^T \mathbf{W}_2 \mathbf{M}_2 \right) \mathbf{M}_1 \mathbf{W}_0^{-1/2} \gamma_k = \lambda_k^2 \gamma_k}$$

- almost same computational cost

- different weighting of term at each time may be desirable

Growth of TESVs: 48h vs. 24h+48h

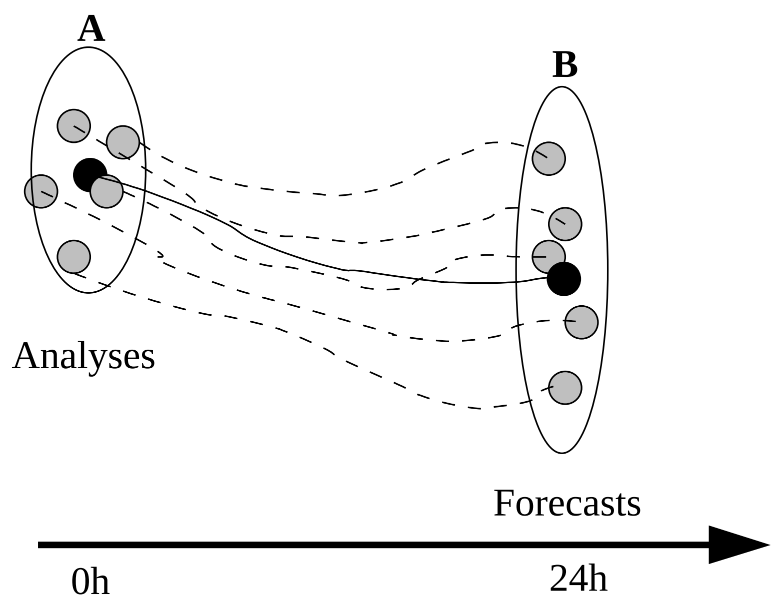


Covariance Propagation with HSVs

- initial time HSVs evolve into (scaled) leading eigenvectors of propagated covariances: $\mathbf{MAM}^T \approx \sum \mathbf{x}_t \mathbf{x}_t^T$ (KF: $\mathbf{P}^f = \mathbf{MAM}^T + \mathbf{Q}$)
- suggests use as basis functions in RRKF
- compare with covariances propagated with Monte-Carlo approach:

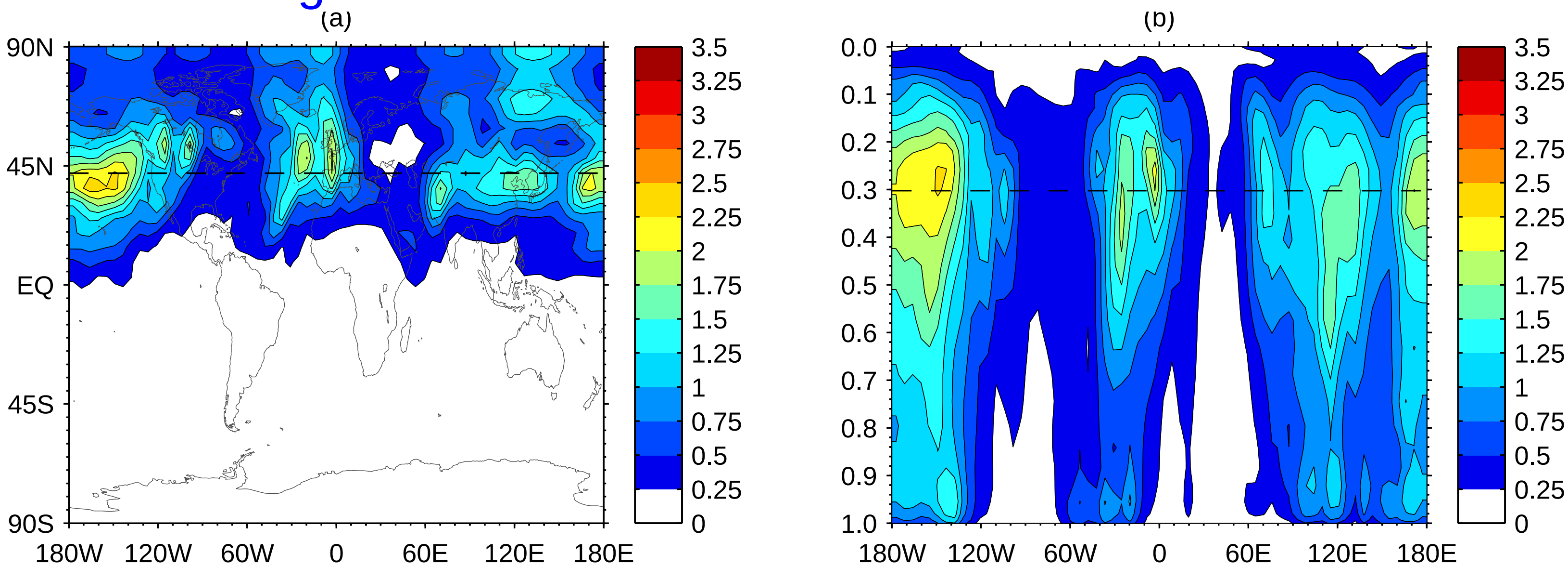
$$\mathbf{MAM}^T \approx \overline{\epsilon_a \epsilon_a^T}\text{, where } \epsilon_a = \mathbf{MA}^{1/2}\epsilon\text{, and } \epsilon \sim N(0, \mathbf{I})$$

- with 25 HSVs, variance appears to be underestimated (especially eastern NA, EUR) vs. 150 random perturbations (similar cost)

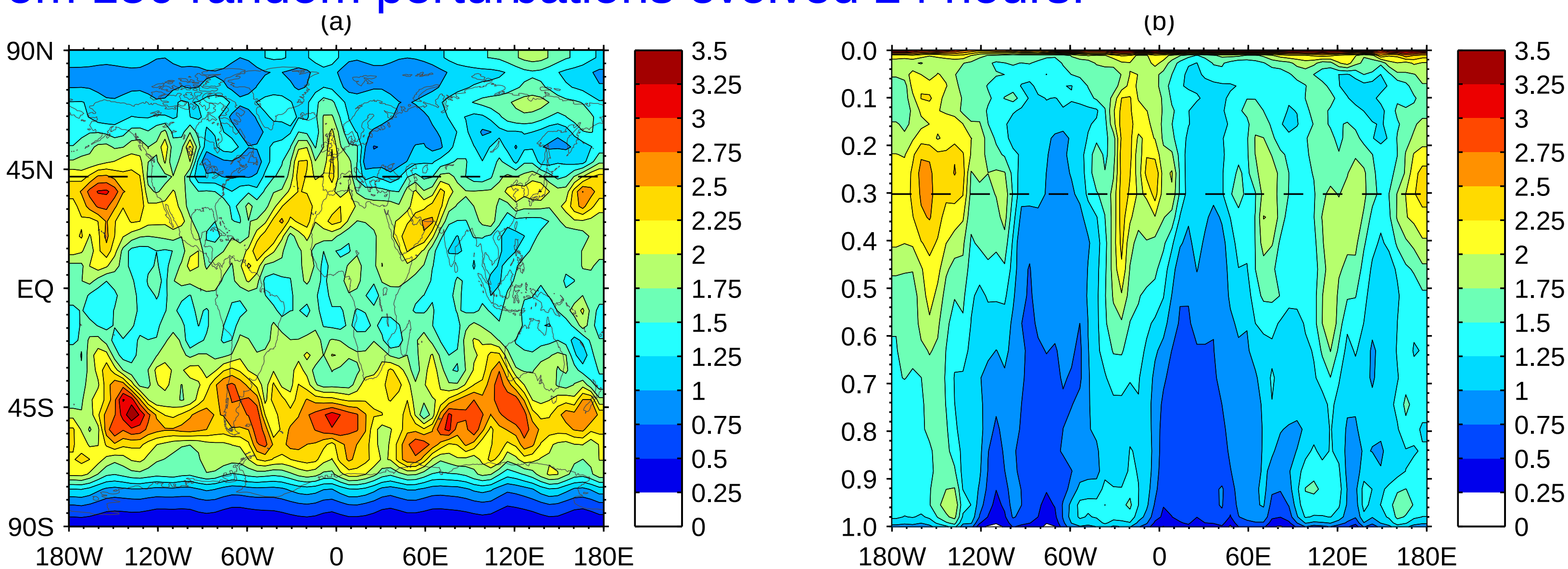


Standard Deviation of Propagated $\mathbf{A}$ (V)

From 25 leading HSVs evolved 24 hours:



From 150 random perturbations evolved 24 hours:



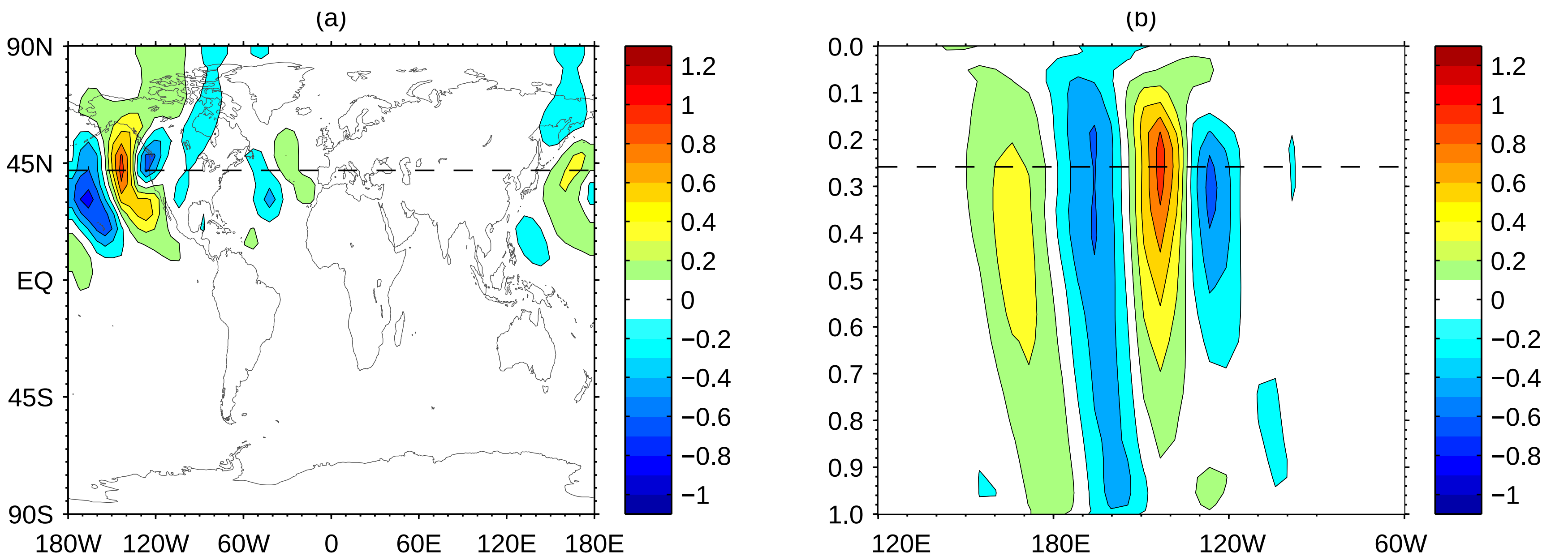
Flow-Dependent $\mathbf{B}$ in 3d-var

- use 25 leading HSVs partially evolved by 6 hours
- covariance matrix composed of two components:

$$\mathbf{B}_{sv} = \alpha_1 \mathbf{B}_{hi} + \alpha_2 \sum \mathbf{x}_{6h} \mathbf{x}_{6h}^T$$

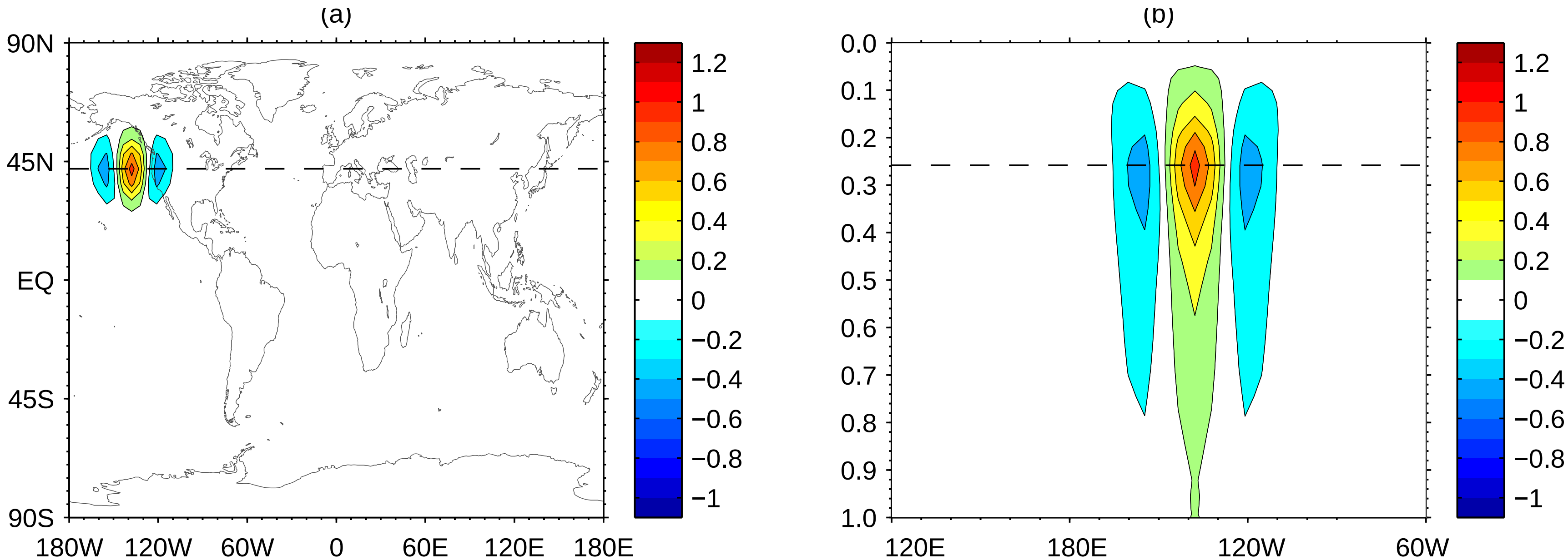
- similar to existing approaches (ENKF: Hamill and Snyder, HSV: Fisher, Bred mode: UKMet)
- 1-obs exp't in 3d-var shows spatial covariance structure of $\mathbf{B}_{sv}$ (V at 250 hPa, 40°N, 140°W):

Only SVs ($\alpha_1 = 0, \alpha_2 = 1$)

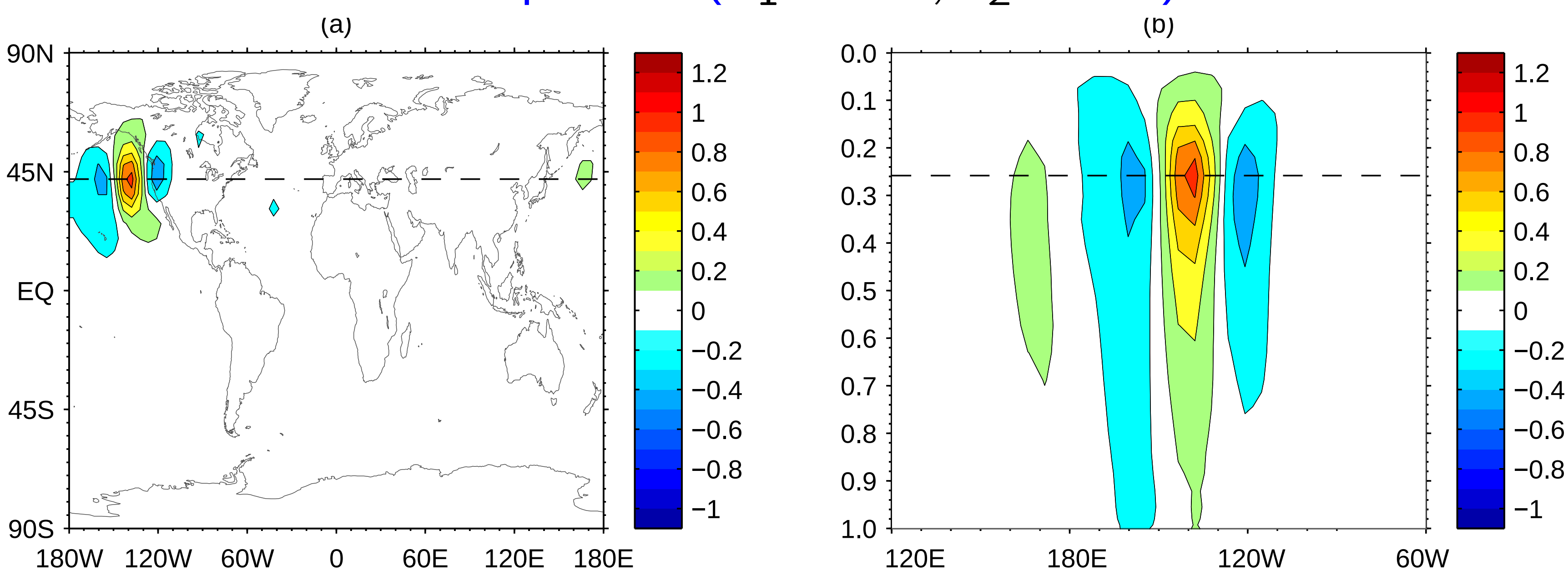


Flow-Dependent $\mathbf{B}$ in 3d-var

Only $\mathbf{B}_{hi}$ ($\alpha_1 = 1, \alpha_2 = 0$)



Combination of two components ($\alpha_1 = 0.1, \alpha_2 = 0.9$)



RRKF using HSVs

- lack of projection of evolved SVs with coincident initial time SVs (Gelaro et. al. 1998) prevents cycling of covariances with HSVs
- possible solution: use "balanced" combination of finite-time (or time-averaged) SVs and evolved SVs (Farrell and Ioannou 2001)
- "balancing" defines a single subspace that contains both the subspace in which the model is most sensitive and the optimal response sub-space $\rightarrow$ should improve overlap
- time-averaged SVs similar to Stochastic Optimals (B. Farrell) defined for <u>autonomous</u> systems as leading eigenvectors of:

$$\tilde{\mathbf{Q}} = \sum_{n=0}^{\infty} (\mathbf{M}^n)^T (\mathbf{M}^n) \Delta t$$

- FI (2001) showed effectiveness of using balanced truncation of $\mathbf{Q}$ and:

$$\tilde{\mathbf{P}} = \sum_{n=0}^{\infty} (\mathbf{M}^n) (\mathbf{M}^n)^T \Delta t$$

  to define the subspace for propagating covariances in a RRKF
- FI (1999) showed good results when approximating $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{P}}$ with a single well-chosen term in sum

Conclusions

Results:

- preconditioning with HEVs reduces required number of iterations in 3d/4d-var by 35%-50% (reuse eigenvectors from 6 or 18UTC)
- SVs useful for identifying unwanted numerical instabilities in GEM
- new type of SV (time-averaged) allows optimizing growth over a range of lead times
- SVs may provide effective way for cycling covariances in 3d/4d-var

Future work:

- more testing of preconditioning in 3d/4d-var (determine effectiveness at earlier cut-off times)
- investigate approaches to eliminate numerical instabilities
- evaluate impact of increasing resolution and including more physics on SVs (moist processes, subgrid scale orography)
- compare SVs (growth rates) from specific case with ECMWF
- more fully compare covariance propagation using HSVs and Monte Carlo approach
- evaluate impact of balancing SVs and evolved SVs for covariance propagation in a cycling context (compare with ENKF)