Properties of Analysis and Forecast Errors Derived from the Adjoint of a 4D-Var System

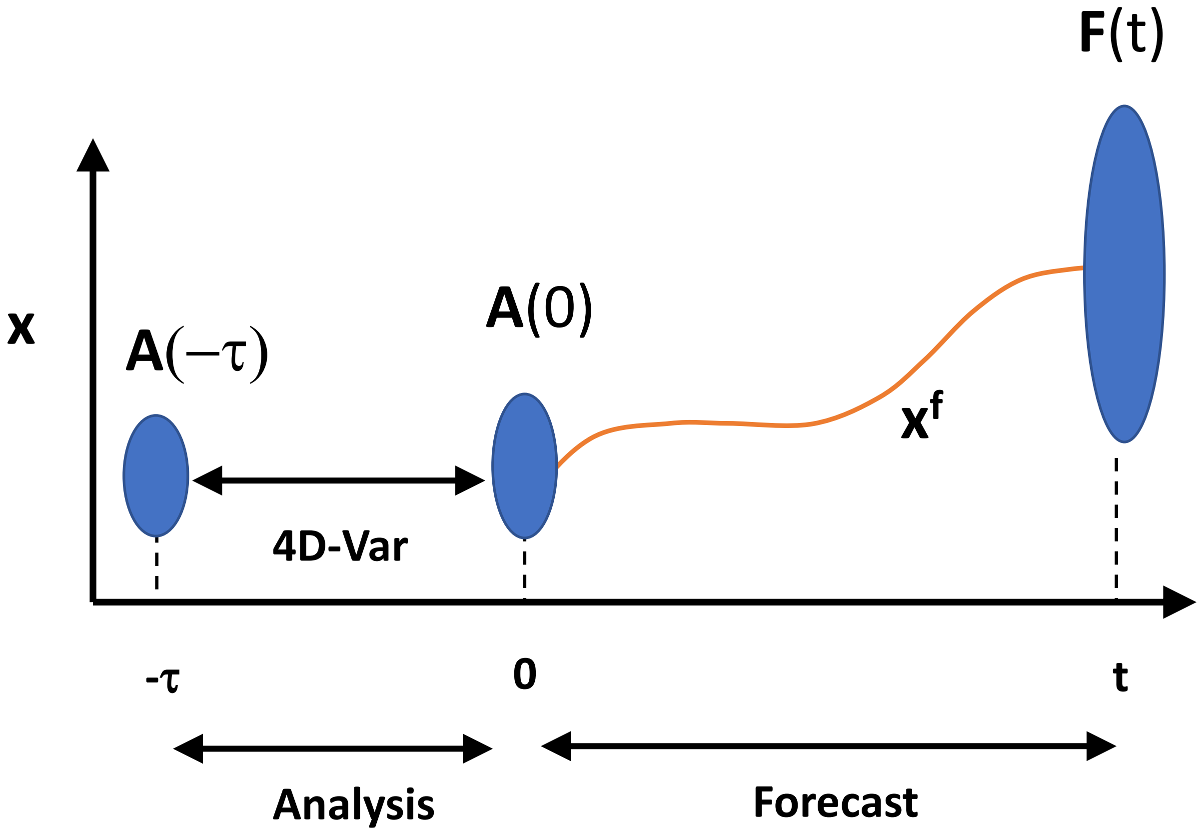
Andy Moore[1] & Hernan Arango[2]
1: UC Santa Cruz
2: Rutgers University

Outline

- **Motivation**
- **Analysis and forecast error covariance – a quick review**
- **An adjoint approach for computing expected error covariance matrix information**
- **Application to ocean fronts – ubiquitous features at all scales**
- **Properties of error covariance for growing and decaying instabilities**
- **Forecast skill**

Motivation

- Quantification of errors and uncertainties of analyses and forecasts is important but challenging.
- Ensemble methods provide a direct means for estimating error covariance matrices:
  - limited sample size necessitates localization and inflation
- Direct estimates of analysis error covariance in variational DA systems are challenging to compute (*e.g. Ngodock et al., 2020*) and are often underestimates (e.g. *Fisher and Courtier, 1995*).
- An approach is described here which employs the tangent linear and adjoint of a 4D-Var DA system to explore the properties and topology of the *expected* analysis and forecast error covariances matrices.



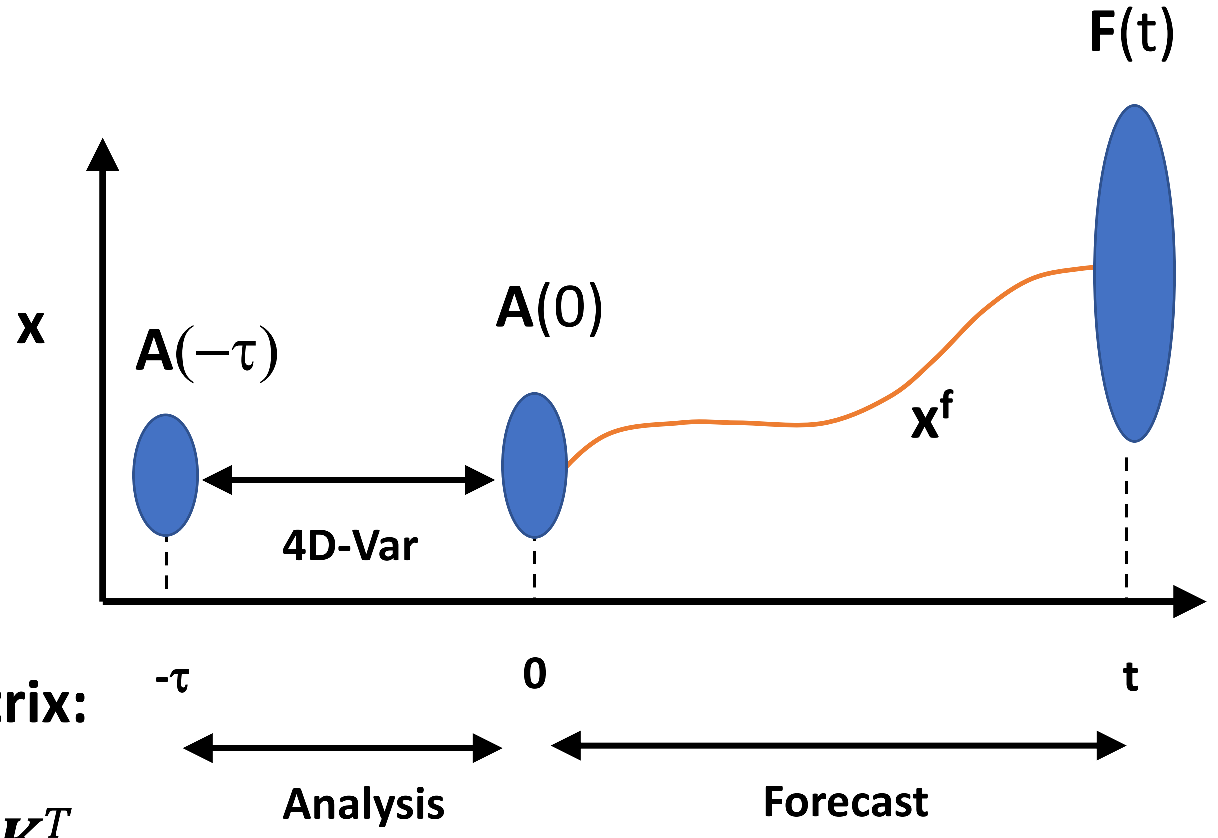
The Expected Analysis Error Covariance

The best, linear, unbiased estimate:

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{K}\boldsymbol{d}$$

$$\boldsymbol{d} = \left(\boldsymbol{y}^o - H\left(\boldsymbol{x}^b\right)\right)$$

$$\boldsymbol{K} = \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}$$



Expected analysis error covariance matrix:

$$\boldsymbol{A}(-\tau) = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{B}(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})^T + \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T$$

$$\boldsymbol{A}(0) = \boldsymbol{M}_b\boldsymbol{A}(-\tau)\boldsymbol{M}_b^T \qquad \boldsymbol{F}(t) = \boldsymbol{M}_f\boldsymbol{A}(0)\boldsymbol{M}_f^T$$

$\boldsymbol{M}_b$ = TL model linearized about $\boldsymbol{x}^b$

$\boldsymbol{M}_f$ = TL model linearized about $\boldsymbol{x}^f$

An Alternative Approach

Treat the DA system as a function of $\boldsymbol{d}$:

$$\boldsymbol{x}^a = \boldsymbol{x}^b + f(\boldsymbol{d})$$

Consider an infinite *ensemble* of analyses computed from perturbed observations $N(\boldsymbol{0}, \boldsymbol{R})$ and background $N(\boldsymbol{0}, \boldsymbol{B})$.

To 1st –order the expected analysis error covariance is given by:

$$\boldsymbol{A}(-\tau) = (\boldsymbol{I} - (\partial f/\partial \boldsymbol{d})\boldsymbol{H})\boldsymbol{B}(\boldsymbol{I} - (\partial f/\partial \boldsymbol{d})\boldsymbol{H})^T + (\partial f/\partial \boldsymbol{d})\boldsymbol{R}(\partial f/\partial \boldsymbol{d})^T$$

(*Moore et al, 2012*)

$(\partial f/\partial \boldsymbol{d}) =$ ***Tangent linearization of DA system***

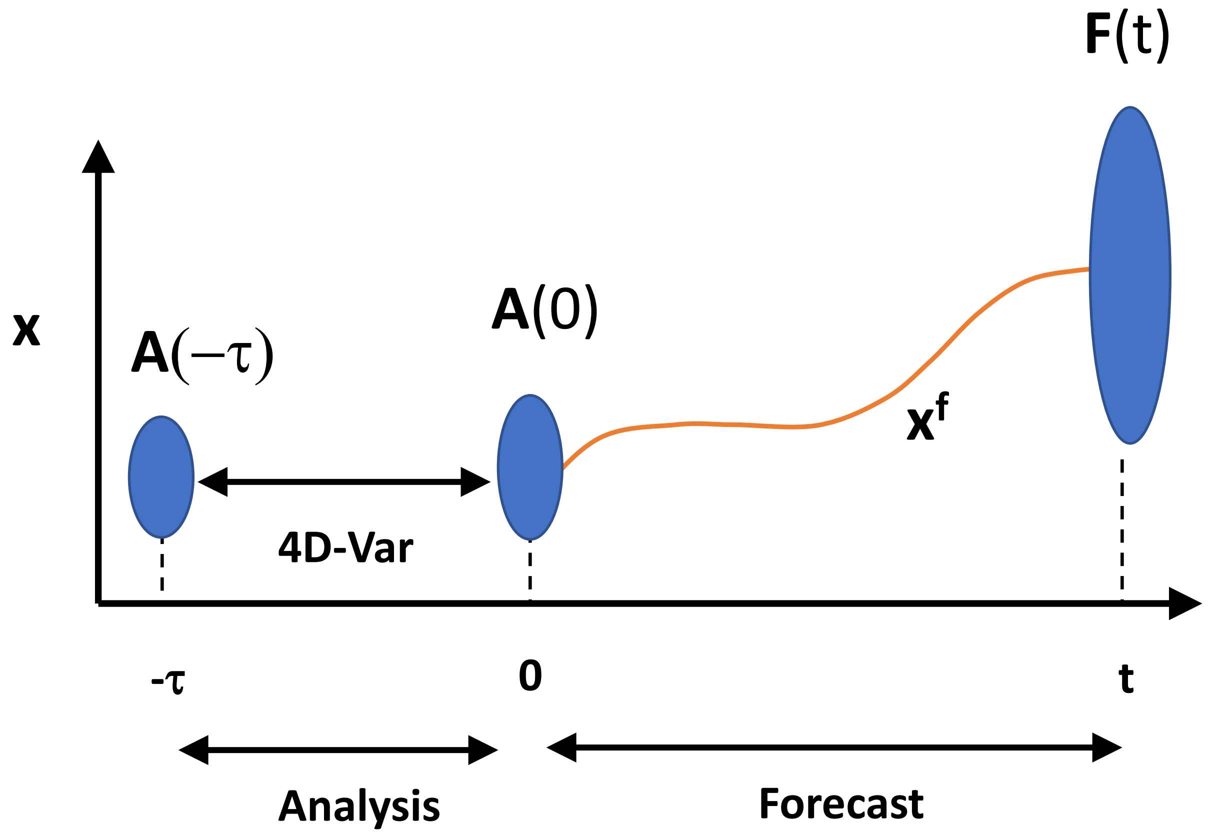
$(\partial f/\partial \boldsymbol{d})^T =$ ***Adjoint of DA system***

$$\boldsymbol{A}(-\tau) = (\boldsymbol{I} - (\partial f/\partial \boldsymbol{d})\boldsymbol{H})\boldsymbol{B}(\boldsymbol{I} - (\partial f/\partial \boldsymbol{d})\boldsymbol{H})^T + (\partial f/\partial \boldsymbol{d})\boldsymbol{R}(\partial f/\partial \boldsymbol{d})^T$$

$$\boldsymbol{A}(0) = \boldsymbol{M}_b \boldsymbol{A}(-\tau)\boldsymbol{M}_b^T$$

$$\boldsymbol{F}(t) = \boldsymbol{M}_f \boldsymbol{A}(0)\boldsymbol{M}_f^T$$

- **Costly to compute.**
- **Provides an explicit *analytical* operator for each covariance matrix -> far more theoretically appealing than an ensemble.**
- **We can exploit iterative Krylov methods and theorems of linear algebra to explore covariance matrix properties and topology.**



Some Underlying Assumptions

- **Perturbations about the unperturbed forecast mimic forecast errors (e.g. Belo-Pereira and Berre, 2006; Berre et al, 2006) => the unperturbed forecast represents the ensemble mean.**

$$\boldsymbol{\varepsilon}(t) = \boldsymbol{x}(t) - \boldsymbol{x}^f(t)$$

- **The tangent linear assumption is invoked for everything:**

$d\boldsymbol{\varepsilon}/dt = \boldsymbol{\Phi}(t)\boldsymbol{\varepsilon}(t)$ $\boldsymbol{\Phi}(t)$= the "resolvent" matrix

$\boldsymbol{\varepsilon}(t) = \boldsymbol{M}_f(0,t)\boldsymbol{\varepsilon}(0)$ $\boldsymbol{M}_f(0,t)$= the "propagator" matrix

$Tr\left(\boldsymbol{\Phi}(t)\right)$ yields information about the flow of probability

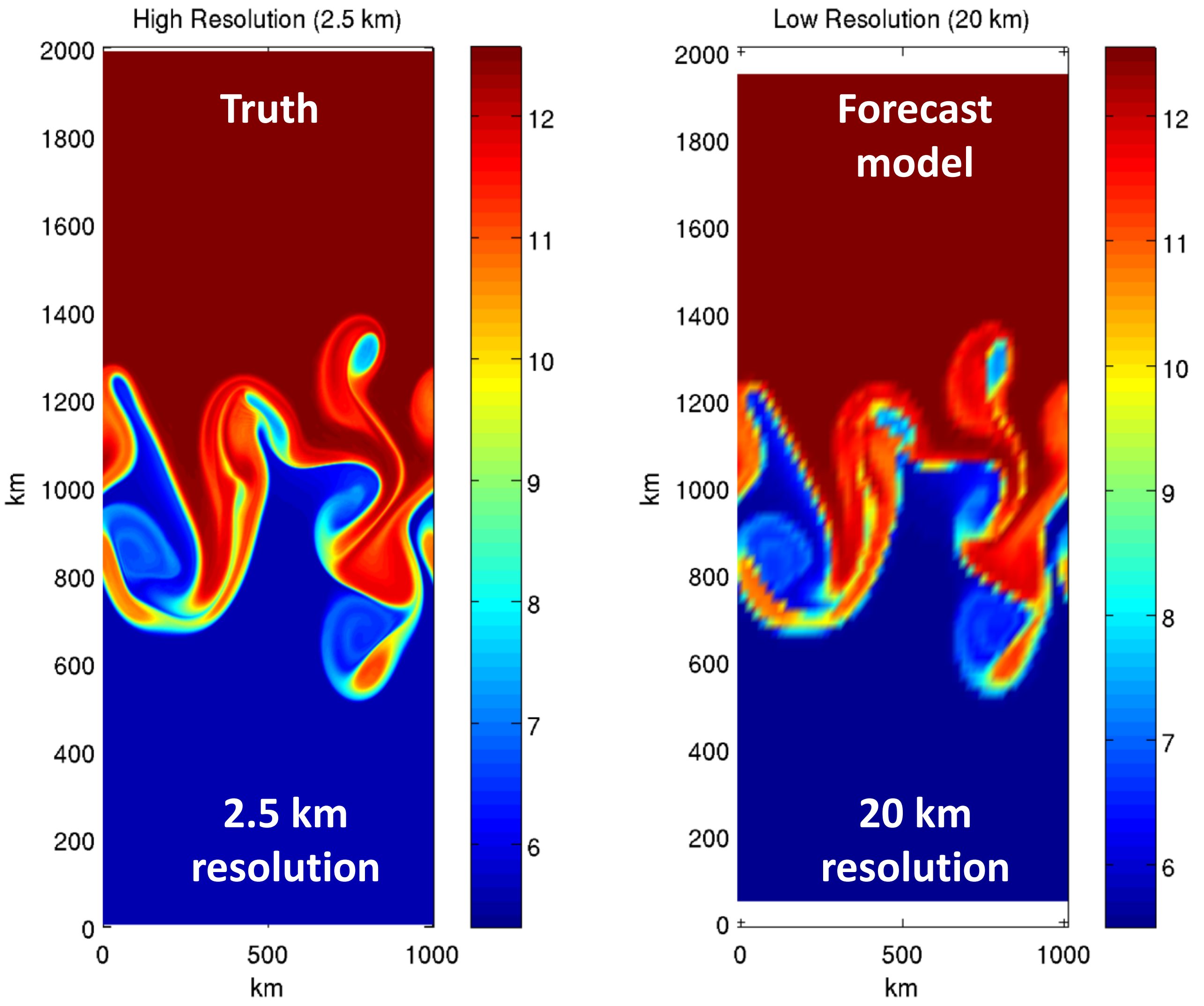
$det\left(\boldsymbol{M}_f(0,t)\right)$ represents the volume occupied by errors in state-space**

** |*det*(**M**)| is the volume of the parallelepiped defined by the *rows* of **M**

Paternal Twin Experiments

- Relaxation of a meridional temperature front
- Unforced
- No salinity

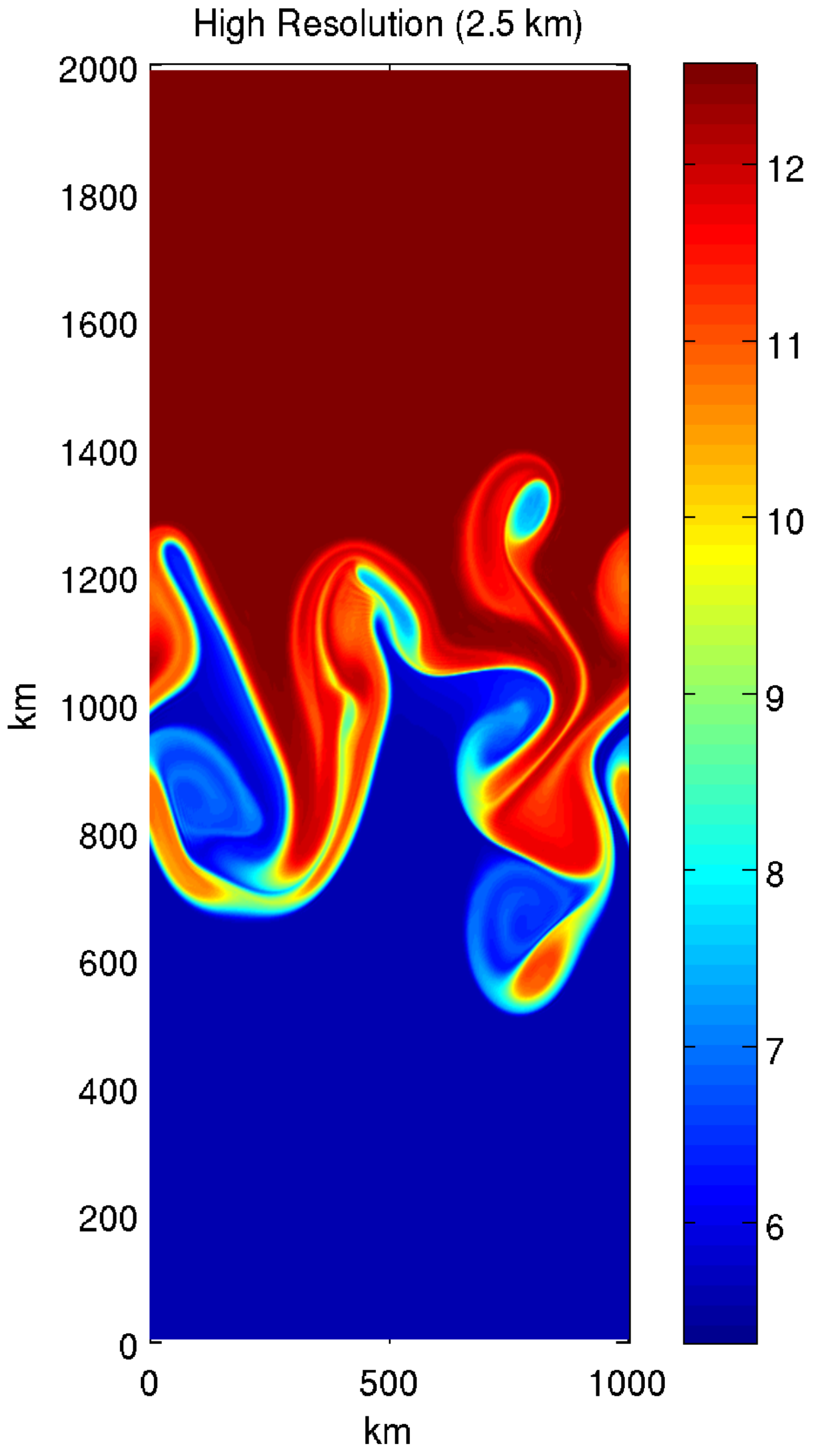
- Fronts are ubiquitous in the ocean at ***all*** scales.

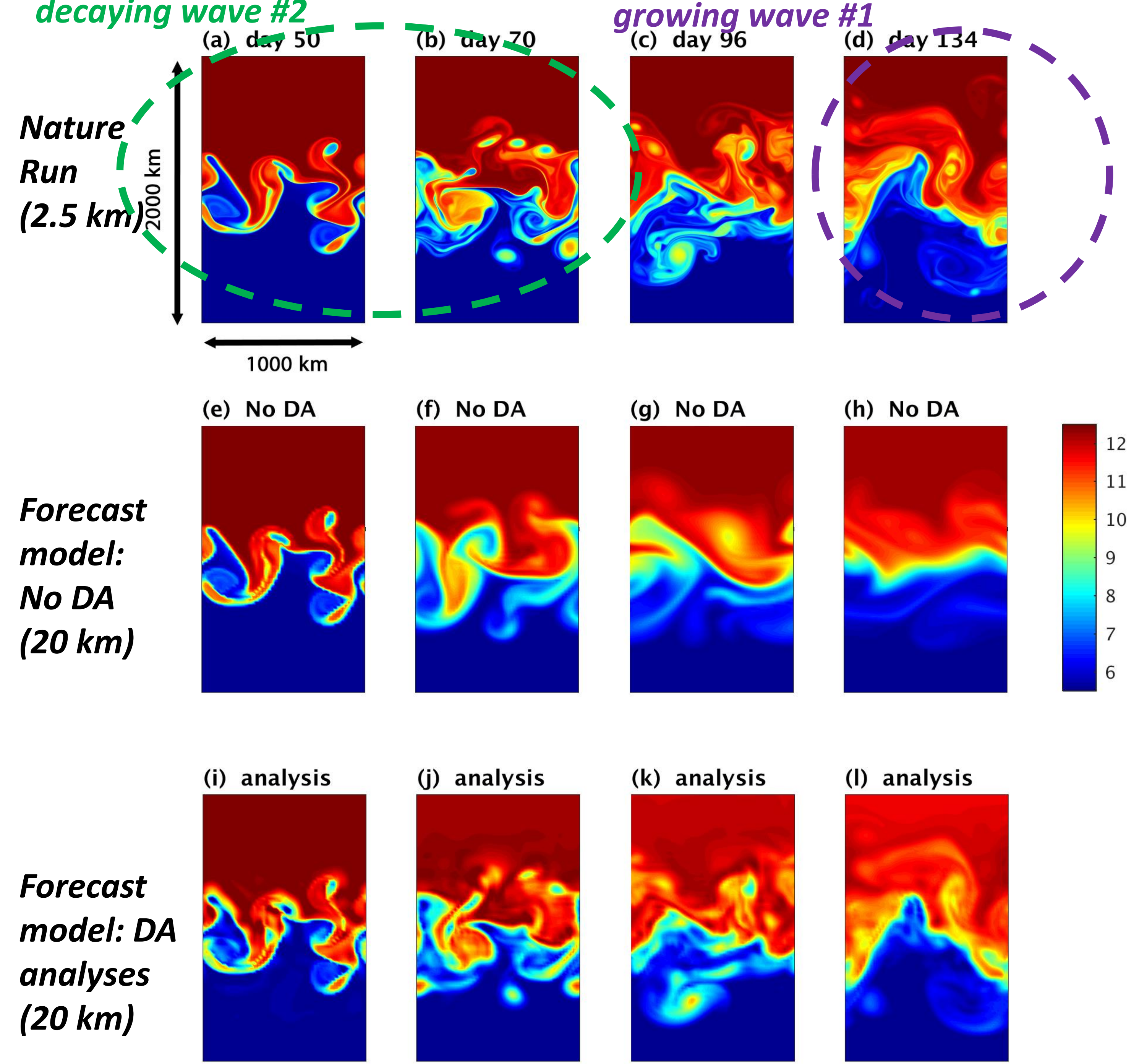


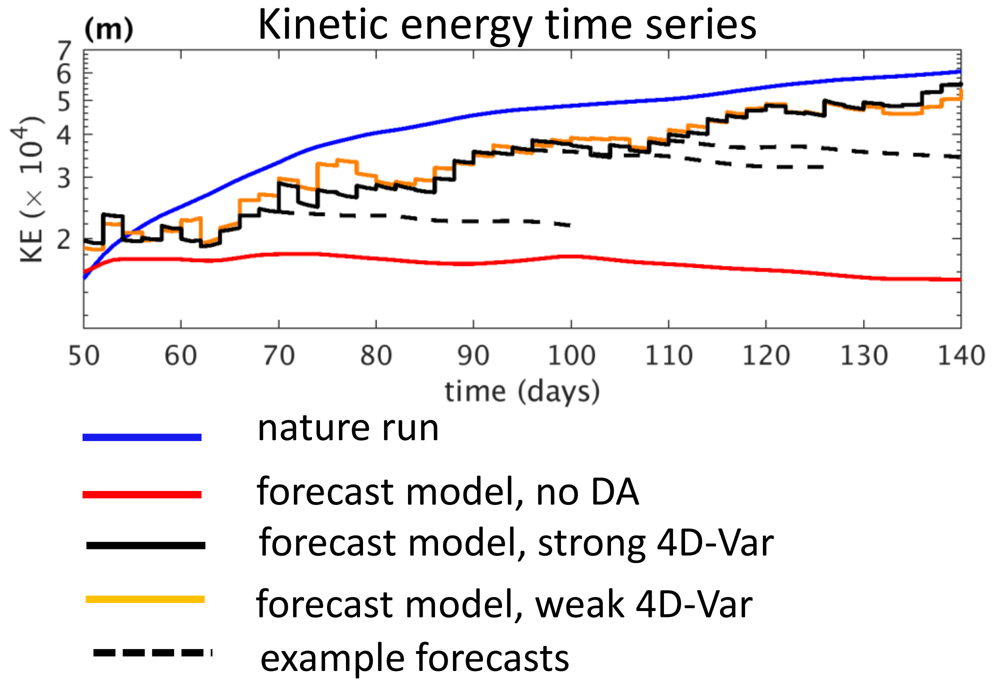
ROMS
Sequential 4D-Var
2 day 4D-Var cycles
Obs of T, 0-1000m
Obs spacing ds=3
Sampled once per day
Obs error=0.3C
Strong & weak constraint
m=25 inner-loops
k=1 outer-loops

A(0) **and** ***F(t)*****:**
$(3\times10^5)\times(3\times10^5)$

Baroclinic instability in a SH re-entrant channel
1000km X 2000km X 5000m


High Resolution (2.5 km)
km
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
500
1000
km
12
11
10
9
8
7
6





Note

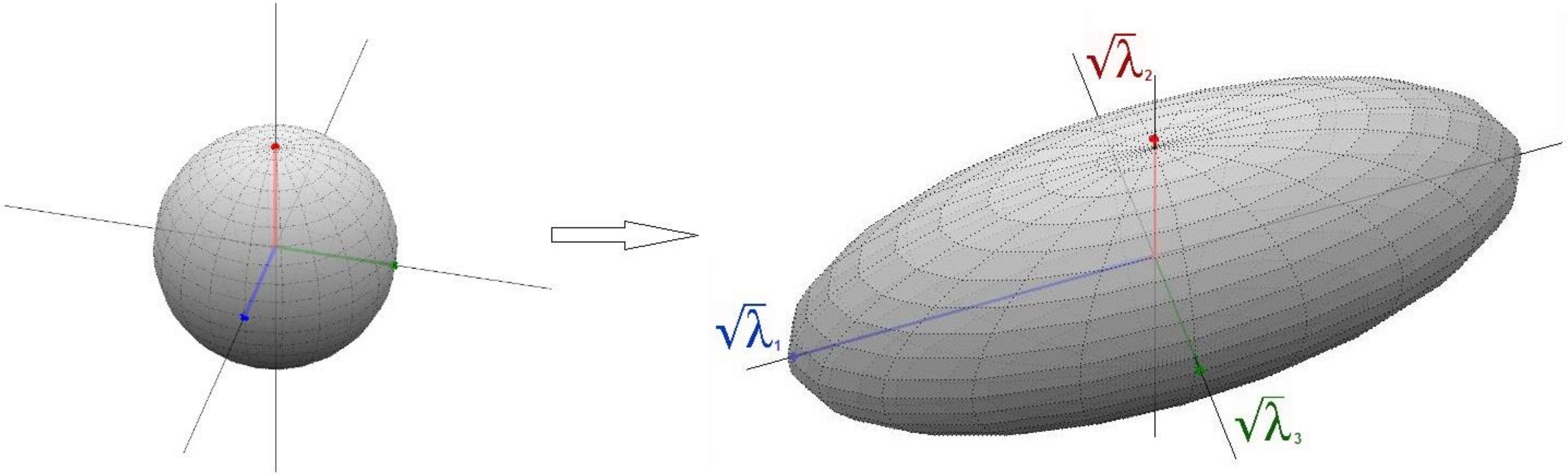
- **Conversion of APE to KE**
- **DA energizes the circulation by "propping up" isopycnals**
- **Weak constraint corrections for model error can further energize the circulation**
- **Define error covariance using energy norm:**

$$\boldsymbol{\varepsilon} = \boldsymbol{x} - \boldsymbol{x}^f$$

$$\boldsymbol{C} = \boldsymbol{U} E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\}\boldsymbol{U}^T$$

where $\boldsymbol{x}^T \boldsymbol{U}^T \boldsymbol{U} \boldsymbol{x}$ is the total energy

A Geometrical Interpretation of *A* and *F*



hypersphere =>
flat EOF spectrum

hyperellipse =>
peaked EOF
spectrum

hyper volume ∝ $(det(\mathbf{UAU}^T))^{1/2}$ or $(det(\mathbf{UFU}^T))^{1/2}$

total variance = $Tr(\mathbf{UAU}^T)$ or $Tr(\mathbf{UFU}^T)$

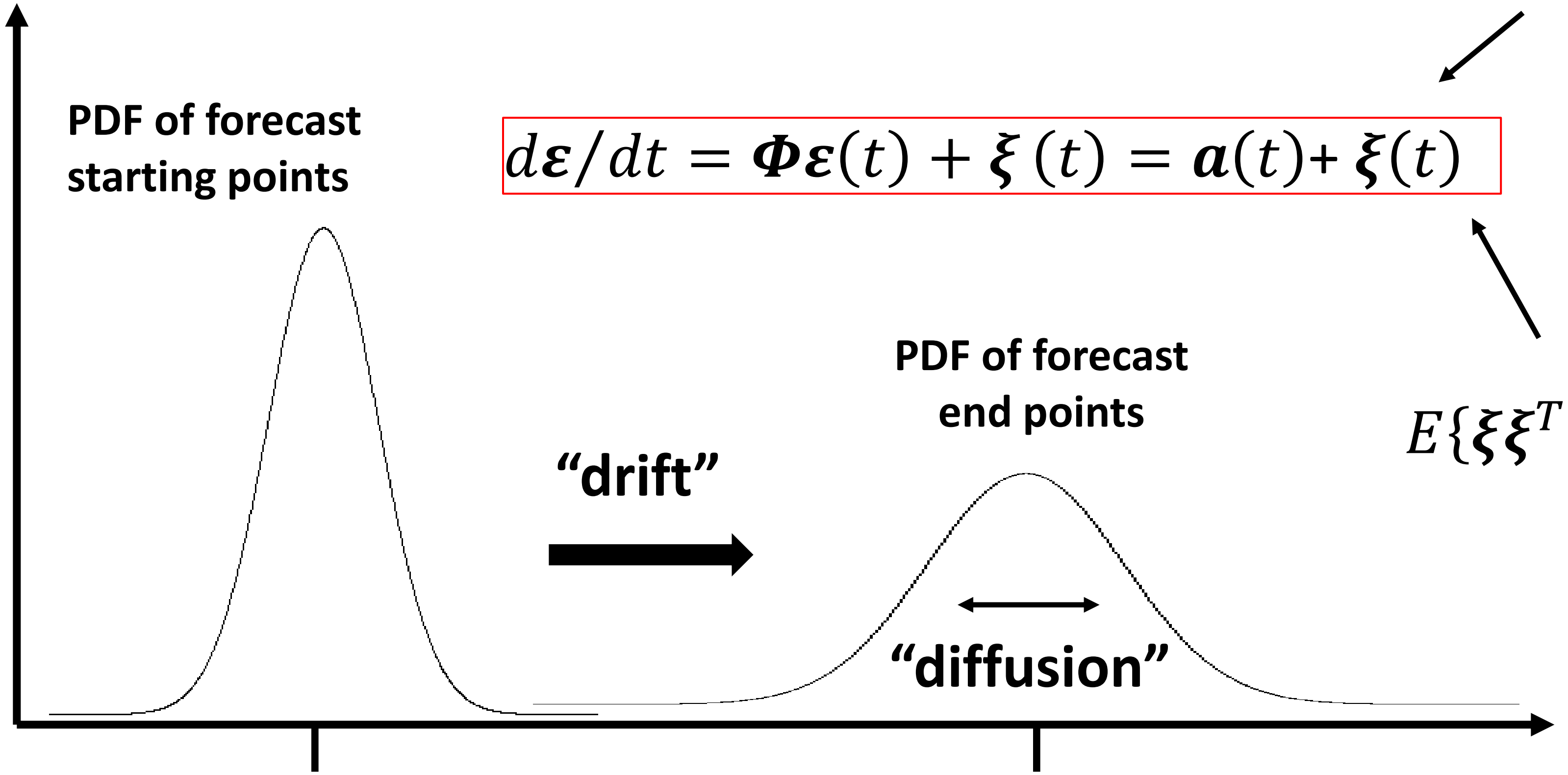
$$det(\mathbf{A}) = \prod_{i=1}^{N} \lambda_i \qquad Tr(\mathbf{A}) = \sum_{i=1}^{N} \lambda_i$$

Forecast Error PDF

Stochastic model error

$$d\boldsymbol{\varepsilon}/dt = \boldsymbol{\Phi}\boldsymbol{\varepsilon}(t) + \boldsymbol{\xi}(t) = \boldsymbol{a}(t) + \boldsymbol{\xi}(t)$$

$$E\{\boldsymbol{\xi}\boldsymbol{\xi}^T\} = \boldsymbol{Q} = (q_{i,j})$$



$\overline{\boldsymbol{\varepsilon}(t_0)}$ $\overline{\boldsymbol{\varepsilon}(t)}$

Conditional probability

$$p \equiv p(\boldsymbol{\varepsilon}, t | \boldsymbol{\varepsilon}_0, t_0)$$

Fokker-Planck Equation

$$\partial p \big/ \partial t = -\sum_{i=1}^{N} \partial(a_i p) \big/ \partial \varepsilon_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \partial^2 (q_{i,j} p) / \partial\varepsilon_i \partial\varepsilon_j$$

"drift" **"diffusion"**

$$\mathcal{C}_i = \boldsymbol{a}(t)p - \frac{1}{2}\sum_{j=1}^{N} \partial(q_{i,j}p)/\partial\varepsilon_i$$

"probability current"

$$\partial p \big/ \partial t = -\sum_{i=1}^{N} \partial \mathcal{C}_i / \partial\varepsilon_i$$

The Fokker-Planck Equation

Forecast errors evolve according to:

$$d\boldsymbol{\varepsilon}/dt = \boldsymbol{\Phi}(t)\boldsymbol{\varepsilon}(t) + \boldsymbol{\xi}(t) = \boldsymbol{a}(t) + \boldsymbol{\xi}(t)$$

The resolvent matrix $\boldsymbol{\Phi}(t)$ "generates" a vector-field $\boldsymbol{a}(t) = (a_i)$ which is the "drift" in the Fokker-Planck equation *i.e.* that translates the mean value of $\boldsymbol{\varepsilon}$ *and* spreads the pdf since, in general, the divergence of the "dirift" does *not* vanish.

$$\partial p/\partial t = -\sum_{i=1}^{N} \partial(a_i p)/\partial \varepsilon_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \partial^2\left((\boldsymbol{g}\boldsymbol{g}^T)_{ij} p\right)/\partial\varepsilon_i \partial\varepsilon_j$$

"drift" **"diffusion"**

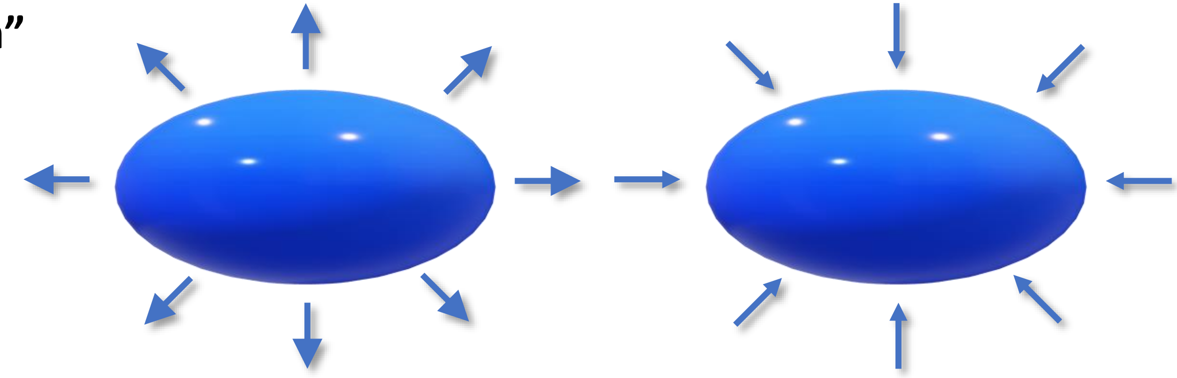
Liouville's equation

From the Jacobi formula:

$$\partial \ln\left(det(\boldsymbol{F}(t))\right)/\partial t = 2Tr(\boldsymbol{\Phi}(t))$$

The divergence of the "drift"

*Note: independent of **U** (i.e. similarity invariant)*



Determinant & Trace Estimation

$$log(det(\boldsymbol{UAU}^T)) = \sum_{i=1}^{N} log(v_i)$$

$$log(det(\boldsymbol{UF}(t)\boldsymbol{U}^T)) = \sum_{i=1}^{N} log(\lambda_i)$$

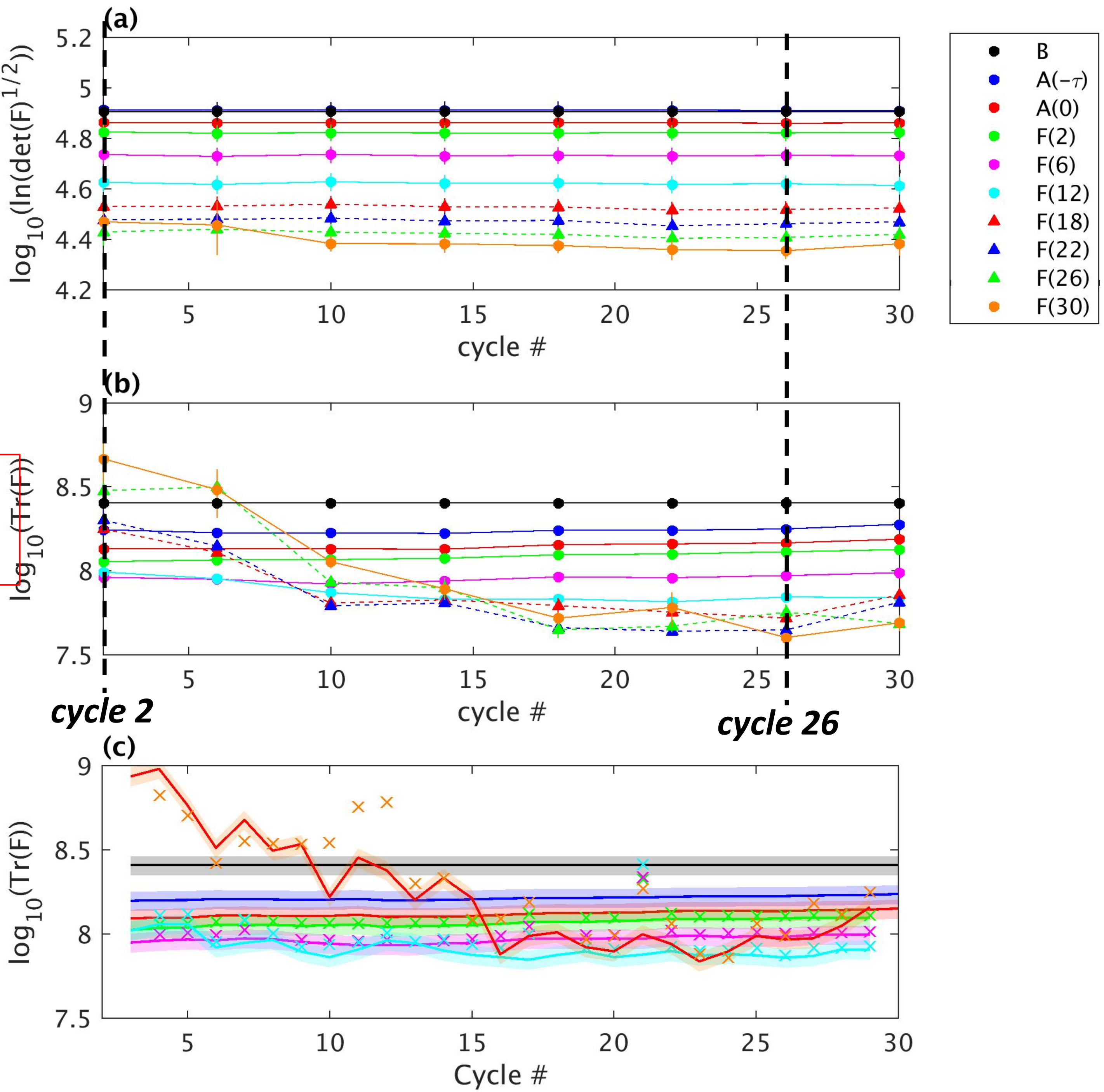
$$Tr(\boldsymbol{UAU}^T) = \sum_{i=1}^{N} v_i$$

$$Tr(\boldsymbol{UF}(t)\boldsymbol{U}^T) = \sum_{i=1}^{N} \lambda_i$$

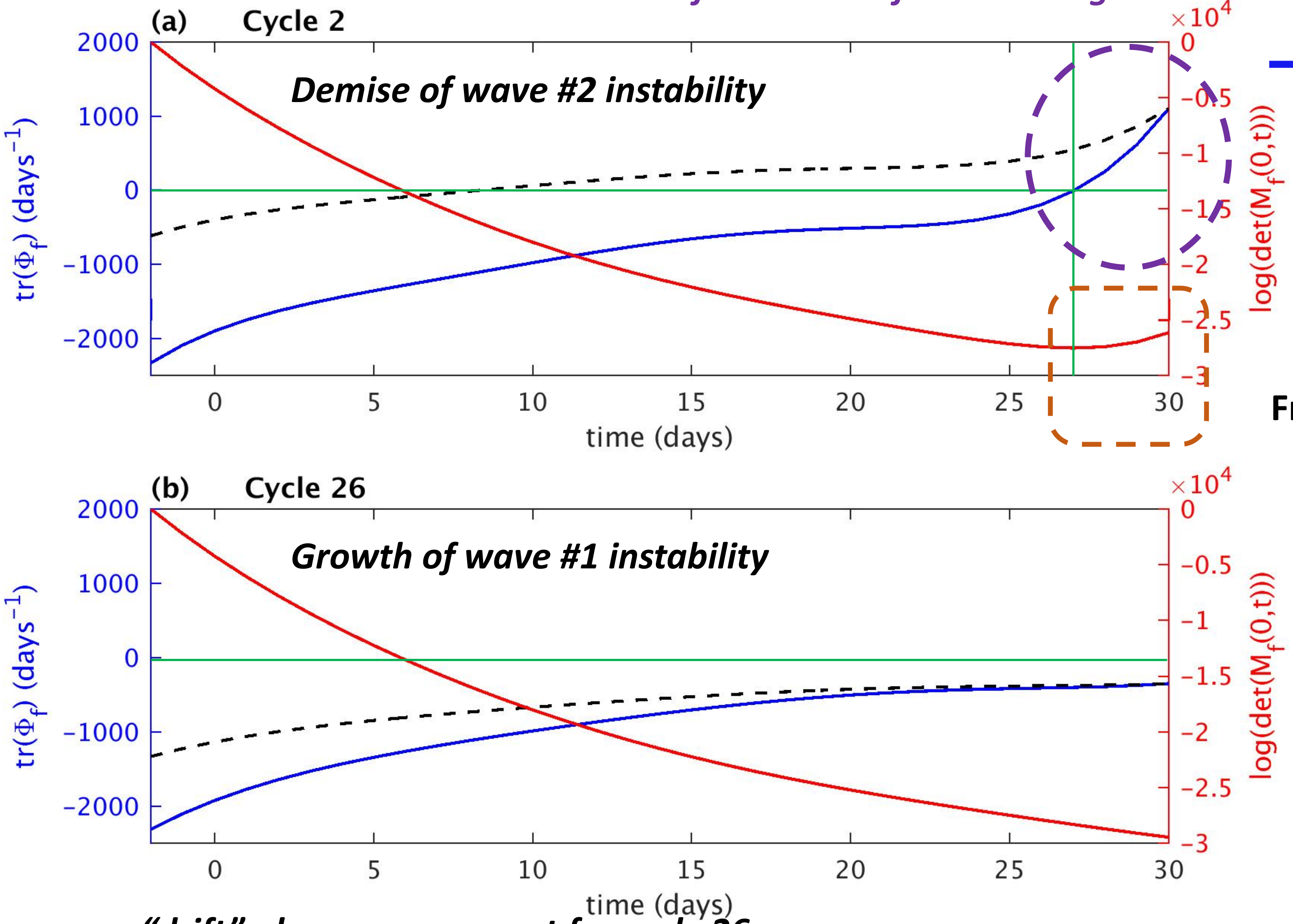
For Gaussian errors:
-1×entropy reduction
(Shannon & Weaver, 1949)

- **Estimates based on Monte Carlo sequence of Lanczos factorizations (Bai et al, 1996)**
- **Yields upper and lower bounds on estimates**
- **Expensive!**

Independent estimate of Tr following Fisher and Courtier (1995)



"drift" switches from convergent to divergent around day 27



$\frac{1}{2}\partial ln\left(det\left(\boldsymbol{F}(t)\right)\right)/\partial t \; = Tr\left(\boldsymbol{\Phi}(t)\right)$

divergence of "drift"

volume of state-space switches from decreasing to increasing

From Liouville's equation, we also have:

$$\ln\left(det\left(\boldsymbol{M}_f(t)\right)\right) = \frac{1}{2}\ln[det(\boldsymbol{F}(t))/det(\boldsymbol{F}(0))]$$

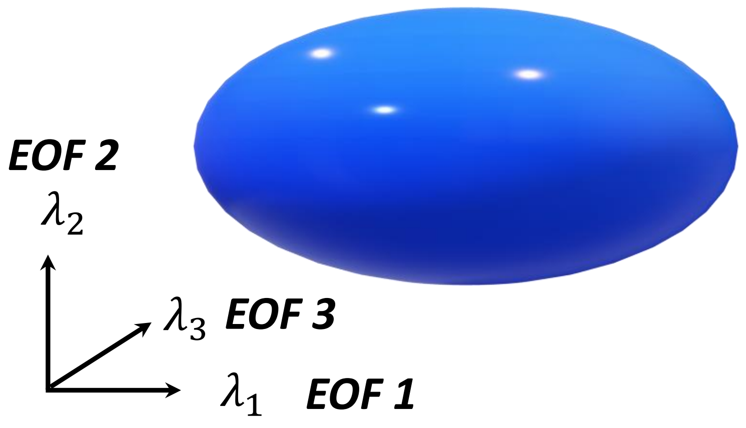
$det\left(\boldsymbol{M}_f(t)\right)$= Volume of state-space occupied by the forecast errors

$\ln\left(det\left(\boldsymbol{M}_f(t)\right)\right)$

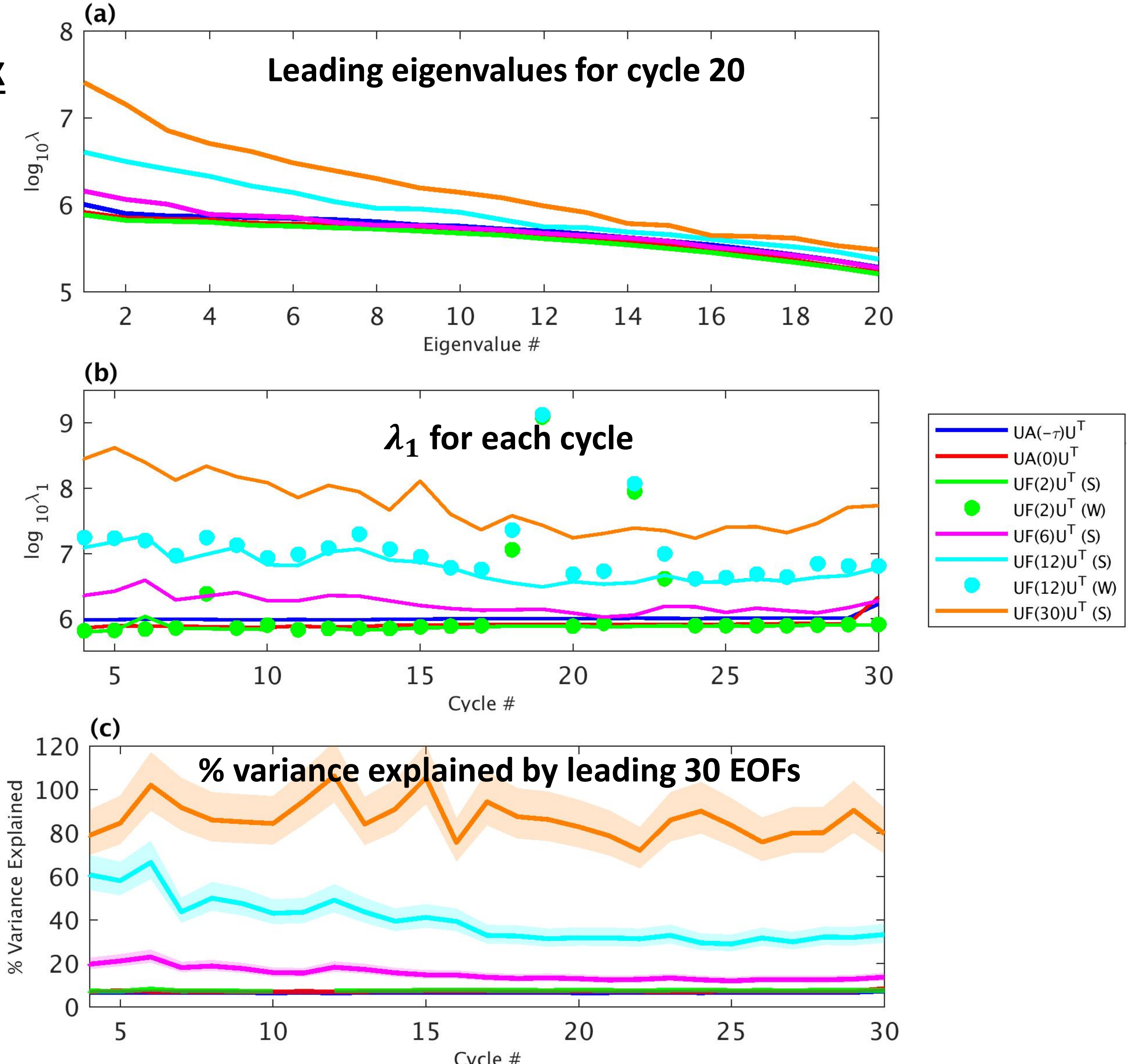
"drift" <u>always</u> convergent for cycle 26

volume of state-space <u>always</u> decreasing for cycle 26

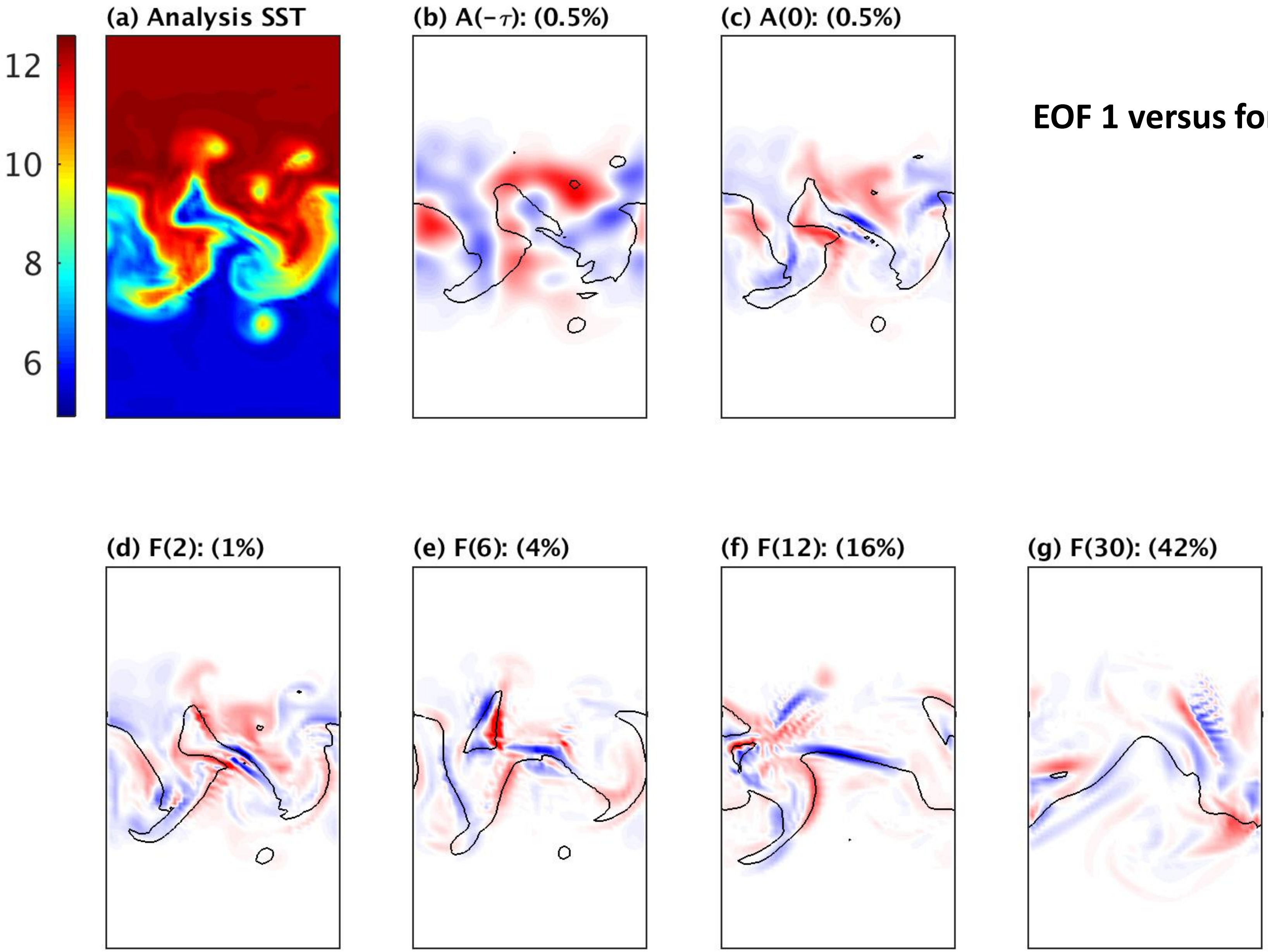
Topology of Error Covariance Matrix



- Error covariance hyper-ellipse becomes more elongated with increasing lead time.
- ~30 EOFS (i.e. 0.01% of the spectrum) account for almost all of the variance at long-lead times.



EOF 1 versus forecast lead-time



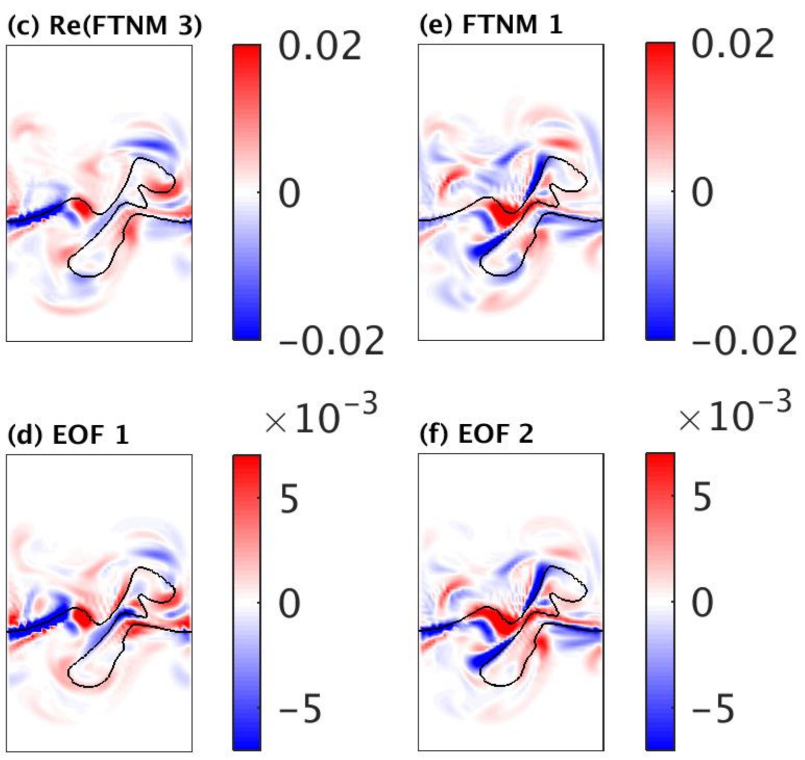
What we know so far:

- We have seen that:
  - the forecast error covariance matrix stretches in a few preferred directions
  - the volume of state-space occupied by forecast errors generally decreases with increasing lead-time
  - the entropy of the errors generally decreases over time
    => self-organization

- Recall the assumption that forecast errors evolve linearly according to:

$$\boldsymbol{\varepsilon}(t) = \boldsymbol{M}_f(0,t)\boldsymbol{\varepsilon}(0)$$

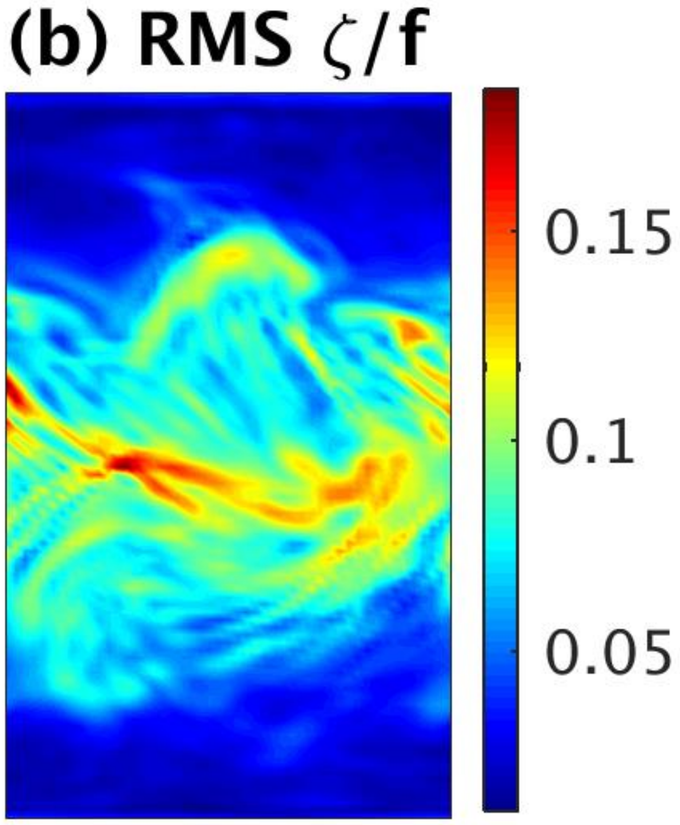
- As $t \rightarrow \infty$ the fastest growing eigenmodes of $\boldsymbol{M}_f(0,t)$ will emerge and determine the leading EOFs



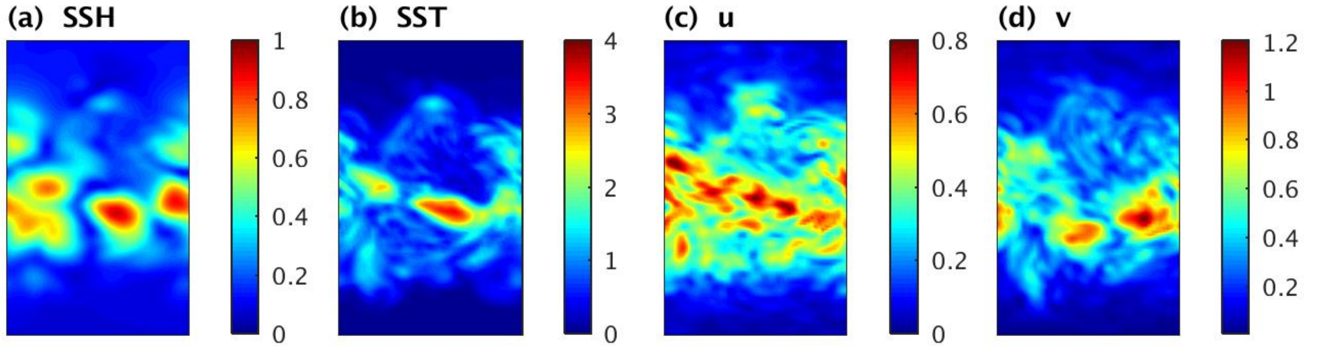
SST errors

Forecast Error Variance as a Predictor of Forecast Skill

- Ense
  for sl
  sprea
  1969;
- This
  the v
  circu
  clima
  Lough
- Theo
  betw
  (Houte
- Also
  metr

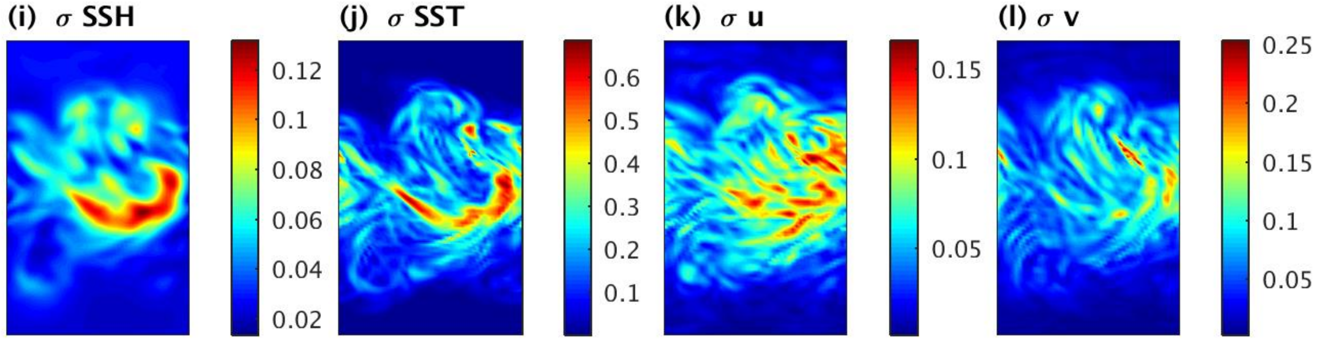


mean Rossby number



Mean forecast error on day 12

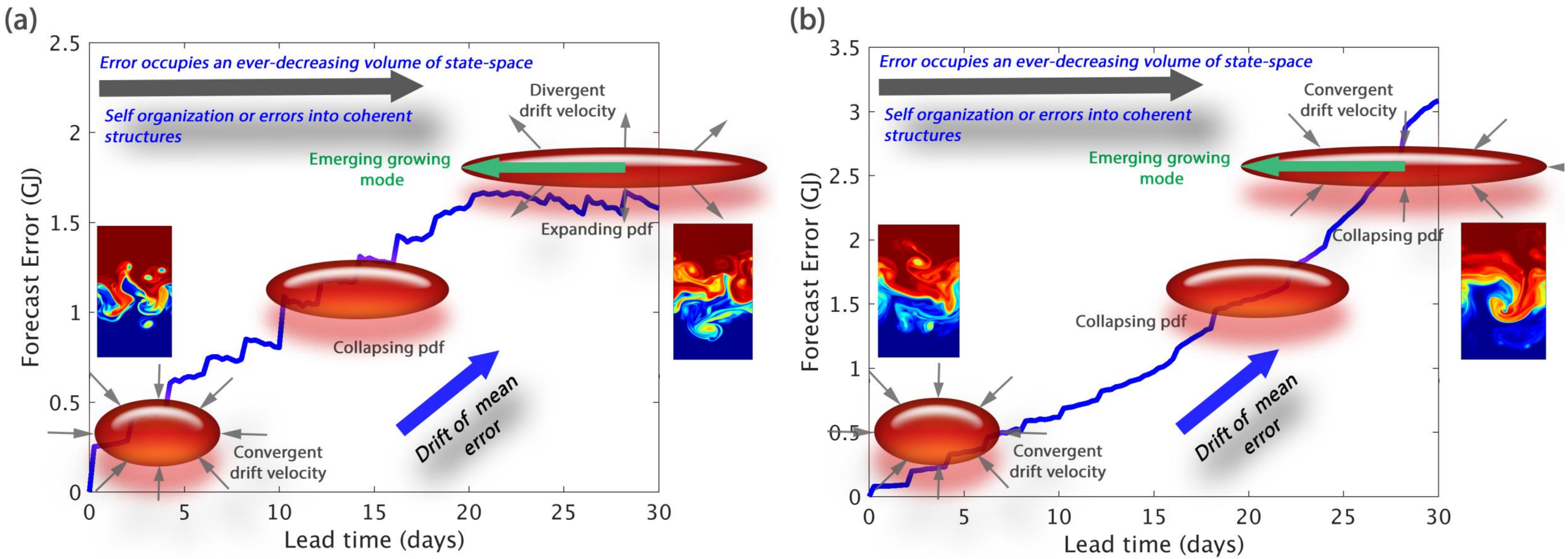
Expected forecast error is an underestimate of actual forecast error, but locations of largest error variance are captured well



Day 12 forecast error std from leading 3o EOFs of F

Summary

Two dynamical regimes with different forecast error behaviors were identified:



Decay of wave number 2 instability **Growth of wave number 1 instability**

Summary

- **An alternative approach for exploring properties of expected error covariances explored.**
- **Provides an explicit operator for the expected analysis error (*A*) and forecast error (*F*) covariance matrices -> provides in-depth information about behavior and topology.**
- **Approach applied to dynamics of baroclinically unstable fronts – two distinct regimes identified.**
- **Future experiments will include surface forcing – can be important for frontogenesis and frontolysis.**
- **Fronts are ubiquitous in the ocean at all scales -> scaling analysis may shed light on how the results presented here apply at other scales (such as the sub-mesoscale).**
- **While computationally demanding, the adjoint-based approach proposed here will become more tractable for systems of larger dimension, especially in light of speed-ups of 4D-Var.**
- **Potential practical utility as an indicator of expected forecast error.**

Moore, A.M. and Arango, H.G., 2021: On the behavior of ocean analysis and forecast error covariance in the presence of baroclinic instability. Ocean Modelling,
https://doi.org/10.1016/j.ocemod.2020.101733