THE VIRTUES OF ENSEMBLE FORECASTING

Zoltan Toth

Global Systems Division, NOAA/OAR/ESRL

Jie Feng, Malaquias Peña, Yuejian Zhu, and Yan Luo

Acknowledgements:
Kevin Kelleher, Mozheng Wei, Yi Wang

THE VIRTUES OF ENSEMBLE FORECASTING

MYTH OR REALITY ?

Zoltan Toth

Global Systems Division, NOAA/OAR/ESRL

Jie Feng, Malaquias Peña, Yuejian Zhu, and Yan Luo

Acknowledgements:
Kevin Kelleher, Mozheng Wei, Yi Wang

ENSEMBLE FORECASTING REVISITED

OR THE UNTOLD STORY OF THE ENSEMBLE MEAN

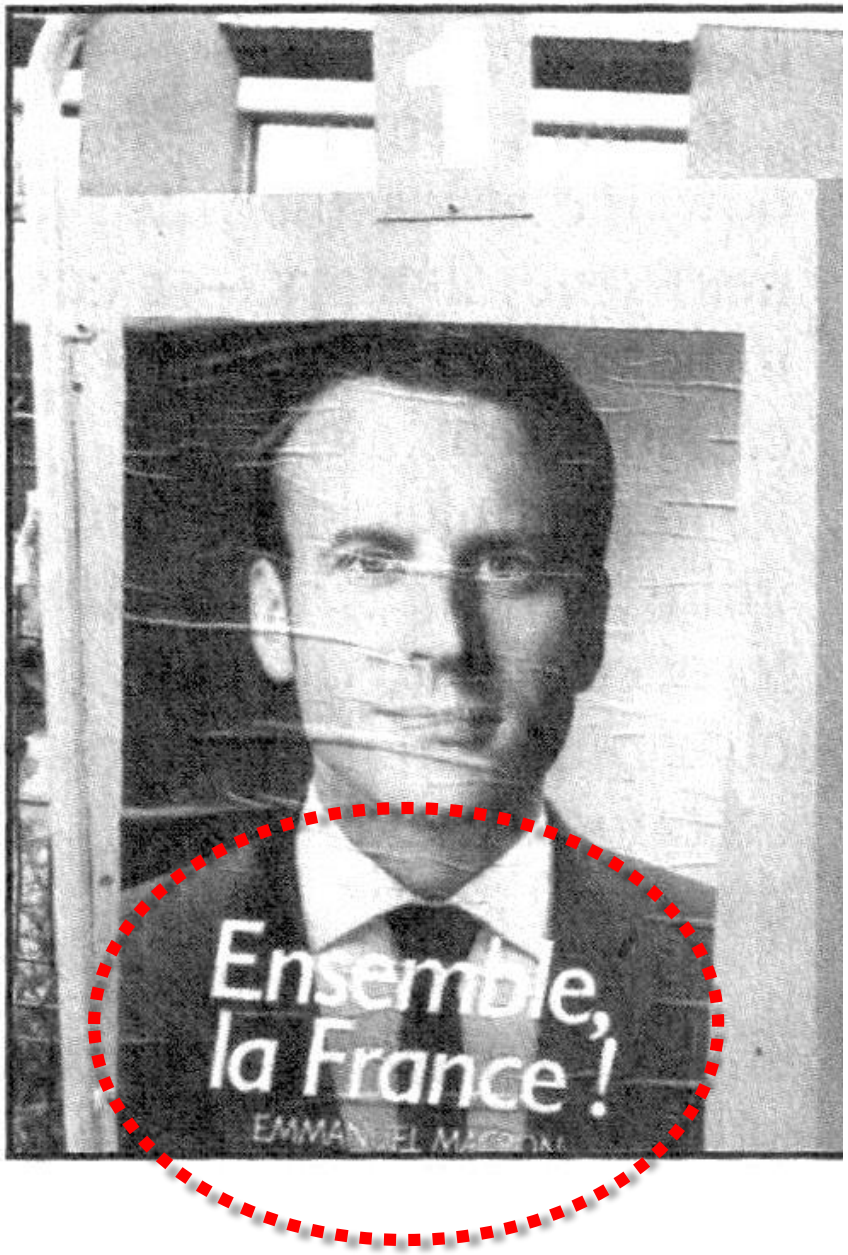
Zoltan Toth

Global Systems Division, NOAA/OAR/ESRL

Jie Feng, Malaquias Peña, Yuejian Zhu, and Yan Luo

Acknowledgements:
Kevin Kelleher, Mozheng Wei, Yi Wang

EMERGING NATIONAL PRIORITIES



OUTLINE

- Historical context - Ensemble Mean basics
- Logistic function to describe
  - Control forecast error and its reduction due to nonlinearities
- Initial value vs saturation related filtering
- Projection of perturbations on control error
- Alternatives to dynamically generated ensembles
- How to choose initial perturbations?

HISTORICAL PERSPECTIVE

- Ensemble Forecasting (EF) emerged along dynamically based Numerical Weather Prediction - *Lewis 2005*
  - Eady, Thompson, Leith, *Lorenz 1965*

The proposed procedure chooses a finite ensemble of initial states, rather than the single observed initial state. Each state within the ensemble resembles the observed state closely enough so that the differences might be ascribed to errors or inadequacies in observation. A system of dynamic equations previously deemed to be suitable for forecasting is then applied to each member of the ensemble, leading to an ensemble of states at any future time. From an ensemble of future states, the probability of occurrence of any event, or such statistics as the ensemble mean and ensemble standard deviation of any quantity, may be evaluated. Between the near future, when all states within an ensemble will look about alike, and the very distant future, when two states within an ensemble will show no more resemblance than two atmospheric states chosen at random, it is hoped that there will be an extended range when most of the states in an ensemble, while not constituting good pin-point forecasts, will possess certain important features in common. It is for this extended range that the procedure may prove useful.

conditions. If distinct régimes are present, however, it may be possible to predict the régime, with a reasonable probability of success, at a considerably longer range than that at which one can hope to predict the state within the régime.

a glob of points each of which would follow its own deterministic path. (E. Epstein 2002, personal communication)

- some vagueness

CONCEPT OF & PRODUCTS FROM EF

- **Ensemble of initial states** around
  - "Observed state" *OR*
  - Best / unperturbed / control analysis
- **State estimate**
  - Control (c) *OR* Ensemble Mean (em)?
  - Initial value, *OR* full nonlinear saturation related filtering?
- **Error estimate**
  - Statistical or ensemble spread?
- **Probabilistic forecasts**
  - Statistical or ensemble derived?

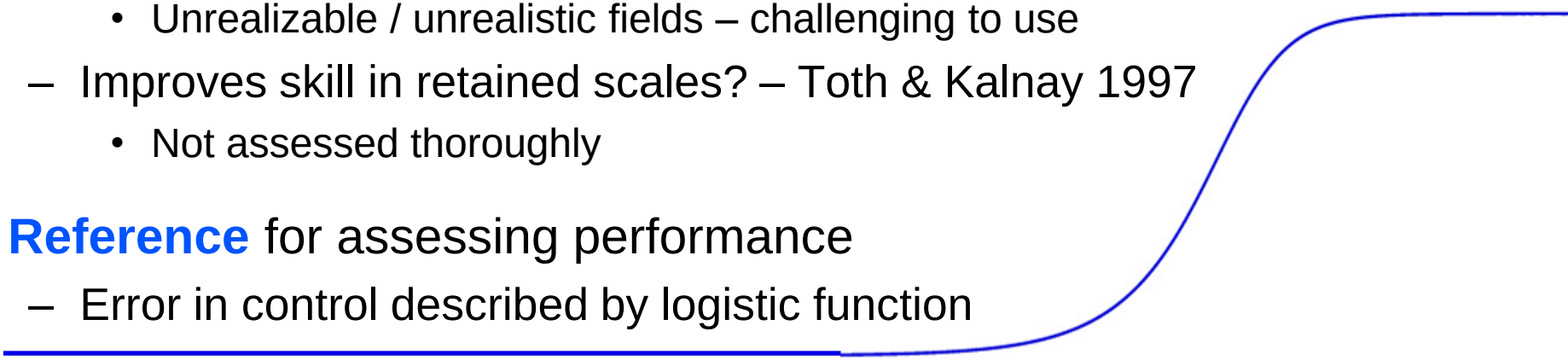
THRUST OF TALK

- **Critical review** of some basic questions about EF
  - Being long in field one may take things granted
  - Some NWP scientists instinctively question logic behind EF
    - *Whose instincts are right?* - Pose & probe questions
- **EF works** - ensemble mean, spread, probabilities used
  - *What are the mechanisms behind? -* Look behind curtain
- **N times higher cost** than single forecast
  - Or must compromise quality by degrading model used
- Any **opportunities for alternatives?**
  - Distinguish between
    - End goal – eg, probabilistic products – we need this, *vs*
    - Means – eg, ensemble or other (statistical?) methods
      - Need *one* of these, there are methods other than ensemble
  - Consider performance & cost of alternatives
    - Pros & cons for EF

Focus on** state estimate – assess **ensemble mean

ENSEMBLE MEAN (EM) BASICS

- **Definition** – Arithmetic mean of members
- **Characteristics**
  - Filters out progressively larger unpredictable scales - Lorenz 1965; TK97
    - Unrealizable / unrealistic fields – challenging to use
  - Improves skill in retained scales? – Toth & Kalnay 1997
    - Not assessed thoroughly
- **Reference** for assessing performance
  - Error in control described by logistic function

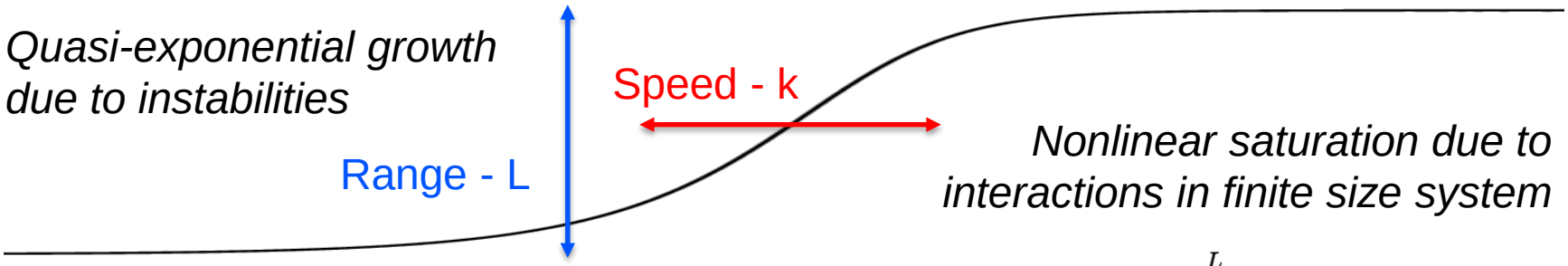


- **Parametric modelling** of error in EM vs control -
  - Initial error variance in control – ***Rms(C-Reality)***
  - Perturbation variance - ***Rms(P-C)***
  - Fraction of perturbation projecting on control error – ***F(P:(C-R))***
  - Number of ensemble members - ***n***
  - Lead time - ***lt***

$$I(EM) = (Rms(C-R) - \mathrm{Rms}(\mathrm{EM}-R))/E(C-R)$$

- **Metric** for impact of EM – % difference btw error in control vs EM

LOGISTIC RELATIONSHIP



$$f(x) = \frac{L}{1 + e^{-k(x - x_0)}}$$

where

- e = the natural logarithm base (also known as Euler's number),
- x_0 = the x-value of the sigmoid's midpoint,
- L = the curve's maximum value, and
- k = the steepness of the curve.[1]

The *standard logistic function* is the logistic function with parameters ($k = 1$, $x_0 = 0$, $L = 1$) which yields

$$f(x) = \frac{1}{1 + e^{-x}}$$

- **Generic relationship** widely used in
  - Biology, chemistry, geosciences, demography, economics, psychology, sociology, political science, linguistics, statistics, etc
- Used to describe **perturbation or error growth**
  - In nonlinear systems like the atmosphere *(Lorenz 1969)*
- We will describe **error in unperturbed "control" forecast**
  - Applied to true error evaluated against *reality*
    - As opposed to "perceived error" evaluated against proxy for reality (analysis)
  - Serves as basic reference

ENSEMBLE MEAN VS. SMOOTHING

TABLE 4. The effect of optimal spatial smoothing on the control and 10-member ensemble mean forecasts for the period 23 May–3 June 1992 with 10%/20% initial perturbations for the Northern and Southern Hemispheres, respectively. For further details, see text.

Lead time (days)	Optimal smoothing (~triangular truncation)		Ensemble advantage over control retained	
	Control	Ensemble	PAC	Percent total
5	T30	T40	0.02	62.5%
7	T25	T35	0.033	63.8%
9	T20	T30	0.042	60.5%

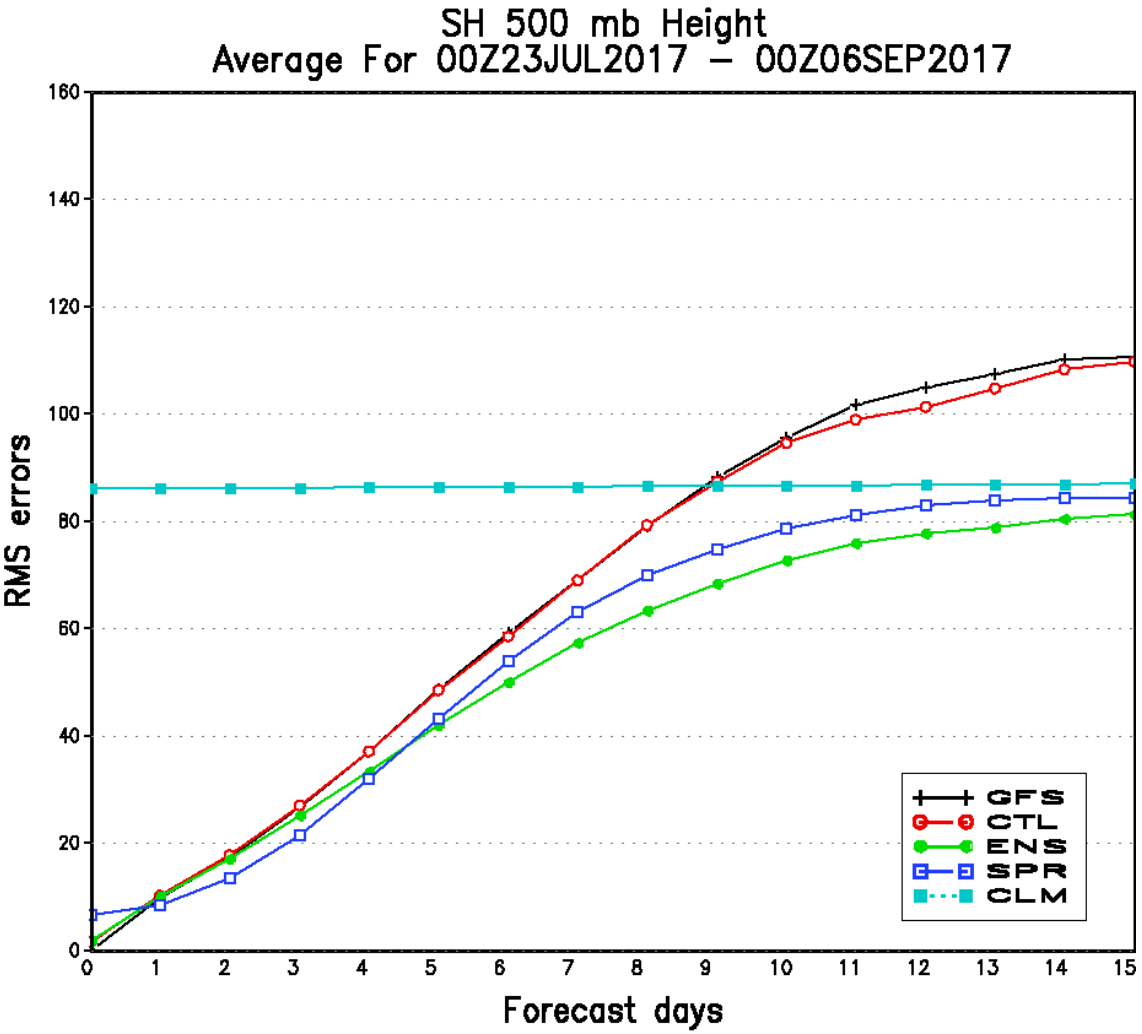
Toth & Kalnay 1997

- **Control & ens mean progressively filtered** w increasing lead time to optimize PAC
  - Stronger filter at longer leads & for control
- **Small sample**, non-exhaustive study
- **Ensemble retains** some **advantage** in PAC

REAL-WORLD EXAMPLE

14-members from NCEP ensemble

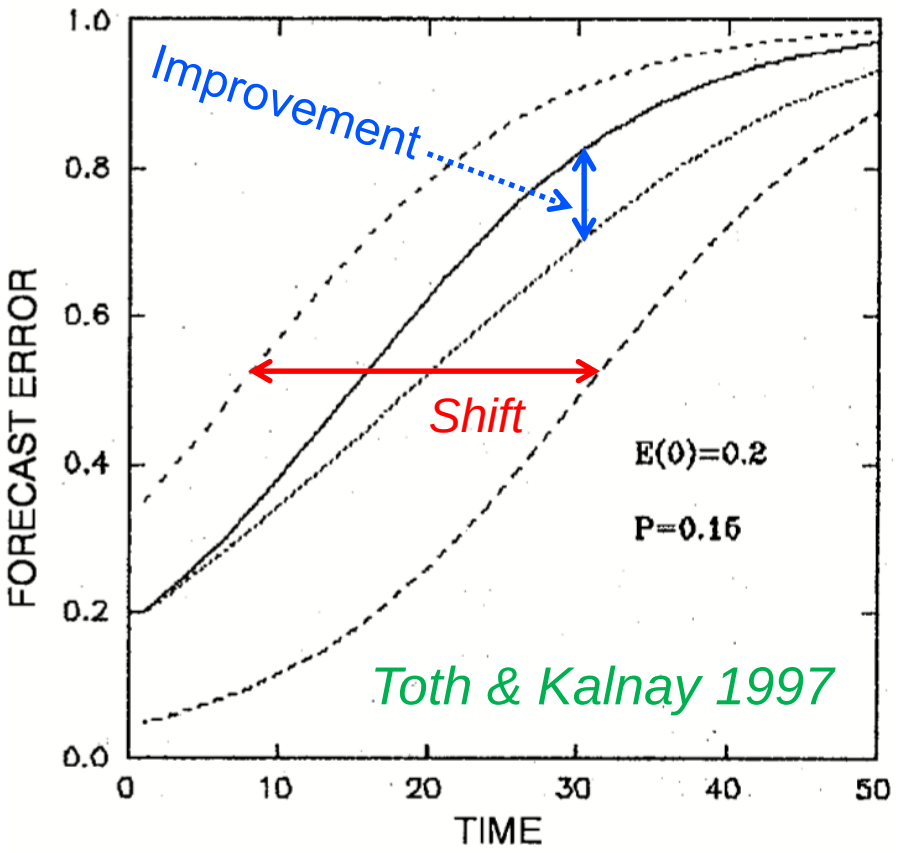
How to explain difference between error in Control vs EM?



- **EM** has lower error than **Control**
- EM saturates at lower level than Control

IMPROVED STATE ESTIMATION?

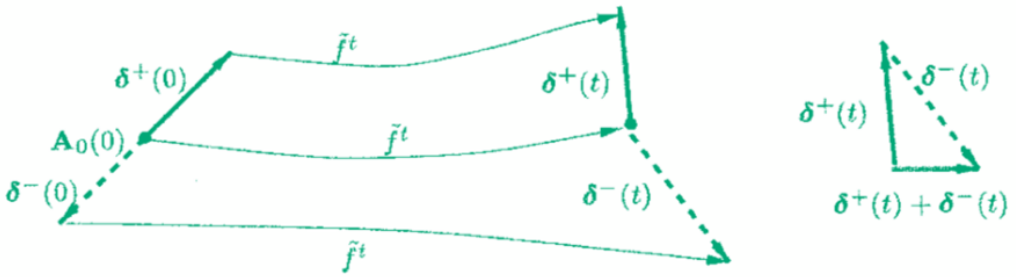
- Control (solid) & perturbed forecast errors (dashed) described by logistic curve
- **Perturbation assumed to project onto error in control**
- **Ens mean error reduced due to nonlinear filtering**



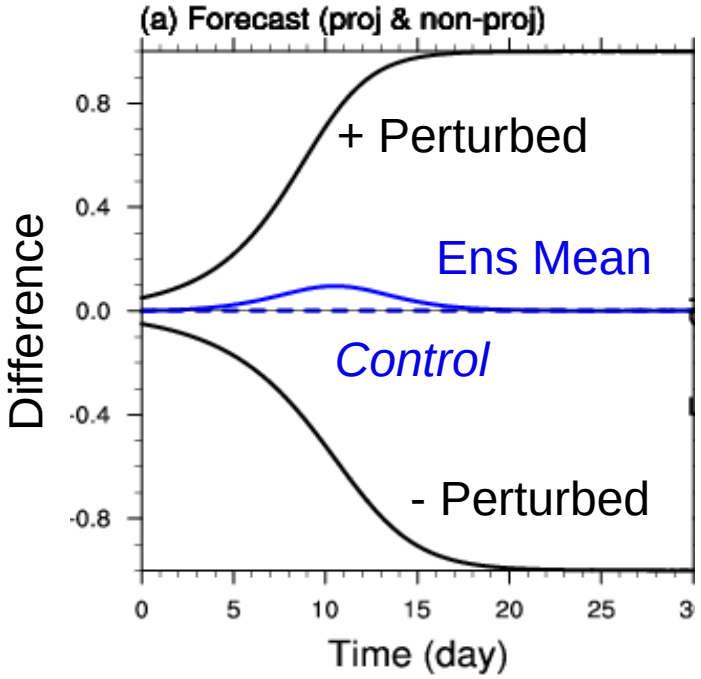
- Assesses **impact from initial perturbations** that project on error
- How much of perturbations do project onto error?
- What is effect of non-projecting perturbations?
  - Not explored yet – How much hurts?
- Effect of full saturation related filtering ignored

ISOLATE *INITIAL VALUE* RELATED NONLINEAR EFFECT

- **Symmetric pair of growing perts** centered at control
- Replace "shift of logistic curve kernel" in TK97 with
- **Differential growth** on two sides of control - *Gilmour et al '01*



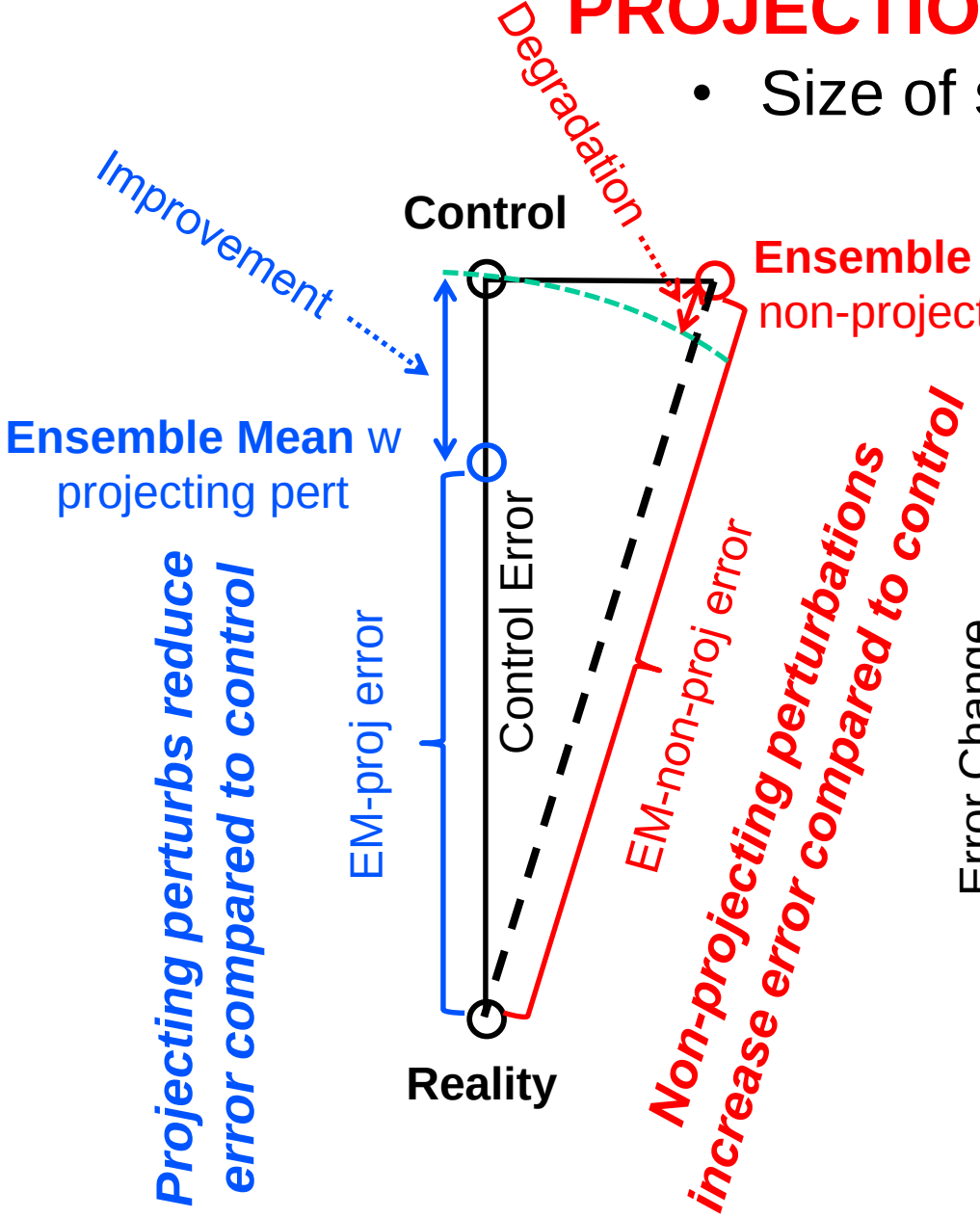
  - Ignore misalignment of pairwise perts due to "rotation" – underest.
- EM deviates from control due to nonlinearities

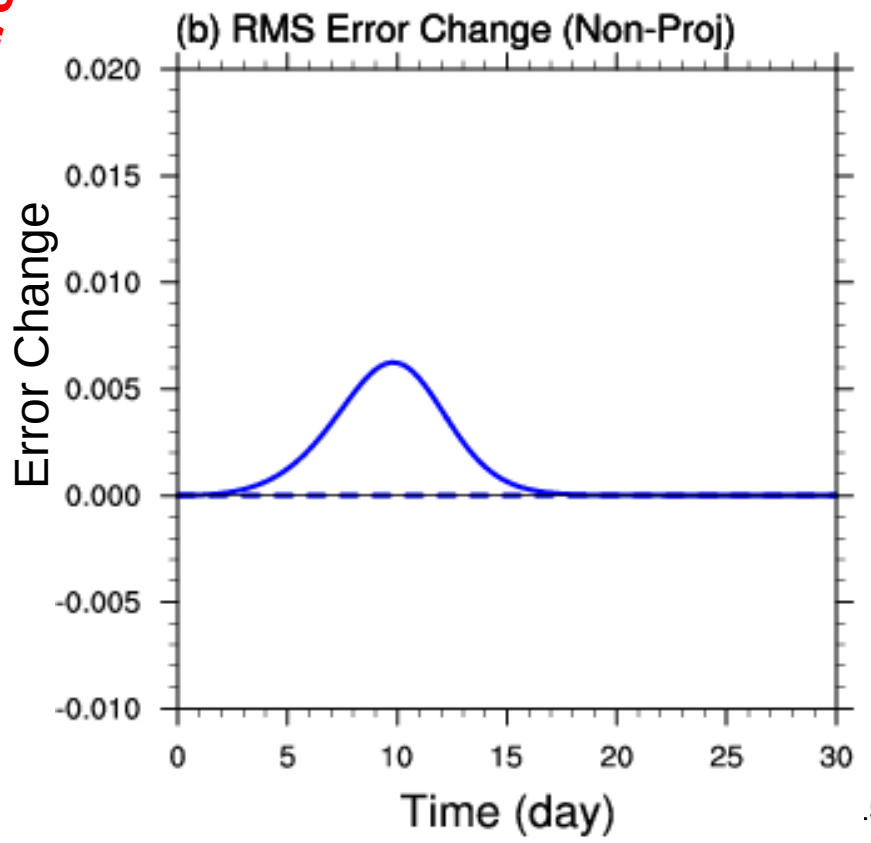


  - Evaluate expected difference connected to initial conditions
    - Ignore differences in saturated phase
- Difference btw control & ens mean related to
  - Error in ensemble mean
- Effect depends on whether perturbations
  - Do or do not project on control error

CHANGE IN CONTROL ERROR DEPENDING ON PROJECTION OF PERTS ON ERROR

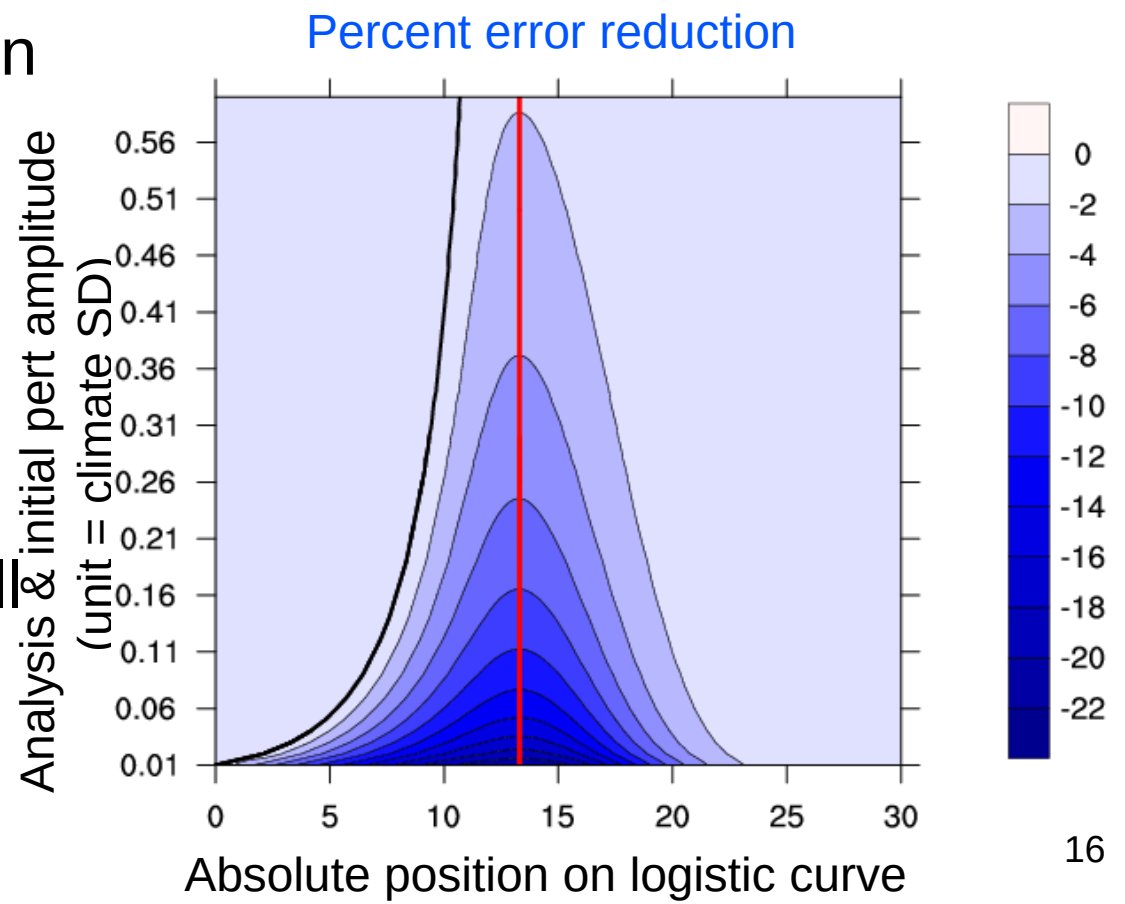
- Size of same change depends if it is
  - Aligned with *OR*
  - Orthogonal to error in control





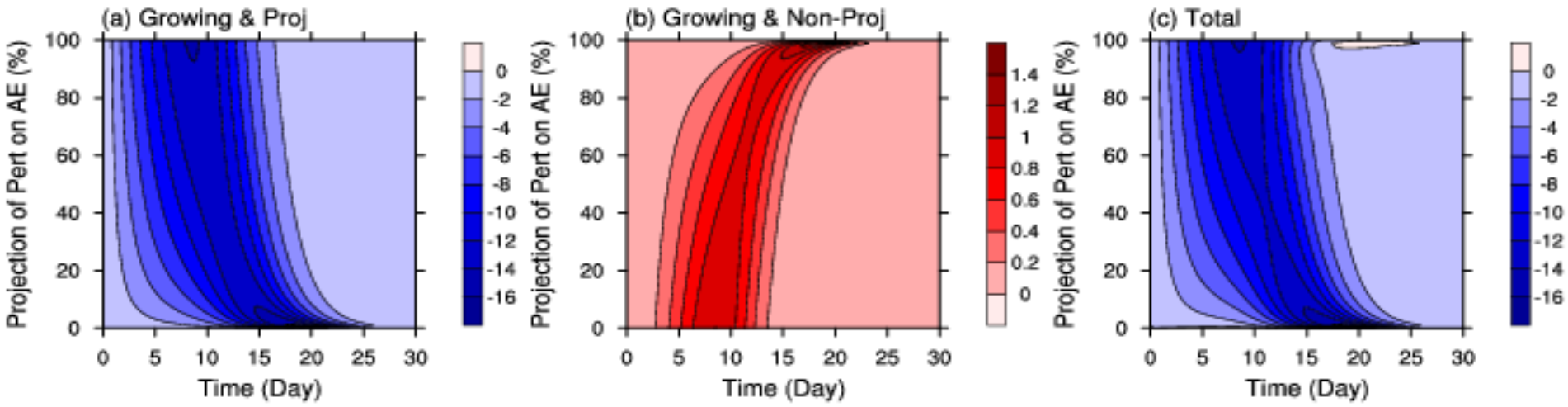
IMPACT OF PERFECT PERTURBATIONS

- Assume a pair of perfect perturbations
  - Projects 100% on error in control
  - Has same amplitude as control error
- Assess % error reduction in ens mean vs control
  - In reference to non-dim position on logistic curve
- Maximum error reduction around midpoint
- Largest error reduction for smallest analysis error
  - More time for impact to amplify
- Impact diminishes as full saturation approached
  - Initial value impact separated



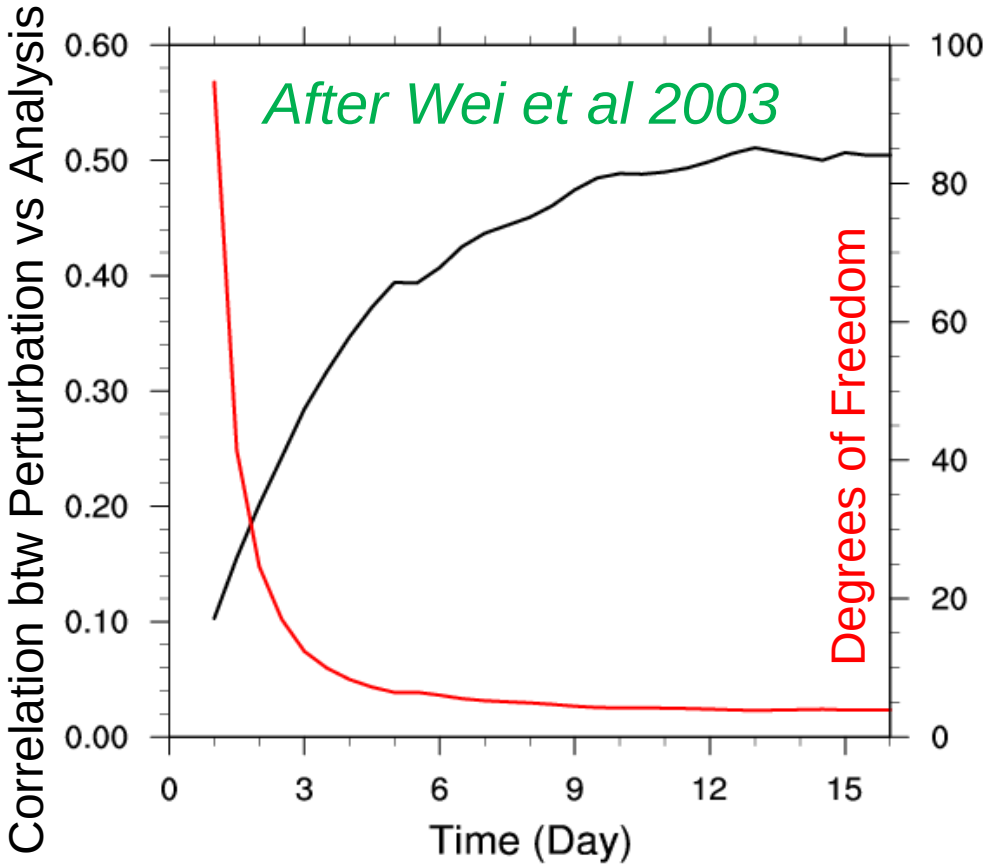
PERFECT PERT. SIZE, IMPERFECT PATTERN

- Consider analysis error amplitude is 0.05 climate SD
  - Vary how much of perturbation projects onto control error
- Assess % change in ens mean error vs control for
  - Projecting, non-projecting, total (sum)
- Error reduction due to projecting component order of magnitude larger than
  - Error increase due to non-projecting component
- Overall impact peaks at ~ 8 % reduction of error in control
  - Scale mislabeled by factor of ~1.3 due to parametric error



HOW PERTRUBATIONS PROJECT ON ERRORS?

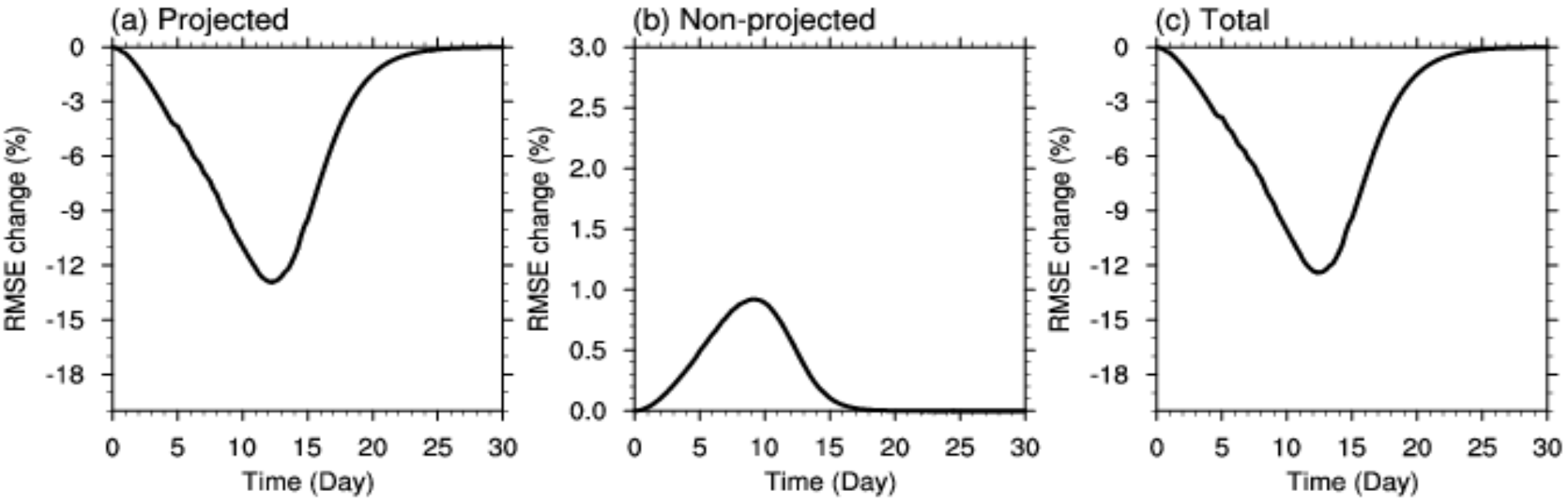
- Evaluate how correlated perturbations are w error
  - Use analysis as proxy for truth
- Projection or explained variance of perts onto error
  - Square of correlation
  - Commensurate with effectiveness of ensemble



- Correlation / projection grows w lead time
  - 1-25% projection D1-15
- Growth due to errors & perturbations "rotating" toward fastest growing directions
  - Congregate in shrinking subspace w diminishing DOF
    - *"Lyapunov effect"*

INITIAL VALUE RELATED FILTERING

- Consider typical projection of perturbations onto errors
  - 1 – 25% from short to longer lead times
- Assess change in control error due to initial value related filtering
- Projecting component of perturbation carries the day
  - 6%+ error reduction btw 9 & 16 days
    - Labels miscalibrated by factor of ~1.3



EFFECT OF MORE MEMBERS?

- So far analyzed effect of a **single pair** of perturbations on error in EM
  - Assume additional pairs statistically identical

- **EM defined** as:

$$EM = \sum_{i=1}^{n} P_i/n = \sum_{i=1}^{n} (C + p_i)/n$$

- As more pairs added, their **individual effect is reduced** by growing denominator =>

- Addition of more members has **ZERO *initial value related* impact** on quality of EM
  - May sound counterintuitive first

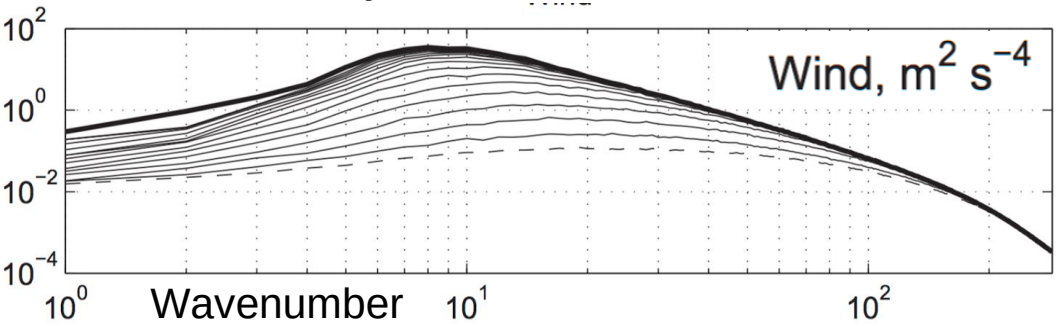
- Will assess **saturation filtering related** effect next

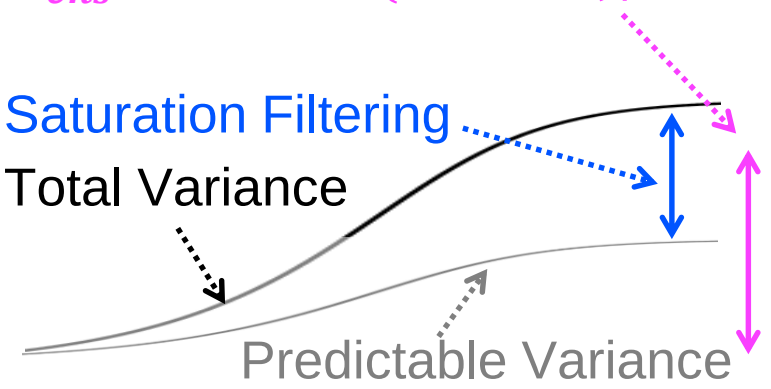
ISOLATING SATURATION RELATED FILTERING

- Perturbation/**error growth in finite systems limited by size of system**
  - Due to nonlinear interactions, error variance saturates at
    - Variance btw 2 randomly chosen states - Twice the climatic variance
- As they approach saturation, **errors become independent of initial conditions**
  - Climatic mean is best forecast at that point w an error of climate variance
- In multiscale systems, **first finest, then progressively coarser scales saturate**
  - Ensemble provides scale dependent saturation (S) related filtering
  - **Heuristic approximation:**

$$S_{ens} = 1 + Var(climate)/n$$

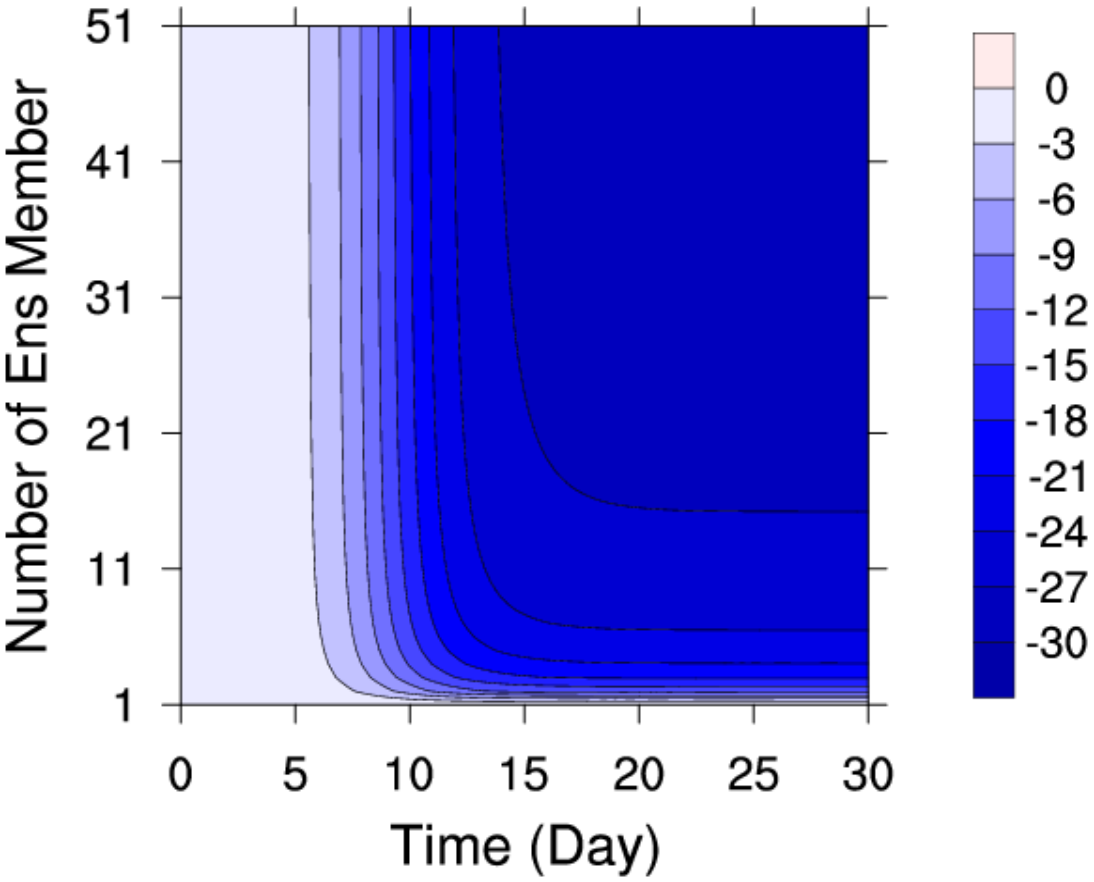
OSSE Analysis of *Prive & Errico 2015*





FILTERING DUE TO FULL SATURATION

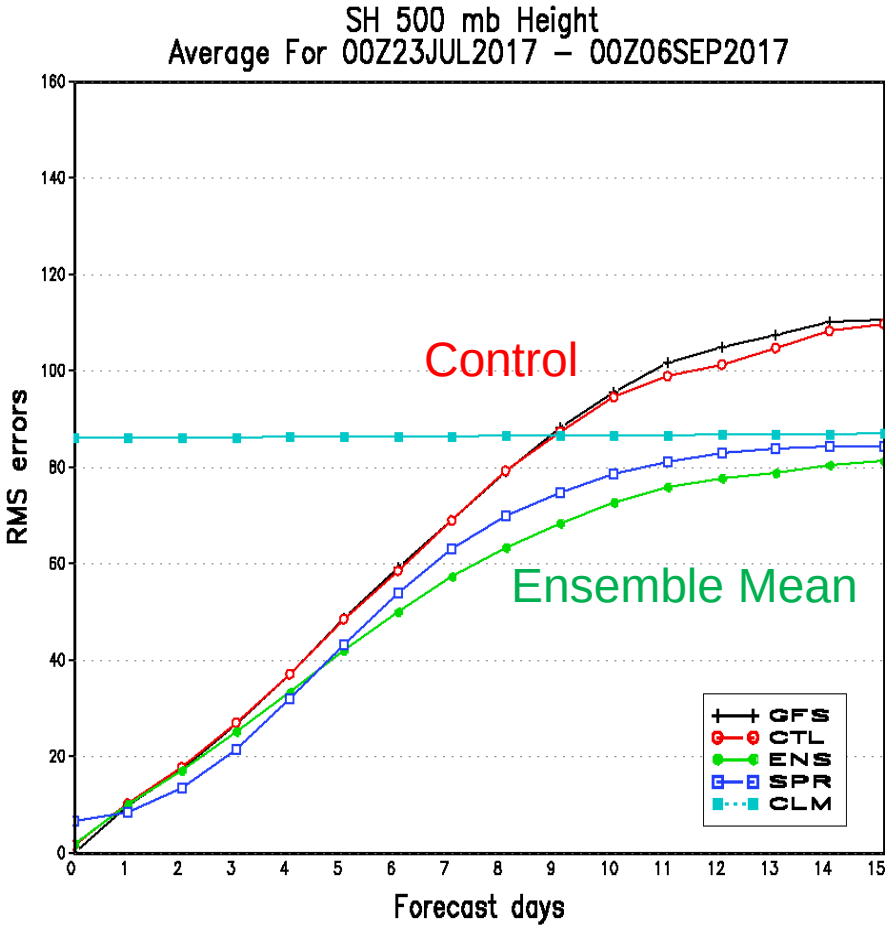
- Assume analysis error of 0.05% climate standard deviation
- Assess error reduction in EM due to elimination of all unpredictible scales



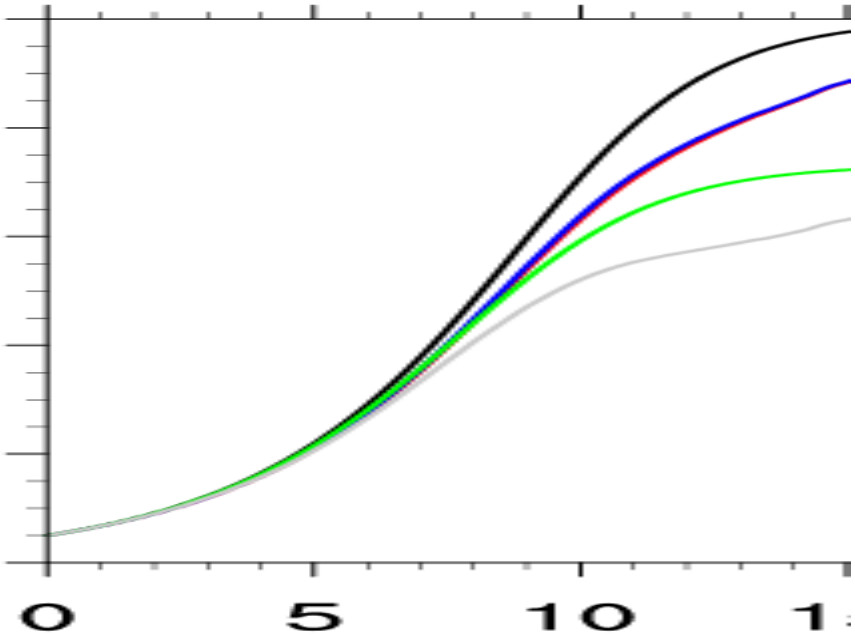
- Benefit approaches 29% maximum theoretical error reduction w 20+ members beyond 15 days
  - Negligible benefit from more than 20 members
- No benefit from more than ~10 members until ~D13
- Much larger gain than from initial value related filtering (max ~8%)

EVALUATION

- Compare predicted vs actual impact of ensemble filtering
- 14 members of NCEP ensemble
- Parameters of model not tuned for selected case
  - Non-dimensional logistic control error and predicted EM error curves stretched for qualitative comparison



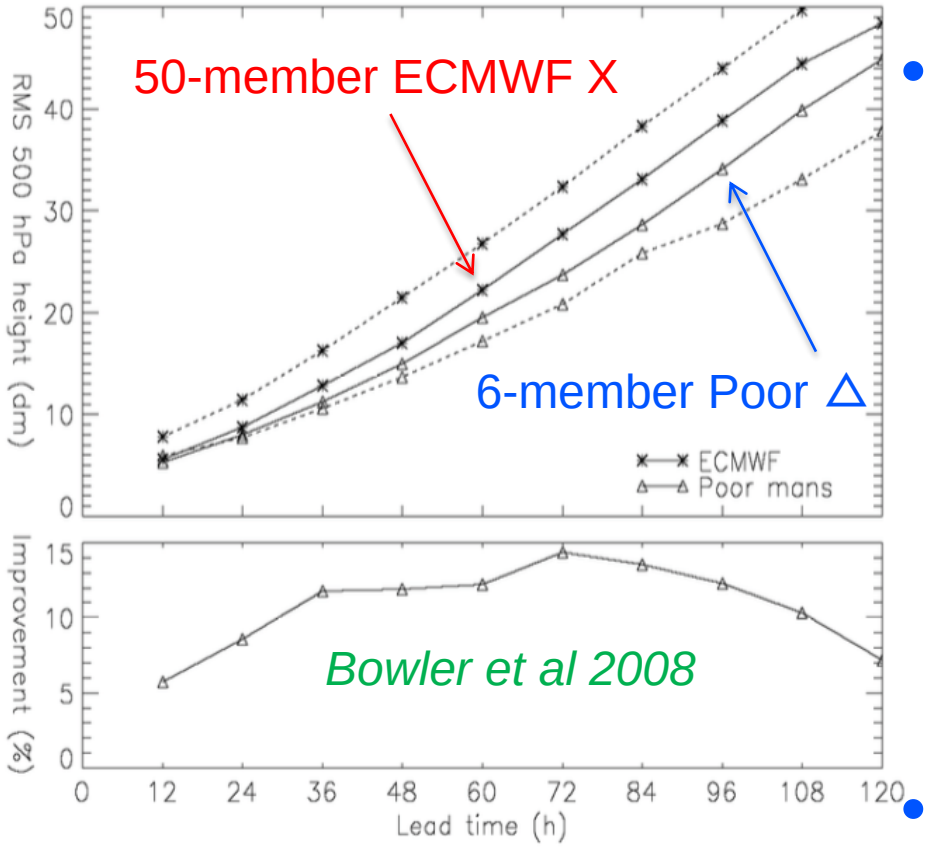
- Saturation related curve explains almost entire error reduction
- Initial value related filtering appears too large
  - Extent of nonlinearity overestimated?



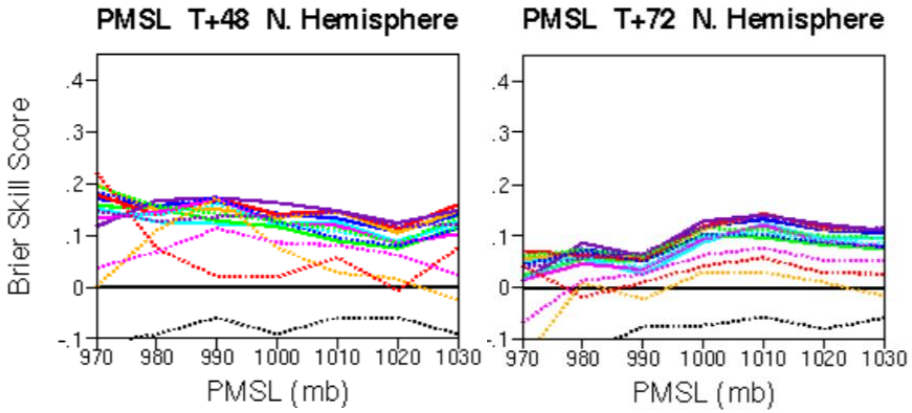
ENSEMBLE AROUND WHAT?

- **Around single best** (control) **analysis**
  - Works only when perturbation projects onto error
  - Yet this concept is considered ”the proper” formation of an ensemble
    - ”Proper” mistaken for “intentional”
      - Not all what’s intended works
- **Around proxy for truth** – “cloud of observations”
  - Set of independently created analysis fields
  - “Perturbations” by definition project onto error =>
    - Mean of initial perturbations closer to reality?
  - “Poor-person’s” ensemble w built-in model diversity
    - Unperturbed forecasts from multiple centers
  - Focus on spread / probabilistic info (the “dress”)
    - State estimate (ens. mean) ignored except one study?

ALTERNATIVE – POOR-PERSON ENSEMBLE



- **State estimation** – Core value
  - 6-member Poor ens. beats 50-member ECMWF
  - Effect of initial values (or models)?



- **Probability of MSLP events**
  - Poor ensemble beats ECMWF for most thresholds

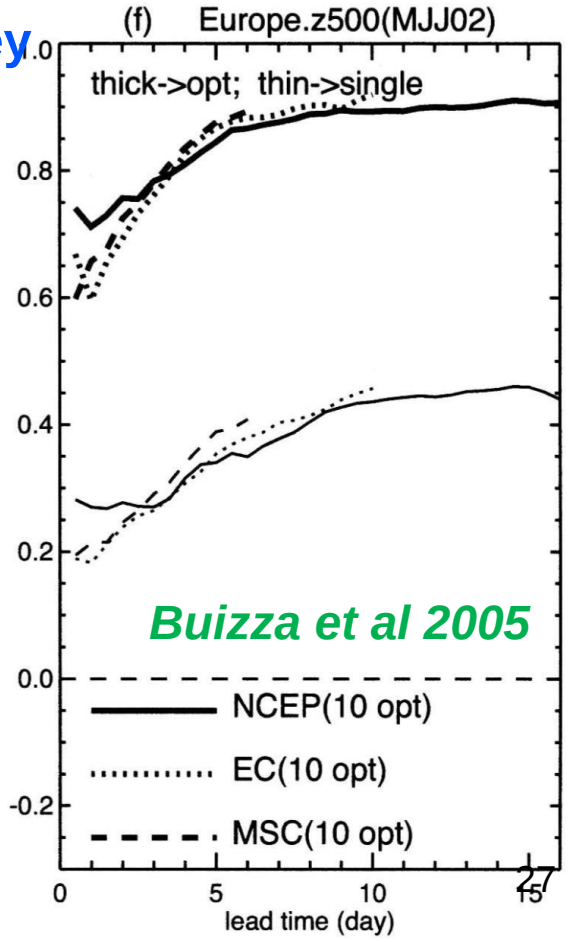
OTHER ALTERNATIVES

Voice of contrarian (at ensemble meeting)

- **State estimation**
  - Scale dependent filtering of control like TK97?
  - Other way of using info from members?
- **Error variance estimation**
  - Statistics of error around control (MOS, etc)
    - Statistically derived “dress” around control
- **Probabilities**
  - Based on “dress”
- **Scenarios / covariances**
  - More advanced statistical methods?

CHOICE OF INITIAL PERTURBATIONS

- EM study highlights benefits from **maximizing projection of perturbations on analysis error**
- Analysis error at any time
  - Instantaneous manifestation of DA-forecast cycle
    - **Dynamical amplification & perpetuation of growing errors**
- Small / large subspace of growing / decaying perturbations
  - **Large projection of perts on growing errors is key**
- **Cycled Perturbation (CP) schemes** such as BV or ETR show higher projection at short lead times than
  - SV or multiple analysis schemes
- Characteristics / potential benefits of CP schemes
  - Minimize noise, maximize growing perts
  - Temporal continuity for downstream ensemble applications
  - Can use SAFE estimates of analysis error variance
    - *Pena & Toth 2014, Feng et al 2017*



SUMMARY

- Attempted **parametric description** of effect of ensemble filtering
- **Separated effects** of
  - Initial value (IV) related filtering of predictable scales
    - Independent of number of members
  - Full saturation (FS) related filtering of unpredictable scales
    - Driven by number of members
- **FS dominates results** and explains most gain in NCEP ensemble
  - Significant effect at mid- and loner ranges
- **IV filtering maxes in mid lead-time range**
  - Only minor degradation from non-projecting perturbations
  - Explained error variance as metric for perturbations
- Reviewed **benefits of cycled perturbation schemes**

DISCUSSION

- **Qualitative similarity** btw paramet. model & NCEP ensemble results
  - Model's parameters not tuned to specific application
- **Deploy dynamically generated ensembles wisely** – Balance btw
  - Costs (N times increase)
  - Benefits – sometime marginal
- **Initial value related benefits** pronounced in mid lead-time range
  - Questionable if use restricted to short or long leads only
- **Error & perturbations evolve in small (~5-dim) subspace**
  - Can large ensembles be justified?
  - How much saturation rel. filtering is reproducible statistically?
- Consider **alternatives if warranted by cost/benefit analysis**
  - Use ensemble at intermediate time scales when nonlinear filter most effective
  - Consider statistical alternatives when focus on short or long lead times
    - Room for innovative approaches

BACKGROUND