The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2050-3

Targeted Training Activity: Predictability of Weather and Climate: Theory and Applications to Intraseasonal Variability

27 July - 7 August, 2009

Errors in Forecasts: Roles of Initial States, Model Errors, and Chaos

David Straus

Center for Ocean-Land-Atmosphere Studies/Goerge Mason University (COLA-GMU)
Institute for Global Environment & Society (IGES)
Calverton, MD
USA

International Centre for Theoretical Physics

Targeted Training Activity:

Predictability of Weather and Climate

David M. Straus

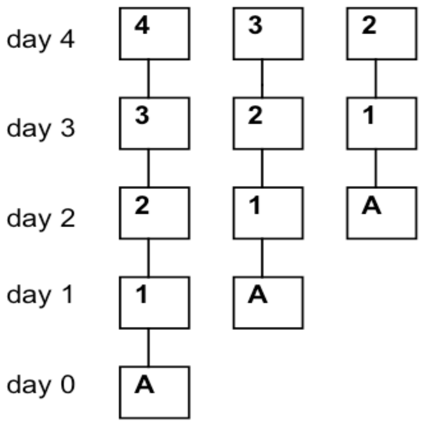
(1) The Global Atmospheric Circulation: Observations

(2) Modeling the Weather and Climate

(3) Errors in Forecasts: Roles of Initial States, Model Errors, and Chaos

(4) Climate Predictability on Seasonal Time Scale: Role of Boundary Forcing

(5) Seasonal Mean Predictability over the Pacific - North American Region



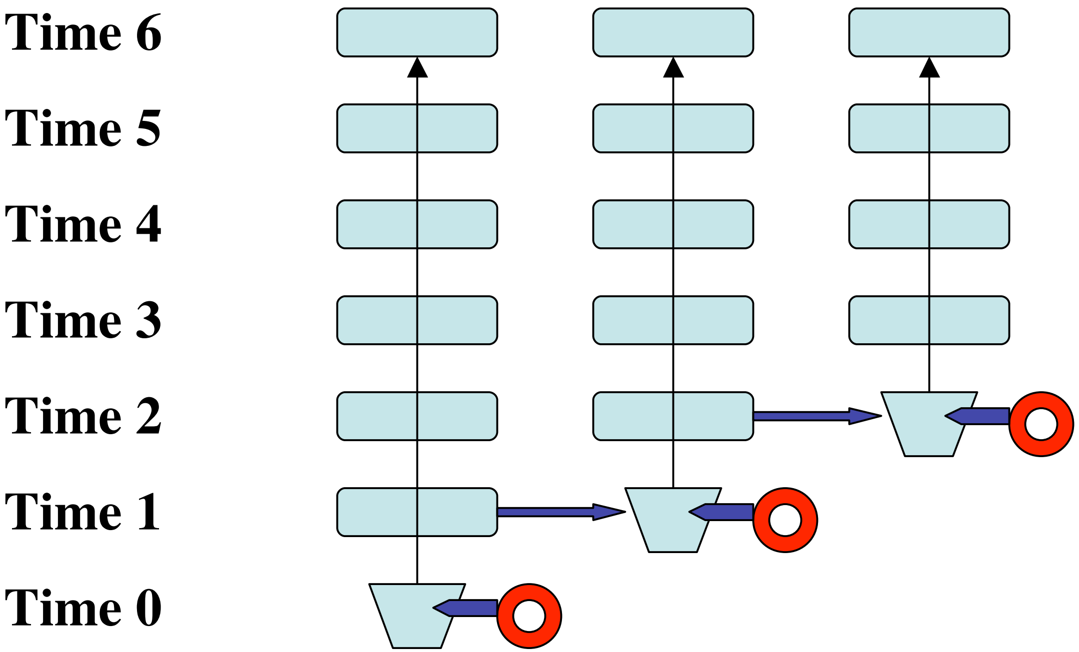
Schematic of Analysis Forecast System. The bold letter A within a box refers to an analysis for the given day, that is the estimate of the real state of the atmosphere on that day. The bold numbers in the boxes refer to the range of the forecast. The days labeled on the left refer to the verifying time

Analysis and Forecast

Analysis is the estimation of the current state of the atmosphere, expressed as a state of a numerical model (denoted by "A" in diagram).* The analysis is expressed in terms of all model prognostic variables, on the model's horizontal and vertical grid.

Forecast is a projection into the future made using a numerical model from an initial state given by the Analysis.

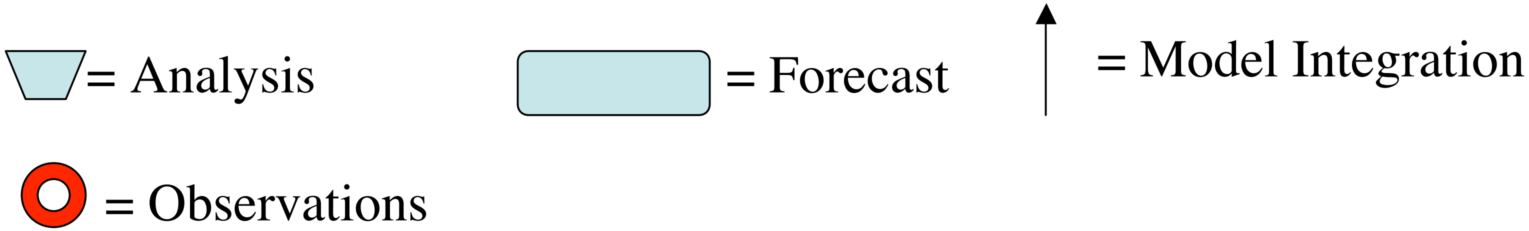
* In practice, the analysis starts from a previous very short range (6 hour) forecast - the model variables are changed to be consistent with the current observations for those areas/levels/ variables which are observed



(1) Model forecast provides "first guess" for use in the analysis

(2) Observations modify first guess of atmospheric state to produce an analysis

(3) The analysis is the initial condition for the subsequent forecast



The difference between a forecast which has been run for N days and the analysis at the end of the N days (the so-called "verifying analysis") is called the forecast error.

Forecast error has several components:
(1) analysis error: The initial conditions obtained from the analysis may have errors that are not small.
(2) model error: The model itself has physical errors.
(3) predictability error: Any (inevitable) small errors in the analysis will amplify with time.

Definition of predictability error:
The difference between two model forecasts started from initial conditions very close to each other. (The predictability error measures the forecast error that would be seen if the models were perfect and the analysis very good.)

Identical twin model configuration:
Run forecasts with the same model, but with initial states close to each other.

It is generally believed that with current NWP forecast models (e.g. ECMWF), the forecast error *for the first few days* is dominated by the analysis error.

Weather Predictability:
Growth of Errors in Operational Forecast Models

Lorenz, E. N., 1982: Atmospheric Predictability Experiments with a Large Numerical Model. *Tellus*, **34,** 505-513.

Dalcher, A., and E. Kalnay, 1987: Error Growth and Predictability in Operational ECMWF Forecasts. *Tellus*, **39A,** 474-491.

Simmons, A., and A. Hollingsworth, 2002: Some Aspects of the Improvement in Skill of Numerical Weather Prediction, *Quart. J. Roy. Meteor. Soc.*, **128,** 647-678.

Operational ECMWF forecasts and analyses for winter of 1980 / 1981

100 - day period starting from 1 December 1980

For each day we have the analysis, and forecasts starting from that analysis for 1 day, 2 days, ... 10 days.

Consider measure of error between two forecasts at any time to be the square root of the global mean of the squared difference between 500 hPa height (Z). This error is called the rms error.

Consider two forecasts one started j days earlier, one started k days earlier, both verifying for (valid for) the current time.

Example:
Forecast A is a 2-day forecast started on 23 December (j=2)
Forecast B is a 3-day forecast started on 22 December (j=3)

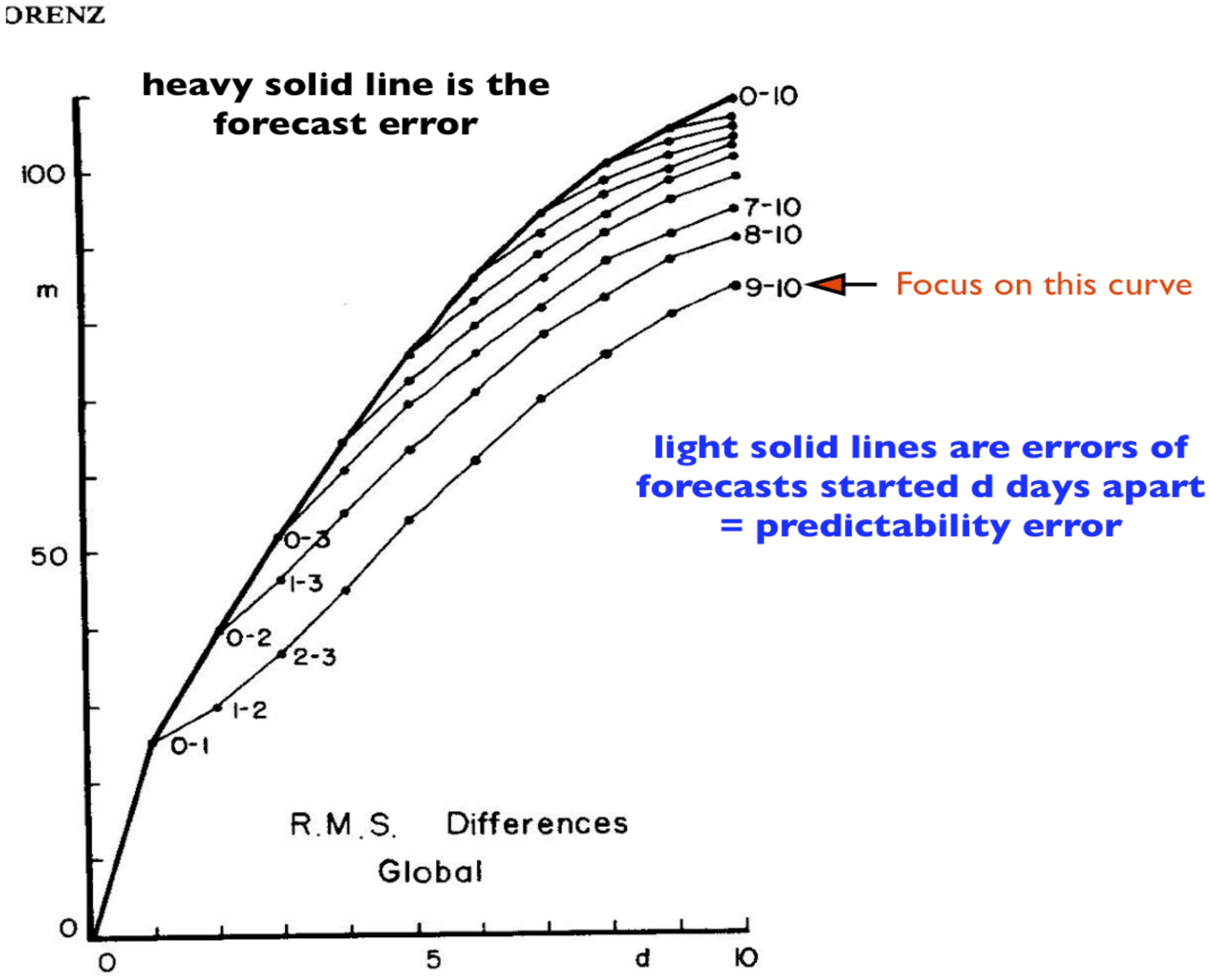
The "error" in height between these two forecasts valid for Dec. 25 is:

$$E_{j,k}^2 = \frac{1}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\phi \int_0^{2\pi} d\lambda \cos(\phi) \ (Z_A - Z_B)^2$$

where ϕ is latitude and λ longitude.

$E_{(j,k)}$

j=0 is the analysis



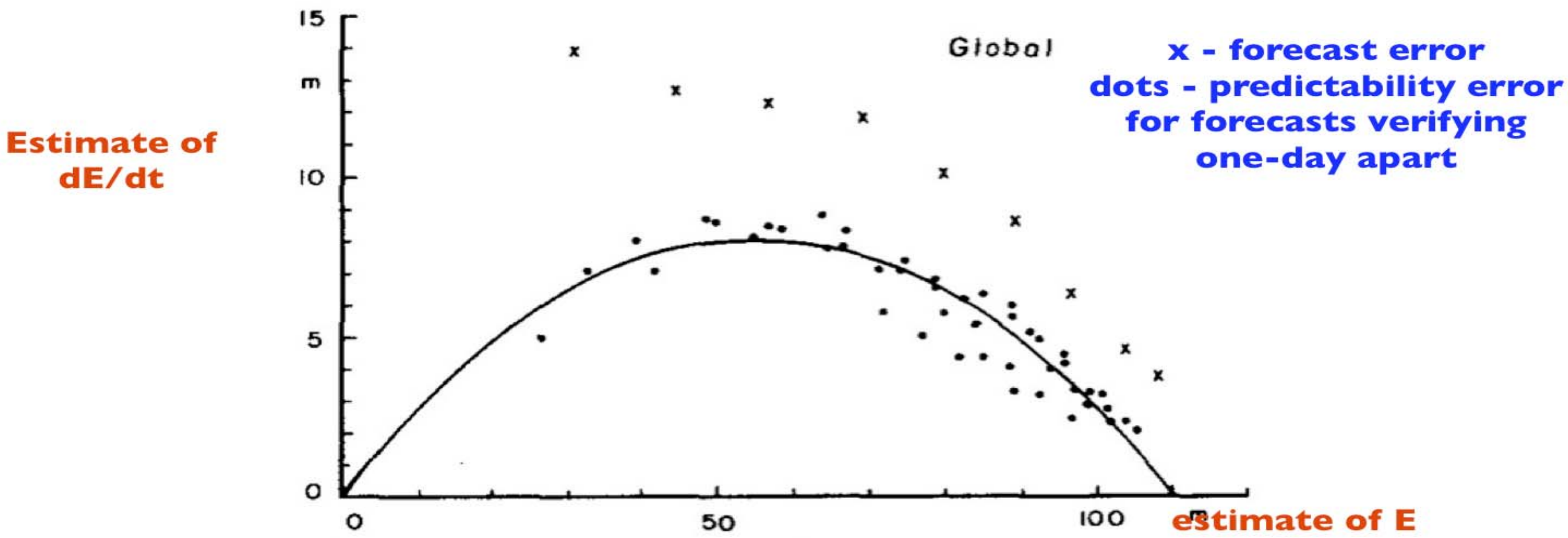


Fig. 2. Increases in global root-mean-square 500-mb height differences, $E_{j+1,k+1} - E_{jk}$, plotted against average height differences $(E_{j+1,k+1} + E_{jk})/2$, in meters, for each one-day segment of each thin curve in Fig. 1 (large dots), and increases $E_{0,k+1} - E_{0k}$ plotted against average differences $(E_{o,k+1} + E_{0k})/2$, for each one-day segment of heavy curve in Fig. 1 (crosses). Parabola of "best fit" to large dots is shown; see text.

previous figure suggests a parabolic form for error growth:

$$\frac{dE}{dt} = aE - bE^2$$

E = rms of 500 hPa z

$$Lim_{t \to \infty} E = E_\infty$$

At large time, errors should saturate

$$aE_\infty = bE_\infty^2 \qquad E_\infty = a/b$$

$$\frac{d\varepsilon}{dt} = a\varepsilon - bE_\infty \varepsilon^2 = a\left(\varepsilon - \varepsilon^2\right) = a\varepsilon\left(1 - \varepsilon\right) \qquad \varepsilon \equiv E/E_\infty$$

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = \left(\frac{1}{1-\varepsilon}\right)\frac{d\varepsilon}{dt} + \frac{\varepsilon}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt} = \frac{1}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt}$$

or

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = a\frac{\varepsilon}{1-\varepsilon}$$

defining: **we have**

$$f \equiv \frac{\varepsilon}{1-\varepsilon} \qquad \frac{df}{dt} = af \qquad f = e^{a(t-t_0)}$$

(here t_0 has a specific meaning)

$$\varepsilon = \frac{f}{1+f} = \frac{e^{a(t-t_0)}}{1+e^{a(t-t_0)}} = \frac{e^{\frac{1}{2}a(t-t_0)}}{e^{-\frac{1}{2}a(t-t_0)} + e^{\frac{1}{2}a(t-t_0)}}$$

t_0 is the time at which E = 1/2

or

$$\varepsilon = \frac{1}{2}\left[1 + \tanh\left(\frac{1}{2}a(t-t_0)\right)\right]$$

Some identities:

$$tanh(x) \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$1 + tanh(x) = \frac{2e^x}{e^x + e^{-x}} \qquad \frac{1}{2}(1 + tanh(x)) = \frac{e^x}{e^x + e^{-x}}$$

for **we have** **or**

$$\varepsilon \ll 1 \qquad \frac{d\varepsilon}{dt} = a\varepsilon \qquad \varepsilon = ce^{a(t-t_0)}$$

the doubling time for small errors τ is defined by:

$$\frac{\varepsilon(t_2)}{\varepsilon(t_1)} = e^{a(t_2-t_1)} = 2 \qquad \text{or} \qquad (t_2 - t_1) \equiv \tau = \ln(2)/a$$

doubling time for small errors is 2.42 days

Application of New Error Growth model to same ECMWF forecast data set used by Lorenz (Dalcher and Kalnay 1987)

Some Differences in Approach

(1) Error refers to globally averaged mean squared difference of 500 hPa height
(2) Model actual forecast errors, not only identical twin forecast errors
(3) Remove model "systematic error", that is the mean climate error.

Let z_f be the forecast height field for a given forecast range (1 to 10 days) and z_v the analysis height field. An average over the whole dataset of forecasts (and analyses) is denoted by an overbar, and a deviation from that average by a prime. Then the systematic error is:

$$\overline{z_f} - \overline{z_a}$$

and the random error is:

$$z_f' - z_a'$$

Then the total mean squared error can be written as:

$$\overline{(z_f - z_a)^2} = (\overline{z_f} - \overline{z_a})^2 + \overline{(z_f' - z_a')^2}$$

where the first term is the square of the systematic error and the second the "random error variance", also referred to as simply the error variance. Note that the error variance defined here is still a function of latitude and longitude.

Note: **A statistical forecast** **of interest is that you forecast climatology, that is:**

$$z_s = \overline{z_a}$$

where the subscript "s" denotes statistical. The mean squared error in this statistical forecast is just the variance of the analyses:

$$\overline{(z_a - z_s)^2} = \overline{(z_a - \overline{z_a})^2} = \overline{z_a'^{\,2}}$$

The internal error variance describes the difference in error between two model forecasts (also called identical twin error) is just:

$$\overline{(z_1' - z_2')^2}$$

where the subscripts 1 and 2 denote two model forecasts started from different analyses. At large time, the two forecasts become statistically independent, so that this error becomes:

$$\overline{(z_1')^2} + \overline{(z_2')^2} = 2\overline{(z')^2}$$

or twice the random error variance.

Solution of three-term error growth model
We rewrite the model in a slightly different form, following Kalnay and Dalcher:

$$\frac{dE}{dt} = (AE + S)\left(1 - \frac{E}{E_\infty}\right)$$

where E_∞ is the value of E as t $\rightarrow \infty$

and S is *hypothesized* by KD to be related to GCM error

with

$$\frac{e}{E_\infty} = \varepsilon$$

$$\frac{S}{E_\infty} = \sigma$$

Note that σ is a growth rate (units of inverse time)

we have:

$$\frac{d\varepsilon}{dt} = (A\varepsilon + \sigma)(1 - \varepsilon)$$

$$\frac{d}{dt}\left(\frac{\varepsilon}{1-\varepsilon}\right) = \frac{1}{(1-\varepsilon)^2}\frac{d\varepsilon}{dt} = \frac{1}{(1-\varepsilon)}(A\varepsilon + \sigma)$$

$$\frac{df}{dt} = A\frac{\varepsilon}{(1-\varepsilon)} + \sigma\frac{1}{(1-\varepsilon)} = Af + \sigma(1+f) = (A+\sigma)f + \sigma$$

with: $f = \frac{\varepsilon}{(1-\varepsilon)}$ $\qquad \varepsilon = \frac{f}{(1+f)}$ $\qquad 1 + f = \frac{1}{1-\varepsilon}$

The solution is given by the sum of the general homogeneous solution plus the particular solution:

$$f = Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}$$

$$\varepsilon = \frac{Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}}{1 + Ce^{(A+\sigma)t} - \frac{\sigma}{A+\sigma}}$$

$$\varepsilon = \frac{Ce^{(A+\sigma)t} + \frac{A}{A+\sigma} - 1}{Ce^{(A+\sigma)t} + \frac{A}{A+\sigma}}$$

$$\varepsilon = 1 - \frac{1}{Ce^{(A+\sigma)t} + \frac{A}{\sigma+A}} = 1 - \frac{A+\sigma}{A + (\sigma+A)Ce^{(A+\sigma)t}}$$

where C can be related to the initial error. If the term S = 0, we just obtain:

$$\varepsilon = 1 - \frac{1}{1 + Ce^{(A+\sigma)t}}$$

Error Growth Model Applied to each total wavenumber n separately

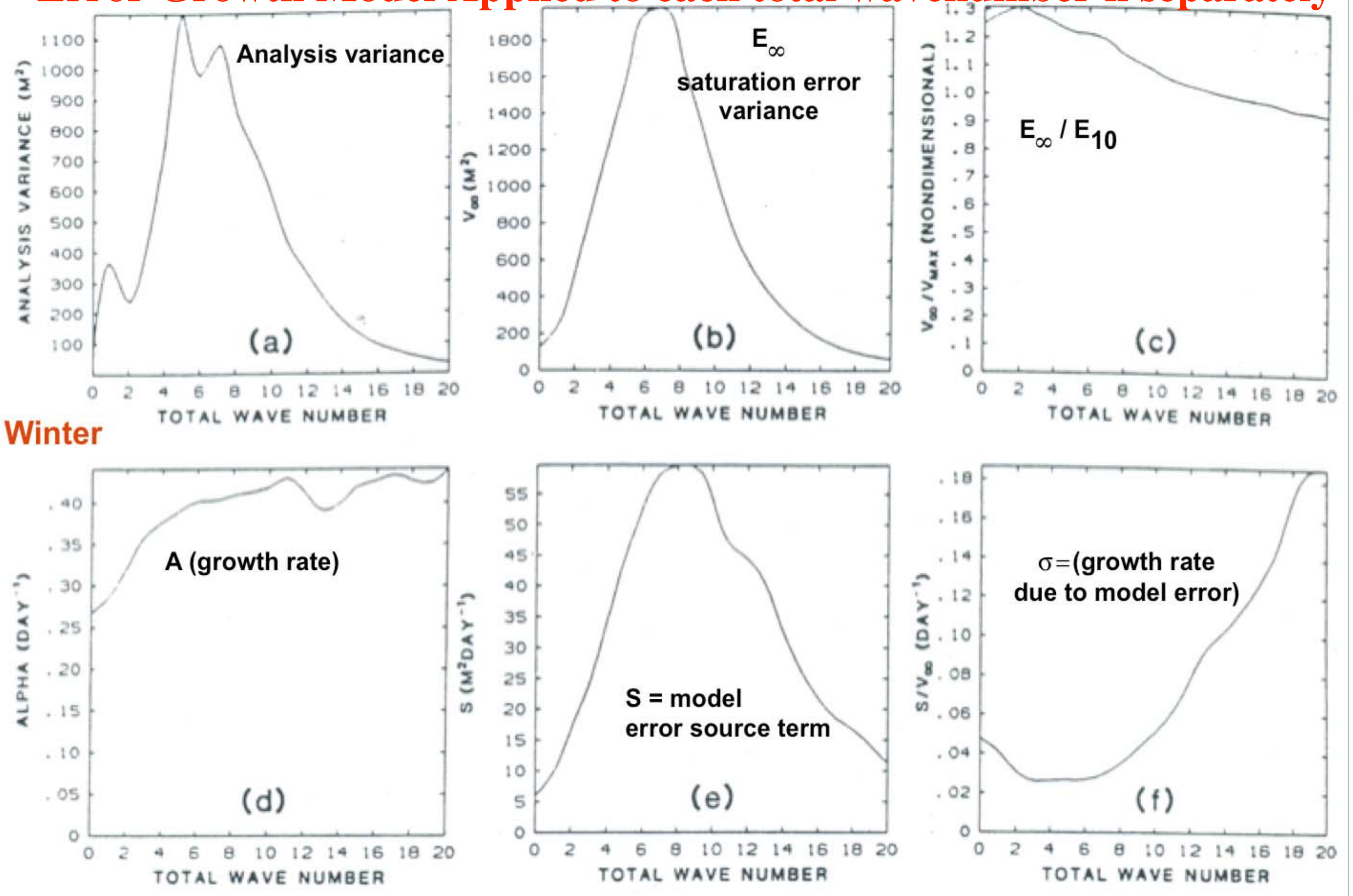


Fig. 13. Fitted parameters (three-parameter model) as a function of total wave number for the winter. (a variance; (b) V_∞, the saturation error variance; (c) V_∞ divided by $V(10)$, the error variance at day 10 growth rate; (e) S, the source of errors due to model deficiencies (f) S/V_∞.

Estimating the Effects of Initial Condition (Analysis) Error and Model Error

-We don't know what the analysis error is!

- We do have a number of analyses available to us for any day: Analyses using difference numerical models for the first guess (and different ways of assimilating the observed data) give us a measure of the uncertainty in the analysis.

- Consider the analysis for a given day produced using model A for the first guess, and one using model B for the first guess. Call these Analysis A and Analysis B.

  -From each of these analyses we can make a forecast using any numerical model

  -Specifying the forecast means specifying both the analysis and the model.

- Thus we can compare the forecast made from **Analysis A using Model B** with the forecast made from **Analysis B using Model B**. This gives one estimate of the **error due to the initial condition.**

- Likewise, we can compare the forecast made from **Analysis A using Model B** with the forecast made from **Analysis A using Model A** to get an estimate of the **error due to the the model.**

(a) Difference between (UK anal + UK model) and (EC anal + EC model) forecast

(b) Difference between (UK anal + EC model) and (EC anal + EC model) forecast

(c) True forecast error: (EC anal + EC model) minus EC Verifying analysis

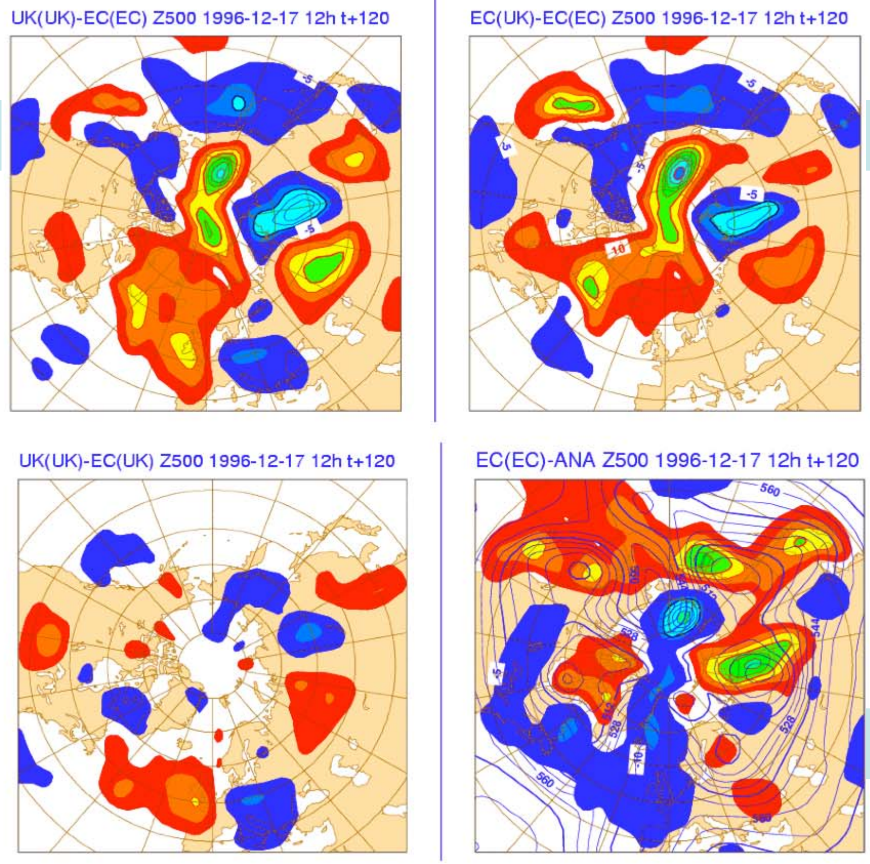
What is the relative contribution of initial and model uncertainties to forecast error?

Richardson (1998, QJRMS) have compared forecasts run with two models (UKMO and ECMWF) starting from either the UKMO or the ECMWF ICs. Results have indicated that initial differences explains most of the differences between ECMWF-from-ECMWF-ICs and UKMO-from-UKMO-ICs forecasts.

5-day forecast of Z500

(a)

(b)



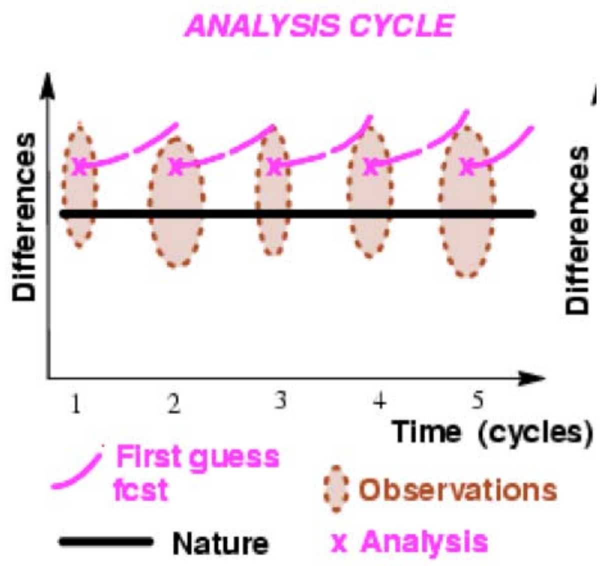
(c)

Ensemble Forecasts: Recognizing that we Do Not Know the True Initial State

-Since the analysis is an approximation to the true atmospheric state, we should also consider other "nearby" states as possible initial conditions for a forecast.

- The idea is to run *a number of forecasts* from initial states that are all consistent with the observed data (but which may differ from the analysis slightly).

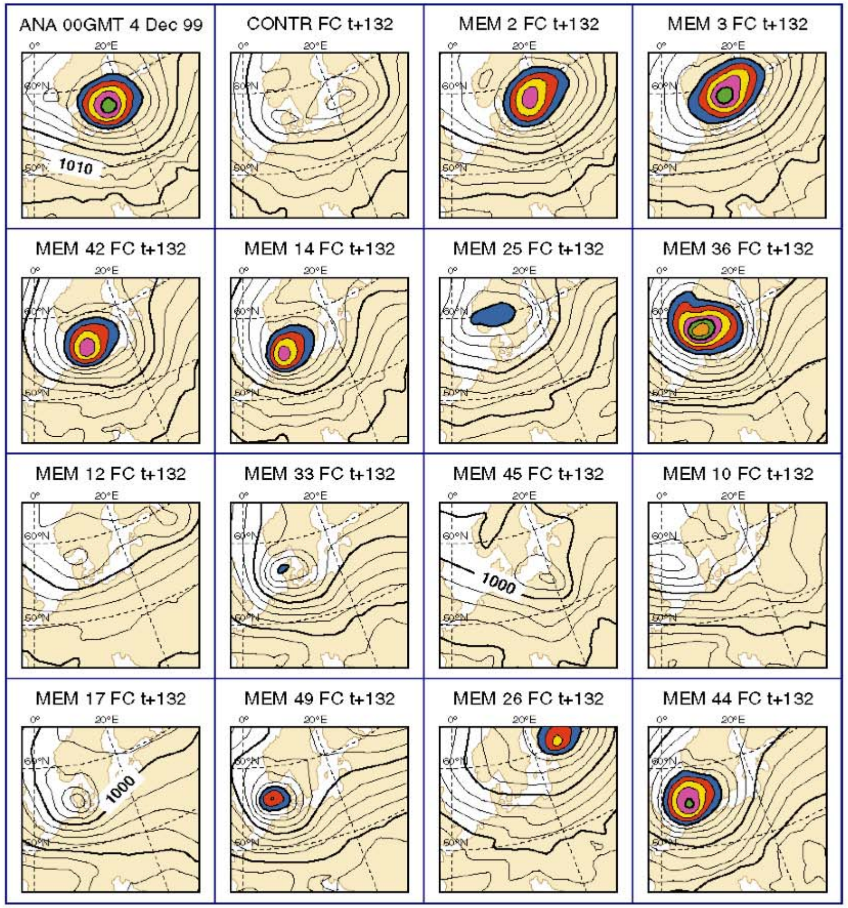
- This will help us to capture dramatic behavior in the real atmosphere that might be missed by a single forecast. We expect sensitivity to small changes in initial state due to chaos!



1. The atmosphere exhibits a chaotic behavior

A dynamical system shows a chaotic behavior if most orbits exhibit sensitivity to initial conditions, i.e. if most orbits that pass close to each other at some point do not remain close to it as time progresses.

This figure shows the verifying analysis (top-left) and 15 132-hour forecasts of mean-sea-level pressure started from slightly different initial conditions (i.e. from initially very close points).



Schematic of ensemble prediction

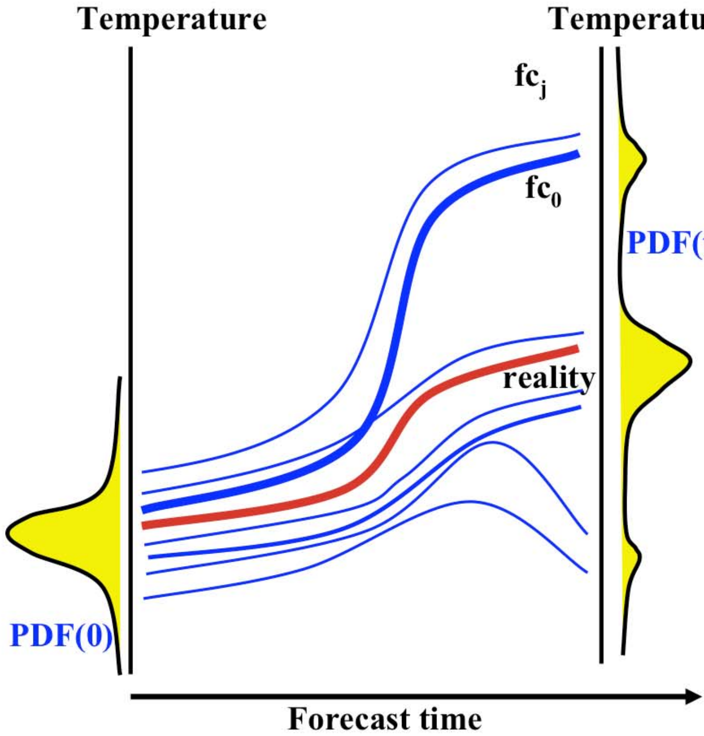
Two are the main sources of error growth: initial and model uncertainties.

Predictability is flow dependent.

A complete description of weather prediction can be stated in terms of an appropriate probability density function (PDF).

PDF(0) = Probability of Real temperature given only the analysis, which is an estimate of the temperature.

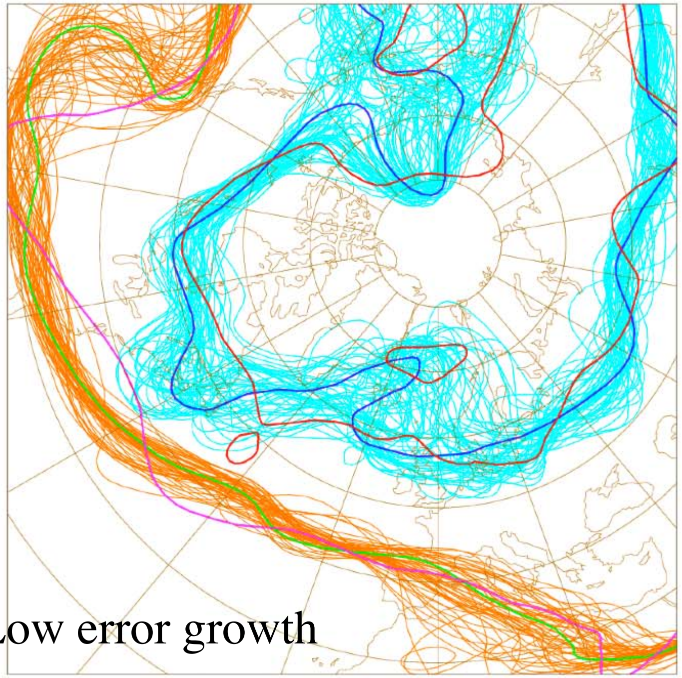
PDF(t) = Probability of Forecast Temperature using an ensemble of forecasts



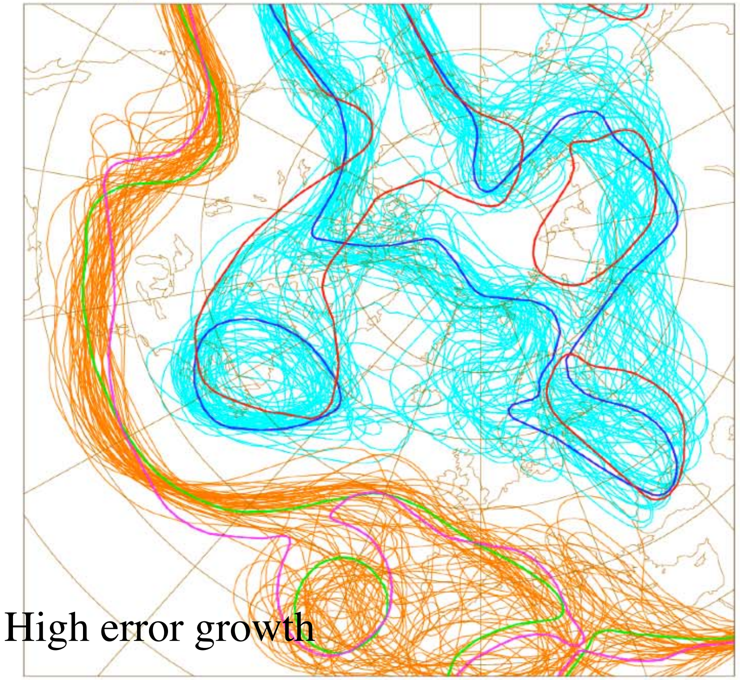
Predictability is flow dependent: 5-day forecasts of Z 500 for 2 cases

5700 m isolines: Analysis **Single control forecast** **Ensemble forecasts**

5200 m isolines: Analysis **Single control forecast** **Ensemble forecasts**



Low error growth



High error growth

ECMWF ICTP Conference & School on Predictability (July 2007) – *Roberto Buizza: Sources of uncertainty (L1)* 11

Defining Initial Conditions for Ensemble Prediction: Singular Vectors

1. The different initial conditions are obtained by adding(small) perturbations to the analysis
2. The perturbations should be those that grow the most rapidly, so that you can sample different possible forecast behaviors
3. One way to choose these perturbations is given by **Singular Vectors**
4. Singular vectors are defined as the perturbations which grow the **most rapidly over a fixed forecast interval** (usually a few days), using the equations of motion *linearized about the current state*
5. Since predictability depends strongly on initial state, the singular vectors must be computed for each initial state (for each analysis)

Defining Initial Conditions for Ensemble Prediction: Bred Vectors

At NCEP a different strategy based on perturbations growing fastest in the analysis cycles (bred vectors, BVs) is followed. The breeding cycle is designed to mimic the analysis cycle.

Each BV is computed by (a) adding a random perturbation to the starting analysis, (b) evolving it for 24-hours (soon to 6), (c) rescaling it, and then repeat steps (b-c).

BVs are grown non-linearly at full model resolution.