INTERCOMPARISON OF THE CANADIAN, ECMWF, AND NCEP ENSEMBLE FORECAST SYSTEMS

Zoltan Toth[3],

Roberto Buizza[1], Peter Houtekamer[2],

Yuejian Zhu[4], Mozheng Wei[5], and Gerard Pellerin[2]

[1] : European Centre for Medium-Range Weather Forecasts, Reading UK (www.ecmwf.int)

[2] : Meteorological Service of Canada, Dorval, Quebec, Canada (www.msc-smc.ec.gc.ca)

[3] : NCEP/EMC, Washington, US (www.emc.ncep.noaa.gov)

[4] : SAIC at NCEP/EMC, Washington, US (www.emc.ncep.noaa.gov)

[5] : UCAR Visiting Scientist, NCEP/EMC, Washington, US

OUTLINE

- MOTIVATION FOR ENSEMBLE/PROBABILISTIC FORECASTING
  - User Needs
  - Scientific needs
- SOURCES OF FORECAST ERRORS
  - Initial value
  - Model formulation
- ESTIMATING & SAMPLING FORECAST UNCERTAINTY
- DESCRIPTION OF ENSEMBLE FORECAST SYSTEMS
  - ECMWF
  - MSC
  - NCEP
- FORECAST EXAMPLE
- COMPARATIVE VERIFICATION
- ONGOING RESEARCH / OPEN QUESTIONS

MOTIVATION FOR ENSEMBLE FORECASTING

- ***FORECASTS ARE NOT PERFECT* - IMPLICATIONS FOR:**
  - **USERS:**
    - Need to know how often / by how much forecasts fail
    - Economically optimal behavior depends on
      - Forecast error characteristics
      - User specific application
        - » Cost of weather related adaptive action
        - » Expected loss if no action taken
      - EXAMPLE: Protect or not your crop against possible frost

    Cost = 10k, Potential Loss = 100k => Will protect if P(frost) > Cost/Loss=0.1

    - ***NEED FOR PROBABILISTIC FORECAST INFORMATION***

  - **DEVELOPERS:**
    - Need to improve performance - *Reduce error in estimate of first moment*
      - Traditional NWP activities (I.e., model, data assimilation development)
    - Need to account for uncertainty - *Estimate higher moments*
      - New aspect – How to do this?
    - Forecast is incomplete without information on forecast uncertainty
    - ***NEED TO USE PROBABILISTIC FORECAST FORMAT***

USER NEEDS – PROBABILISTIC FORECAST INFORMATION FOR MAXIMUM ECONOMIC BENEFIT

ECONOMIC VALUE OF FORECASTS

Given a particular forecast, a user either does or does not take action (eg, protects its crop against frost) *Mylne & Harrison, 1999*

OBSERVATION \ FORECAST	*YES*	*NO*
YES	H(its) ***Mitigated Loss***	M(isses) ***Loss***
NO	F(alse alarms) ***Cost***	C(orrect rejections) ***No Cost***

$$Mean\ Expense_{fc} = hML + mL + fC$$

$$Mean\ Expense_{perf} = oML$$

$$ME_{cl} = min[oL,\ oML + (1-o)C]$$

$$Value = \frac{ME_{cl} - ME_{fc}}{ME_{cl} - ME_{perf}}$$

o=climatological frequency

Optimum decision criterion for user action: P(weather event)=C/L

(Murphy 1977)

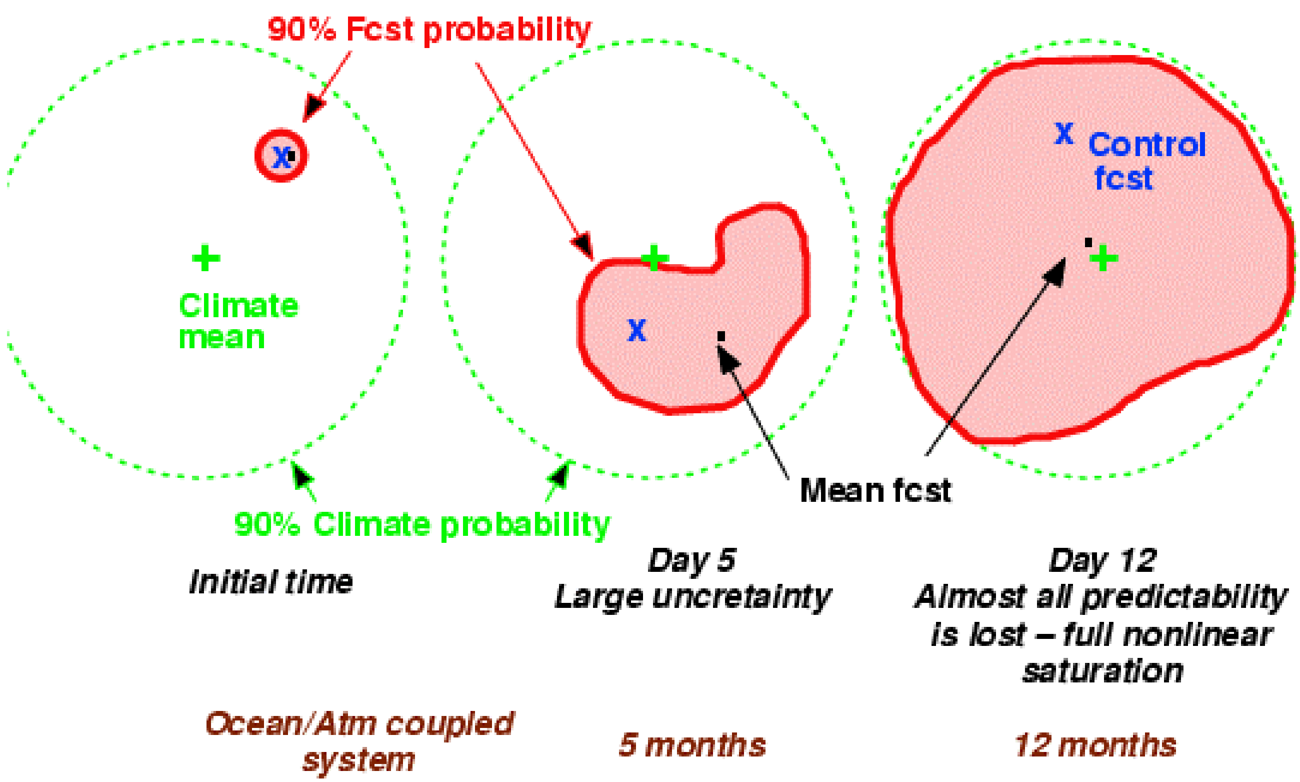
SCIENTIFIC NEEDS - DESCRIBE FORECAST UNCERTAINTY ARISING DUE TO CHAOS

ORIGIN OF FORECAST UNCERTAINTY

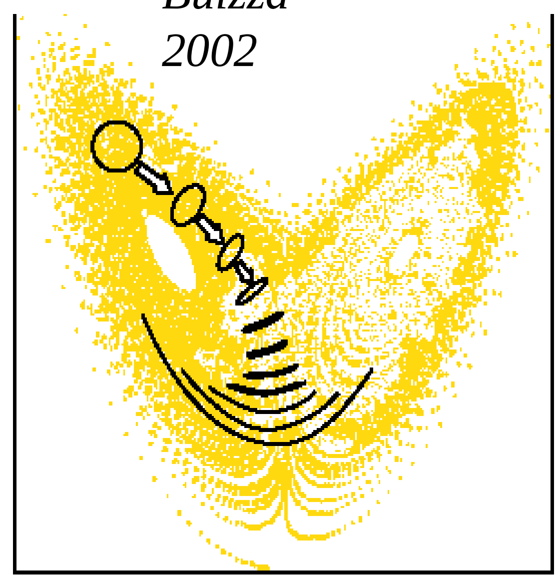
1) The atmosphere is a **deterministic system** *AND* has at least one direction in which **perturbations grow**

2) **Initial** state (and model) has **error** in it ==>

Chaotic system + Initial error =(Loss of) Predictability



Buizza 2002



FORECASTING IN A CHAOTIC ENVIRONMENT

DETERMINISTIC APPROACH - PROBABILISTIC FORMAT

SINGLE FORECAST - ***One integration with an NWP model***

- Is not best estimate for future evolution of system
- Does not contain all attainable forecast information
- Can be combined with past verification statistics to form probabilistic forecast
    - Gives **no estimate of flow dependent variations in forecast uncertainty**

PROBABILISTIC FORECASTING - ***Based on Liuville Equations***

- Initialize with probability distribution function (pdf) at analysis time
- Dynamical forecast of pdf based on conservation of probability values
- **Prohibitively expensive** -
    - Very high dimensional problem (state space x probability space)
    - Separate integration for each lead time
    - Closure problems when simplified solution sought

FORECASTING IN A CHAOTIC ENVIRONMENT - 2

DETERMINISTIC APPROACH - PROBABILISTIC FORMAT

MONTE CARLO APPROACH – *ENSEMBLE FORECASTING*

- ***IDEA*: Sample sources of forecast error**
  - Generate initial ensemble perturbations
  - Represent model related uncertainty

- ***PRACTICE*: Run multiple NWP model integrations**
  - Advantage of perfect parallelization
  - Use lower spatial resolution if short on resources

- ***USAGE*: Construct forecast pdf based on finite sample**
  - Ready to be used in real world applications
  - Verification of forecasts
  - Statistical post-processing (remove bias in 1st, 2nd, higher moments)

CAPTURES FLOW DEPENDENT VARIATIONS

IN FORECAST UNCERTAINTY

SOURCES OF FORECAST ERRORS

IMPERFECT KNOWLEDGE OF

INITIAL CONDITIONS

- Incomplete observing system (not all variables observed)
- Inaccurate observations (instrument/representativeness error)
- Imperfect data assimilation methods
  - Statistical approximations (eg, inaccurate error covariance information)
  - Use of imperfect NWP forecasts (due to initial and model errors) –
  - Effect of cycling (forecast errors "inherited" by analysis – use breeding)

GOVERNING EQUATIONS:

- Imperfect model
  - Structural uncertainty (eg, choice of structure of convective scheme)
  - Parametric uncertainty (eg, critical values in parameterization schemes)
  - Closure/truncation errors (temporal/spatial resolution; spatial coverage, etc)

NOTES:

- Two main sources of forecast errors hard to separate =>
- Very little information is available on model related errors
- Tendency to attribute all forecast errors to model problems

SAMPLING FORECAST ERRORS =

REPRESENTING ERRORS ORIGINATING FROM TWO MAIN SOURCES

INITIAL CONDITION RELATED ERRORS – *"Easy"*

- Sample initial errors
- Run ensemble of forecasts
- It works
  - Flow dependent variations in forecast uncertainty captured (show later)
  - Difficult or impossible to reproduce with statistical methods

MODEL RELATED ERRORS – *No theoretically satisfying approach*

- Change structure of model (eg, use different convective schemes, etc, MSC)
- Add stochastic noise (eg, perturb diabatic forcing, ECMWF)
- Works? Advantages of various approaches need to be carefully assessed
  - Are flow dependent variations in uncertainty captured?
  - Can statistical post-processing replicate use of various methods?
- Need for a
  - more comprehensive and
  - theoretically appealing approach

SAMPLING INITIAL CONDITION ERRORS

CAN SAMPLE ONLY WHAT'S KNOWN – FIRST NEED TO

ESTIMATE INITIAL ERROR DISTRIBUTION

THEORETICAL UNDERSTANDING – THE MORE ADVANCED A SCHEME IS (e. g., 4DVAR, Ensemble Kalman Filter)

- The lower the overall error level is
- The more the error is concentrated in subspace of Lyapunov/Bred vectors

PRACTICAL APPROACHES **–**

ONLY SOLUTION IS MONTE CARLO (ENSEMBLE) SIMULATION

- ***Statistical approach*** (dynamically growing errors neglected)
  - Selected estimated statistical properties of analysis error reproduced
    - Baumhefner et al – Spatial distribution; wavenumber spectra
    - ECMWF – Implicite constraint with use of Total Energy norm
- ***Dynamical approach*** – Breeding cycle (NCEP)
  - Cycling of errors captured
  - Estimates subspace of dynamically fastest growing errors in analysis
- ***Stochastic-dynamic approach*** – Perturbed Observations method (MSC)
  - Perturb all observations (given their uncertainty)
  - Run multiple analysis cycles
  - Captures full space (growing + non-growing) of analysis errors

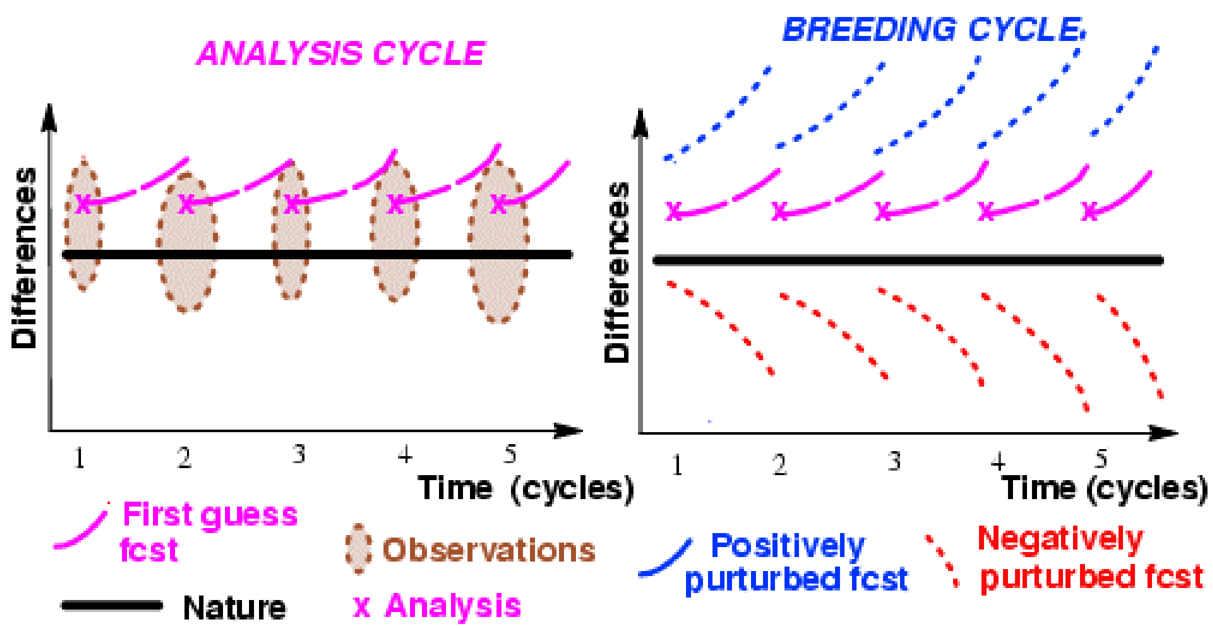
SAMPLING INITIAL CONDITION ERRORS

THREE APPROACHES – SEVERAL OPEN QUESTIONS

- **RANDOM SAMPLING** – **Perturbed observations method** (MSC)
  - Represents all potential error patterns with realistic amplitude
  - Small subspace of growing errors is well represented
  - Potential problems:
    - Much larger subspace of non-growing errors poorly sampled,
    - Yet represented with realistic amplitudes
- **SAMPLE GROWING *ANALYSIS* ERRORS** – **Breeding** (NCEP)
  - Represents dynamically growing analysis errors
  - Ignores non-growing component of error
  - Potential problems:
    - May not provide "wide enough" sample of growing perturbations
    - Statistical consistency violated due to directed sampling? Forecast consequences?
- **SAMPLE FASTEST GROWING *FORECAST* ERRORS** – **SVs** (ECMWF)
  - Represents forecast errors that would grow fastest in linear sense
  - Perturbations are optimized for maximum forecast error growth
  - Potential problems:
    - Need to optimize for each forecast application (or for none)?
    - Linear approximation used
    - Very expensive

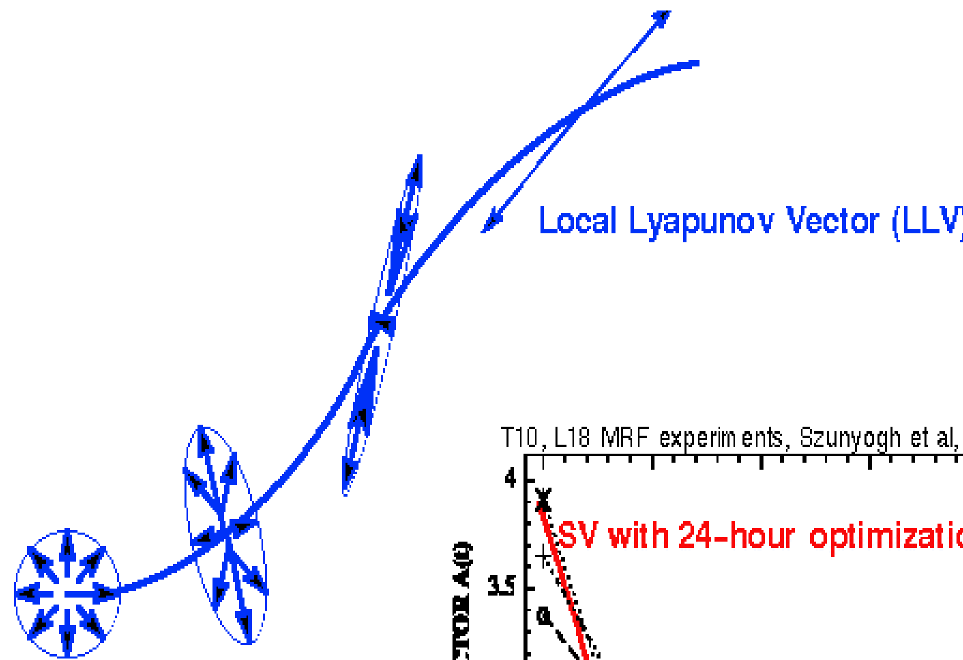
ESTIMATING AND SAMPLING INITIAL ERRORS: THE BREEDING METHOD

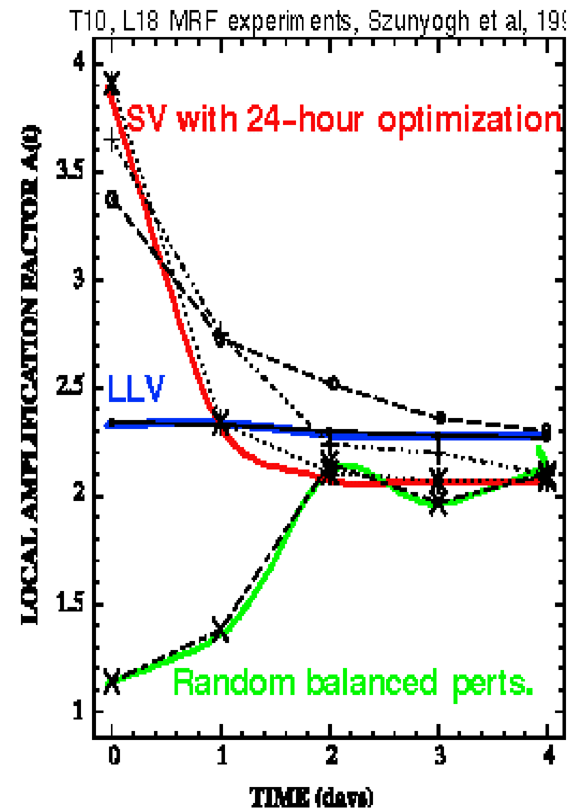
- **DATA ASSIM:** Growing errors due to cycling through NWP forecasts
- **BREEDING:** - Simulate effect of obs by rescaling nonlinear perturbations
  - Sample subspace of most rapidly growing analysis errors
    - Extension of linear concept of Lyapunov Vectors into nonlinear environment
    - Fastest growing nonlinear perturbations
    - Not optimized for future growth –
      - Norm independent
      - Is non-modal behavior important?



LYAPUNOV, SINGULAR, AND BRED VECTORS

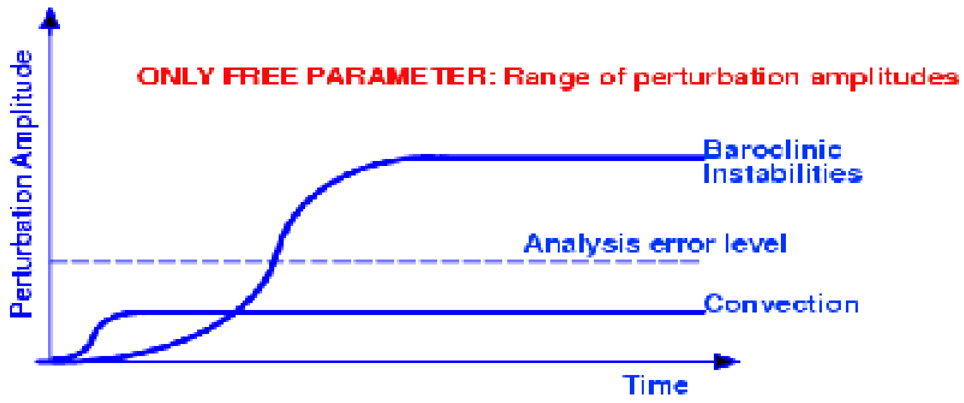
- **LYAPUNOV VECTORS (LLV):**
  - Linear perturbation evolution
  - Fast growth
  - Sustainable
  - Norm independent
  - Spectrum of LLVs
- **SINGULAR VECTORS (SV):**
  - Linear perturbation evolution
  - Fastest growth
  - Transitional (optimized)
  - Norm dependent
  - Spectrum of SVs
- **BRED VECTORS (BV):**
  - Nonlinear perturbation evolution
  - Fast growth
  - Sustainable
  - Norm independent
  - Can orthogonalize (Boffeta et al)





PERTURBATION EVOLUTION

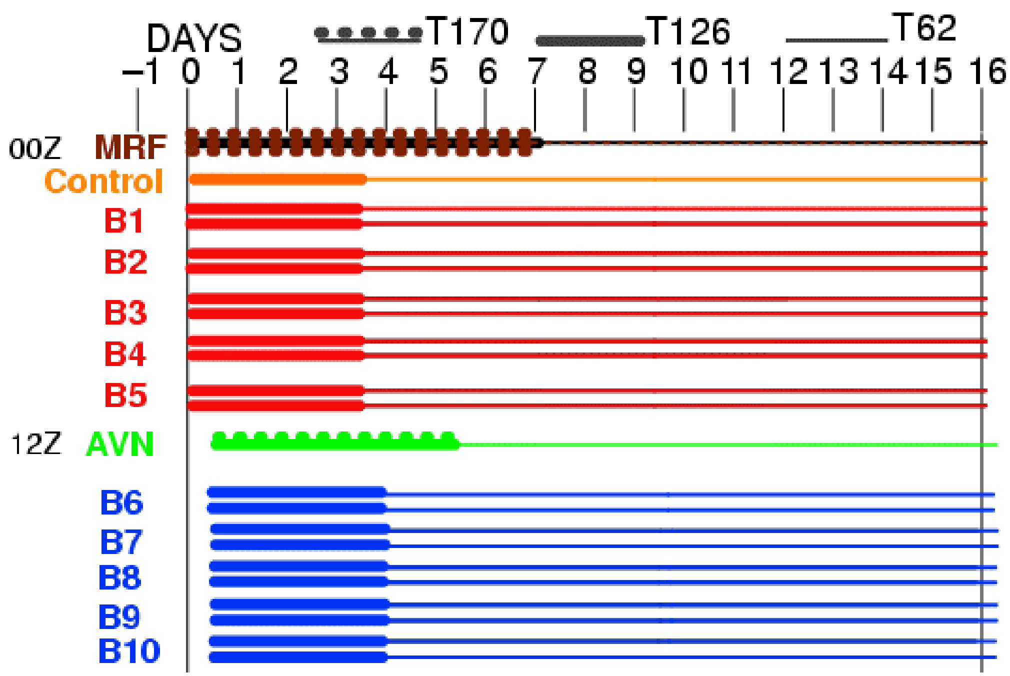
- **PERTURBATION GROWTH**
  - Due to effect of instabilities
  - Linked with atmospheric phenomena (e.g, frontal system)
- **LIFE CYCLE OF PERTURBATIONS**
  - Associated with phenomena
  - Nonlinear interactions limit perturbation growth
  - Eg, convective instabilities grow fast but are limited by availability of moisture etc
- **LINEAR DESCRIPTION**
  - May be valid at beginning stage only
  - If linear models used, need to reflect nonlinear effects at given perturb. Amplitude
- **BREEDING**
  - Full nonlinear description
  - Range of typical perturbation amplitudes is only free parameter



NCEP GLOBAL ENSEMBLE FORECAST SYSTEM

CURRENT *(APRIL 2003)* SYSTEM

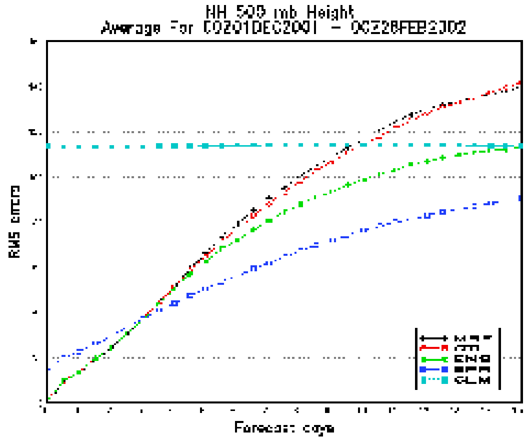
- 10 members out to 16 days
- 2 *(4)* times daily
- T126 out to 3.5 *(7.5)* days
- Model error not yet represented



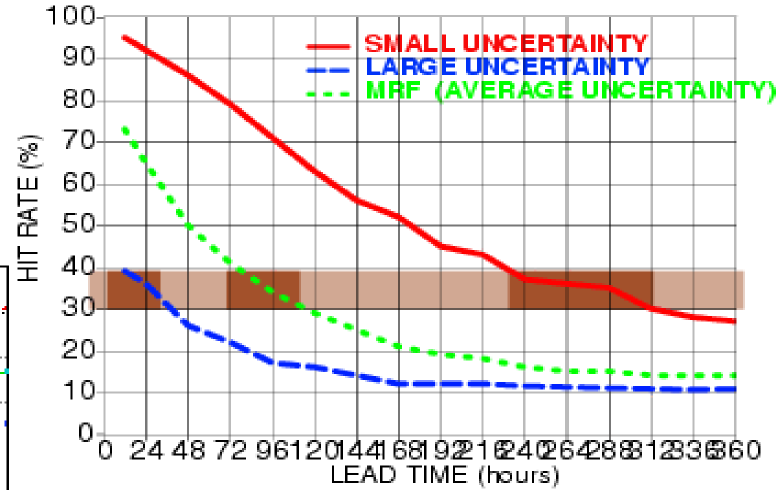
- **PLANS**
- ***Initial perturbations***
  - Rescale bred vectors via ETKF
  - Perturb surface conditions
- ***Model errors***
  - Push members apart
  - Multiple physics (combinations)
  - Change model to reflect uncertainties
- ***Post-processing***
  - Multi-center ensembles
  - Calibrate 1st & 2nd moment of pdf
  - Multi-modal behavior?

ADVANTAGES OF USING ENSEMBLE (VS. CONTROL) FCSTS

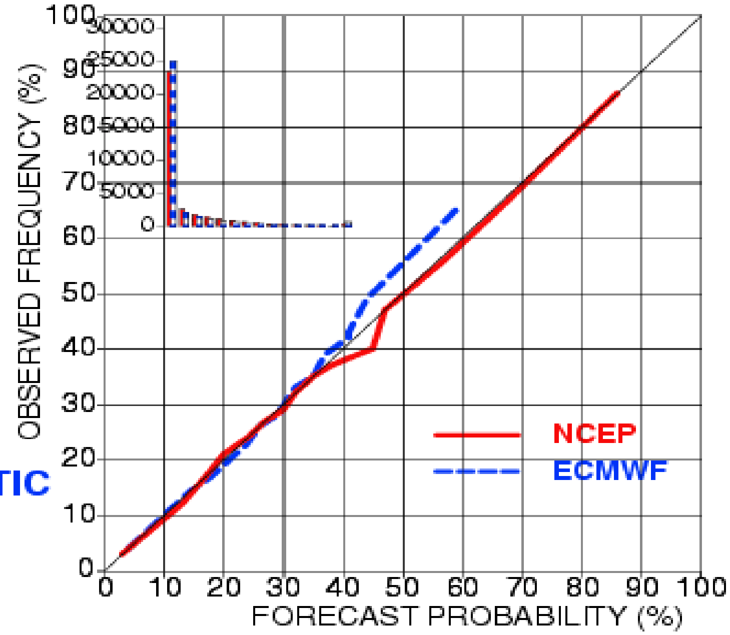
1) IMPROVED EXPECTED VALUE FORECAST



2) CASE DEPENDENT UNCERTAINTY ESTIMATE



3) DETAILED PROBABILISTIC FORECAST



RESOLUTION OF ENSEMBLE BASED PROB. FCSTS

QUESTION:

What are the typical **variations in foreseeable forecast uncertainty?**

What variations in predictability can the ensemble resolve?

METHOD:

Ensemble mode value to distinguish high/low predictability cases

Stratify cases according to ensemble mode value –

Use 10–15% of cases when ensemble is highest/loewest

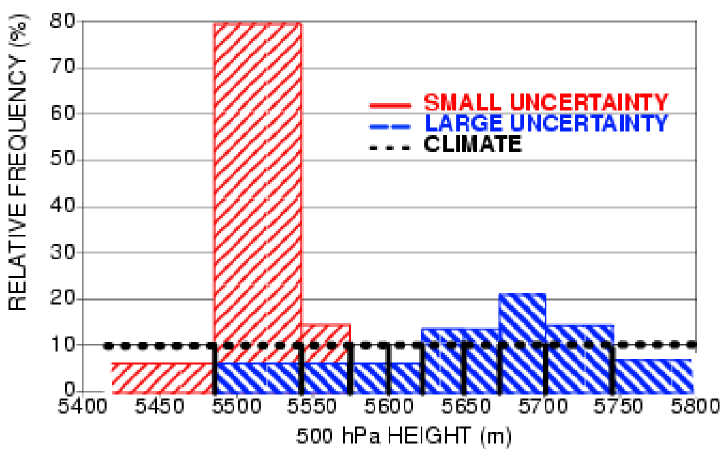
DATA:

NCEP **500 hPa NH extratropical ensemble fcsts** for March–May 1997

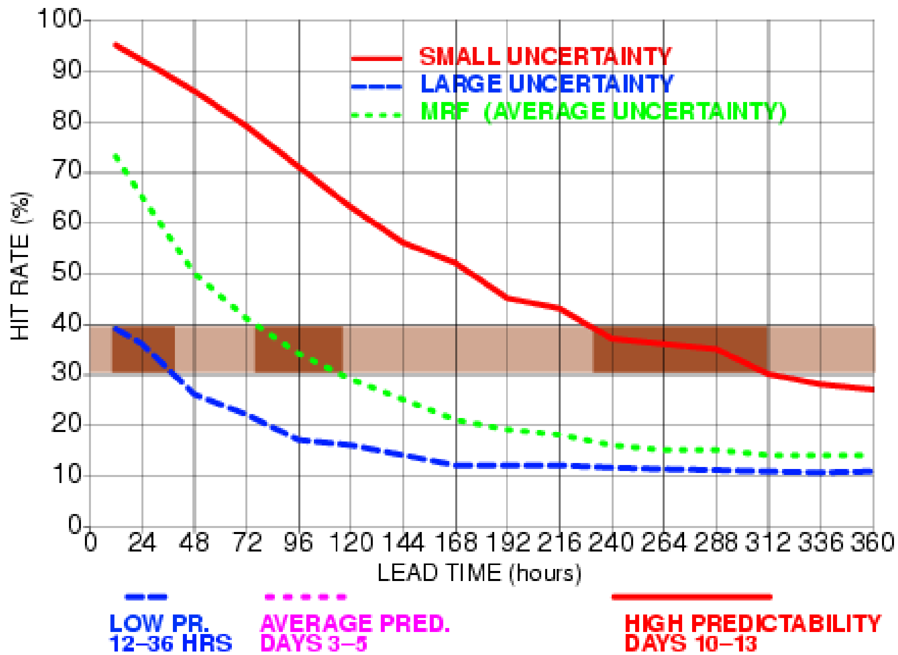
14 perturbed fcsts and high resolution control

VERIFICATION:

Hit rate for ensemble mode and hires control fcst



SEPARATING HIGH VS. LOW UNCERTAINTY FCSTS



THE ***UNCERTAINTY OF FCSTS*** *CAN BE* ***QUANTIFIED IN ADVANCE***

HIT RATES FOR 1–DAY FCSTS

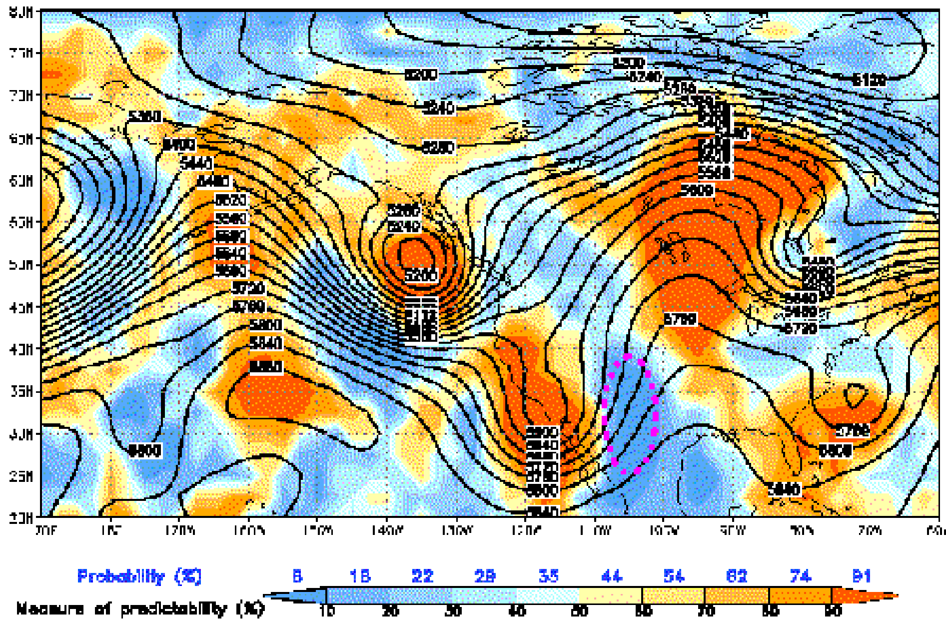
CAN BE ***AS LOW AS 36%****, OR* ***AS HIGH AS 92%***

10–15% OF THE TIME A 12–DAY FCST CAN BE AS GOOD, OR A 1–DAY FCST CAN BE AS POOR AS AN AVERAGE 4–DAY FCAST

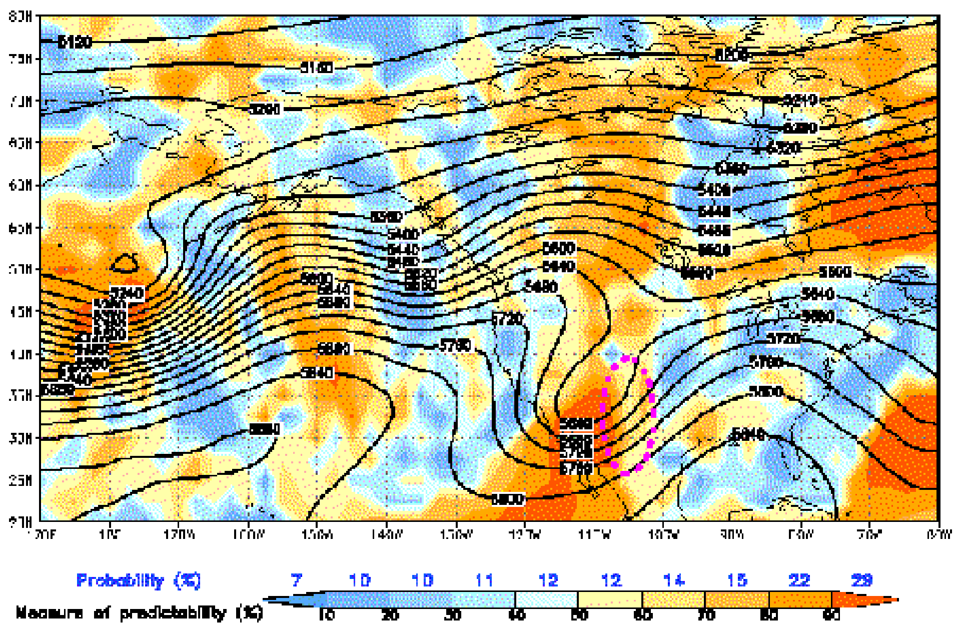
1–2% OF ALL DAYS THE 12–DAY FCST CAN BE MADE WITH MORE CONFIDENCE THAN THE 1–DAY FCST

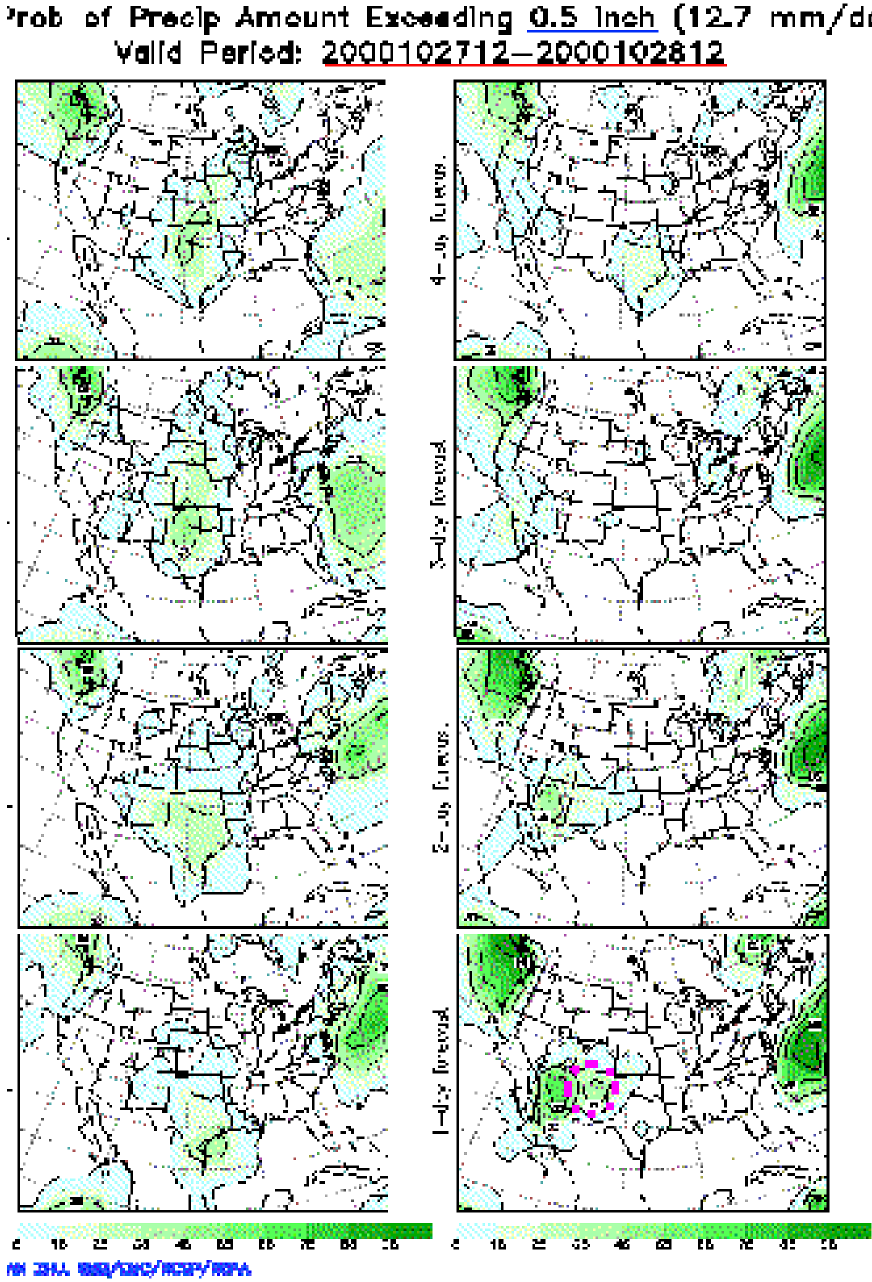
AVERAGE HIT RATE FOR EXTENDED–RANGE FCSTS IS LOW –
VALUE IS IN KNOWING WHEN FCST IS RELIABLE

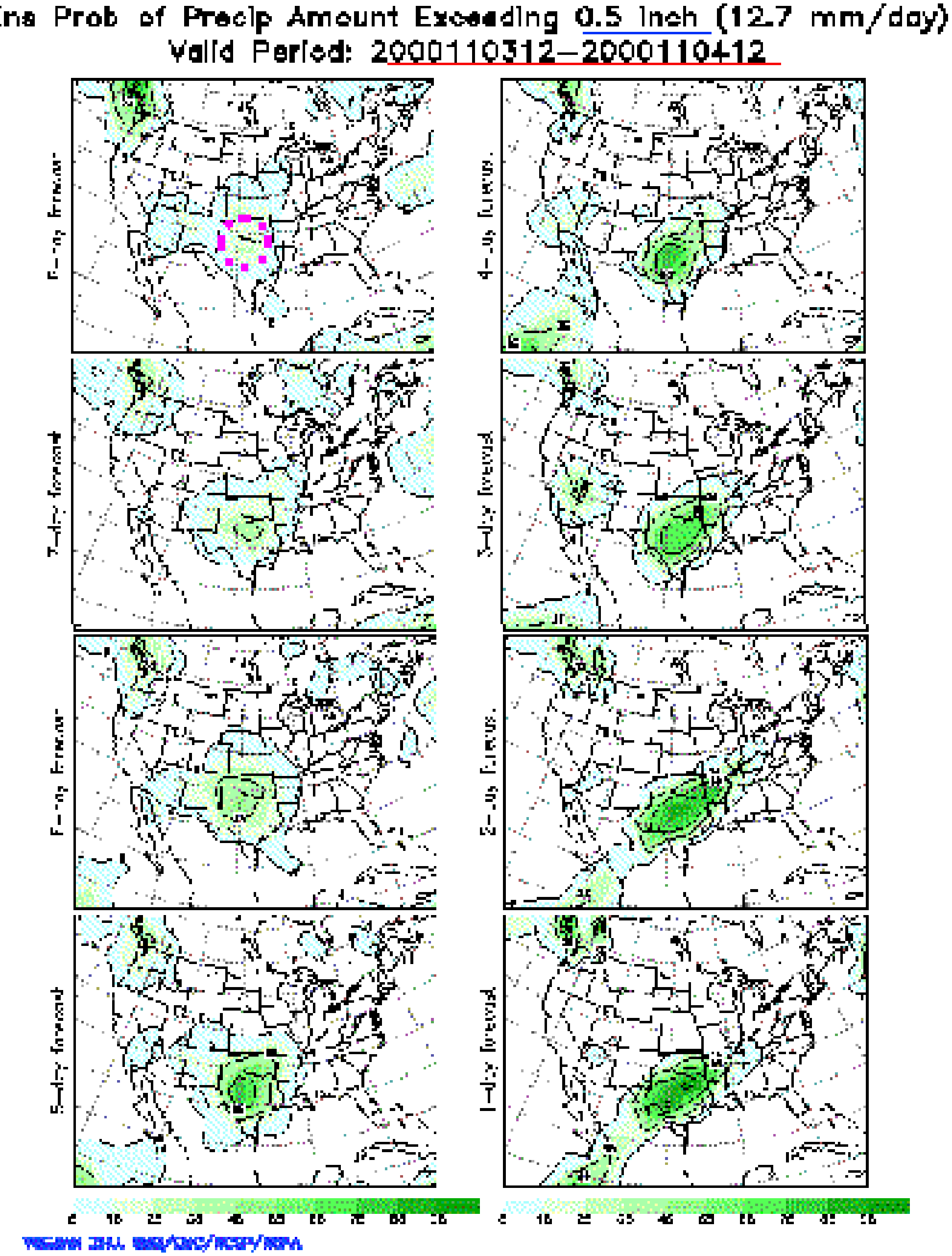
Relative measure of predictability (colors)
for ensemble mean forecast (contours) of 500 hPa height
ini: 2000102700 valid: 2000102800 fcst: 24 hours

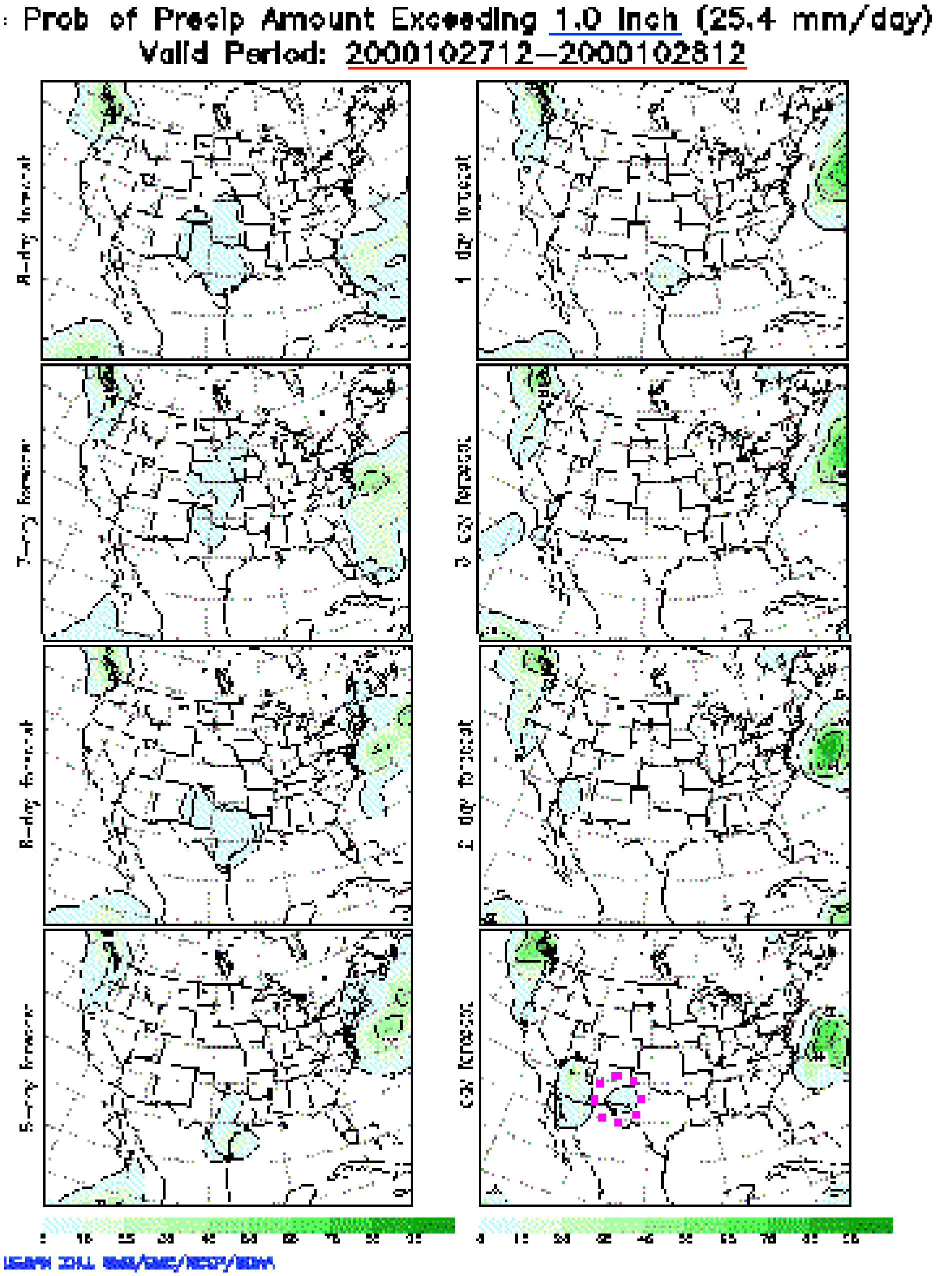


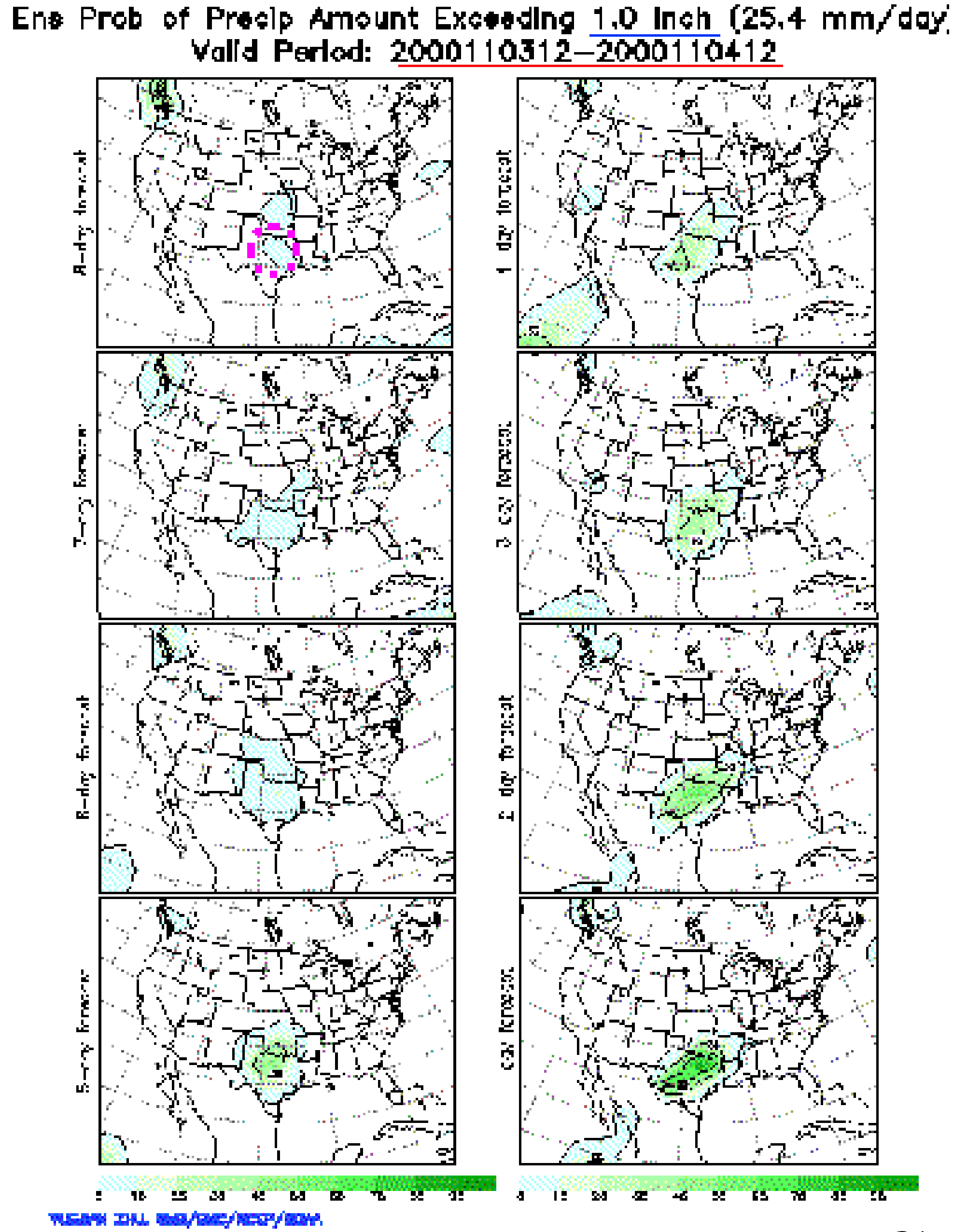
Relative measure of predictability (colors)
for ensemble mean forecast (contours) of 500 hPa height
ini: 2000102700 valid: 2000110400 fcst: 192 hours




Prob of Precip Amount Exceeding 0.5 inch (12.7 mm/day)
Valid Period: 2000102712–2000102812
4-day forecast
3-day forecast
2-day forecast
1-day forecast


Ens Prob of Precip Amount Exceeding 0.5 inch (12.7 mm/day)
Valid Period: 2000110312–2000110412
4-day forecast
3-day forecast
2-day forecast
1-day forecast

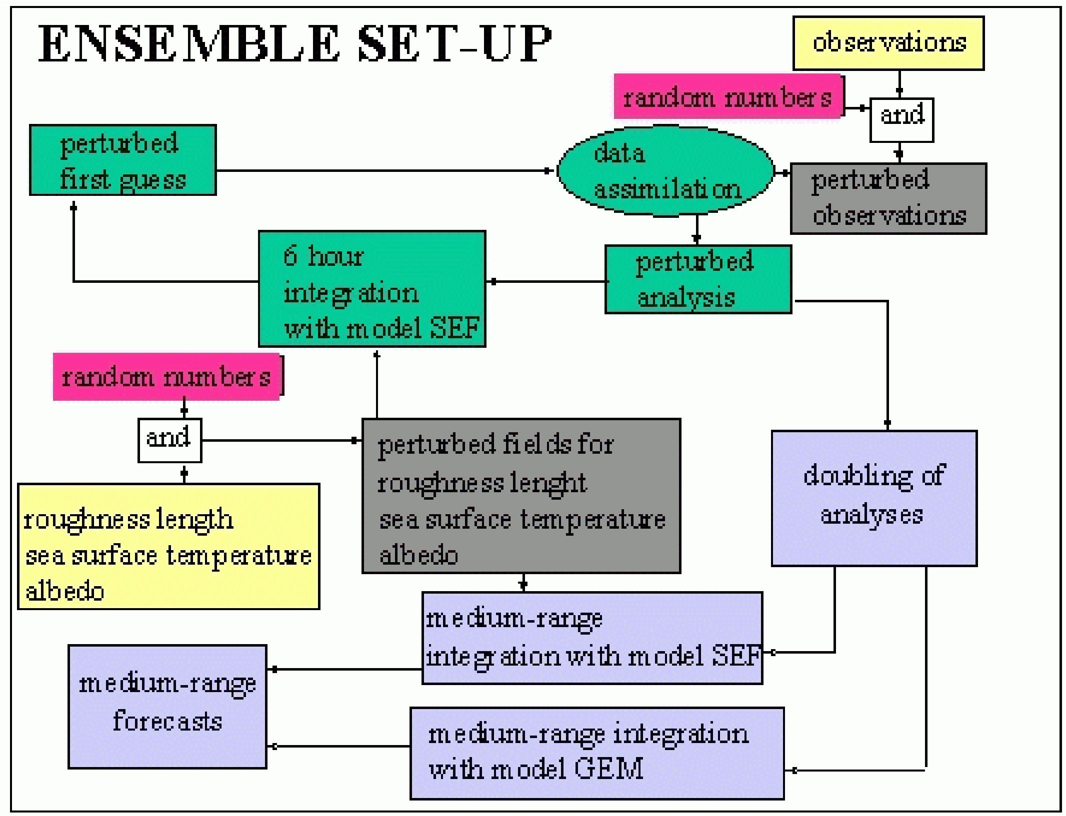

Prob of Precip Amount Exceeding 1.0 inch (25.4 mm/day)
Valid Period: 2000102712–2000102812


Ens Prob of Precip Amount Exceeding 1.0 inch (25.4 mm/day
Valid Period: 2000110312–2000110412

Monte Carlo approach (MSC): all-inclusive design

The MSC ensemble has been designed to simulate:

- observation errors (random perturbations);
- imperfect boundary conditions;
- model errors (2 models and different parameterisations).



Simulation of initial uncertainties: selective sampling

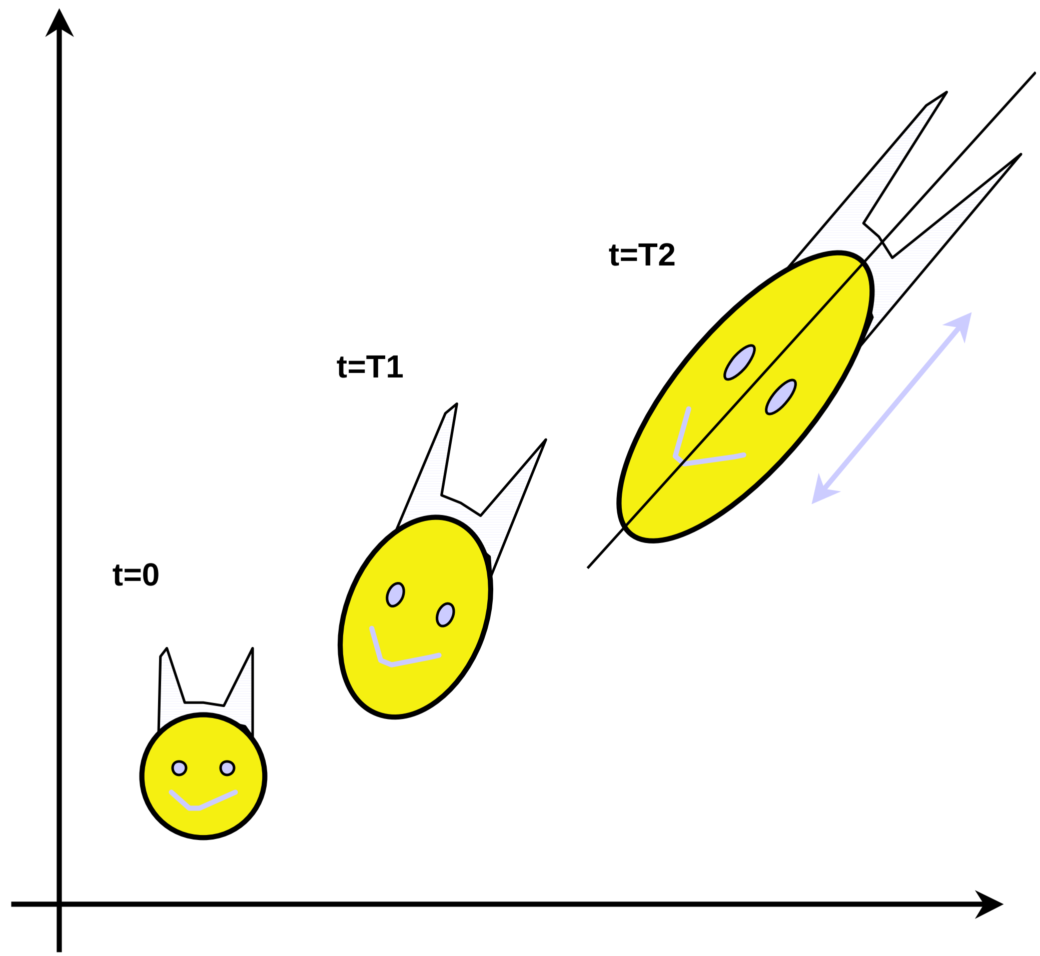
At MSC, the perturbed initial conditions are generated by running an ensemble of assimilation cycles that use perturbed observations and different models (Monte Carlo approach).

At ECMWF and NCEP the perturbed initial conditions are generated by adding perturbations to the unperturbed analysis generated by the assimilation cycle. The initial perturbations are designed to span only a subspace of the phase space of the system (selective sampling). These ensembles do not simulate the effect of imperfect boundary conditions.

Selective sampling: singular vectors (ECMWF)

Perturbations pointing along different axes in the phase-space of the system are characterized by different amplification rates. As a consequence, the initial PDF is stretched principally along directions of maximum growth**.**

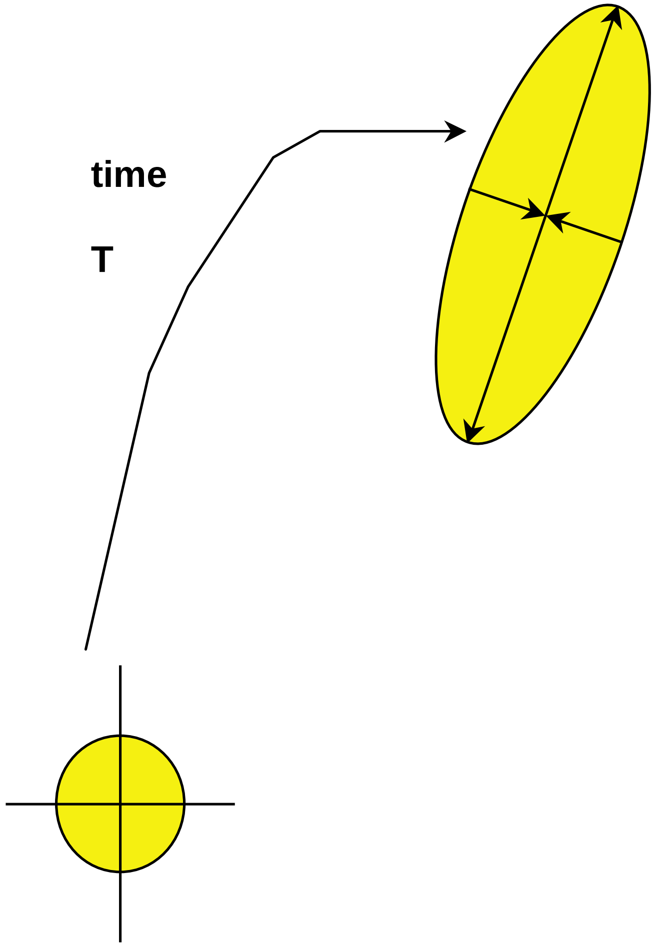
The component of an initial perturbation pointing along a direction of maximum growth amplifies more than components pointing along other directions.



Selective sampling: singular vectors (ECMWF)

At ECMWF, maximum growth is measured in terms of total energy. A perturbation time evolution is linearly approximated:

The adjoint of the tangent forward propagator with respect to the total-energy norm is defined, and the singular vectors, i.e. the fastest growing perturbations, are computed by solving an eigenvalue problem:



Description of the ECMWF, MSC and NCEP systems

The three ensembles differ also in size, resolution, daily frequency and forecast length.

	MSC	ECMWF	NCEP
Pj (model uncertainty)	2 models + Diff. Ph. Par.	Pj=P0 (single model)	Pj=P0 (single model)
dPj (random mod err)	2 models + Diff. Ph. Par.	dPj=rj*Pj (stoch. physics)	dPj=0
Aj	2 models	Aj=A0 (single model)	Aj=A0 (single model)
oj (obs error)	Random perturbations	-	-
ej (initial uncertainty)	ej from Anal. Cycles	ej=e0+dej(SV)	ej=e0+dej(BV)
hor-res HRES control	-	-	T170(d0-7)>T126(d7-16)
hor-res control	TL149	TL255 (d0-10)	T126(d0-3.5)>T62(d3.5-16)
hor-res pert members	TL149	TL255 (d0-10)	T126(d0-3.5)>T62(d3.5-16)
vertical levels (c&pf)	23 and 41, 28	40	28
top of the model	10hPa	10hPa	3hPa
perturbed members	16	50	10
forecast length	10 days	10 days	16 days
daily frequency	00 UTC	12 UTC (00 UTC exp)	00 and 12 UTC
operational impl.	February 1998	December 1992	December 1992

Some considerations on model error simulation

The MSC multi-model approach is very difficult to maintain. On the contrary, the ECMWF stochastic approach is easy to implement and maintain

The disadvantage of the ECMWF approach is that it only samples uncertainty on short-scales and it is not designed to simulate model biases

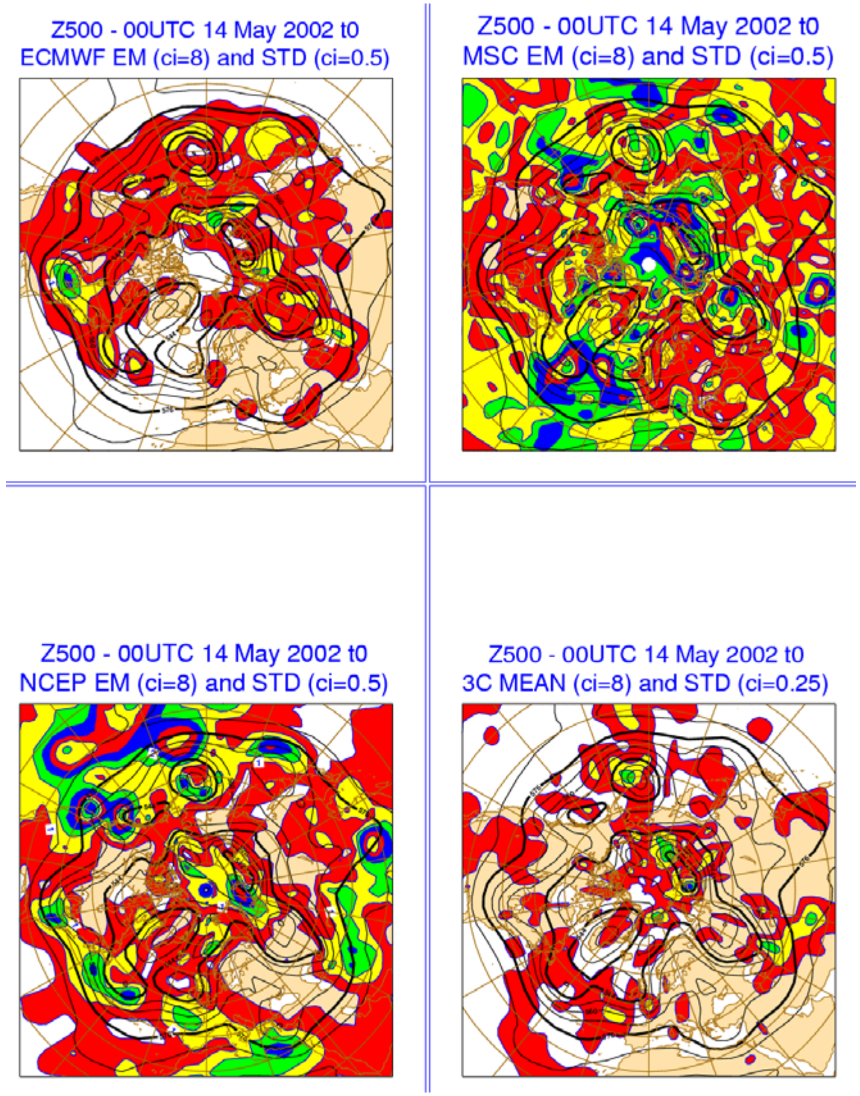
A possible way forward is to use one model but use different sets of tuning parameters in each perturbed member (NCEP plans)

Similarities/differences in EM & STD: 14 May 2002, t=0

Due to the different methodologies, the ensemble initial states are different.

This figure shows the ensemble mean and standard deviation at initial time for 00UTC of 14 May 2002. The bottom-right panel shows the mean and the std of the 3 centers' analyses.

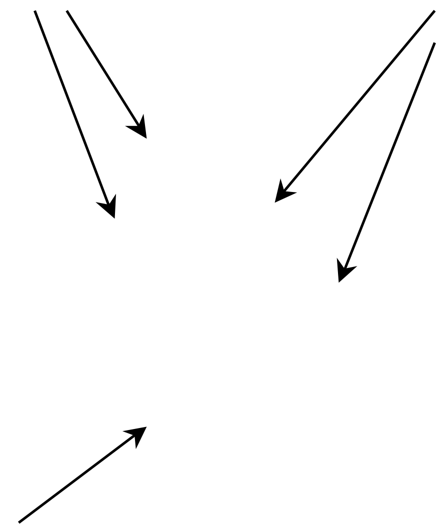
- **Area: the three ensembles' put emphasis on different areas; EC has the smallest amplitude over the tropics.**
- **Amplitude: the ensembles' stds are larger than the std of the 3-centers' analyses (2 times smaller contour interval); EC has ~2 times lower values over NH.**



Similarities/differences in EM & STD: 14 May 2002, t+48h

This figure shows the t+48h ensemble mean and standard deviation started at 00UTC of 14 May 2002. The bottom-right panel shows the 3-centers' average analysis and root-mean-square error.

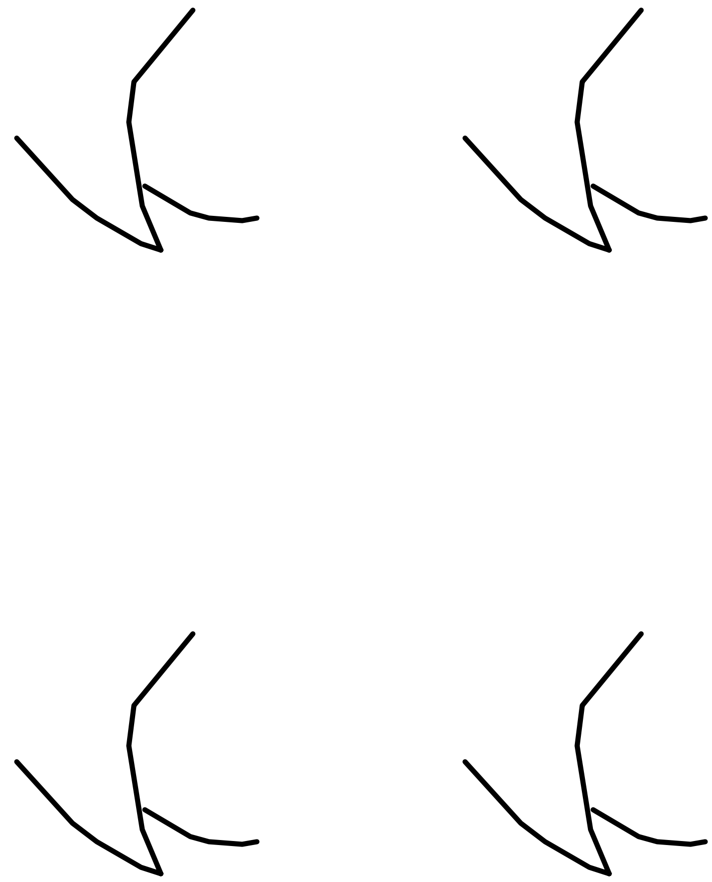
- **Area**: there is some degree of similarity among the areas covered by the evolved perturbations.
- **Amplitude**: similar over NH; EC smaller over tropics.
- **Std-vs-rmse**: certain areas of large spread coincide with areas of large error.



Similarities/differences in EM & STD: 14 May 2002, t+120h

This figure shows the t+120h ensemble mean and standard deviation started at 00UTC of 14 May 2002. The bottom-right panel shows the 3-centres' average analysis and average forecast root-mean-square error.

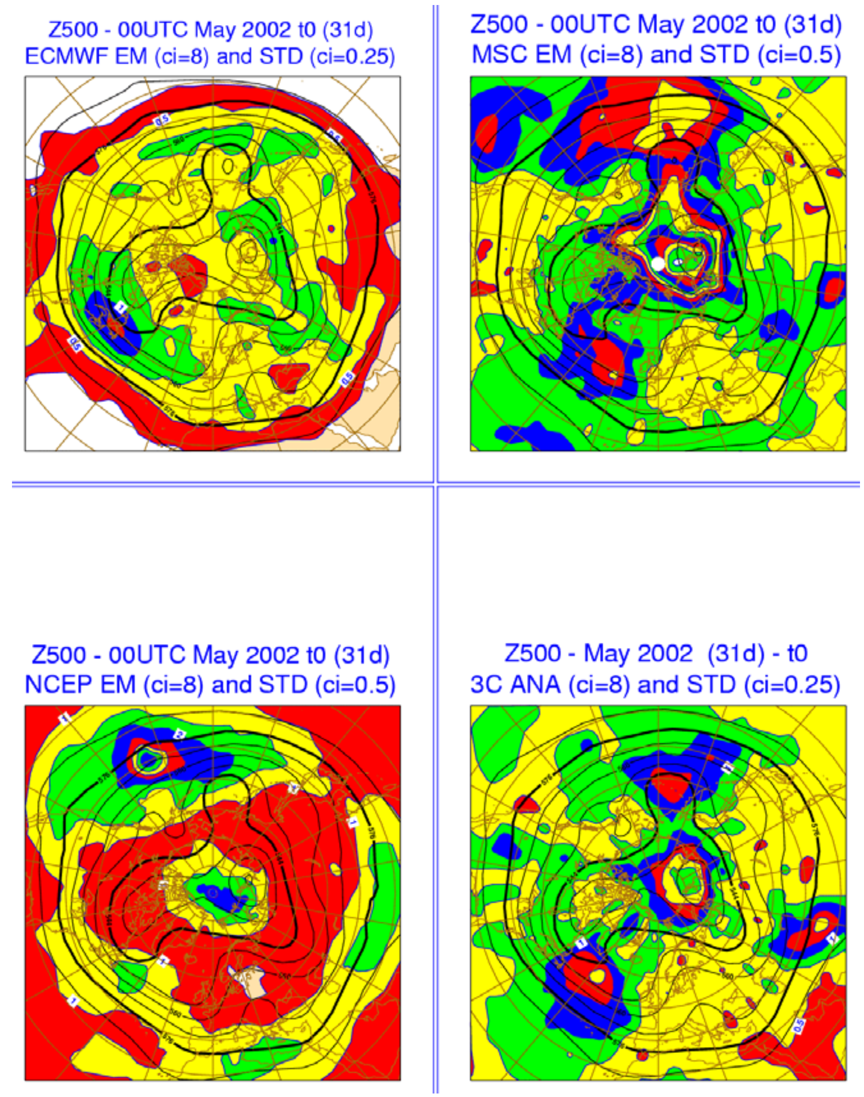
- **<u>Area</u>: perturbations show maximum amplitude in similar regions.**
- **<u>Amplitude</u>: EC perturbations have larger amplitude.**
- **<u>Std-vs-rmse</u>: there is a certain degree of agreement between areas of larger error and large spread.**



Similarities/differences in EM & STD: May 2002, t=0

This figure shows the May02-average ensemble mean and standard deviation at initial time (10 members, 00UTC). The bottom-right panel shows the average and the std of the 3-centres' analyses.

- **Area**: NCEP and MSC peak over the Pacific ocean and the Polar cap while EC peaks over the Atlantic ocean; MSC shows clear minima over Europe and North America.
- **Amplitude**: MSC and NCEP are ~2 times larger than the std of the 3 centres' analyses (2-times larger contour interval); EC has amplitude similar to 3C-std over NH but has too small amplitude over the tropics.



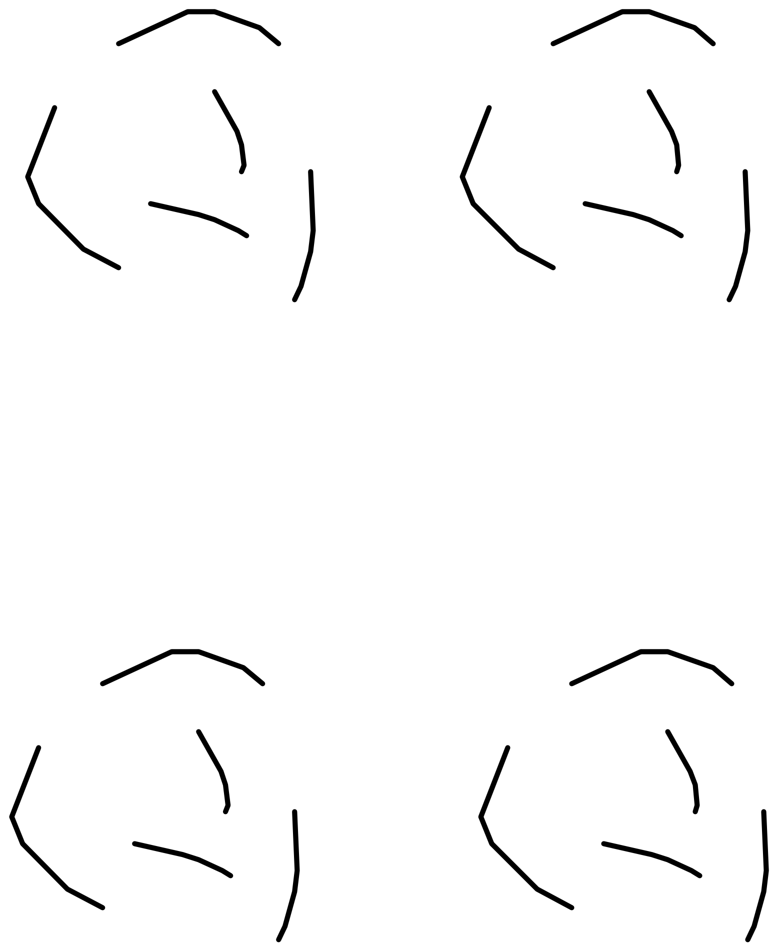
Similarities/differences in EM & STD: May 2002, t=0

This figure shows the May02-average ensemble mean and standard deviation at initial time (10 members, 00UTC).

The bottom-right panel shows the EC analysis and the Eady index (*Hoskins and Valdes* 1990), which is a measure of baroclinic instability:

(the static stability N and the wind shear have been computed using the 300- and 1000-hPa potential temperature and wind).

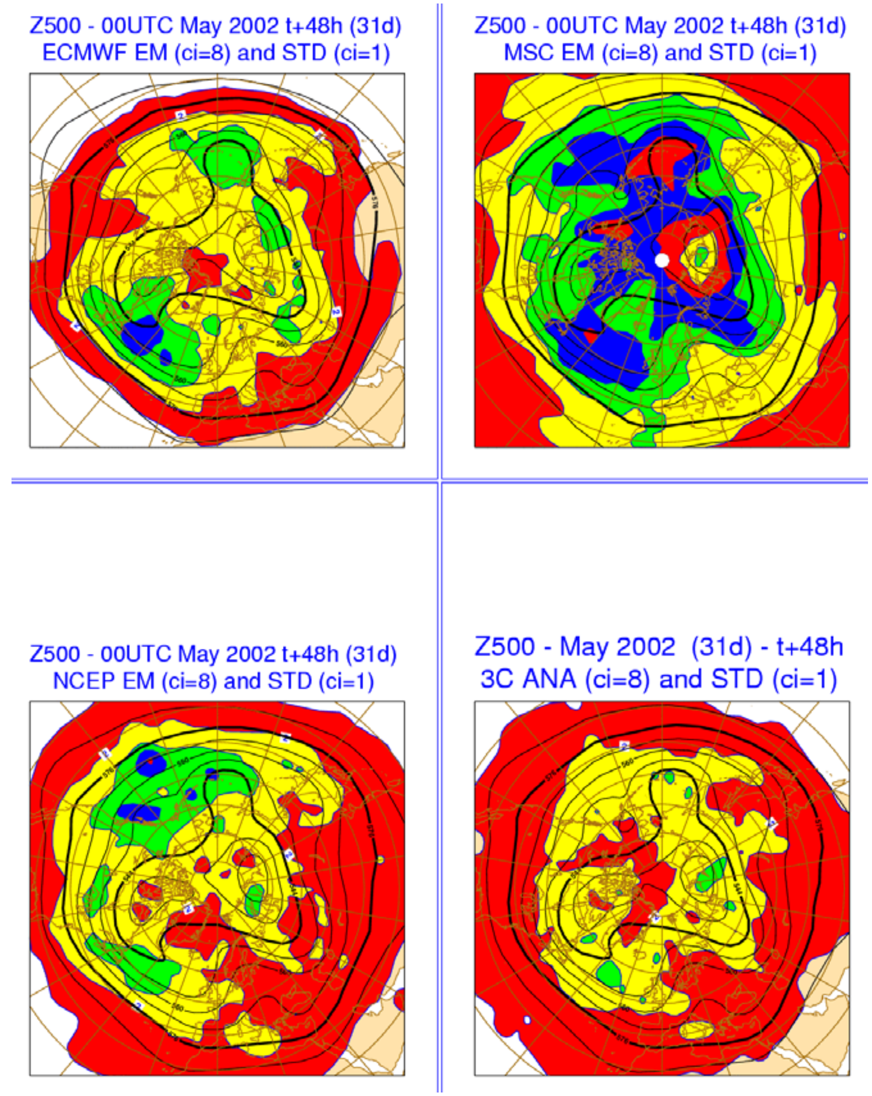
EC std shows a closer agreement with areas of baroclinic instability.



Similarities/differences in EM & STD: May 2002, t+48h

This figure shows the May02-average ensemble mean and standard deviation at t+48h (10 members, 00UTC) The bottom-right panel shows the average and the std of the 3-centres' analyses.

- **Area: NCEPS and MSC give more weight to the Pacific while EC gives more weight to the Atlantic; NCEP initial relative maximum over the North Pole cap has disappeared; MSC shows still a large amplitude north of Siberia.**
- **Amplitude: MSC has the largest amplitude over NH; EC has the smallest amplitude over the tropics.**



The test period and the verification measures

- **The test period is May-June-July 2002 (MJJ02).**

- **Scores for Z500 forecasts over NH (20:80°N) are shown.**
- **All forecasts data are defined on a regular 2.5-degree latitude-longitude grid.**
- **Each ensemble is verified against its own analysis.**

- **For a fair comparison, only 10 perturbed members are used for each ensemble system (from 00UTC for MSC and NCEP and from 12UTC for ECMWF).**

- **Probability forecasts' accuracy has been measured using the Brier skill score (BSS), the area under the relative operating characteristic curve (ROC) and the ranked probability skill score (RPSS). Probabilistic forecasts are average scores computed considering 10 climatologically equally likely events *(see talk by Z. Toth for a definition).***

PATTERN ANOMALY CORRELATION (PAC)

METHOD: Compute standard PAC for

- Ensemble mean & Control fcsts

EVALUATION

Higher control score due to bette

- Analysis + NWP model

Higher ensemble mean score du

- Analysis, NWP model, ANI
- Ensemble techniques

RESULTS

CONTROL

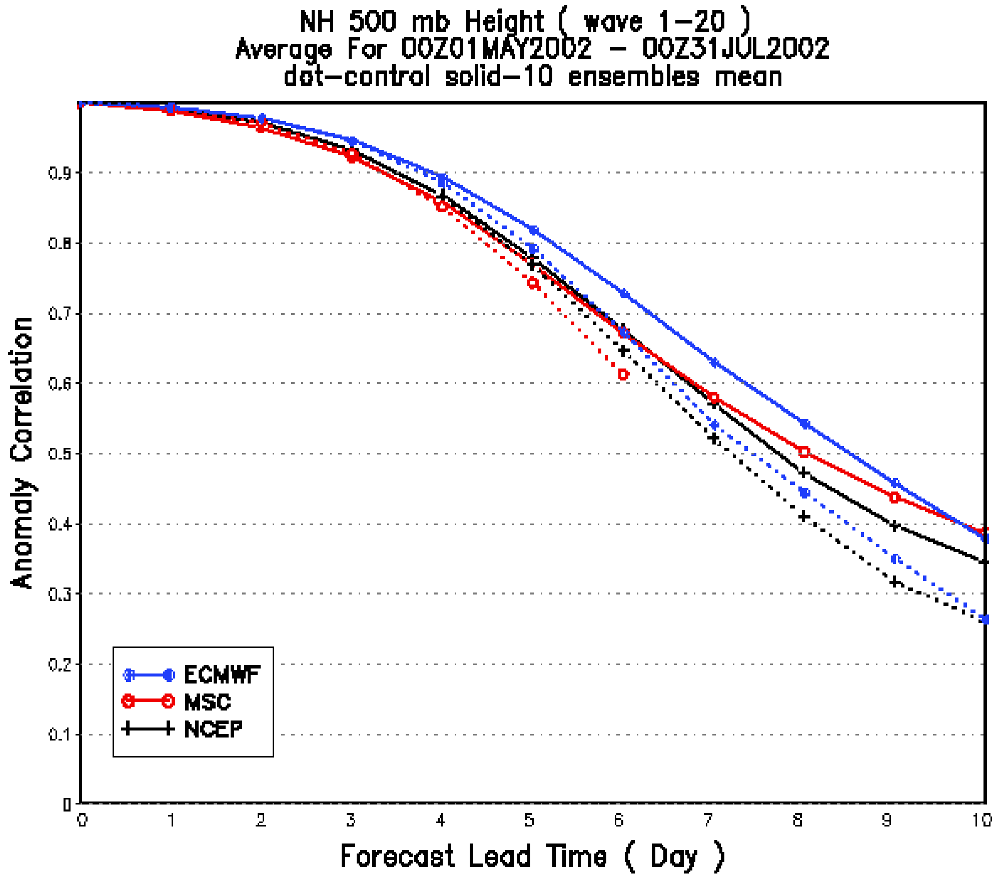
- **ECMWF** best throughout
  - Good analysis/model

ENSEMBLE VS. CONTROL

- **CANADIAN** poorer days 1-3
  - Stochastic perturbations?
- **NCEP** poorer beyond day 3
  - No model perturbations?

ENSEMBLE

- **ECMWF** best throughout
  - Good analysis/model?



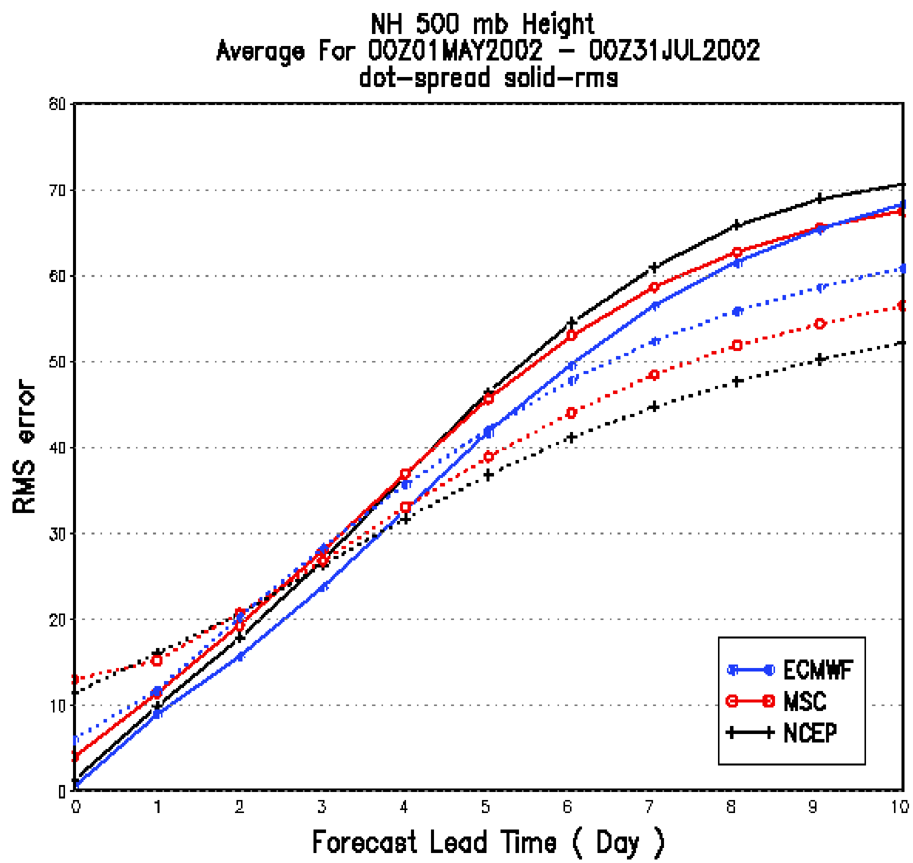
RMS ERROR AND ENSEMBLE SPREAD

RMS ENSEMBLE MEAN ERROR

- **ECMWF** best overall
  - Good analysis/model?
- **NCEP** competitive till day 1
  - Decent initial perturbations?
- **CANADA** best day 10
  - Model divers. helps reduce bias?

RMS ENSEMBLE SPREAD

- **CANADA**, **NCEP** highest days 1-2
  - Too large initial perturbation?
- **ECMWF** highest days 3-10
- **ECMWF** perturbation growth hiest
  - Stochastic perturbations help?



OUTLIER STATISTICS

METHOD:

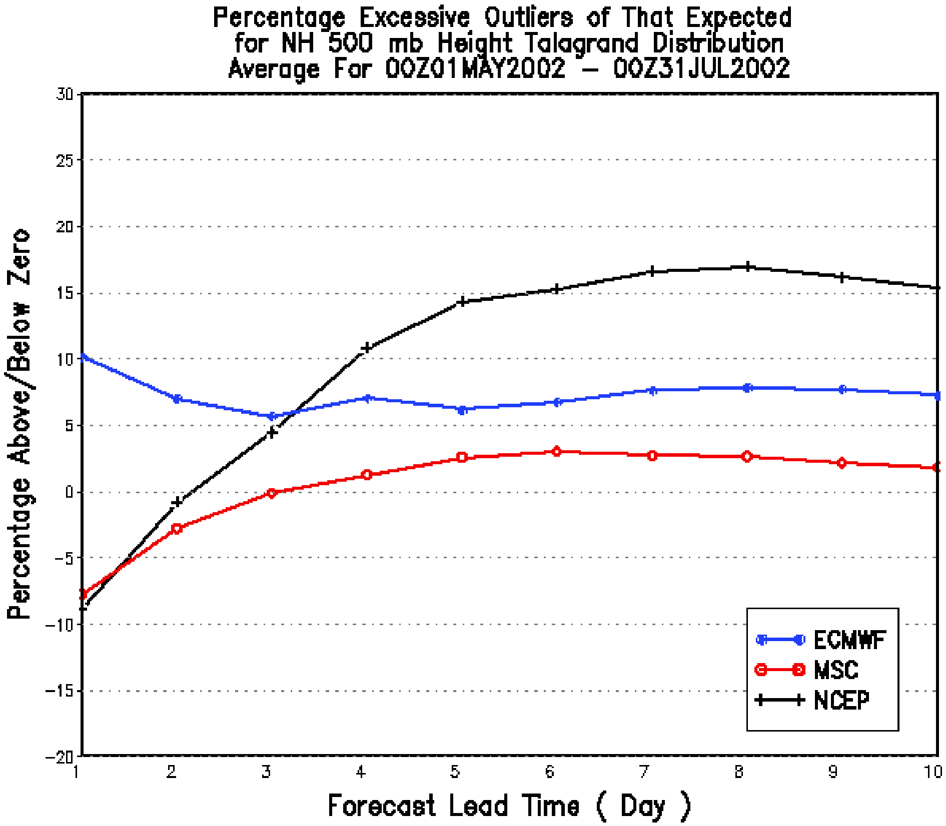
- Assess how often verifying analysis falls outside range of ensemble

EVALUATION:

- Perfect statistical consistency:
  - 2/N+1 is expected number
  - Excessive values above expected value shown

RESULTS

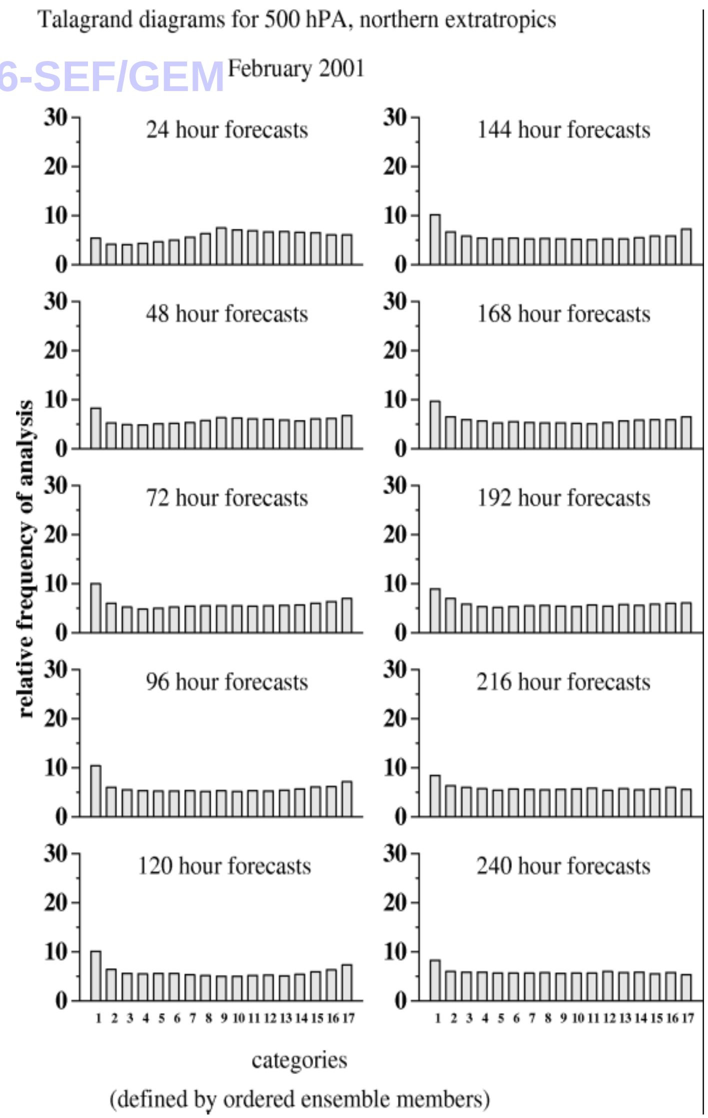
- **CANADIAN**
  - Best overall performance
- **NCEP, CANADIAN**
  - Too large spread at day 1
- **NCEP**
  - Too small spread days 5-10
- **ECMWF**
  - Too small spread (especially at day 1)



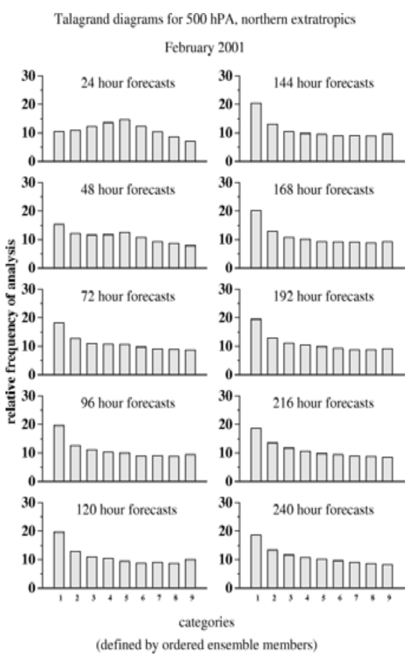
The impact of using a second model at MSC

The warm bias was reduced substantially and the U-shape disappeared by combining the two ensembles into the 16-SEF/GEM ensemble.

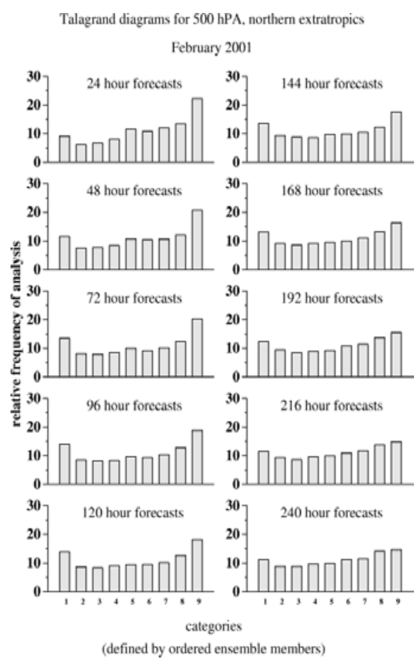
16-SEF/GEM



8-SEF



8-GEM



TIME CONSISTENCY OF ENSEMBLES

METHOD:

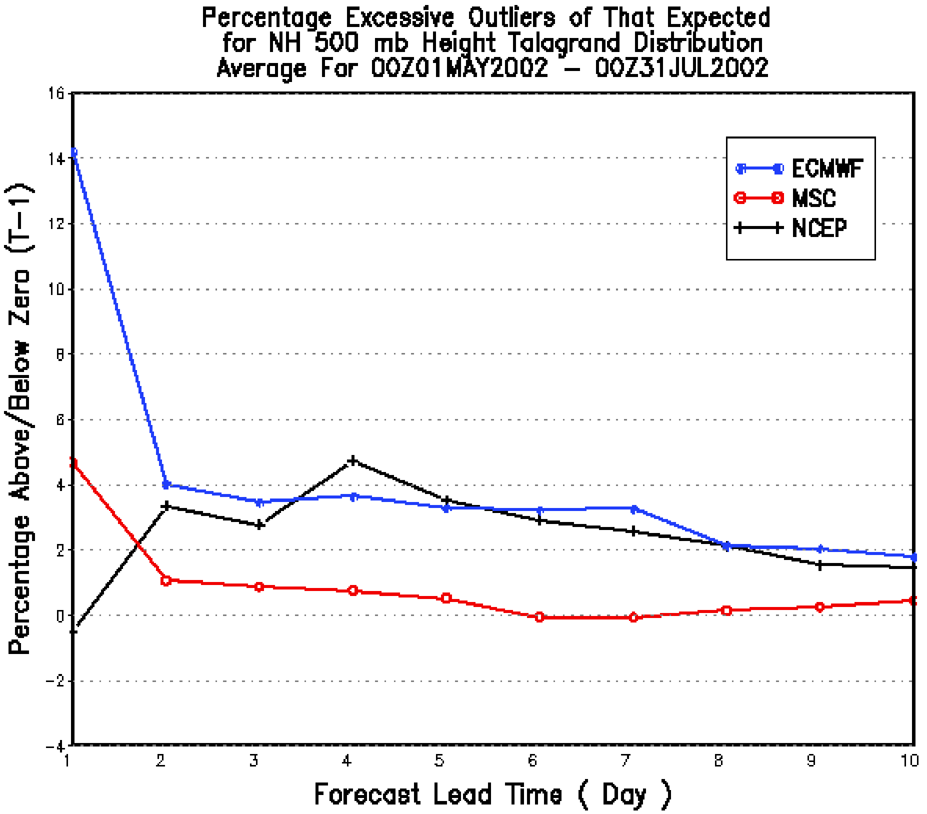
- Assess how often next-day ensemble members fall outside current ensemble

EVALUATION:

- Perfect time consistency:
  - 2/N+1 is expected number
  - Excessive values above expected value shown

RESULTS

- All systems good (except 1-d **EC**)
- **NCEP** best at 1-day lead
- **CANADIAN** best afterward



BRIER SKILL SCORE (BSS)

METHOD:

Compares pdf against analysis

- Resolution (random error)
- Reliability (systematic error)

EVALUATION

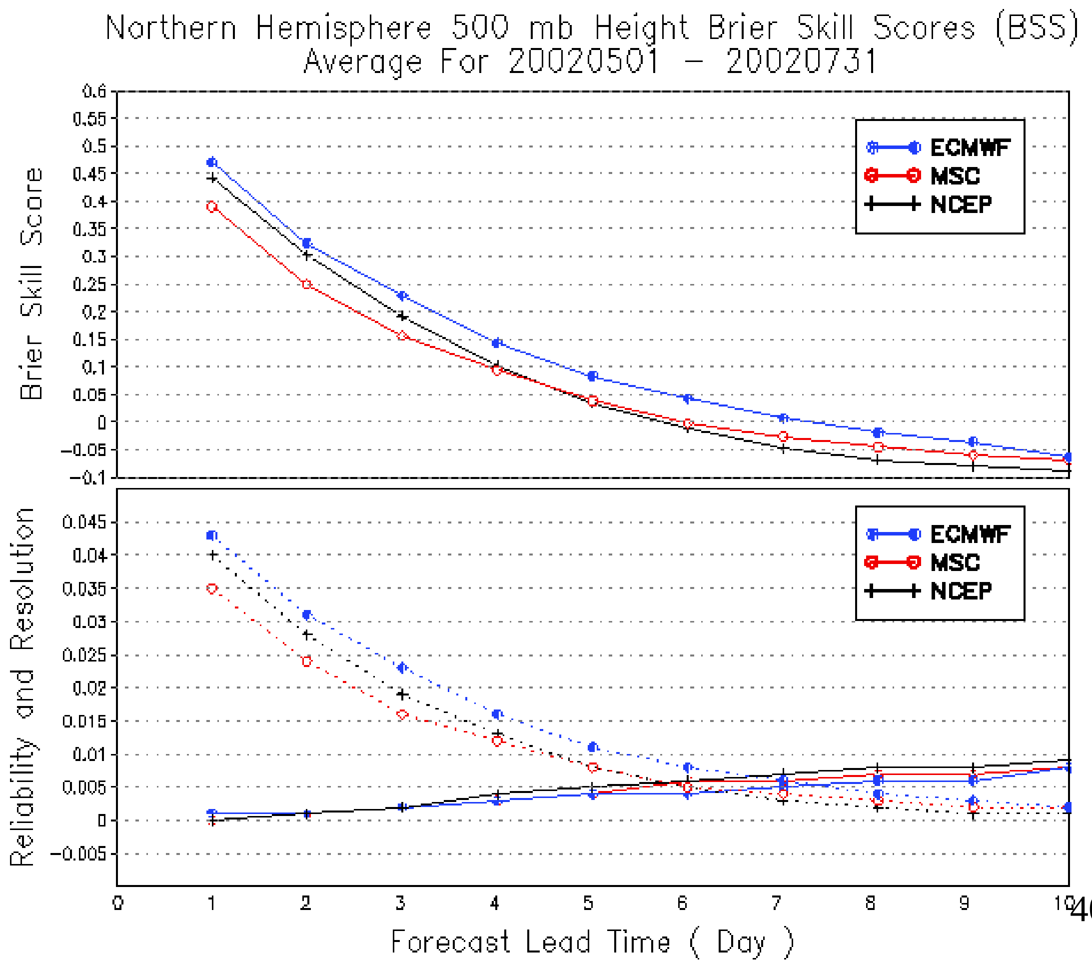
BSS	Higher better
Resolution	Higher better
Reliability	Lower better

RESULTS

Resolution dominates initially

Reliability becomes important later

- **ECMWF** best throughout
  - Good analysis/model?
- **NCEP** good days 1-2
  - Good initial perturbations?
  - No model perturbation hurts?
- **CANADIAN** good days 8-10
  - Model diversity helps?



RELATIVE OPERATING CHARACTERISTICS (ROC)

METHOD:

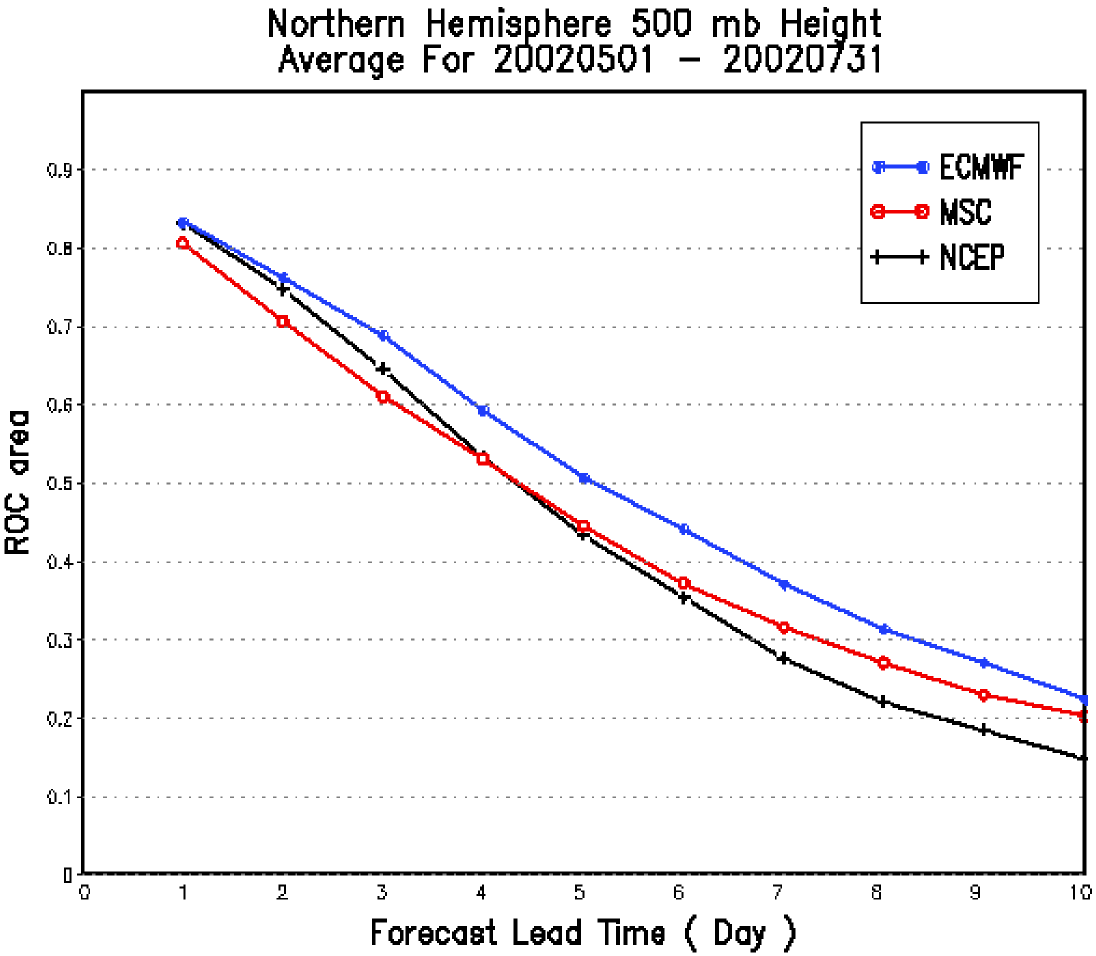
- Plot hit vs. false alarm rate
- Goal:
  - High hit rate &
  - Low false alarm rate
- Measure area under curve

EVALUATION

Larger ROC area better

RESULTS

- **ECMWF** best throughout
  - Better analysis/model?
- **NCEP** very good days 1-2
  - Good initial perturbations?
  - No model perturbation hurts?
- **CANADIAN** good days 8-10
  - Multimodel approach helps?



PERTURBATION VS. ERROR CORRELATION ANALYSIS (PECA)

METHOD: Compute correlation between ens perturbtns and error in control fcst fc

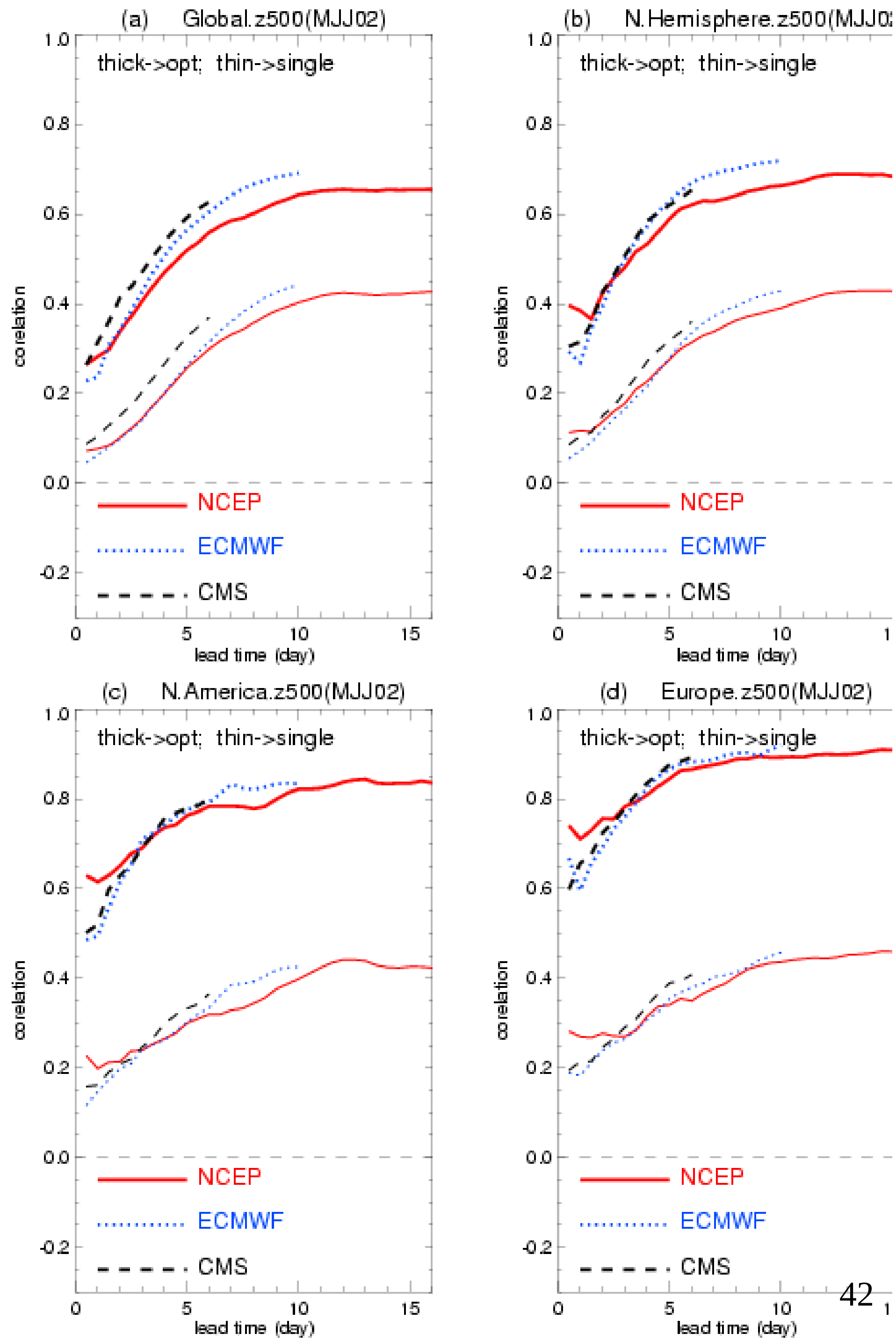
- Individual members
- Optimal combination of members
- Each ensemble
- Various areas, all lead time

EVALUATION: Large correlation indicates ens captures error in control forecast

- Caveat – errors defined by analysis

RESULTS:

- **Canadian** best on large scales
  - Benefit of model diversity?
- **ECMWF** gains most from combinations
  - Benefit of orthogonalization?
- **NCEP** best on small scale, short term
  - Benefit of breeding (best estimate initial error)?
- PECA increases with lead time
  - Lyapunov convergence
  - Nonlilnear saturation
- Higher values on small scales

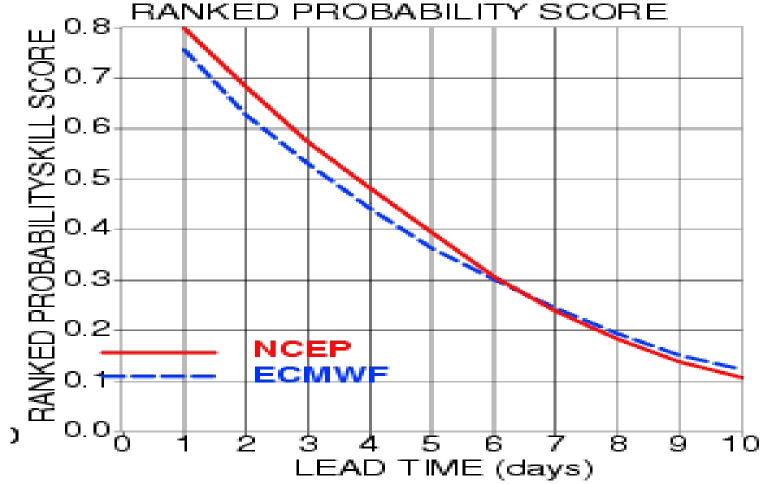


SUMMARY OF FORECAST VERIFICATION RESULTS

Results reflect summer 2002 status

CONTROL FOREST

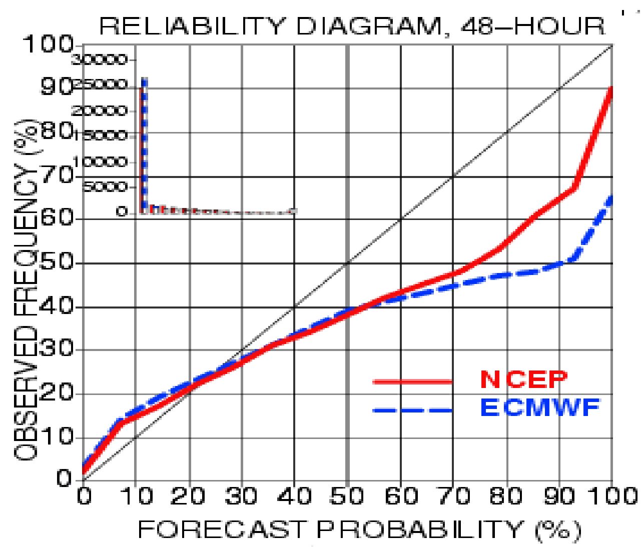
- **ECMWF** best overall control forecast
  - Best analysis/forecast system



Ranked probability skill score for December 1995 – February 1996.

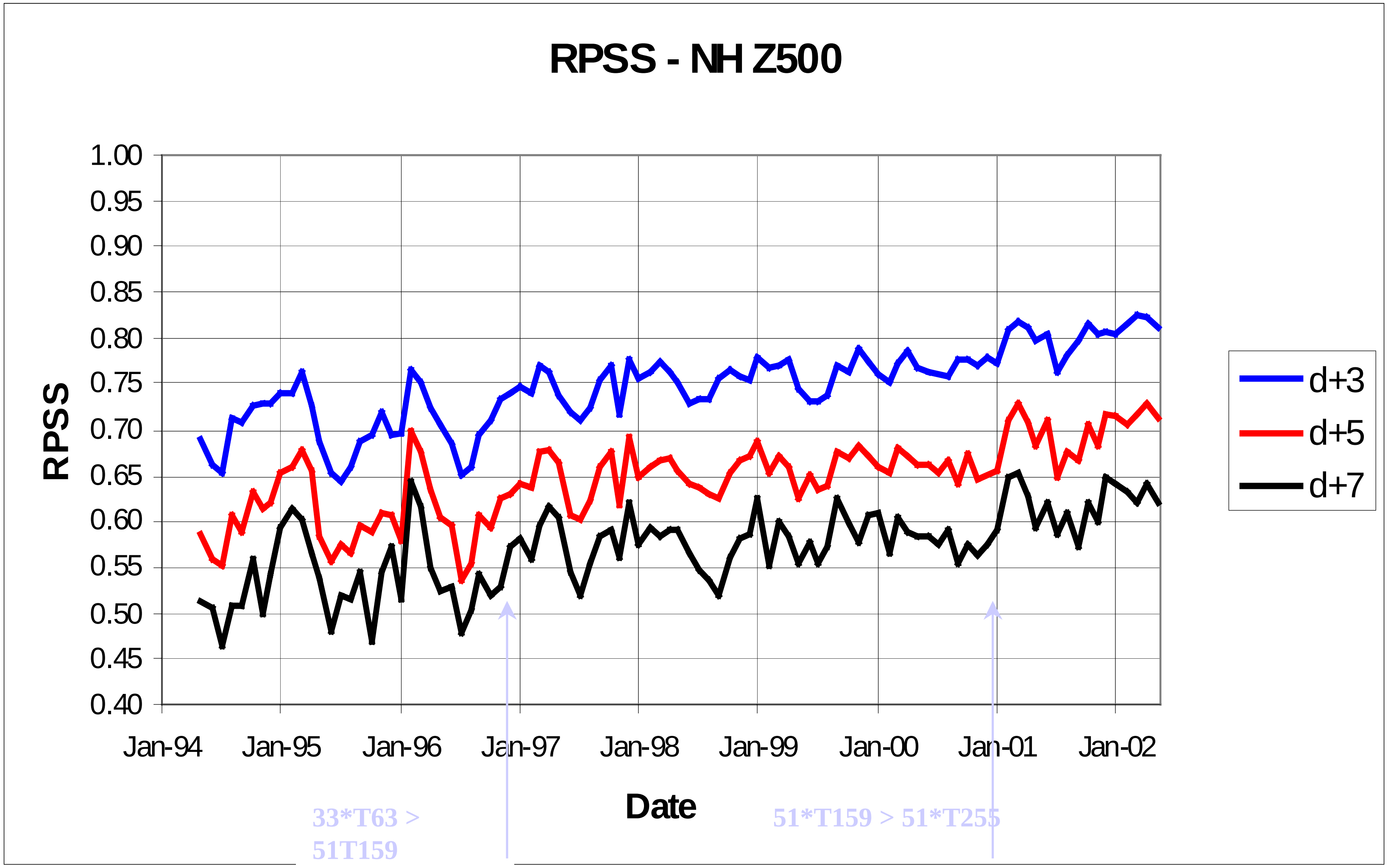
ENSEMBLE FORECAST SYSTEM

- Difficult to separate effect of analysis/model quality
- **ECMWF** best overall performance
- **NCEP**
  - Days 1-3 - Very good (best for PECA)
    - Value of breeding?
  - Beyond day 3 – Poorer performance
    - Lack of model perturbations
- **CANADIAN**
  - Days 6-10 – Better than NCEP
    - Value of model diversity?



Reliability diagram for 48–hour lead time. Forecast probabilities are based on how many ensemble members fell in any particular climate bin at each gridpoint. Insert in upper left corner shows in how many events a particular forecast probability was used. December 1995– February 1996.

EC-EPS: RPSS over NH - d+3, d+5 and d+7



Ongoing research

- MSC:
  - Initial conditions: from an ensemble Kalman filter;
  - Model: development of a sustainable method to perturb the model;
  - Products: automatic generation of ensemble-based worded forecasts.
- ECMWF:
  - Initial conditions: SVs with moist processes, higher resolution, different norm; ensemble data assimilation;
  - Model: higher, possibly variable, resolution; revised stochastic physics;
  - Increased frequency (50 members, 2 times a day).
- NCEP:
  - Initial conditions: use of ETKF for rescaling in breeding method;
  - Model: increased resolution (T126 up to 180h instead of 84h); simulation of model errors;
  - Increased frequency (10 members, 4 times a day).

Open issues

- Is random or selective sampling more beneficial?

  Possible convergence into coupling of data-assimilation and ensemble *(see also T. Hamill's talk).*

- How can an ensemble of first guess fields be used to produce an analysis, or an ensemble of analysis?

  Area of intense research.

- Is optimisation necessary?

  Area of discussion *(see also B. Farrell's talk).*

- How should model error be simulated?

  Need for simulating both random and systematic errors.

- Is having a larger ensemble size or a higher resolution model more important?

  Practical considerations, user needs, post-processing will determine the answer *(see D. Richardson's talk).*