ENSEMBLE PREDICTION: STRENGTHS AND LIMITATIONS

Zoltan Toth[1]

Environmental Modeling Center, NCEP

Acknowledgements:

Yuejian Zhu	EMC[1]
Richard Wobus	EMC[1]
Tim Marchok	GFDL[2]
Hua–Lu Pan	EMC
Stephen Lord	EMC
Istvan Szunyogh	Univ. MD
Eugenia Kalnay	Univ. MD

http://sgi62.wwb.noaa.gov:8080/ens/enshome.html

1. SAIC at NCEP/EMC; 2 SAIC at GFDL

OUTLINE

ENSEMBLE FORECASTING: STRENGTHS AND LIMITS

1) DEFINITION OF CHAOS

2) INITIAL VALUE RELATED PREDICTABILITY/UNCERTAINTY

3) BOUNDARY CONDITION RELATED PREDICT./UNCERTAINTY

4) BRIDGING THE GAP BETWEEN WEATHER AND CLIMATE

5) ADVANTAGES OF USING ENSEMBLE FCSTS

6) PROBLEMS WITH ENSEMBLES

7) VERIFICATION

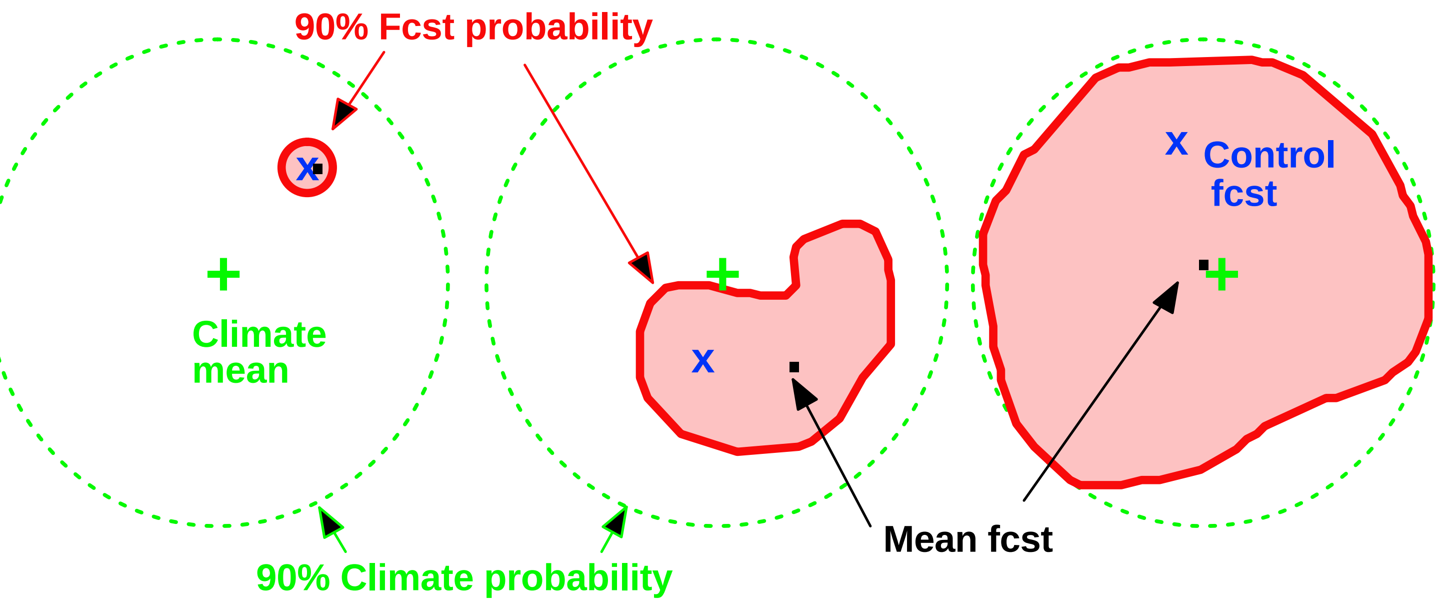
8) POSTPROCESSING

INITIAL VALUE RELATED PREDICTABILITY

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

2) **Initial** state has **error** in it ==>

Chaotic system + Initial error =(Loss of) Predictability



Initial time	Day 5 Large uncretainty	Day 12 Almost all predictability is lost – full nonlinear saturation
Ocean/Atm coupled system	5 months	12 months

Even with perfect model, fcsts diverge and fail – example

CONSEQUENCE:

Single control forecast does not contain all information

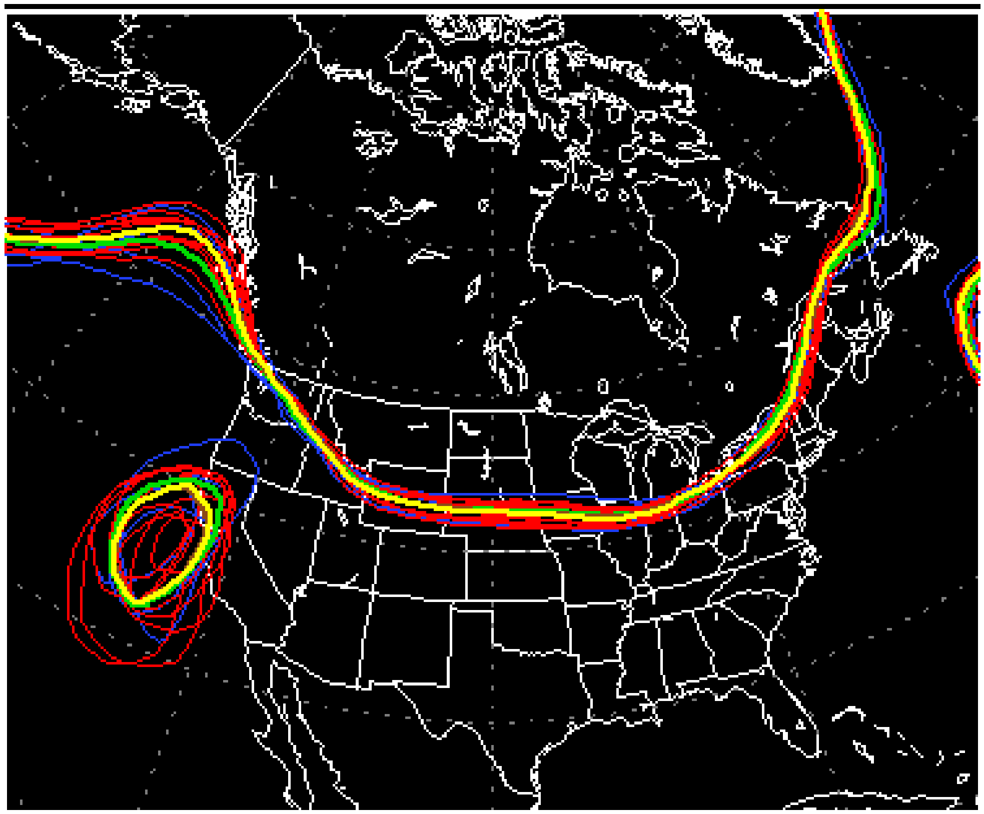
a) How likely this scenario is?

b) What other solutions possible?

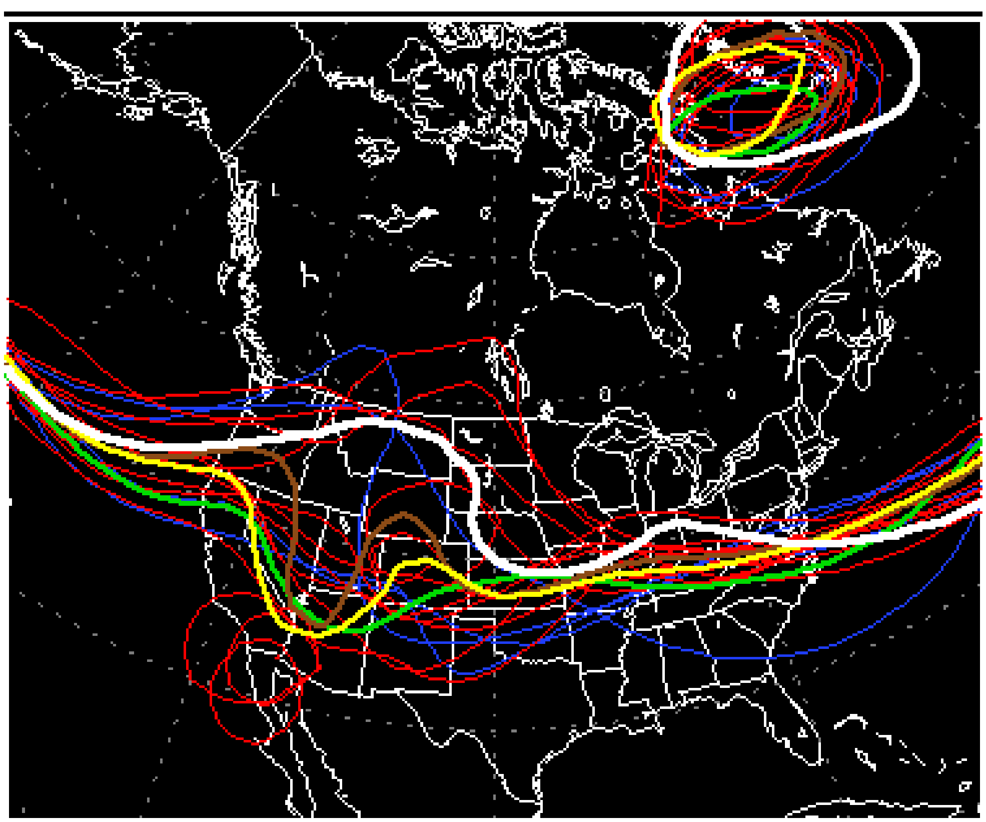
Due to nonlinearities, control fcst is not best estimate

USE ENSEMBLES

NCEP ENSEMBLE, INITIAL TIME, 20010226 0000 UTC



96–HR FORECAST

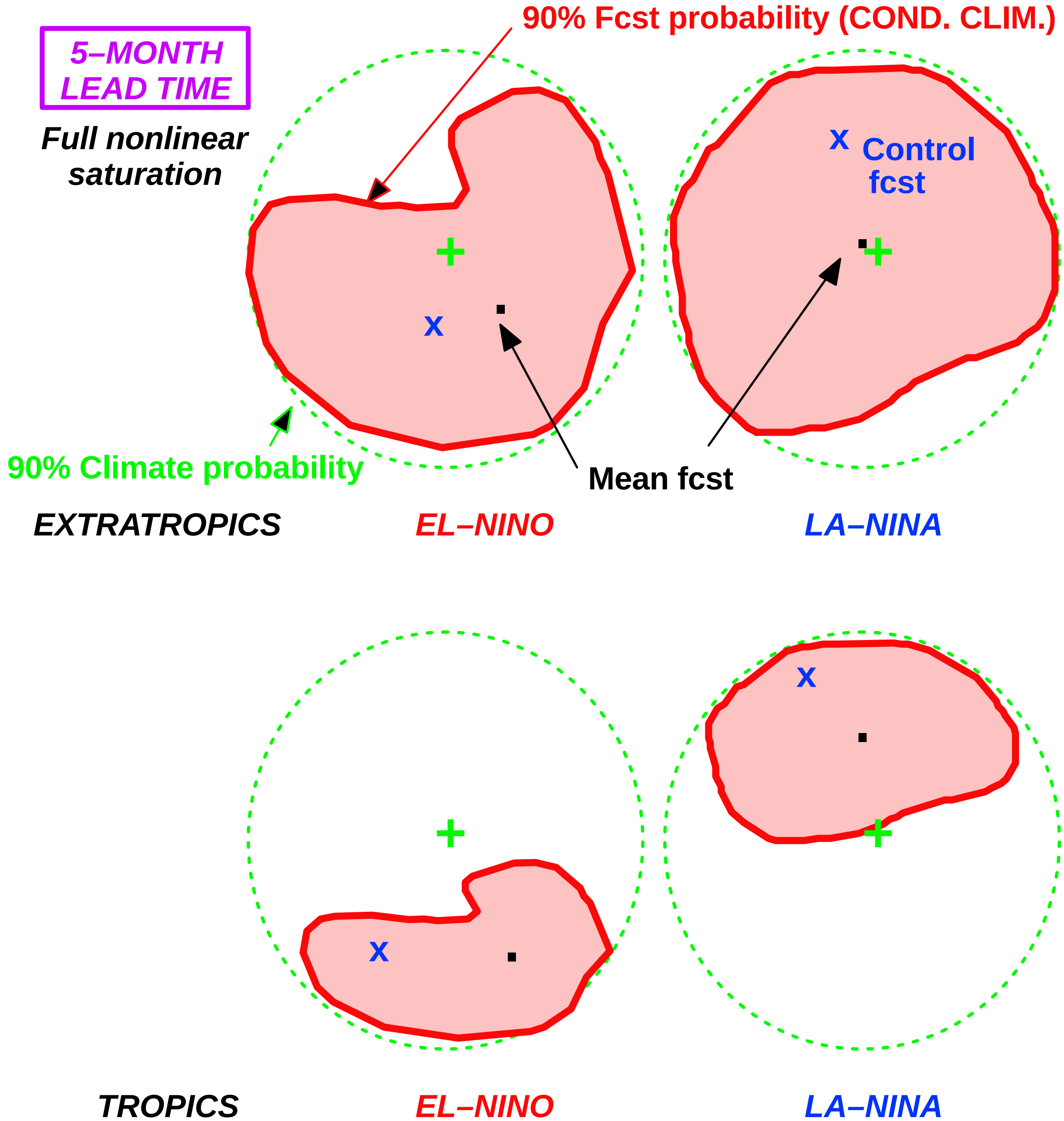


BOUNDARY FORCING RELATED PREDICTABILITY

1) **Different boundary conditions** (eg, SSTA) => Different climates – **Conditional Climatology**

2) **Predict** slowly varying **boundary conditions** – **Run** atmosph. **ensemble** to find ***corresponding climate***

2nd kind of predictability due to boundary forcing



WEATHER FORECASTING, DAY 10

1) **Unconditional climate distribution**

No forecast made

2) **Conditional climate (ENSO forcing)**

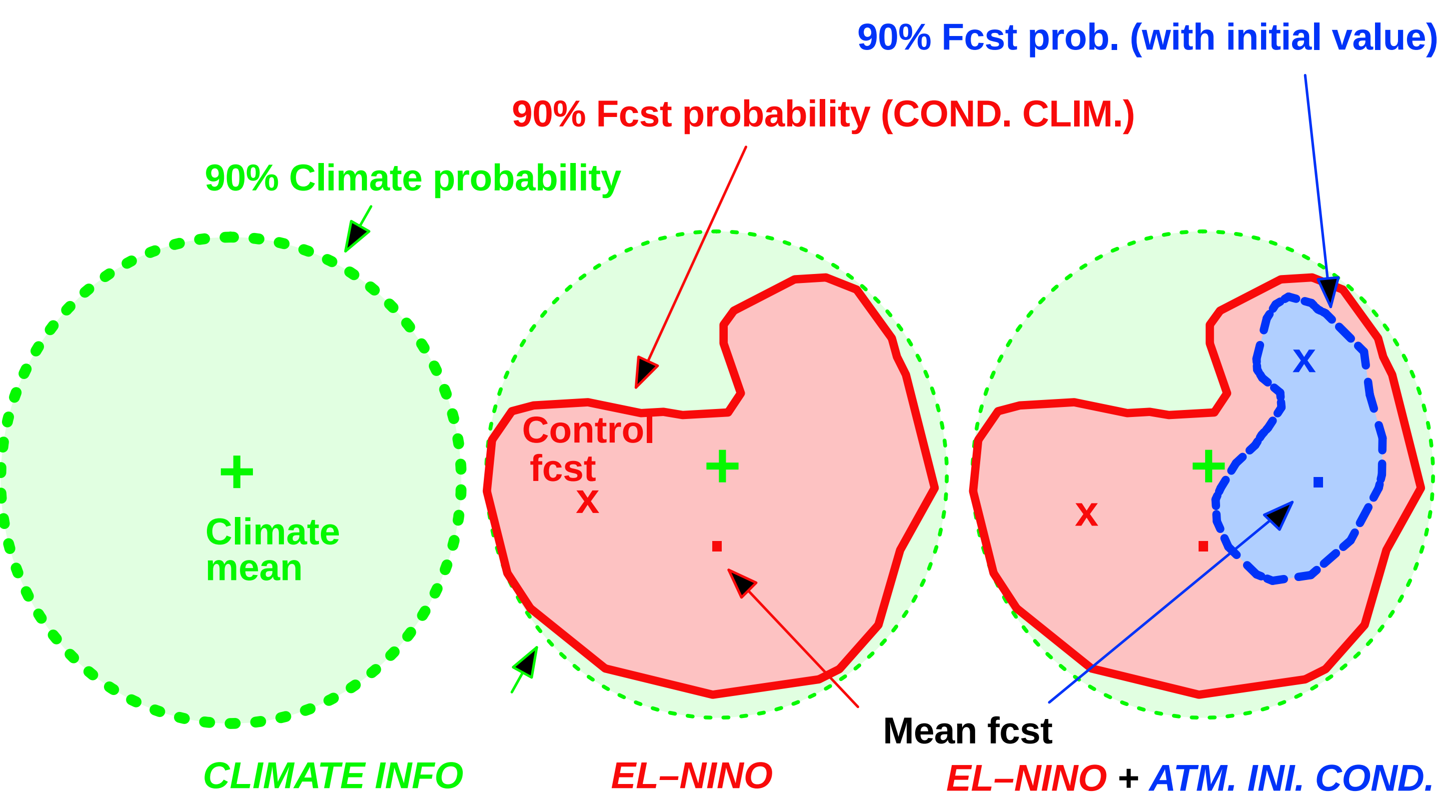
AGCM ensemble, El–Nino forcing –

Second kind of predictability

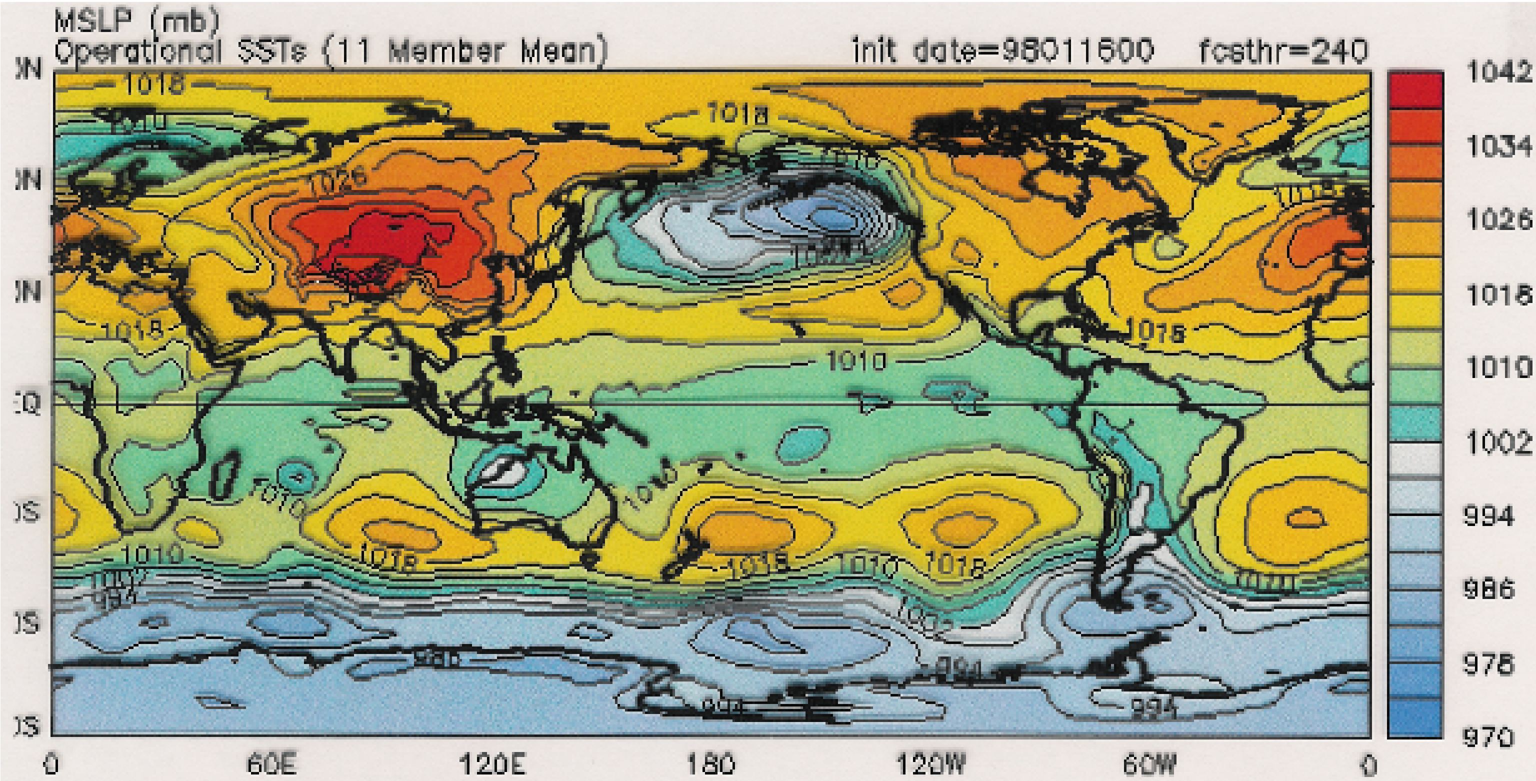
3) **Initial + Boundary condition forecast**

Initial value + boundary forcing

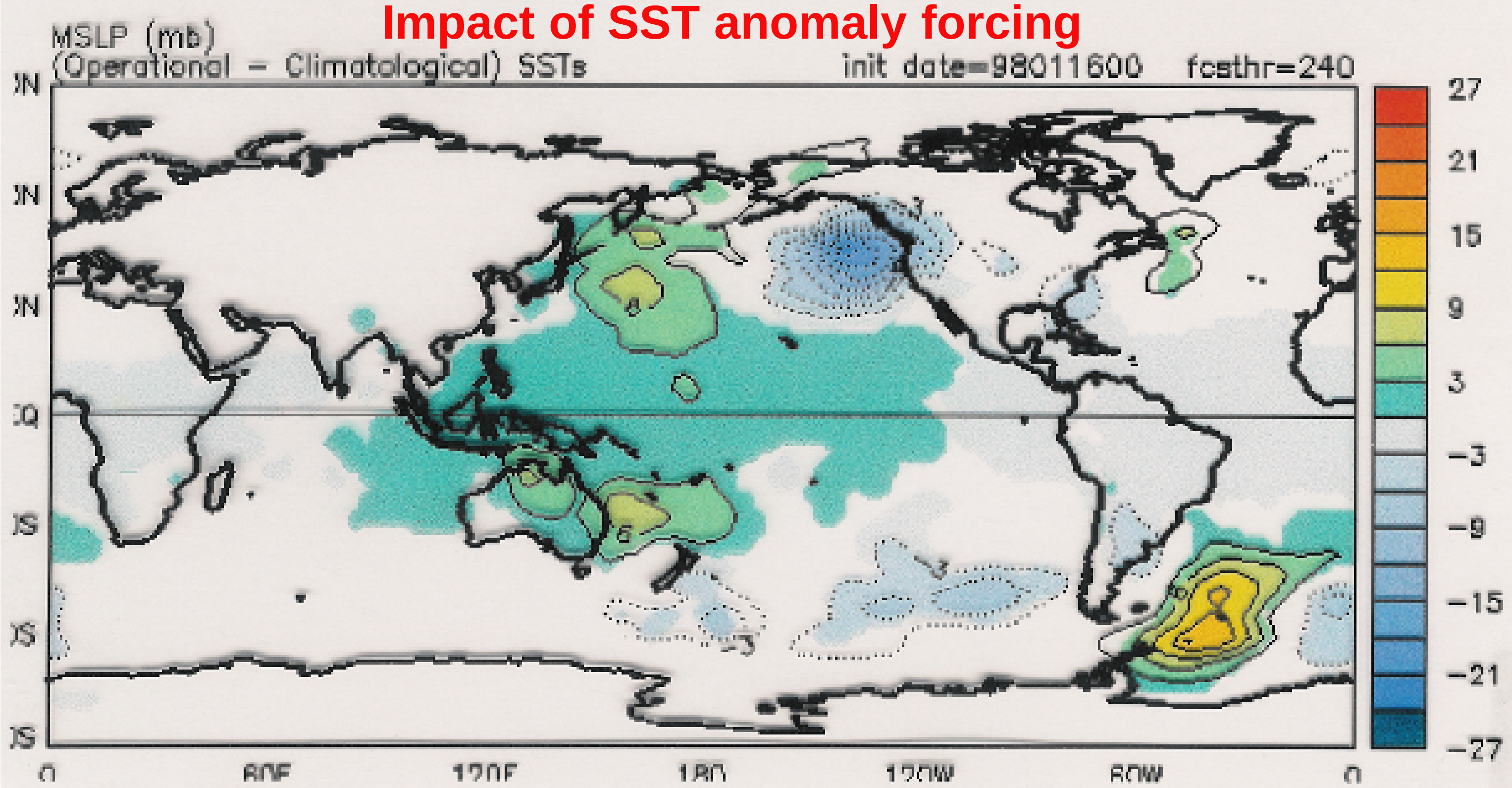
Full predictability (First & second kinds)



Operational 10–day fcst, persisted SST forcing



Impact of SST anomaly forcing



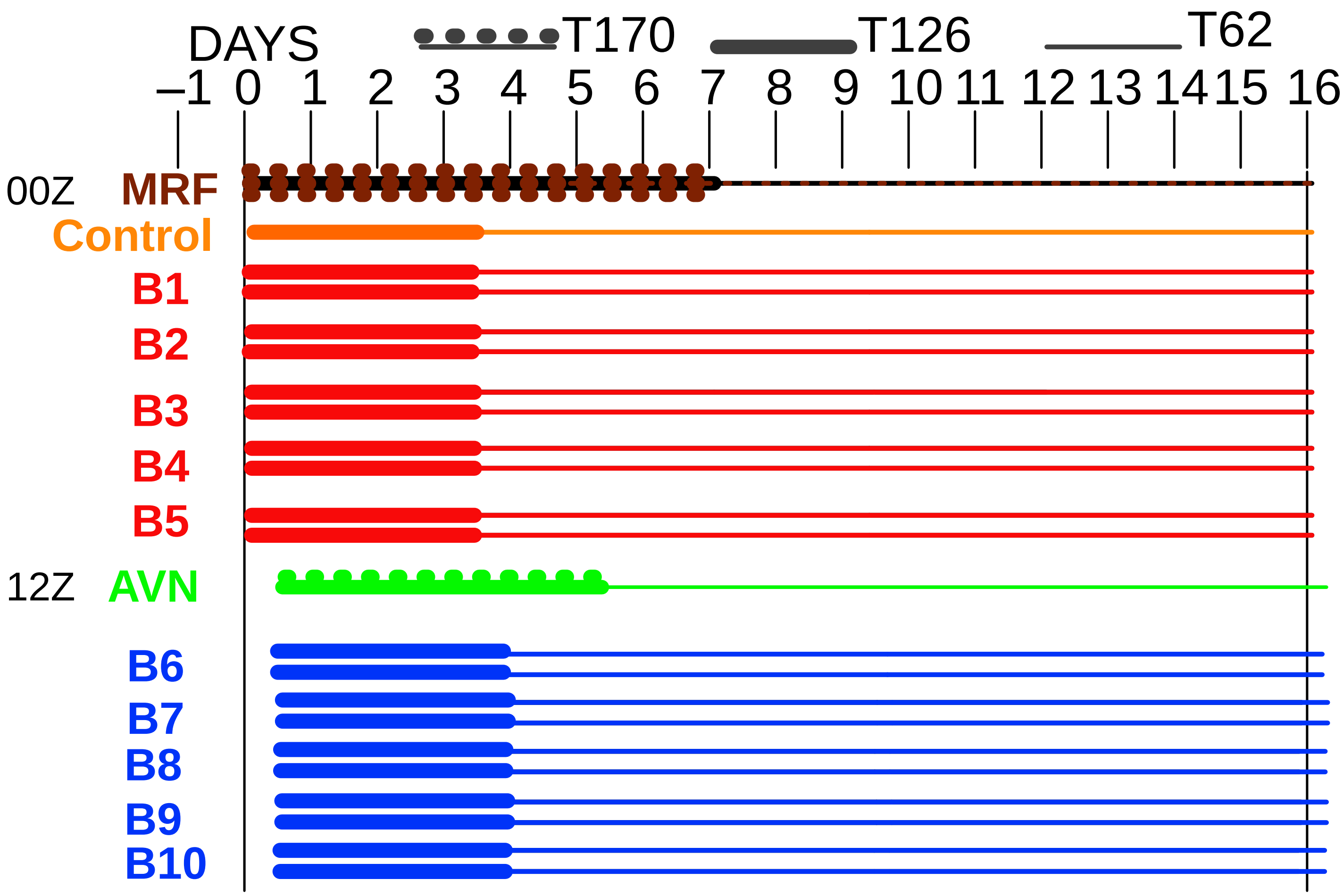
Barsugli et al.

BRIDGING THE GAP BETWEEN WEATHER AND CLIMATE

CURRENT NWS PRACTICE

1) "WEATHER" ENSEMBLE:

With damped persistence SST boundary forcing



PLANNED CHANGES FOR ENSEMBLE IN 2002:

a) 10 forecasts four times per day

b) T126 resolution extended to 180 hrs

SHORTCOMING:

No forecast SST information used from coupled system

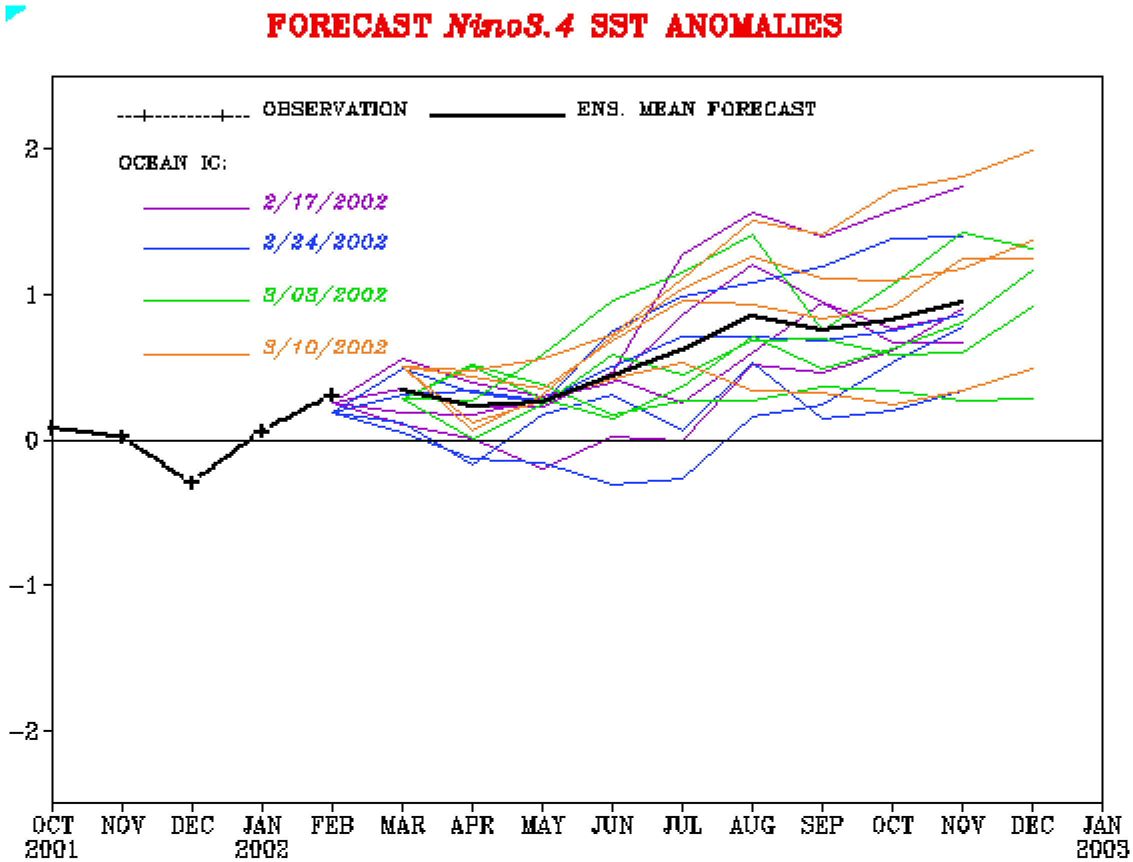
BRIDGING THE GAP BETWEEN WEATHER AND CLIMATE

CURRENT NWS PRACTICE

2) "CLIMATE" ENSEMBLE:

a) 12–months coupled ocean–atm fcsts

b) Average the SST fcsts



c) Run AGCM ensemble forced by average SST fcst

STRENGTH:

Ensemble approach used both for coupled and AGCM model fcsts for enhancing (weak) signal

SHORTCOMINGS:

a) Coupled ensemble (lagged fcst) perturbations not optimal

b) Uncertainty information related to SST fcst is discarded

c) Initial condition information from atmosphere not used

BRIDGING THE GAP BETWEEN WEATHER AND CLIMATE

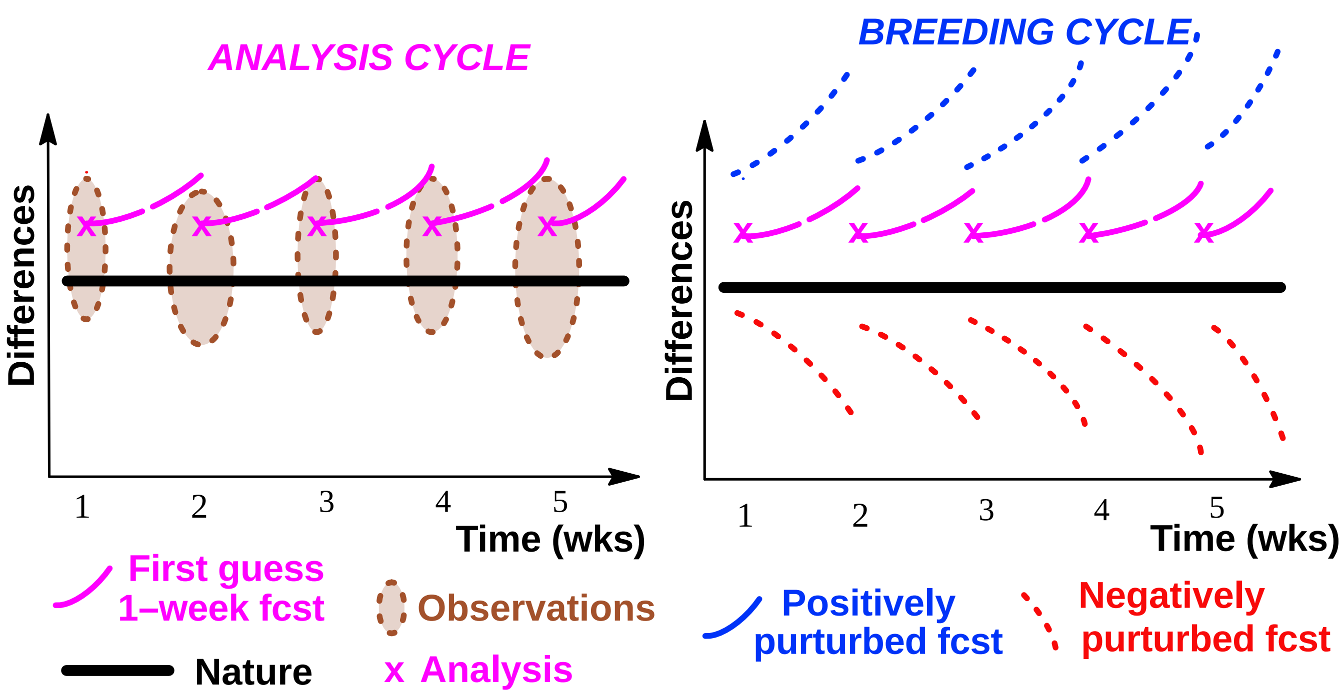
PLANS

3) POSSIBLE FUTURE SYSTEM:

"WEATHER *AND* CLIMATE" ENSEMBLE?

COUPLED MODEL ENSEMBLE –

Use dynamically constructed perturbations



a) *Nonlinear bred perturbations capture dominant ENSO instability*

b) *Initial error present in analysis dominated by same instability*

c) *Symmetrically placed perturbed fcsts provide optimal ensemble*

AGCM ENSEMBLE –

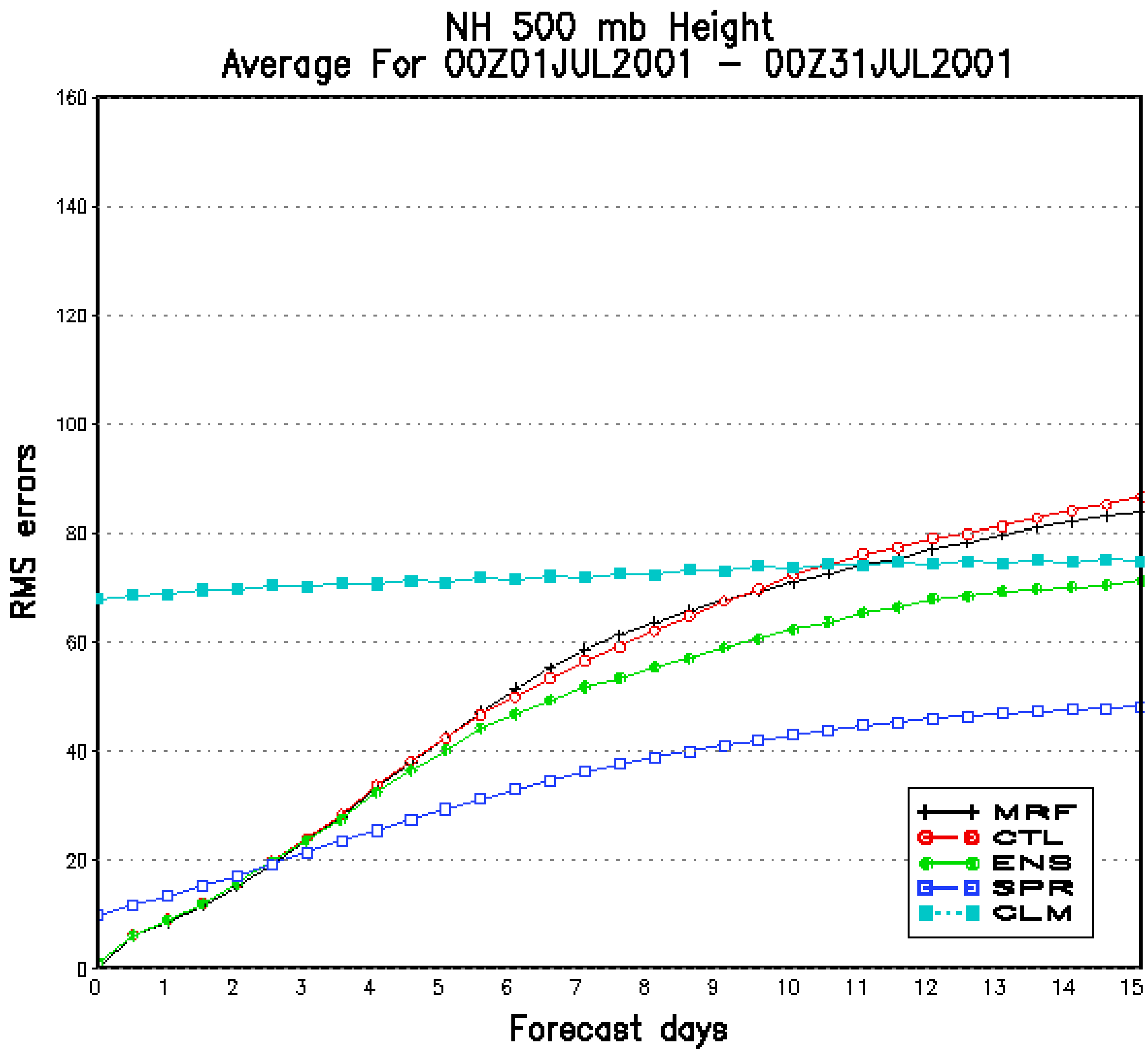
i) Use ensemble SST fcsts as various boundary scenarios

ii) Single set of AGCM fcsts for all time ranges *(D1–climate)*

ONE–TIER SYSTEM – If possible, with new ocean model

ENSEMBLE APPROACH – ADVANTAGES

1) Improved estimate of first moment (ens mean)



2) Estimates fcst uncertainty (probabilistic fcsting)

3) "Traces" of weather provided by ensemble members – Most queries can be answered – eg:

What is the probability of 3 consecutive days with frost next week?

4) Offers flow dependent filter – *no need for time means*

PROBABIL. DAILY WEATHER FCSTS – DAY 1 through CLIMATE

PDF varies slower in time/space as predictability diminishes

CAN REGIME TRANSITIONS BE PREDICTED BEYOND DAY 10?

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:

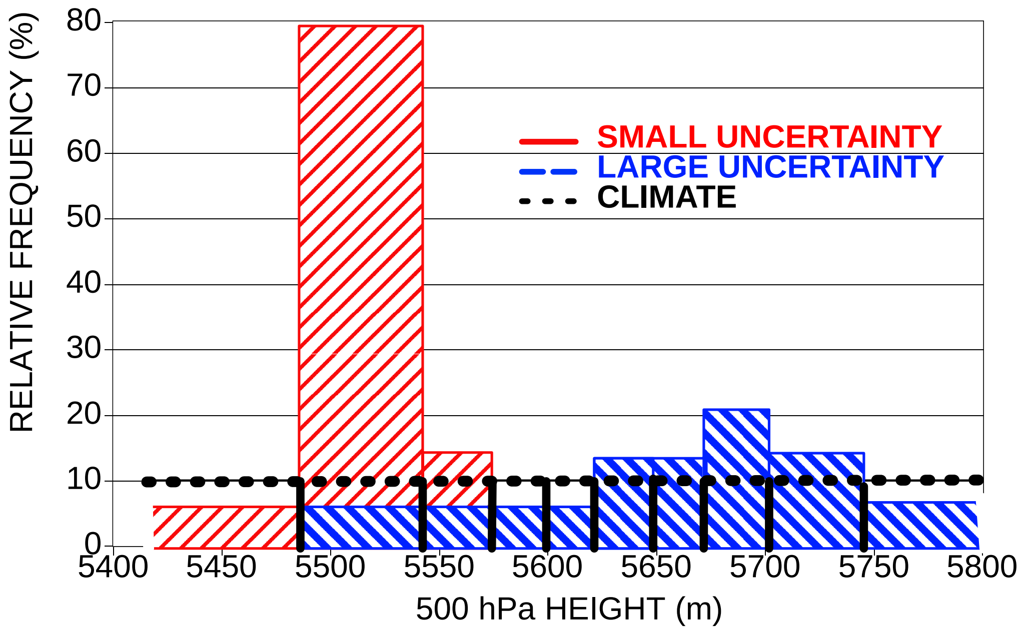
Ensemble mode value to distinguish high/low predictability cases
Stratify cases according to ensemble mode value –
Use 10–15% of cases when ensemble mode is highest/lowest

DATA:

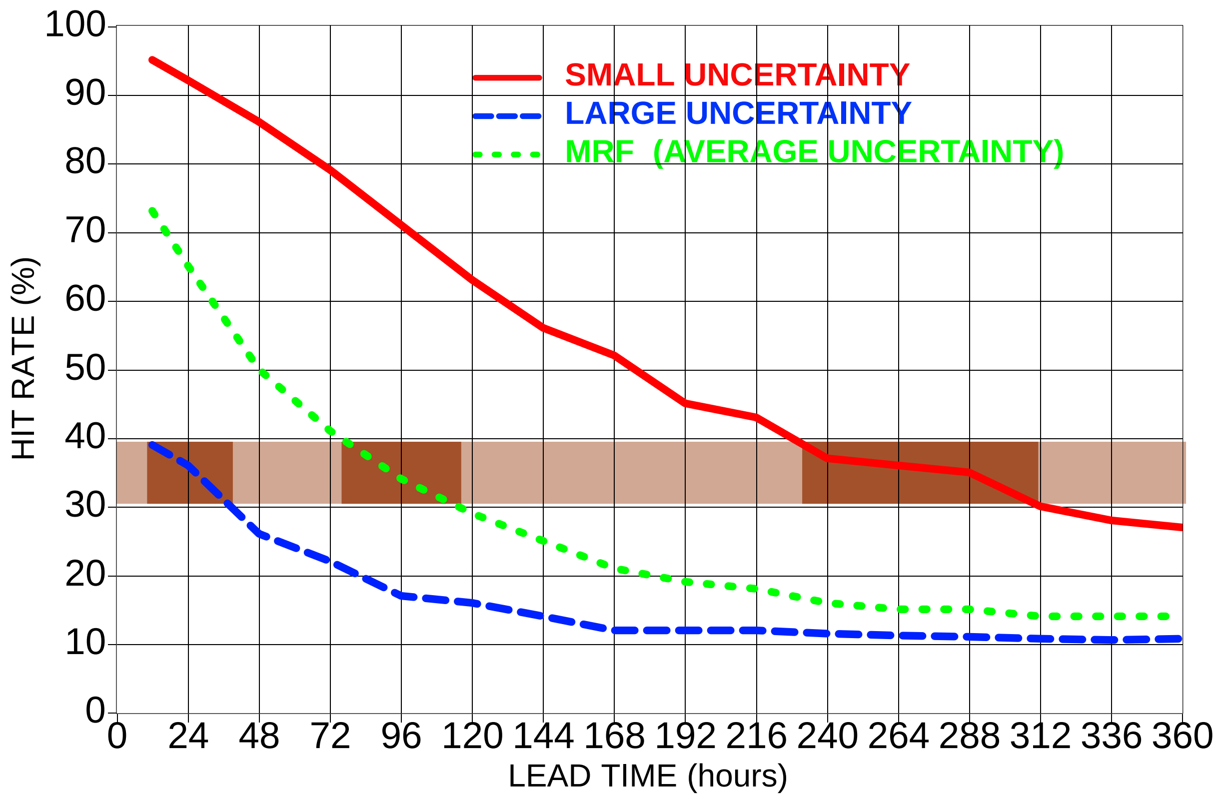
NCEP **500 hPa NH extratropical ensemble fcsts** for March–May 1997
14 perturbed fcsts and high resolution control

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***

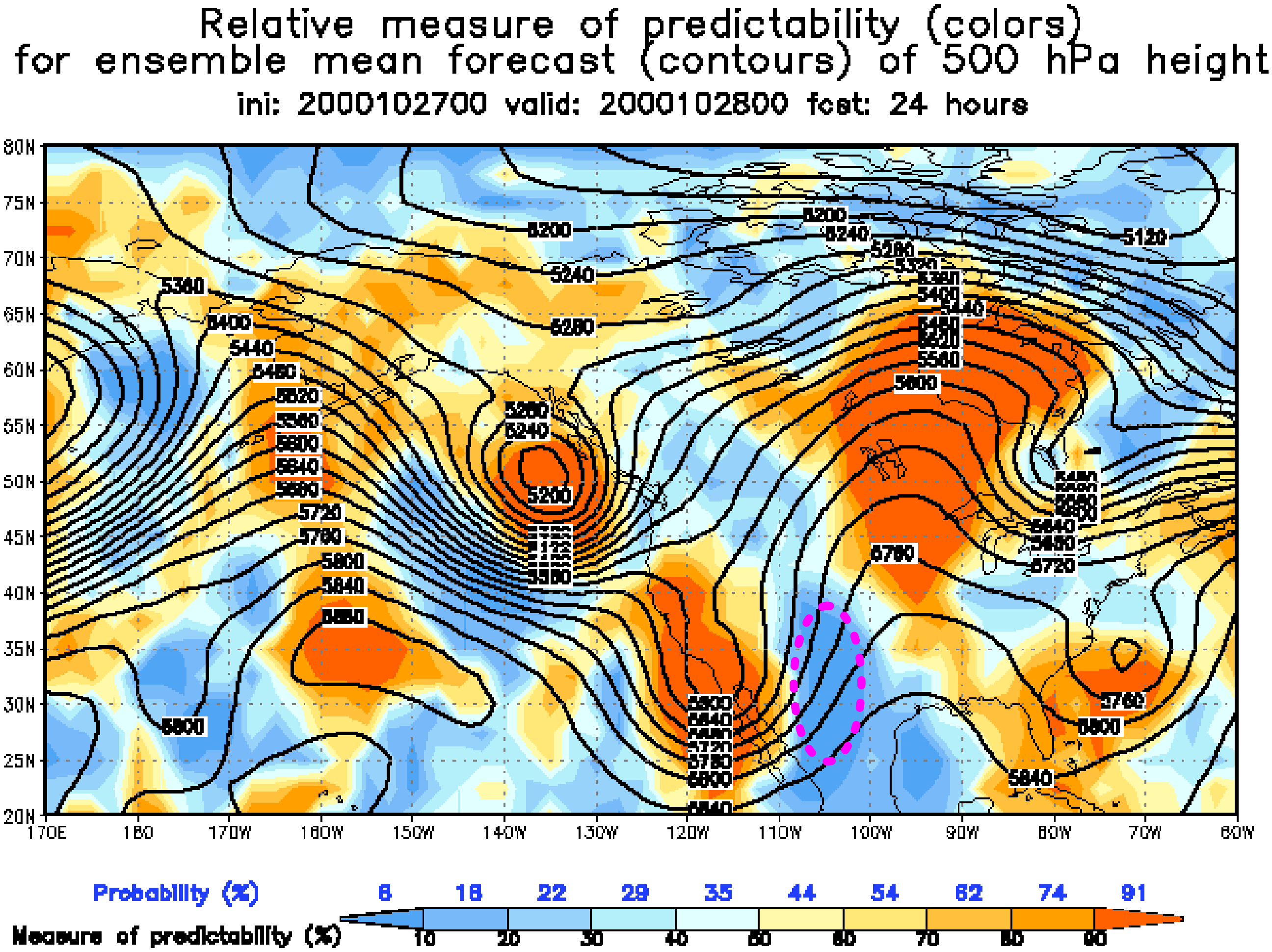
HIT RATES FOR 1–DAY FCSTS

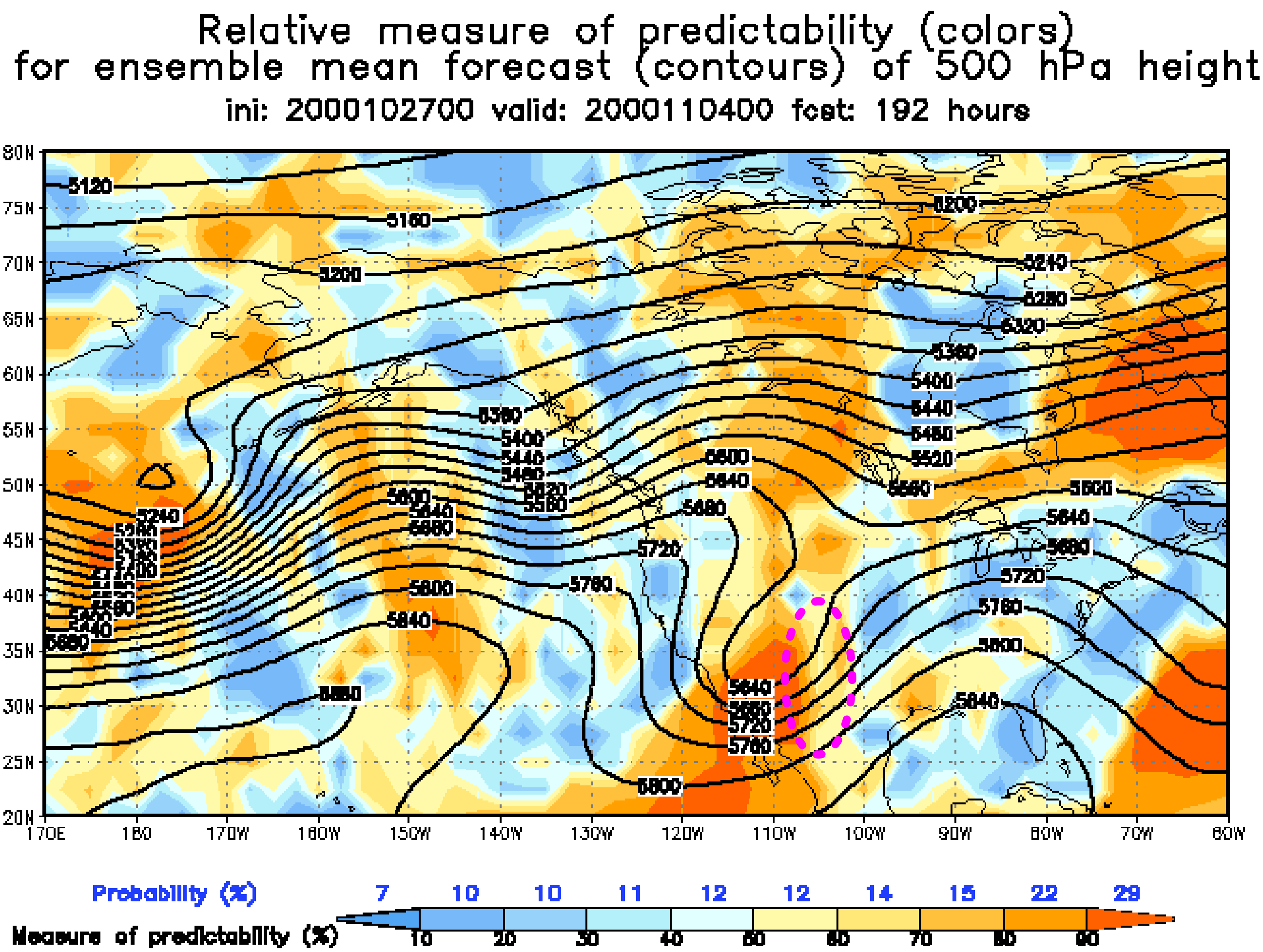
*CAN BE **AS LOW AS 36%**, OR **AS HIGH AS 92%***

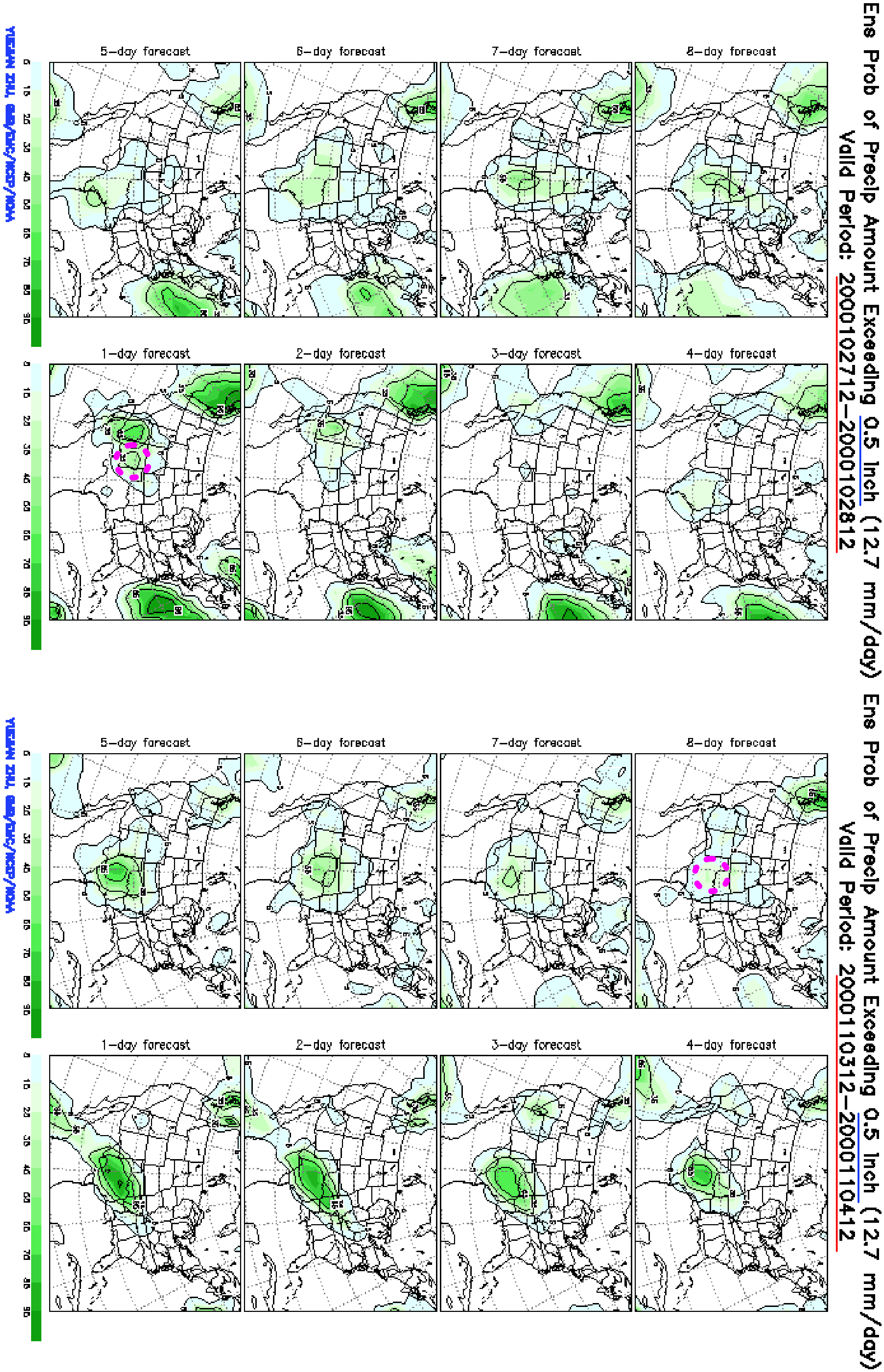
10–15% OF THE TIME A 12–DAY FCST CAN BE AS GOOD, OR A 1–DAY FCST CAN BE AS POOR AS AN AVERAGE 4–DAY FCAST

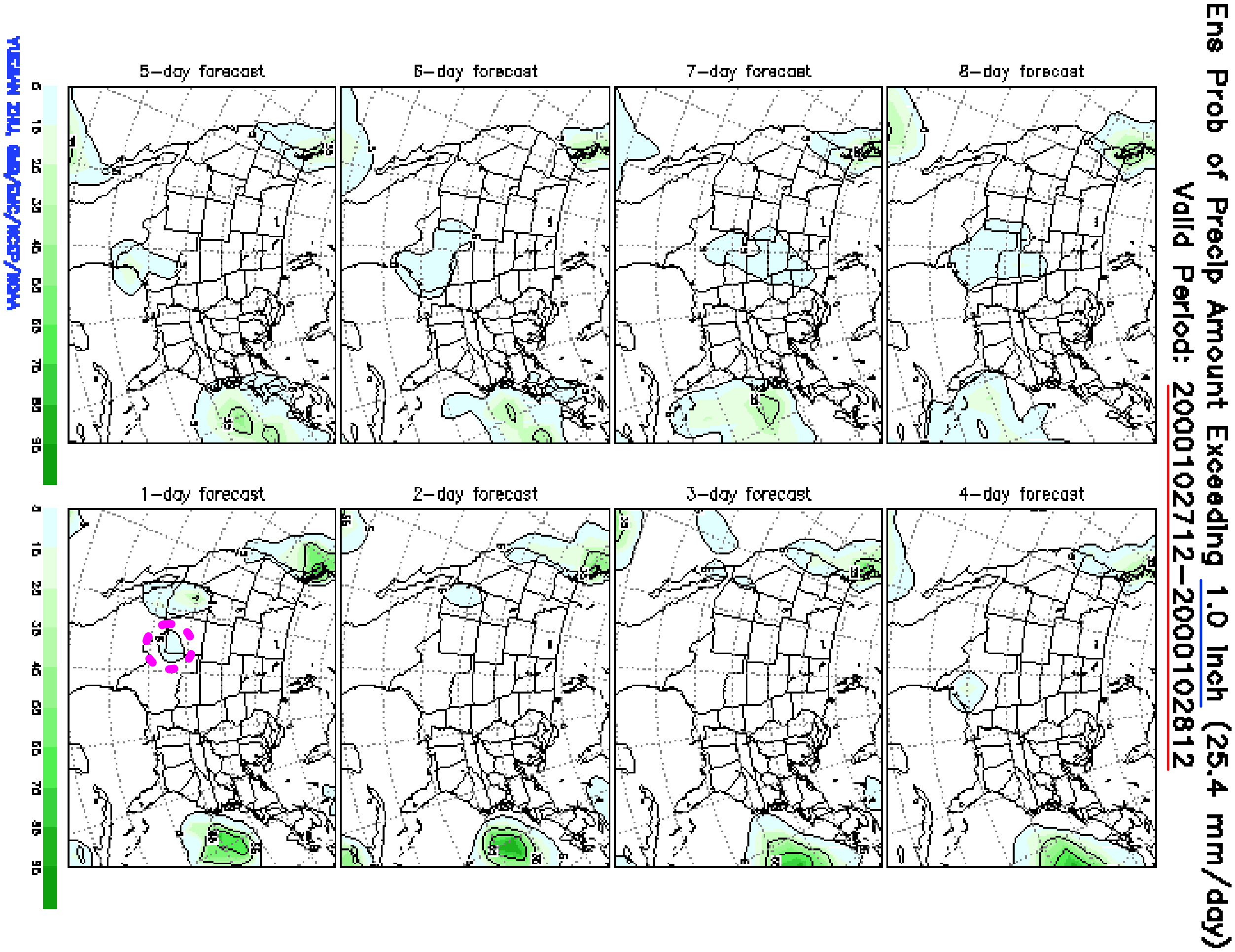
1–2% OF ALL DAYS THE 12–DAY FCST CAN BE MADE WITH MORE CONFIDENCE THAN THE 1–DAY FCST

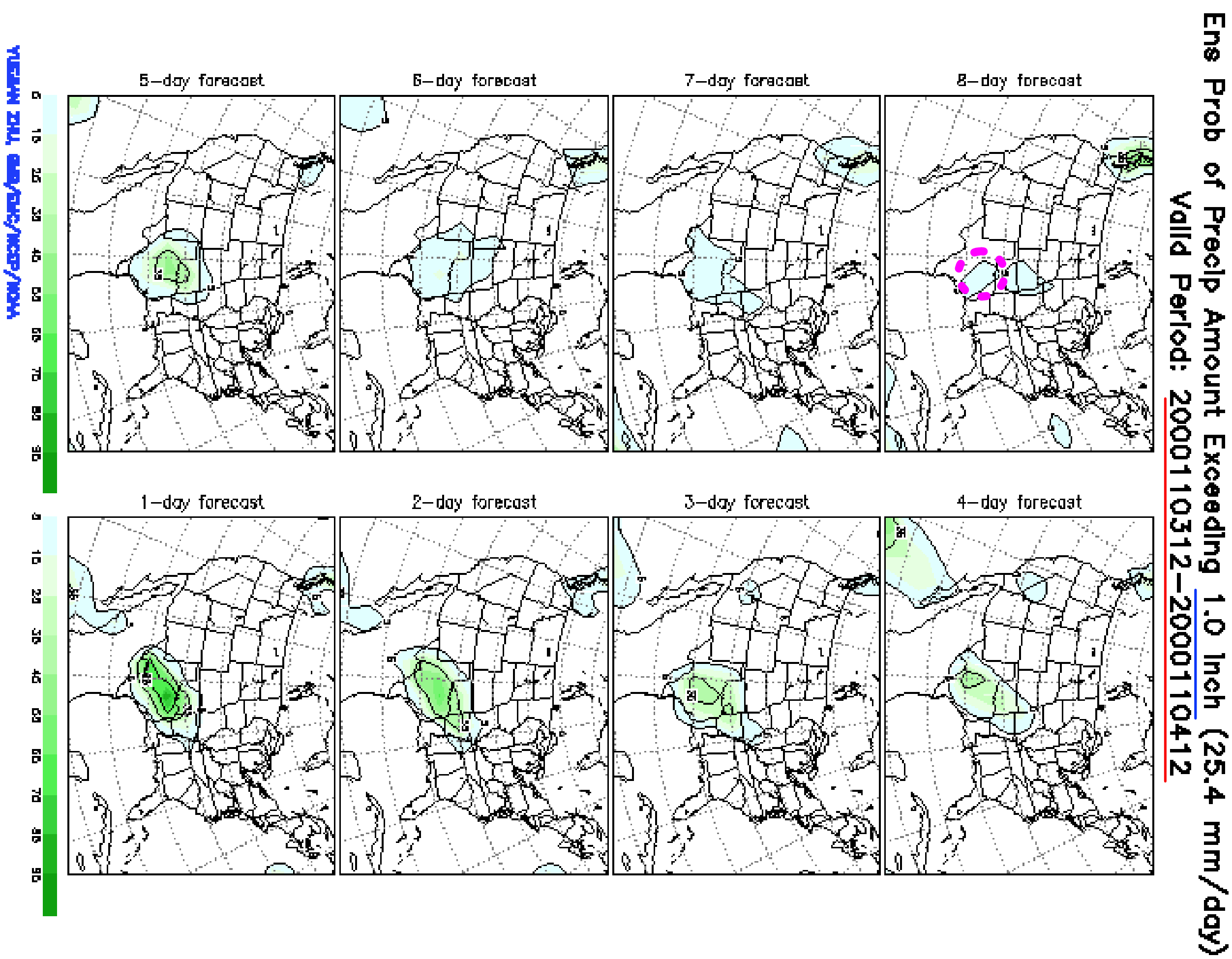
AVERAGE HIT RATE FOR EXTENDED–RANGE FCSTS IS LOW –
VALUE IS IN KNOWING WHEN FCST IS RELIABLE

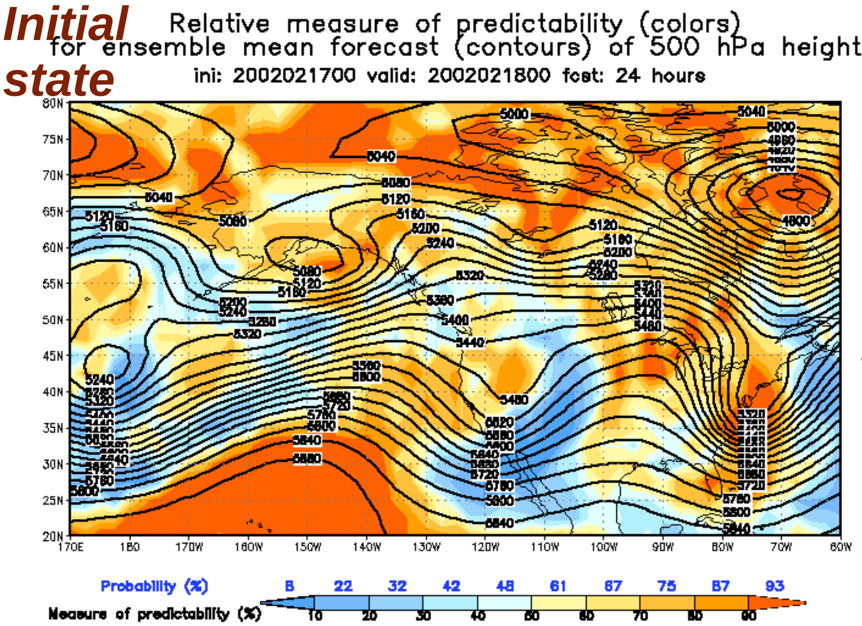

Relative measure of predictability (colors)
for ensemble mean forecast (contours) of 500 hPa height
ini: 2000102700 valid: 2000102800 fcst: 24 hours
Probability (%) 6 16 22 29 35 44 54 62 74 91
Measure of predictability (%) 10 20 30 40 50 60 70 80 90

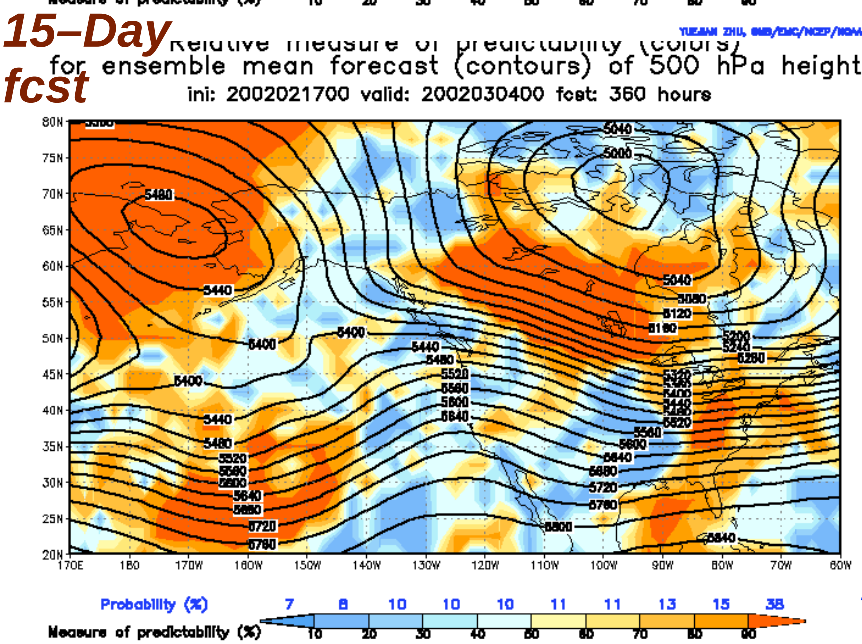

Relative measure of predictability (colors)
for ensemble mean forecast (contours) of 500 hPa height
ini: 2000102700 valid: 2000110400 fcst: 192 hours
Probability (%) 7 10 10 11 12 12 14 15 22 29
Measure of predictability (%) 10 20 30 40 50 60 70 80 90

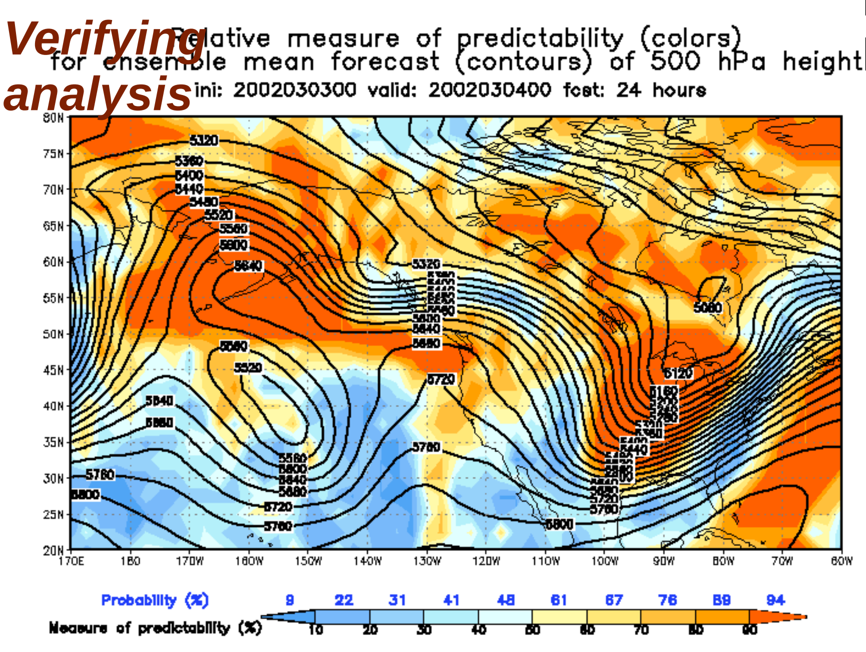

Ens Prob of Precip Amount Exceeding 0.5 inch (12.7 mm/day)
Valid Period: 2000102712–2000102812
1–day forecast
2–day forecast
3–day forecast
4–day forecast
5–day forecast
6–day forecast
7–day forecast
8–day forecast
Ens Prob of Precip Amount Exceeding 0.5 inch (12.7 mm/day)
Valid Period: 2000110312–2000110412


Ens Prob of Precip Amount Exceeding 1.0 inch (25.4 mm/day)
Valid Period: 2000102712–2000102812
1–day forecast
2–day forecast
3–day forecast
4–day forecast
5–day forecast
6–day forecast
7–day forecast
8–day forecast


Ens Prob of Precip Amount Exceeding 1.0 inch (25.4 mm/day)
Valid Period: 2000110312–2000110412
1–day forecast
2–day forecast
3–day forecast
4–day forecast
5–day forecast
6–day forecast
7–day forecast
8–day forecast







GREAT CHALLANGE:

PREDICTING REGIME TRANSITION AT EXTENDED RANGE

ENSEMBLE APPROACH:

CAN BE DONE SUCCESSFULLY

(FROM TIME TO TIME)

LOW LEVEL OF SKILL:

USE PROBABILISTIC FCSTS

+/– 24 HR TIMING ACCURACY:

USE DAILY FCST DATA

(NOT TIME MEAN)

FALSE ALARMS:

NEED TO SIMULATE NATURE'S DIVERSITY IN NWP MODEL

MISSED EVENTS:

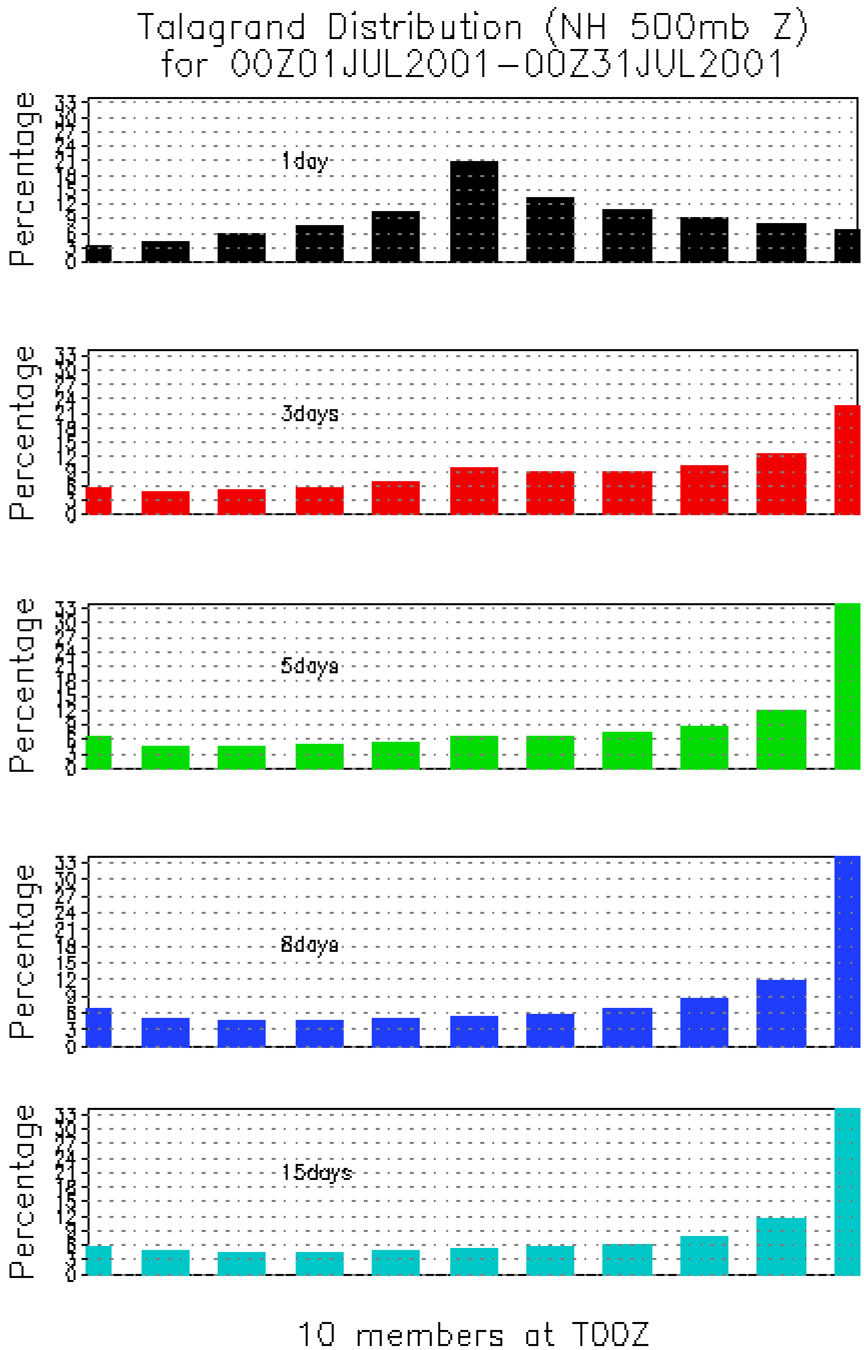
1) LIMIT OF PREDICTABILITY –
NO PROBLEM
2) MODEL FAILURE –
NEED TO INCREASE MODEL DIVERSITY
(NOT "ACCURACY")

ENSEMBLE APPROACH – PROBLEM

UNCERTAINTY DUE TO MODEL ERROR IS NOT ACCOUNTED FOR YET

=> Bias in first and second moment of ensemble

=> 10–25% of time ensemble misses verifying analysis



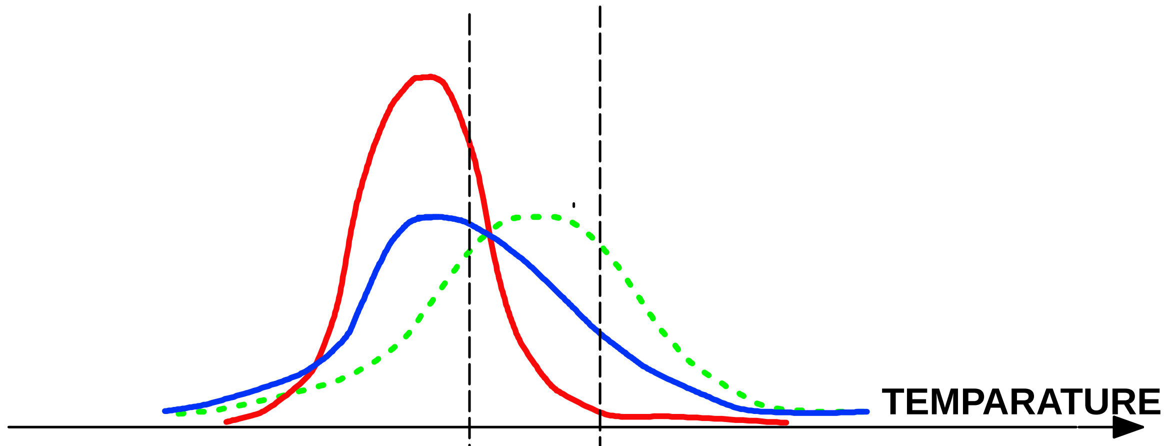
NEED TO INCORPORATE MODEL RELATED UNCERTAINTY –

Vary components/parameters in model

UNTIL THEN:

STATISTICAL POSTPROCESSING – Need large sample

WHY SHOULD USERS (AND NOT ONLY *FCSTERS*) CARE ABOUT UNCERTAINTIES IN WEATHER FCSTS?



CLIMATE PROB	33%	33%	33%
FCST WITH LARGE UNCERTAINTY	48%	29%	23%
MORE RELIABLE FCST	85%	12%	3%

BOTH FCSTS CALL FOR BELOW AVERAGE TEMPERATURES

ECONOMIC EXAMPLE: COMPANY SELLING "WEATHER DERIVATIVES" (INSURANCE THAT FCST IS CORRECT)

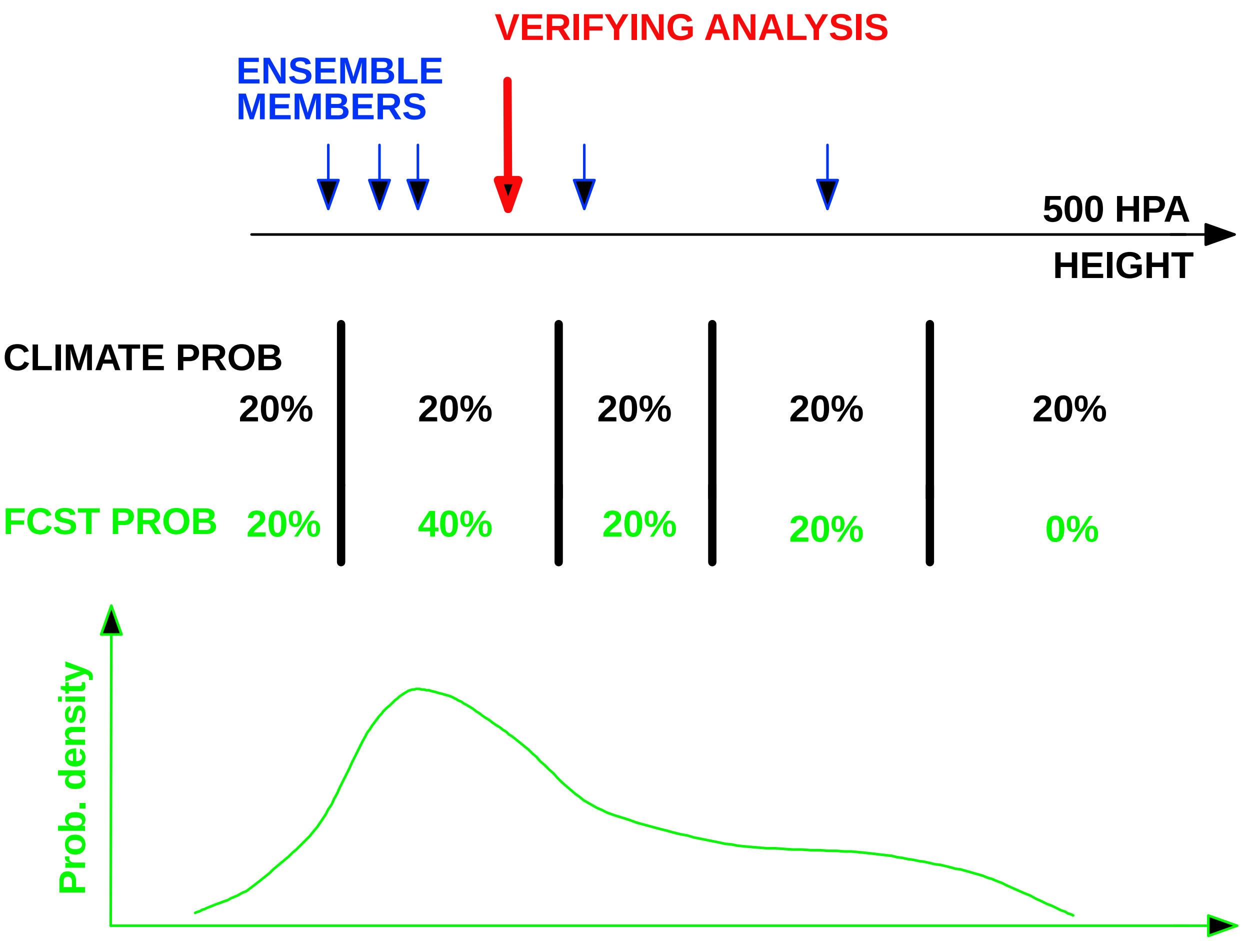
FOR UNCERTAIN FCST, INSURANCE PREMIUM MUST BE HIGHER

IT IS ESSENTIAL THAT THE USERS KNOW ABOUT THE UNCERTAINTIES ASSOCIATED WITH THE WEATHER FCSTS

HOW TO CONSTRUCT PROB FCSTS?

Ensemble gives forecast probability estimates based on finite sample

NEED TO ***INTERPOLATE/EXTRAPOLATE*** probabilities

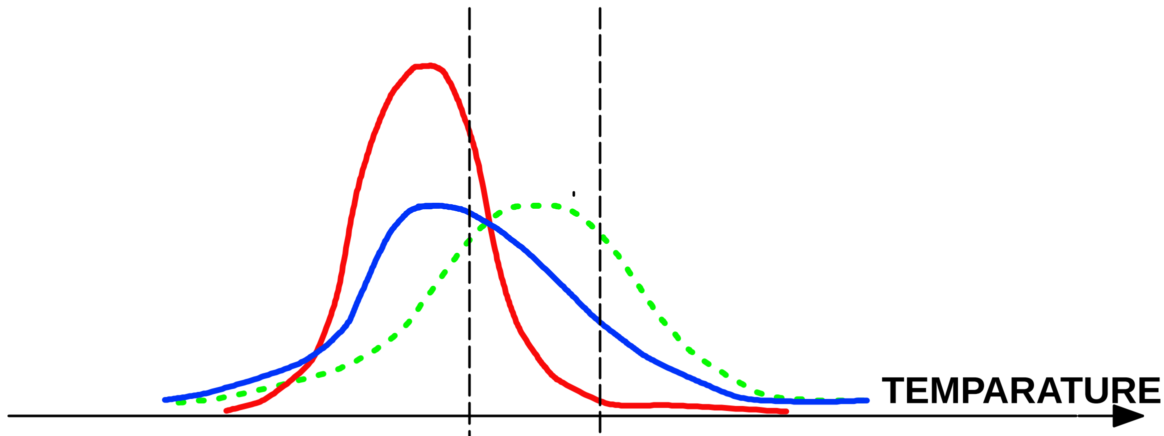


METHODOLOGY:
Fit an analytical probability distribution to finite sample from ensemble

QUESTION: Finite ens sample has sampling noise in it –
How much detail can we trust?

Need to analyze prob fcst performance – ***VERIFICATION***

MAIN CHARACTERISTICS OF PROB FCSTS



CLIMATE PROB	33%	33%	33%
FCST WITH LARGE UNCERTAINTY	48%	29%	23%
FCST WITH MORE INFO	85%	12%	3%

VERIFYING ANALYSIS

1) Must be different from climatology – SHARPNESS

A N D

2) Must be consistent with observations – RELIABILITY

(ie fcst probabilities match observed frequencies)

Sharpness *itself is* ***not a virtue*** *unless fcsts are also* ***reliable***

If both the fcst & obs systems are stationary,

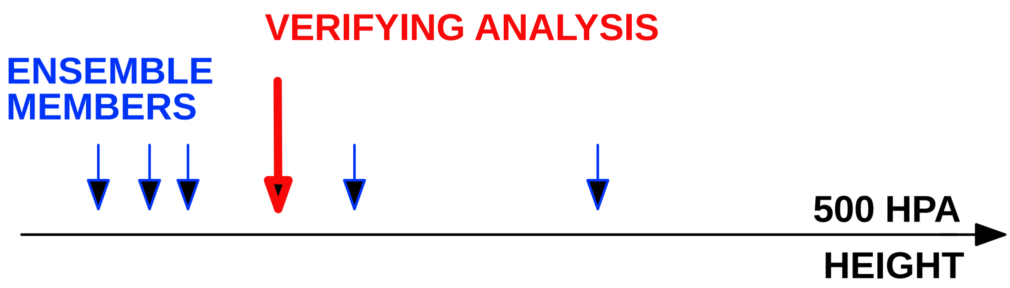
fcsts can be made perfectly reliable through calibration

Remaining sharpness is **RESOLUTION** –

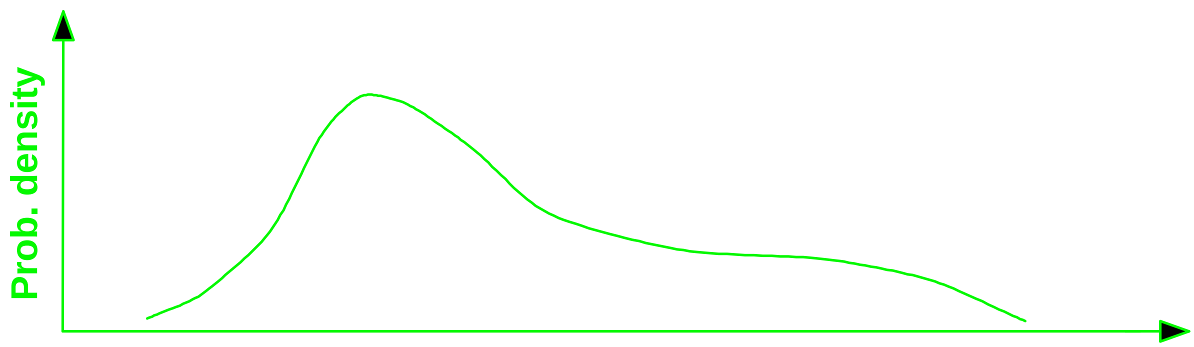
Real virtue of prob fcsts –

Cannot be increased through trivial postprocessing

ENSEMBLE BASED PROBABILISTIC FORECASTS AND THEIR CALIBRATION



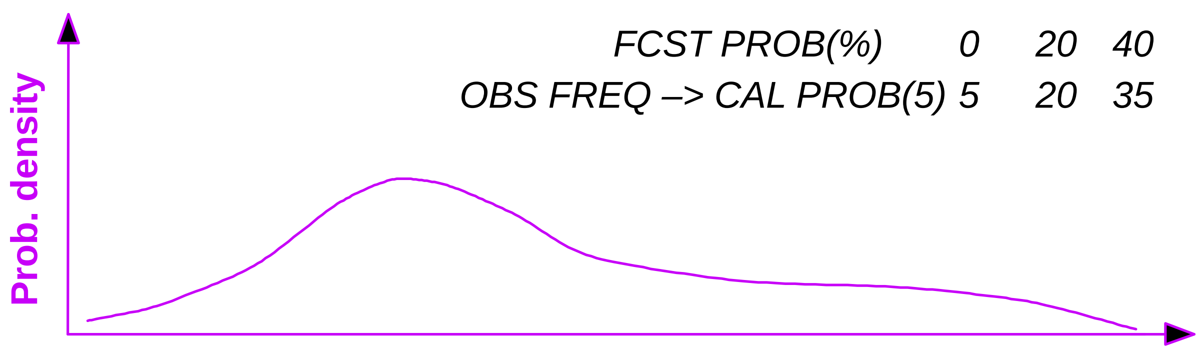
CLIMATE PROB	20%	20%	20%	20%	20%
FCST PROB	20%	40%	20%	20%	0%

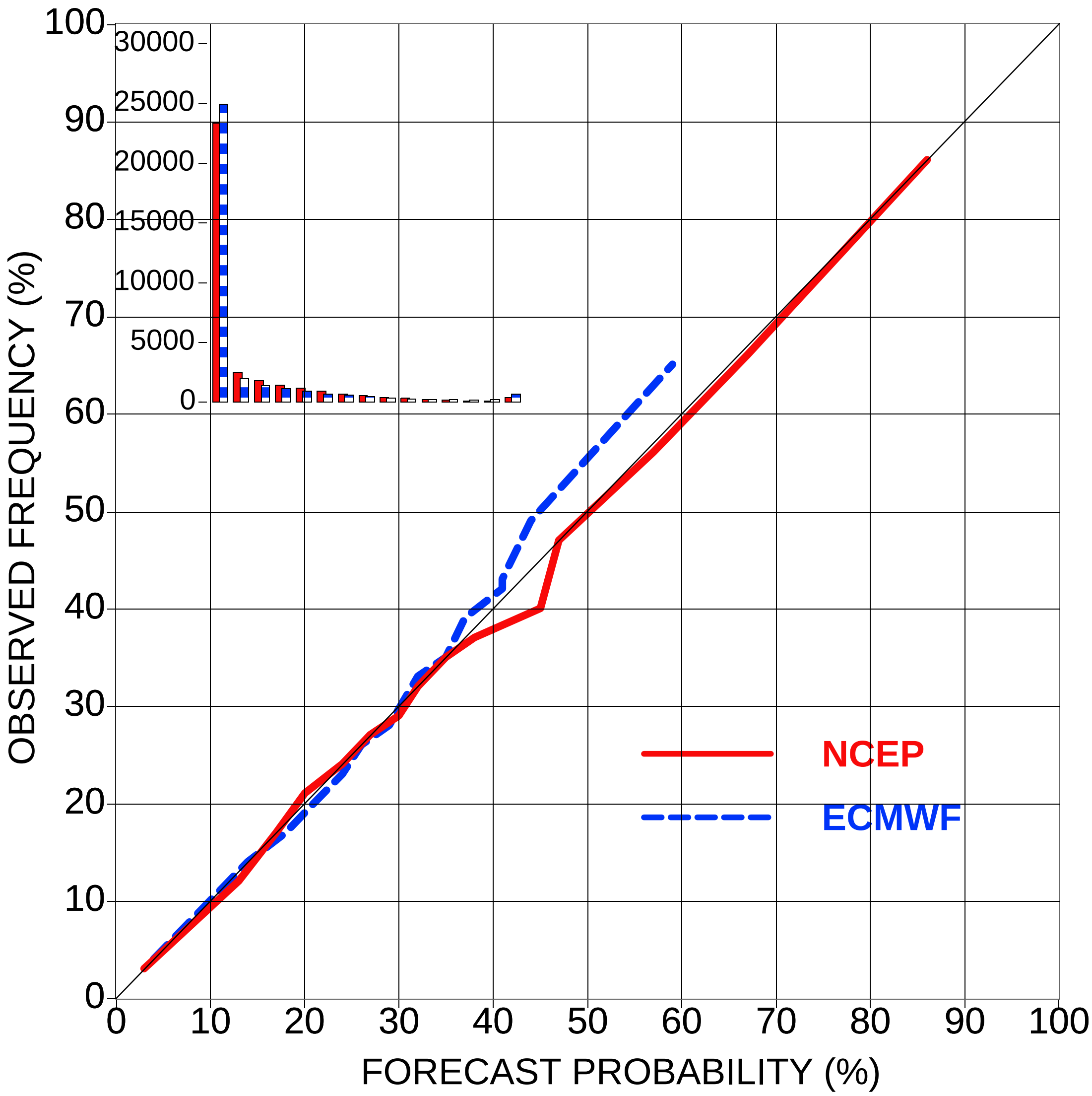


CALIBRATION, based on observed frequency of each fcst prob. value:

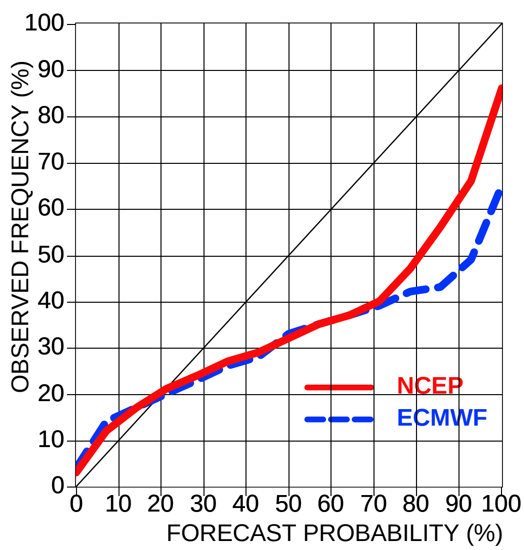
CAL. PROB.	20%	35%	20%	20%	5%

FCST PROB(%)	*0*	*20*	*40*
OBS FREQ –> CAL PROB(5)	*5*	*20*	*35*





Reliability diagram for 3–day lead time ensembles for January 1996. Forecast probabilities are based on observed frequencies associated with the same number of ensemble members falling in a particular bin during December 1–20, 1995. The diagram for uncalibrated forecasts is shown on the right.



VERIFICATION OF PROB FCSTS

1) TRADITIONAL MEASURES:

Categorical fcsts: Average success rate for modus

Point fcsts: Error in expected value, median, etc
(RMS, Pattern Anomaly Correlation)

2) DISTRIBUTIONAL MEASURES:

a) Talagrand (Verification Rank) diagram
– measures reliability/consistency only

b) Reliability diagrams

c) Brier Skill Score

d) Ranked Probability Skill Score

e) Relative Operating Characteristics

f) Information content

g) Economic Value

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*

		FORECAST: *YES*	FORECAST: *NO*
OBSERVATION	YES	H(its) ***Mitigated Loss***	M(isses) ***Loss***
OBSERVATION	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

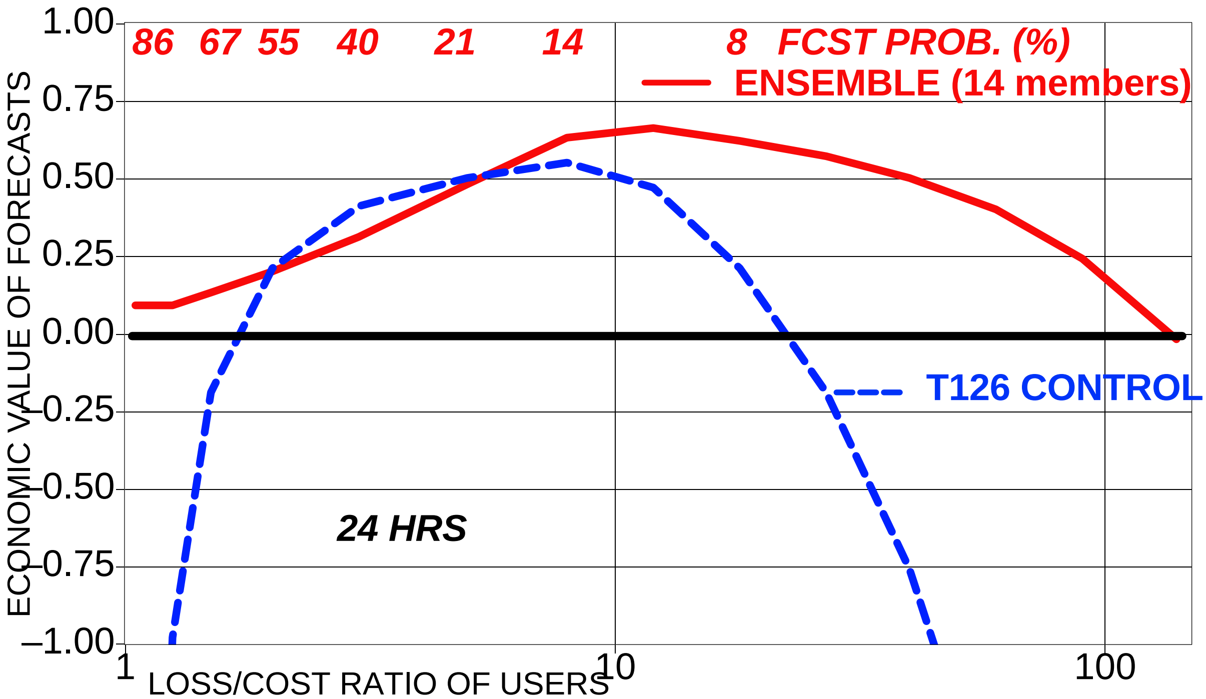
Use 10 climatologically equally likely bins to define events

Hi–res control forecast: If MRF control falls in a given climate bin, forecast is YES and NO otherwise

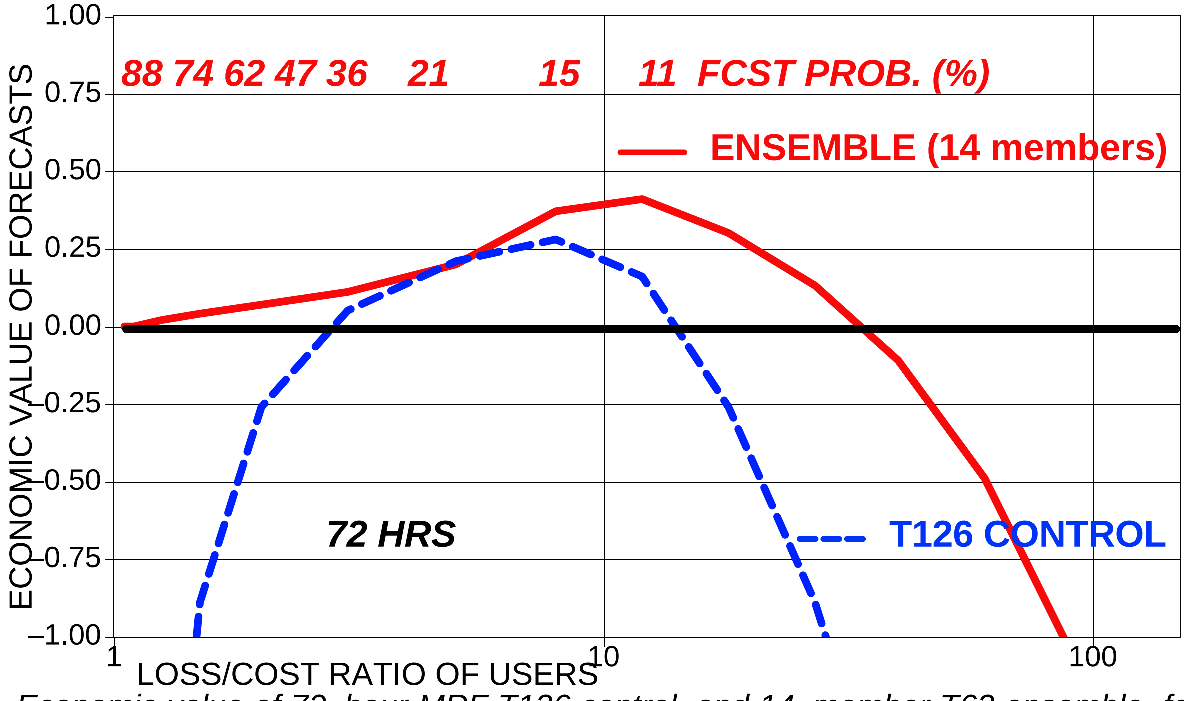
Lo–res ensemble forecast: Probabilities converted to a categorical fcst given the probability exceeds a certain threshold. Eg., all 30% or higher probabilities count as YES. Among different threshold probabilities one can select the one that results in largest economic value.

Results: For majority of users ensemble is more useful

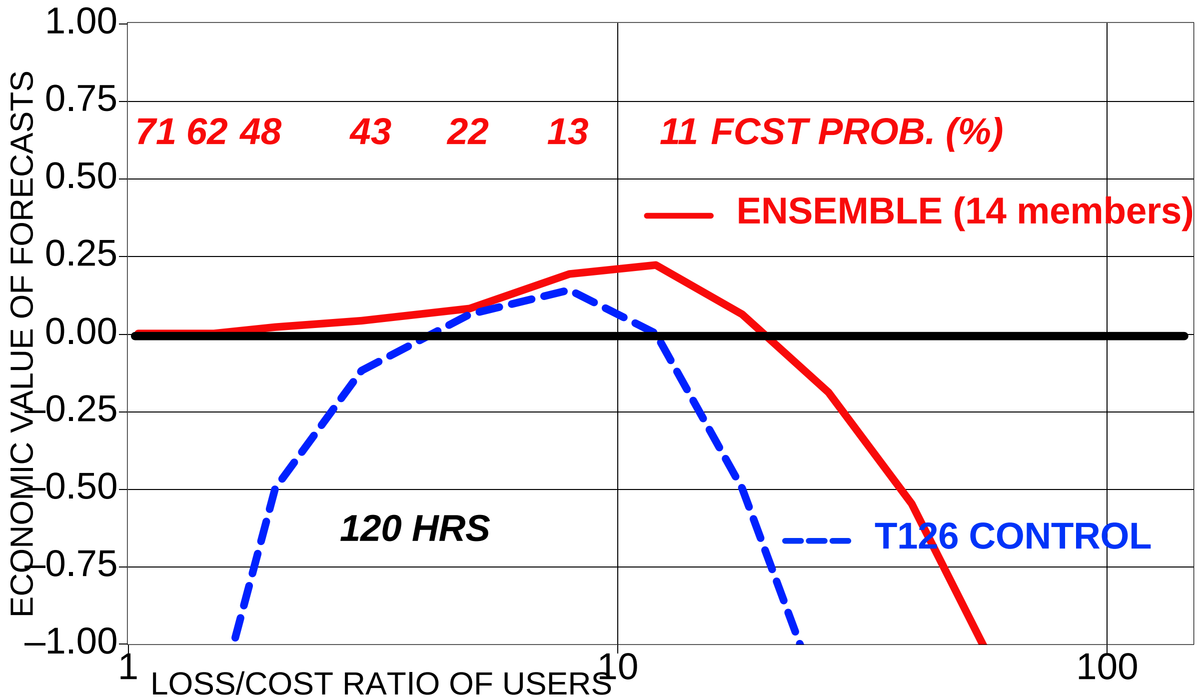
Question: Is it because MRF is dichotomous, while ensemble provides full probability distribution?



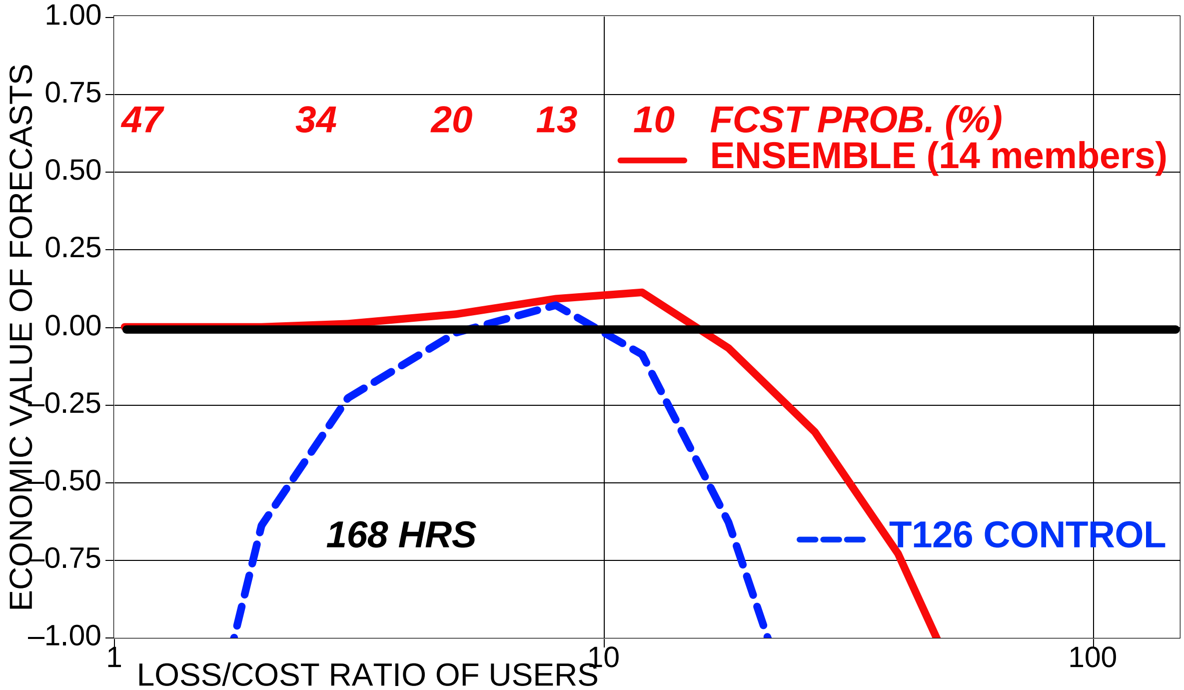
Economic value of 24–hour MRF T126 control, and 14–member T62 ensemble forecasts in predicting events defined in terms of 10 climatologically equally likely bins for 500 hPa height over the NH extratropics, for April–June 1999, for users characterized by different loss/cost ratios (horizontal axis, logarithmic scale). A value of 1.0 stands for using perfect forecasts while values below zero indicate that climatological forecasts are more valuable.



Economic value of 72–hour MRF T126 control, and 14–member T62 ensemble forecasts in predicting events defined in terms of 10 climatologically equally likely bins for 500 hPa height over the NH extratropics, for April–June 1999, for users characterized by different loss/cost ratios (horizontal axis, logarithmic scale). A value of 1.0 stands for using perfect forecasts while values below zero indicate that climatological forecasts are more valuable.



Economic value of 120–hour MRF T126 control, and 14–member T62 ensemble forecasts in predicting events defined in terms of 10 climatologically equally likely bins for 500 hPa height over the NH extratropics, for April–June 1999, for users characterized by different loss/cost ratios (horizontal axis, logarithmic scale). A value of 1.0 stands for using perfect forecasts while values below zero indicate that climatological forecasts are more valuable.



Economic value of 168–hour MRF T126 control, and 14–member T62 ensemble forecasts in predicting events defined in terms of 10 climatologically equally likely bins for 500 hPa height over the NH extratropics, for April–June 1999, for users characterized by different loss/cost ratios (horizontal axis, logarithmic scale). A value of 1.0 stands for using perfect forecasts while values below zero indicate that the use of climatological forecasts are more valuable.

SUMMARY

ENSEMBLE FORECASTING: STRENGTHS AND LIMITS

1) DEFINITION OF CHAOS

   DETERMINISM + INITIAL VALUE SENSTIVITY

2) INITIAL VALUE RELATED PREDICTABILITY/UNCERTAINTY

   ATMOSPH.: DAYS; COUPLED SYSTEM: MONTHS

3) BOUNDARY CONDITION RELATED PREDICT./UNCERTAINTY

   ATMOSPH.: DAYS–MONTHS

4) BRIDGING THE GAP BETWEEN WEATHER AND CLIMATE

   COMBINE INITIAL VALUE AND BOUNDARY FORCING

5) ADVANTAGES OF USING ENSEMBLE FCSTS

   MORE ACCURATE FCST

   CASE DEPENDENT UNCERTAINTY ESTIMATE

   DAILY PROB FCSTS AT ALL TIME RANGES

6) PROBLEMS WITH ENSEMBLES

   MODEL RELATED ERRORS NOT ACCOUNTED FOR:

   BIAS IN FIRST & SECOND MOMENTS

   NEED TO BUILD VARIABILITY INTO MODELS –

   DIFFICULT TASK – IF SUCCESSFUL,

   CAN IMPROVE RESOLUTION

7) VERIFICATION

   RELIABILITY & RESOLUTION

8) POSTPROCESSING

   IMPROVES RELIABILITY, NOT RESOLUTION

   MODEL BIAS PROBLEM GREATLY REDUCED