

# VALIDATION OF PROBABILISTIC FORECASTS

**Zoltan Toth<sup>1</sup>**

Environmental Modeling Center, NCEP

ASSESSING THE VALUE OF PROBABILISTIC FORECASTS FROM  
A SCIENTIFIC PERSPECTIVE

*Acknowledgements:*

<b>Yuejian Zhu<sup>1</sup></b>	EMC
<b>Olivier Talagrand</b>	LDM
<b>Mozheng Wei<sup>1</sup></b>	EMC
<b>Gopal Iyengar</b>	MCMRWF

<http://sgi62.wwb.noaa.gov:8080/ens/enhome.html>

1. SAIC (Beltsville, MD) at NCEP, Zoltan.Toth@noaa.gov; (301) 763–8000/ext. 7268

# OUTLINE

- 1) TYPES OF WEATHER FORECASTS WRT UNCERTAINTY
- 2) TYPES OF WEATHER FORECASTS WRT GENERATION
- 3) ATTRIBUTES OF (PROBABILISTIC) FORECASTS
- 4) PROBABILISTIC VERIFICATION MEASURES
- 5) VALUE OF ENSEMBLE VS. CONTROL FCST
- 6) HOW MUCH DETAIL CAN ENSEMBLES FITHFULLY DEFINE IN  
PDF?

## **CONCLUSIONS**

### **1) Types of weather forecasts wrt uncertainty**

***DICHOTOMOUS (CATEGORICAL) VS. PROBABILISTIC***

### **2) Types of weather forecasts wrt generation**

***STATSTICAL; DYNAMICAL – SINGLE VS. ENSMEBLE  
IN ANY CASE, PDF IS DESIRED, GENERAL FORMAT  
ONLY LIOUVILLE EQS PROVIDE THAT – NEED FOR POSTPROC.***

### **3) Attributes of (probabilistic) forecasts**

***RELIABILITY (NO BIAS) &  
RESOLUTION (SMALL RANDOM ERROR)***

### **4) Probabilistic verification measures**

***BRIER, RANKED PROB., ROC, INFO CONTENT, ETC.***

### **5) Value of ensemble vs. control fcst**

***1ST MOMENT BETTER;  
TEMOPRAL VARIATIONS IN 2ND MOMENT  
BETTER DEFINED PDF***

### **6) How much detail can ensembles faithfully define in pdf?**

***BIMODALITY CAPTURED –  
JUMPS IN CONSECUTIVE CONTROL FCSTS ASSOCIATED WITH IT***

# FORMAT OF FORECASTS RELATED TO UNCERTAINTY

## FORECAST FORMAT

	<b>PROBABILISTIC</b>	<b>DICHOTOMOUS (CATEGORICAL)</b>
<i>UNCERTAINTY</i>	Substantial	Little or no
<i>FORMAT</i>	0–100%	Yes or No
<i>EXAMPLE</i>	<i>Precip above 5 mm</i>	
	80%	Yes
<i>TYPE</i>	General	Special

Resolution in probability space can be set at different levels:

**Very high:** Continuous values

Intermediate: Every 10% (0, 10, 20, etc)

**Very low:** 0% (No) and 100% (Yes)

Dichotomous and probabilistic forecasts are fundamentally not different

=> Quasi-continuous transition

**EXAMPLE:** User wants to know if min temp will be below 5 C

*IF* expected value is below -5 C *OR* above +15C *AND*

expected error less than +/-5 C *THEN*

use of dichotomous format justified (ie, no fcst uncertainty)

*ELSE* use of dichotomous format TRUNCATES fcst info

**PROBABILITY DENSITY FUNCTION** (PDF) is the complete format,  
allowing all queries to be answered

We must **CONDENSE ALL KNOWLEDGE** on future weather into **PDFs**

# TYPES OF WEATHER FORECASTS WRT GENERATION

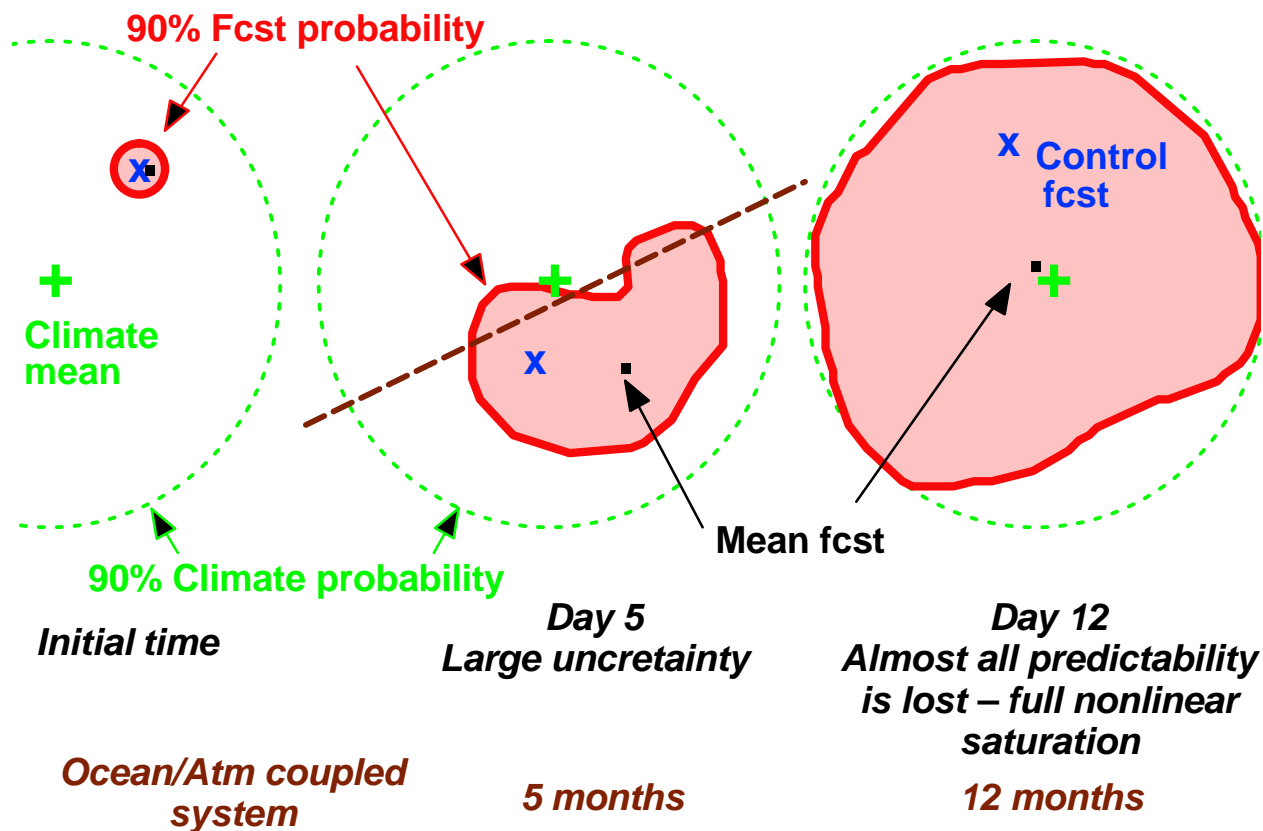
Fcsts in probabilistic & dichotomous format

can be generated by same methods:

- 1) **STATISTICAL** (based on observations)
- 2) **DYNAMICAL**, based on NWP model integration:
  - Single* (combined with past verification statistics)
  - Ensemble* (sample of multiple realizations)
- 3) **LIOUVILLE EQS** (in prob space) – not practical

## FCST 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**



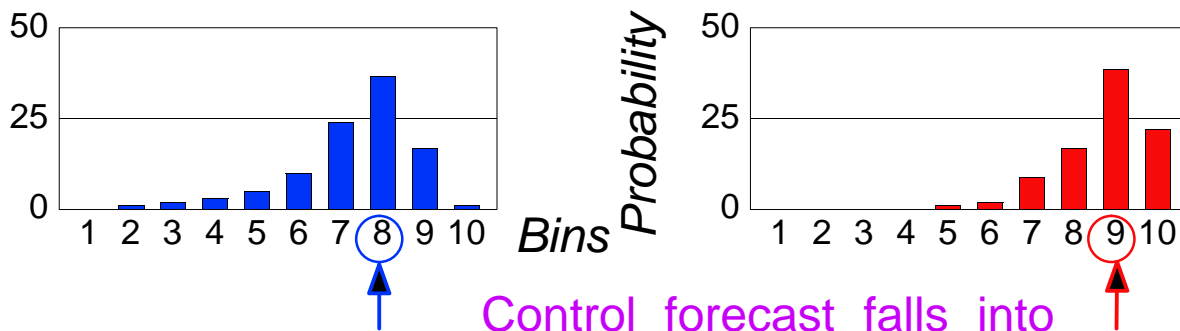
*Ensemble approach potentially offers more fcst info – how do we tell?*

# PDF ESTIMATION BASED ON FINITE SAMPLE

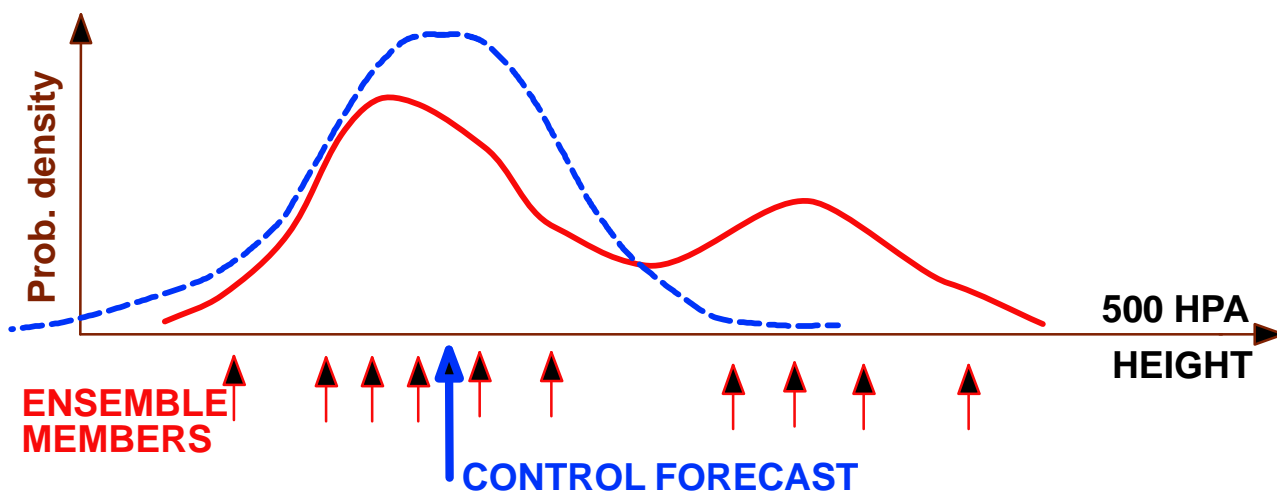
Dynamical method (single or multiple realizations) – not proper format

Must **INTER-/EXTRAPOLATE** probabilities to generate PDF format

**Single integration** (combined with past verification statistics)



**Ensemble integration** (sample of multiple realizations)



	ENSEMBLE (size $m$ ) →	SINGLE (size 1)
EXPECTED VALUE	Mean of sample	Random realization
SHAPE OF PDF	Flow dependent	Statistical average
DETAIL IN PDF	Yes, depends on $m$	No

*Ensemble approach potentially offers more fcst info – how do we tell?*

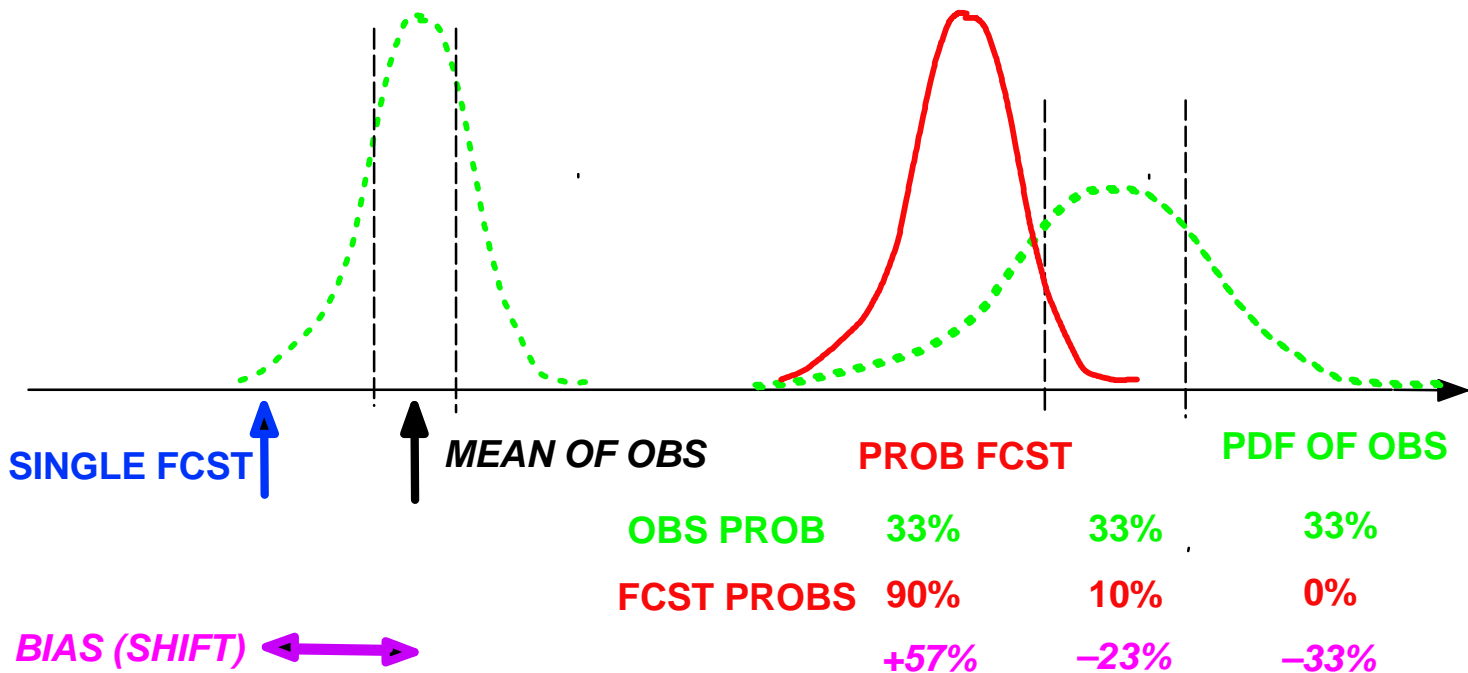
HOW DO WE DECIDE WHICH METHOD IS BEST?

NEED VERIFICATION MEASURES

WHAT TO MEASURE?

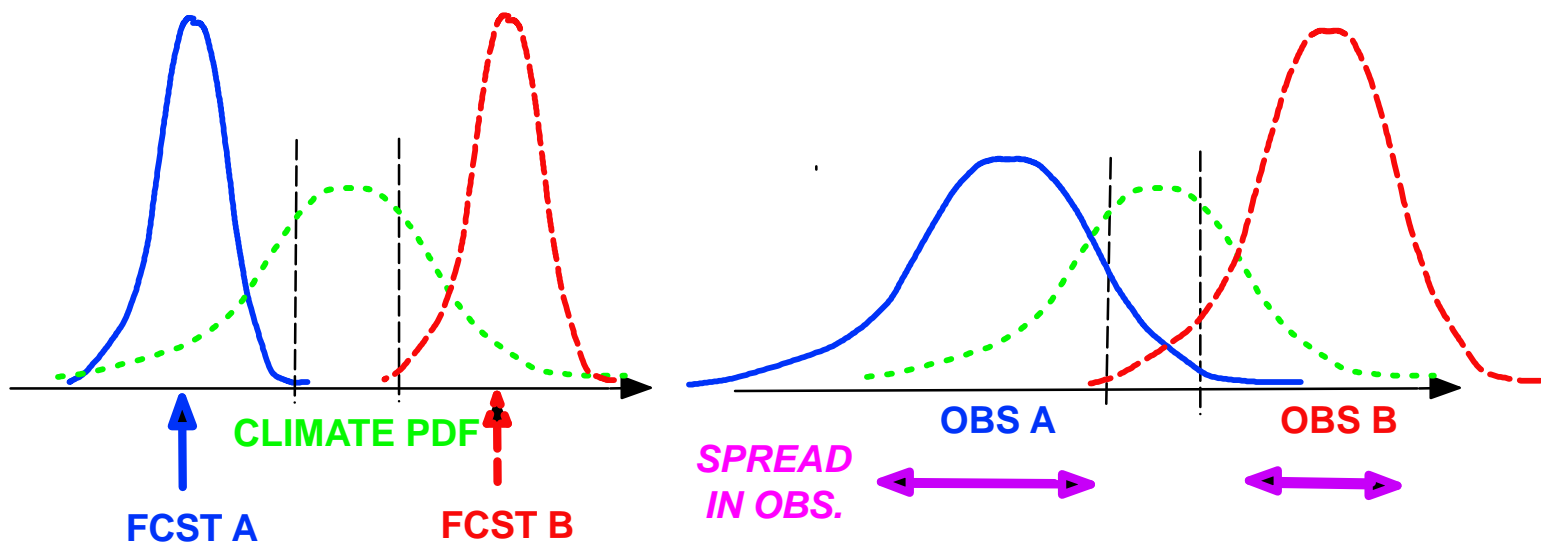
# MAIN STATISTICAL ATTRIBUTES OF FORECASTS

1) **RELIABILITY** – Lack of systematic error (no conditional bias)



Reliability can be statistically corrected (assuming stationary processes)

2) **RESOLUTION** Different fcsts precede different observations



Resolution **CANNOT** be statistically corrected –

**INTRINSIC VALUE** of fcst systems

If fcsts perfectly reliable, resolution = spread in ens. = spread in obs. =>

Perfect forecast system uses 0 & 100% probs & always correct

# VERIFICATION MEASURES FOR PROBABILISTIC FCSTS

## 1) RELIABILITY

Reliability diagram (graphical)

Reliability component of Brier Score

*RELATED ENSEMBLE FCST MEASURE:*

Analysis Rank Histogram (Talagrand diagram)

## 2) RESOLUTION

Reliability/Attributes diagram (graphical)

Resolution component of the Brier Score

Relative Operating Characteristics

Economic Value (D. Richardson's presentation)

*FOR PERFECTLY RELIABLE FCSTS:*

Brier Skill Score

Ranked Probability Skill Score

Information content

*RELATED ENSEMBLE FCST MEASURE:*

RMS error of ensemble mean ( = ensemble spread)

Verifying analysis indistinguishable from ens members

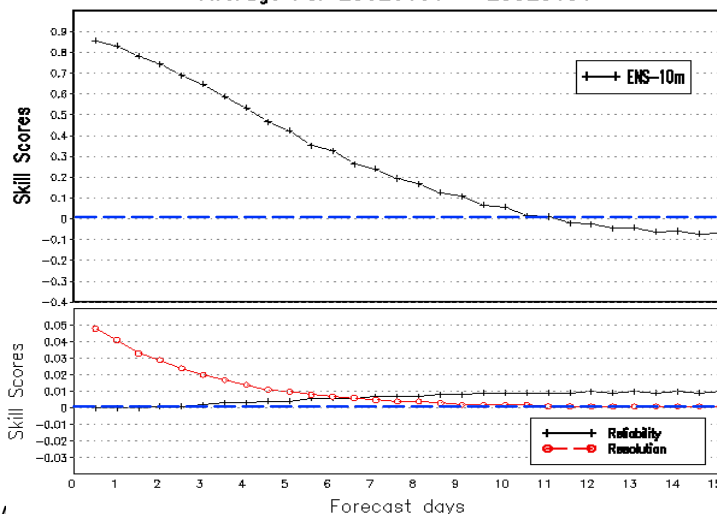
Smaller spread = fcst problem better resolved

## 3) RELIABILITY + RESOLUTION

Brier Skill Score

Ranked Probability Skill Score

Northern Hemisphere 500 mb Height Brier Skill Scores (BSS)  
Average For 20020101 - 20020131



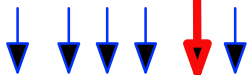


# BRIER SCORE (BS) and BRIER SKILL SCORE (BSS)

For verifying categorical probability forecasts (event occurs or not)

## VERIFYING ANALYSIS

ENSEMBLE MEMBERS



500 HPA  
HEIGHT →

OBSERVATION

$d_i$                       0      1

FCST PROB  
 $p_i$

20%      80%

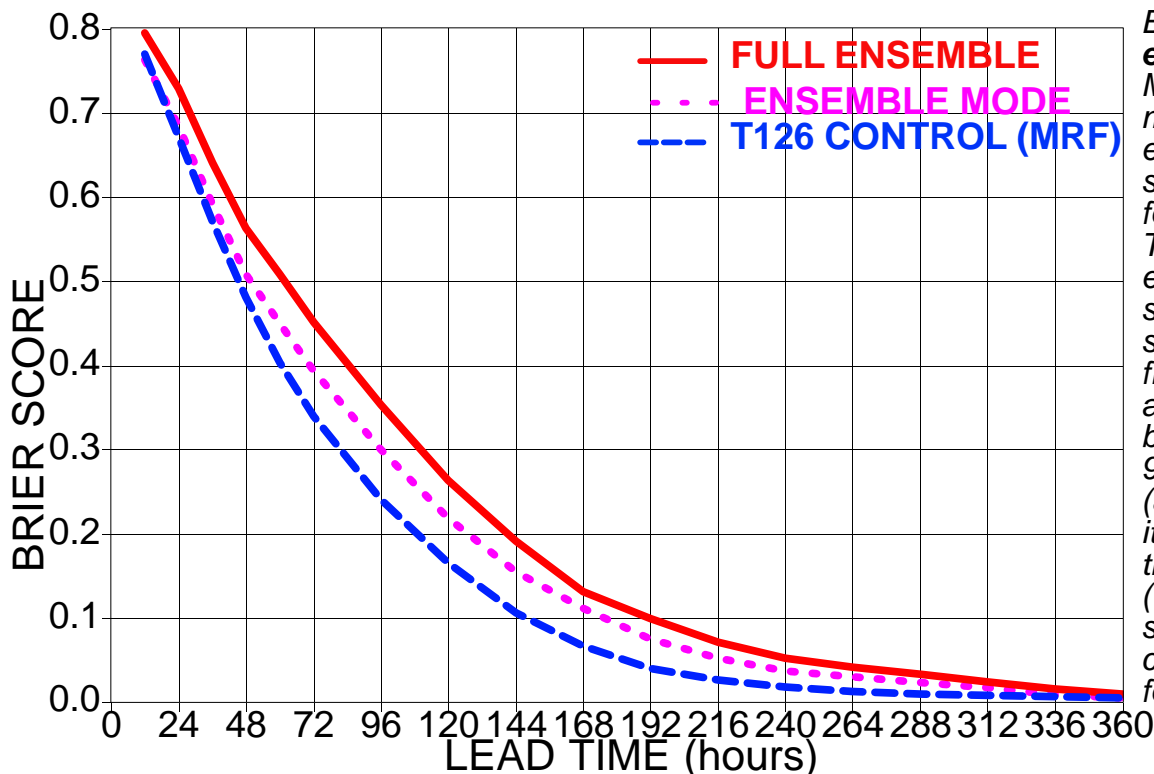
$$BS(p, d) = \frac{1}{n} \left[ \sum_{i=1}^n (p_i - d_i)^2 \right]$$

Total of  $n$  pairs of cases  
 $N_k$  cases with  $p_k$  probability

$$\bar{d}_k = \frac{1}{N_k} \sum_{i \in N_k} d_i$$

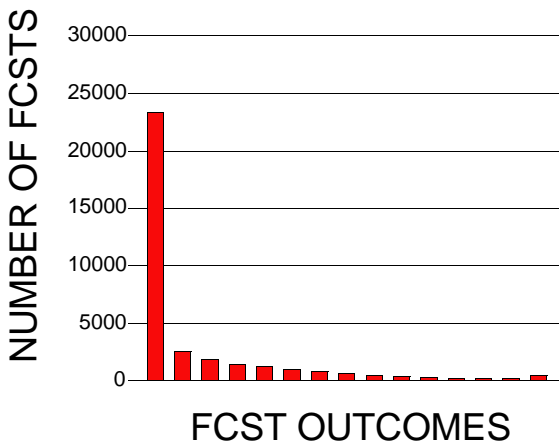
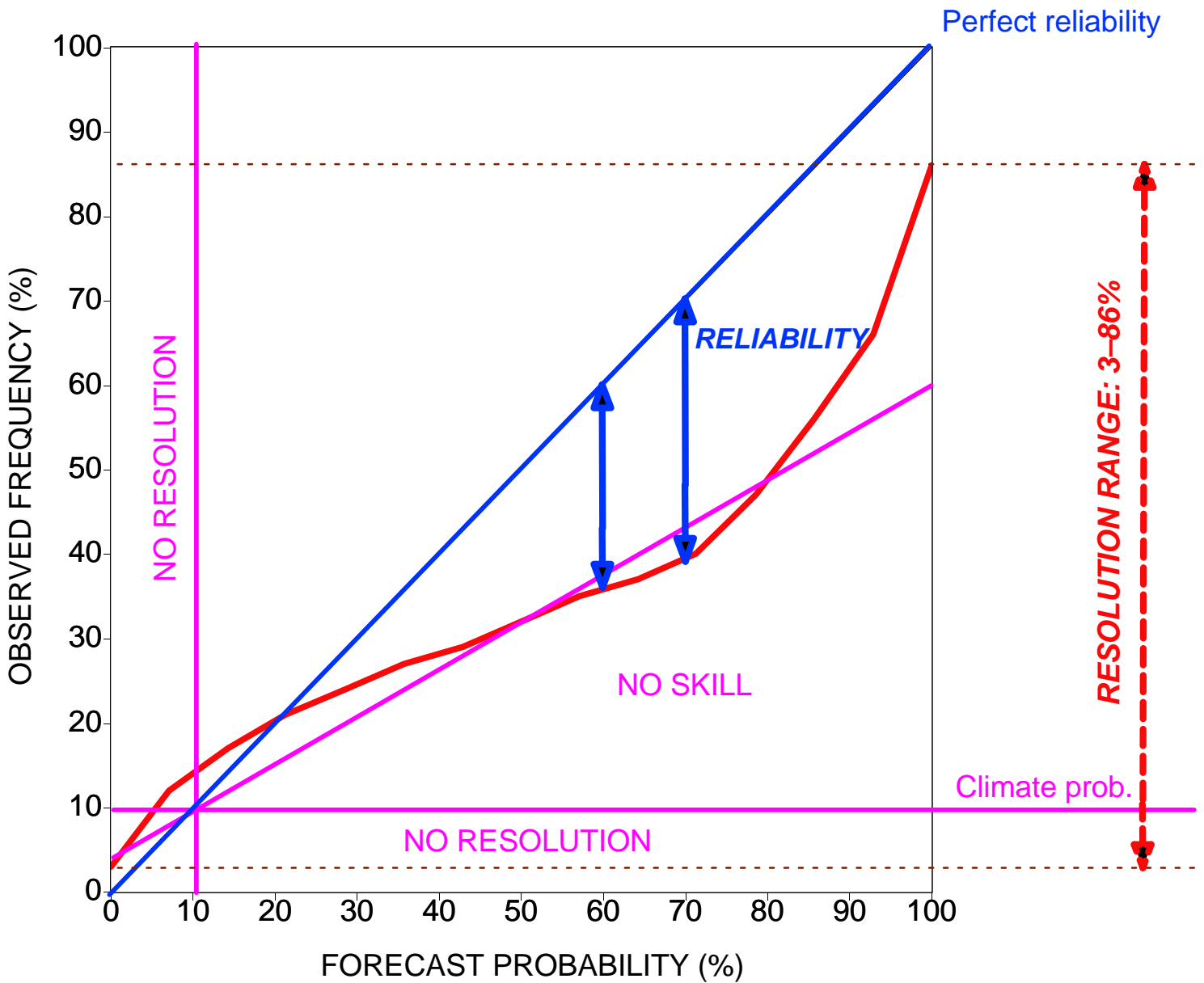
$$BS = \underbrace{\frac{1}{n} \left[ \sum_{k=1}^K N_k (p_k - \bar{d}_k)^2 \right]}_{\text{Reliability}} - \underbrace{\frac{1}{n} \left[ \sum_{k=1}^K N_k (\bar{d}_k - \bar{d})^2 \right]}_{\text{Resolution}} + \underbrace{\bar{d}(1 - \bar{d})}_{\text{Uncertainty}}$$

$$BSS = 1 - \frac{BS(\text{forecast})}{BS(\text{climatology})}$$



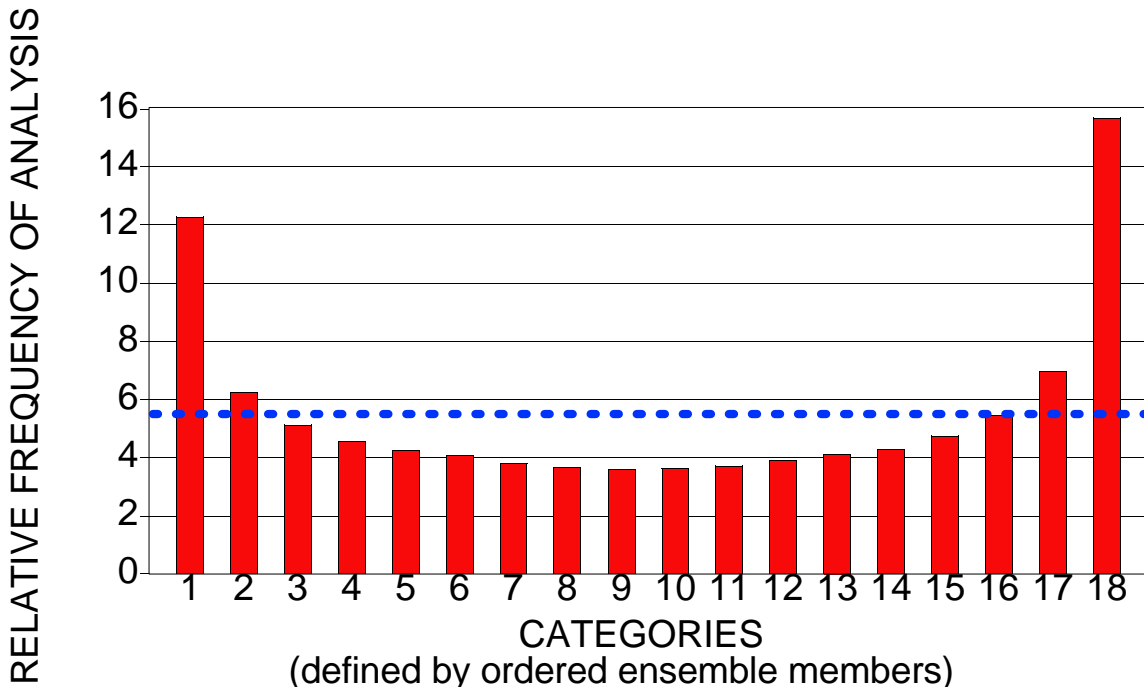
Brier Skill Score for the **NH extratropics**, for March–May 1997. Forecasts are made for 10 climatologically equally likely bins; results shown here are the average for the two extreme bins. The bin where the control or ensemble mode falls is assigned a probability corresponding to the observed frequency of the verifying analysis falling into the same bin ( $P$ ), while the remaining 9 bins are assigned  $(1-P)/9$  (assuming perfect reliability). Note that depending on the value of the mode ( $1 \leq M \leq 10$ ), the corresponding observed frequency for the ensemble (but not for the control) varies widely.

TOTH, Z.: VALIDATION OF PROBABILISTIC FORECASTS  
**RELIABILITY / ATTRIBUTES DIAGRAM**

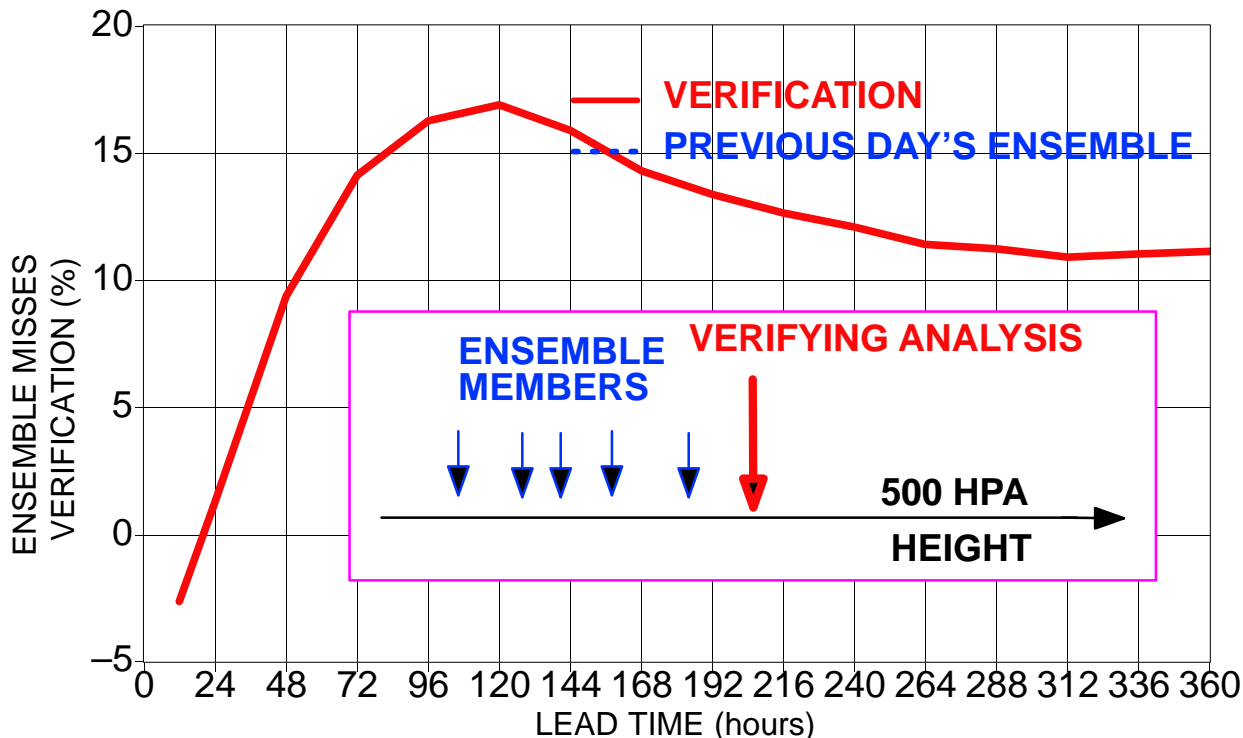


*Reliability diagram for 3-day lead time ensembles for January 1996. Forecast probabilities are based on observed frequencies associated with the same number of ensemble members falling in a particular bin during December 1-20, 1995.*

# ANALYSIS RANK HISTOGRAM (TALAGRAND DIAGRAM)



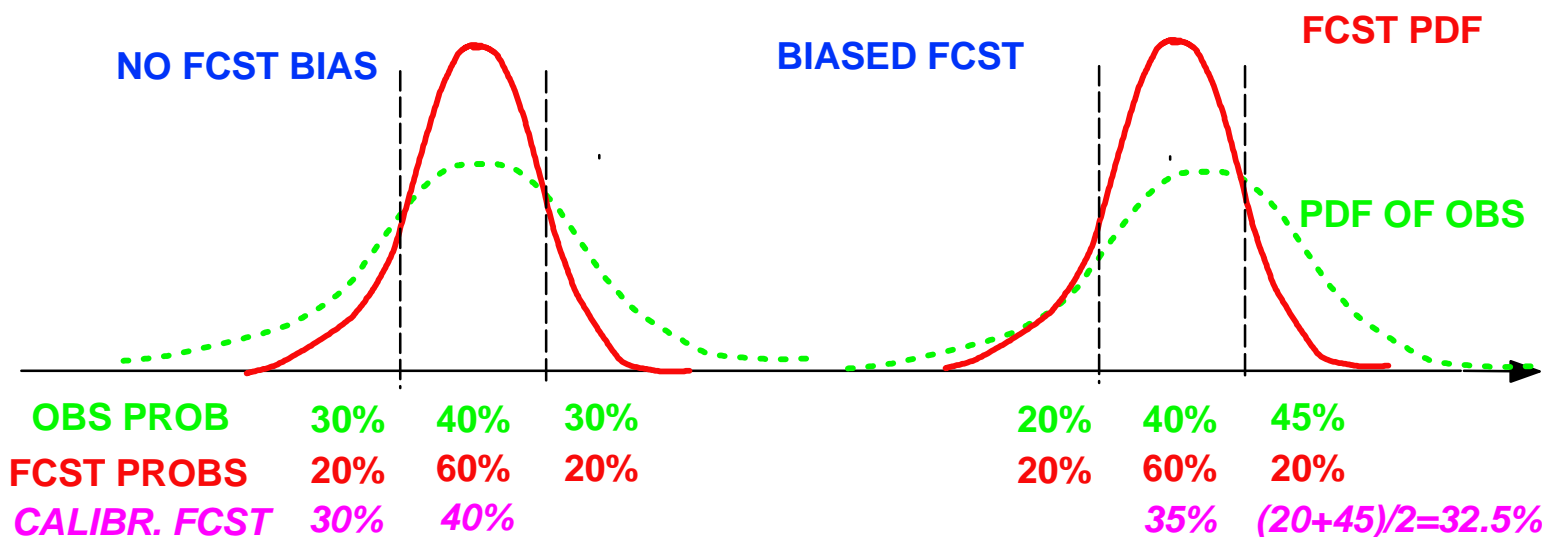
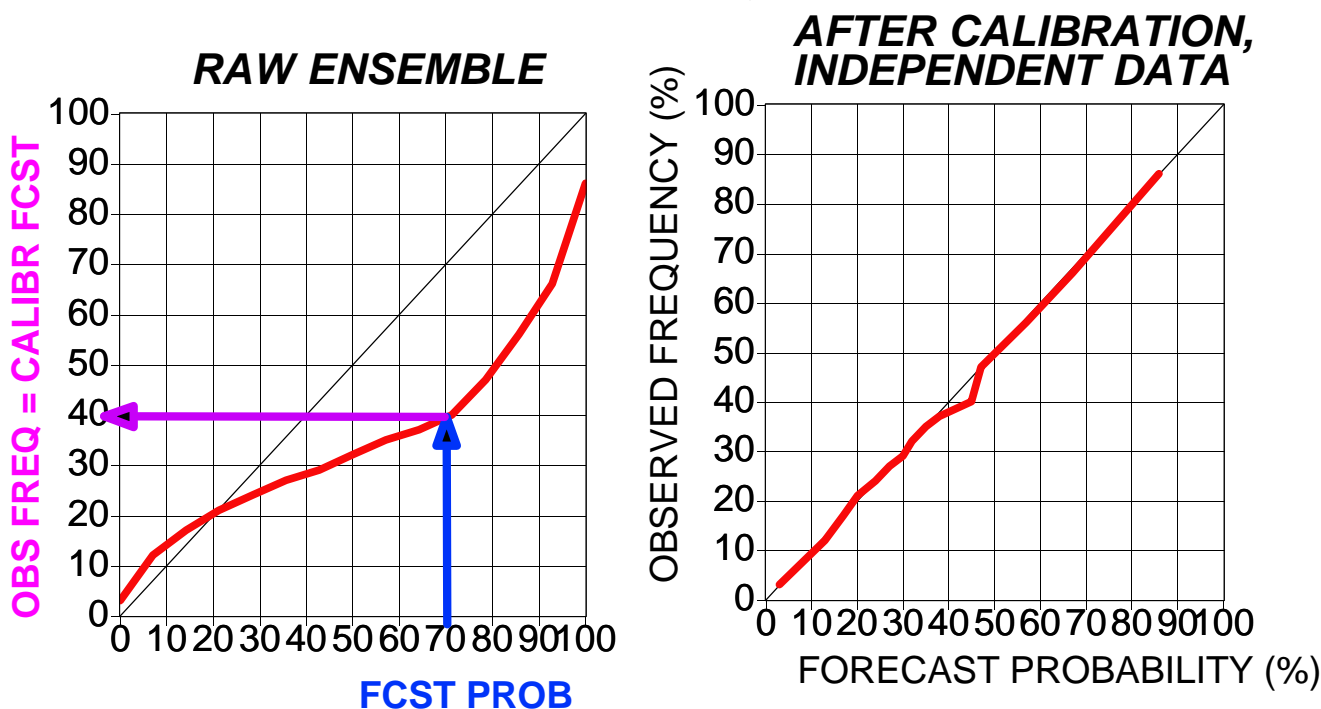
Percentage of cases when the verifying analysis falls in any of the bins defined by the ordered series of the 17 ensemble members, at 120-hour lead time for April–June 1999. The expected value next to one particular forecast in the 17-member ensemble at each grid point at 120 hours lead time for March–May 1997, 500 hPa height over the NH extratropics. The expected value (5.55 %) is marked as a dotted blue line.



Percentage of cases when the 17-member ensemble does not encompass the 500 hPa height verifying analysis over the NH extratropics (in excess of the 11.1% that is expected due to the limited size of the ensemble.) April–June 1999.

# CALIBRATION OF PROBABILISTIC FORECASTS

- 1) **“Relabel” fcst prob by observed frequency** associated with fcst  
 Corrects for bias in spread (if ensemble spread too low)  
 Bias in 1st moment not corrected, just accounted for



- 2) a) **Correct for conditional bias** (shift) in distribution  
 b) **Correct for error in spread**

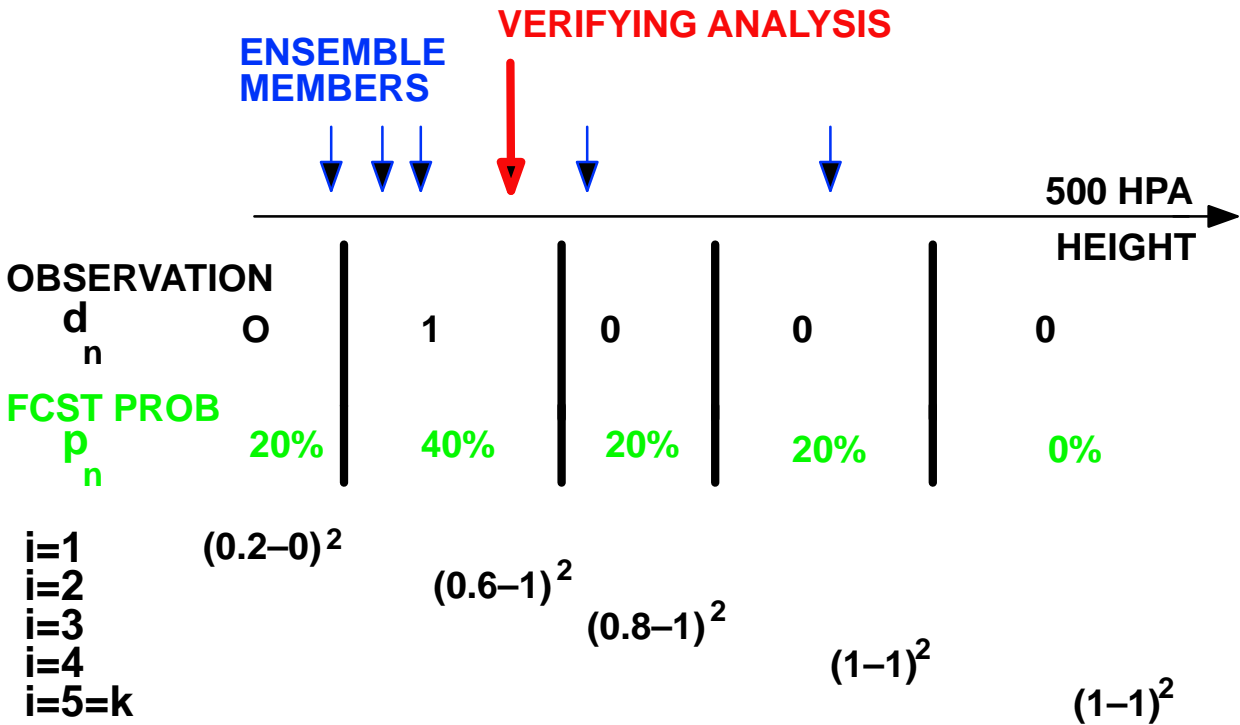
*Balance must be found between*

- a) Details sought in calibration (time, space, meteor. & prob values)
- b) Available fcst–obs archive

## RANKED PROBABILITY SCORE (RPS)

For verifying multicategory probability forecasts in case categories can be ranked or ordered (like temperature)

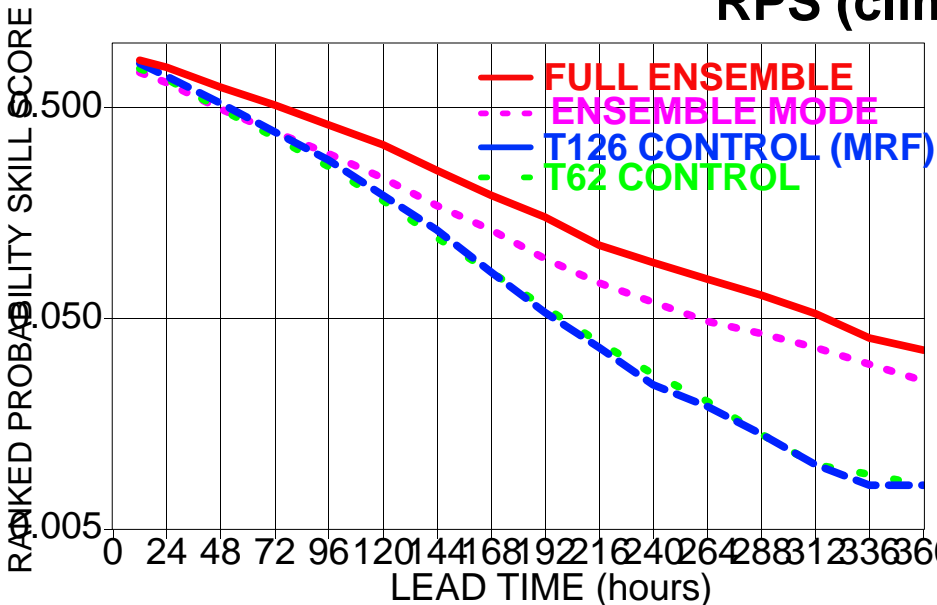
Generalization of Brier score (used for non-ranked classes)



**k = number of categories**

$$RPS(p, d) = \frac{1}{k-1} \left[ \sum_{i=1}^k \left( \sum_{n=1}^i P_n - \sum_{n=1}^i d_n \right)^2 \right]$$

$$RPS \text{ Skill Score (RPSS)} = 1 - \frac{RPS(\text{forecast})}{RPS(\text{climatology})}$$



Ranked probability skill score for a T62 and T126 control and a 10-member ensemble forecast for the 500 hPa height, **NH extratropics**, March–May 1997. Forecast probabilities are made for 10 climatologically equally likely bins and are based on verification statistics from previous month (calibrated forecasts). Control forecasts have two probabilities depending on whether the forecast is in or not in a bin whereas the ensemble probabilities vary depending on how many ensemble members fall in a bin.

## RELATIVE OPERATING CHARACTERISTICS (ROC)

Application of signal detection theory for measuring discrimination between two alternative outcomes

Worded, categorical and probab. forecasts can be compared  
Missed events not considered directly

Stratification according to observations – reliability **NOT** measured

OBSERVATION  
YES  
NO

		FORECAST	
		YES	NO
OBSERVATION	YES	H(its)	M(isses)
	NO	F(false alarms)	C(orrect rejections)

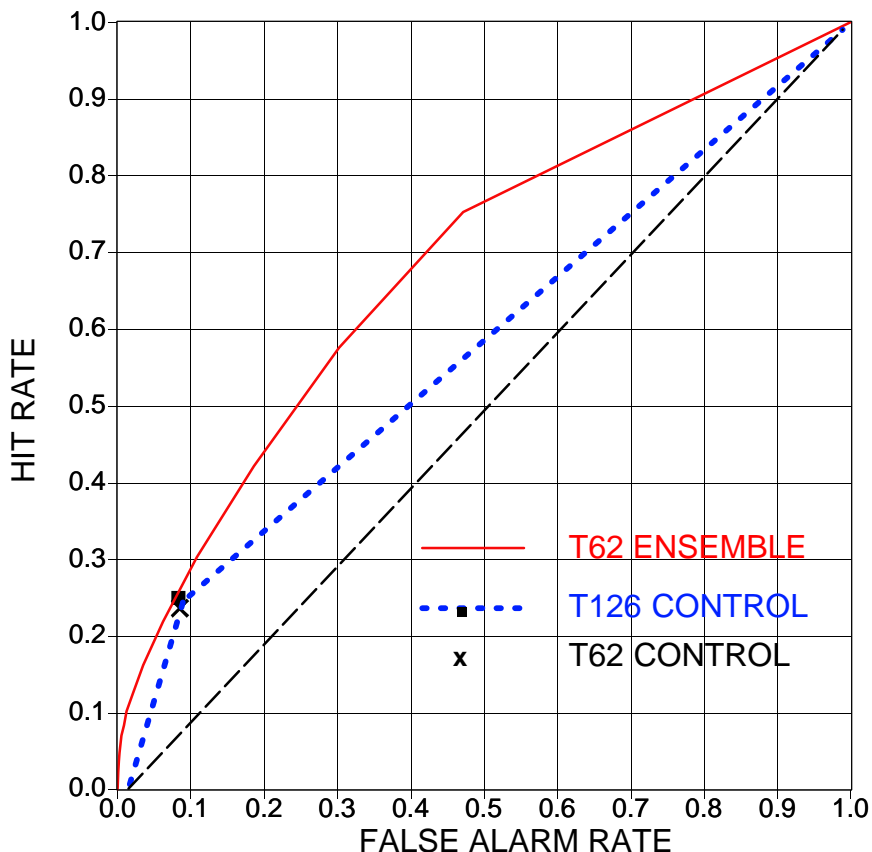
$$\text{Hit Rate (HR)} = \frac{H}{H + M}$$

$$\text{False Alarm Rate (FAR)} = \frac{F}{F + C}$$

Use 10 climatologically equally likely bins to define events

**Categorical forecast:** If control falls in a given climate bin, forecast is YES and NO otherwise

**Ensemble forecast:** Probabilities converted to a categorical fcst given the probability exceeds a certain threshold. Eg., all 30% or higher probabilities count as YES. Using different threshold probabilities yield an HR/FA diagram.



**Measures:** 1) Area between HR–FAR curve and diagonal

2) How different forecast probabilities are given different observations

ROC (Relative Operating Characteristics) curve for a 5-day lead time 14-member T62 ensemble of forecasts and for the T126 and T62 control forecasts predicting events defined in terms of 10 climatologically equally likely bins for the 500 hPa height, NH extratropics, April–June 1999.

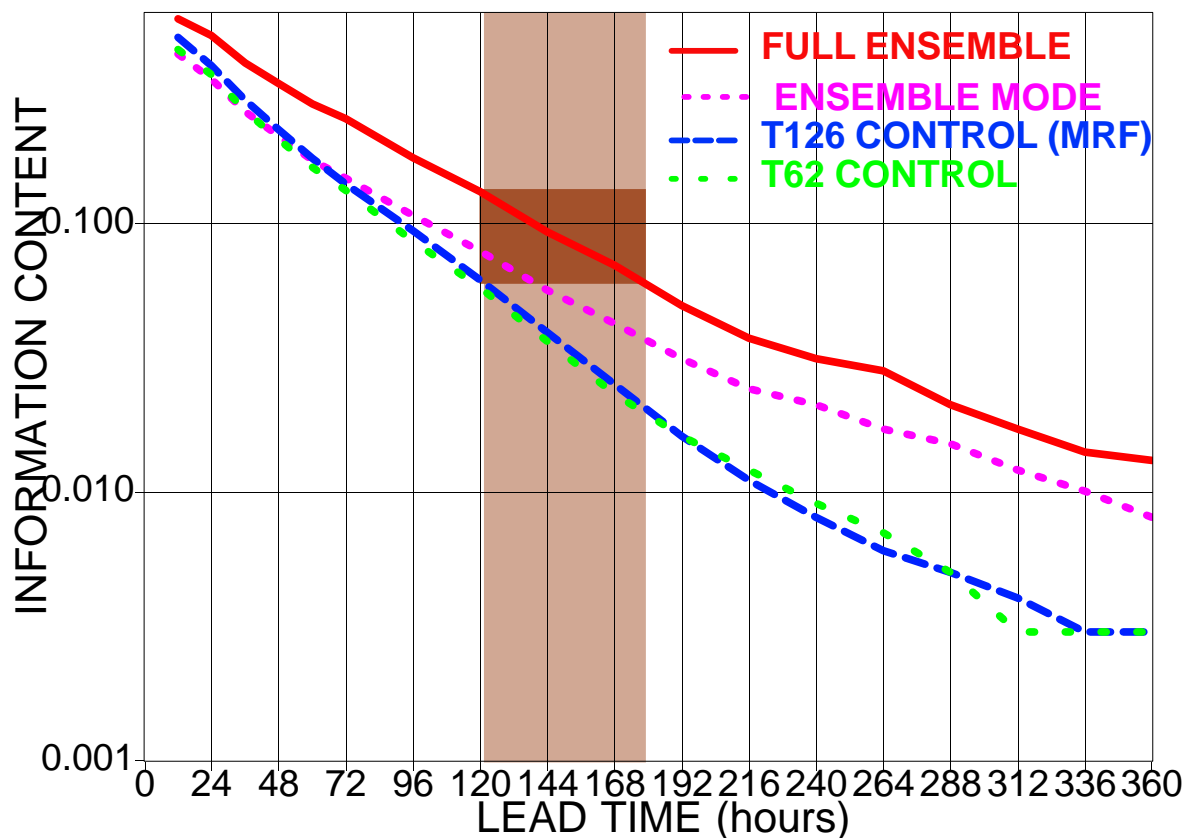
## INFORMATION CONTENT

Use 10 climatologically equally likely bins to define events

$$\text{Entropy} = P \log_2 P:$$

$$\text{Information in one forecast} = I = 1 - \sum_{i=1}^{10} P_i \log_{10} P_i$$

$$\text{Average info in } n \text{ independent fcsts} = I_{\text{ave}} = \frac{1}{n} \sum_{i=1}^n I_i$$



Information content of probabilistic forecasts based on the full ensemble distribution (red continuous line), the mode (most frequent value) of a 10-member ensemble (purple dotted), and the T62 (green short dash) and T126 (blue long dash) control forecasts for the NH extratropics, for March–May 1997. Forecasts are made for 10 climatologically equally likely bins. The bin where the control or ensemble mode falls is assigned a probability corresponding to the observed frequency of the verifying analysis falling into the same bin ( $P$ ), while the remaining 9 bins are assigned  $(1-P)/9$  (assuming perfect reliability that is close to be satisfied when using calibrated forecasts). Probabilities for the full ensemble are based on the number of ensemble members falling into the various bins. Note that the ensemble-based forecast probabilities can vary widely from case to case, depending on how the ensemble members spread while they are fixed for the control forecasts. The advance knowledge of the case dependent reliability of the forecasts translates into substantial gains in terms of the information content the forecasts carry.

ON AVERAGE A 7.5-DAY FULLY PROBABILISTIC FORECAST OR A 6-DAY CATEGORICAL FORECAST ASSOCIATED WITH CASE DEPENDENT RELIABILITY ESTIMATES HAS AS MUCH INFORMATION CONTENT AS A 5-DAY CATEGORICAL FORECAST

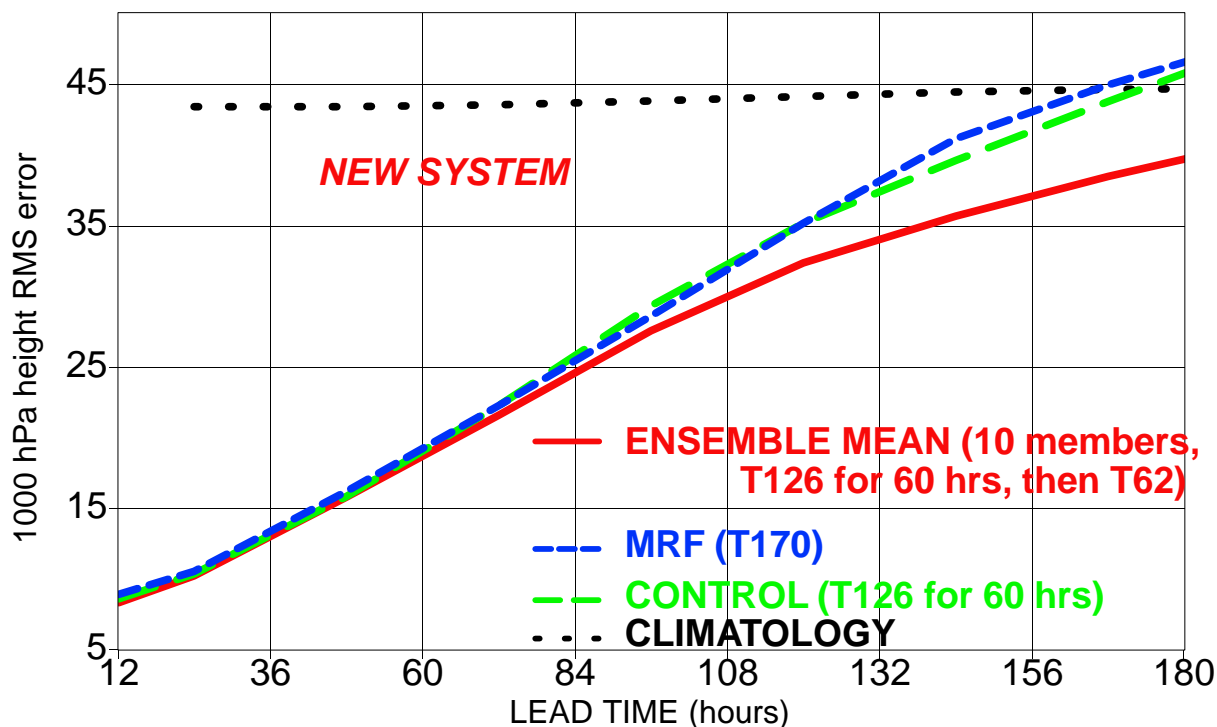
A 7.5-DAY FULLY PROBABILISTIC FORECAST HAS MORE THAN TWICE AS MUCH INFORMATION CONTENT THAN A 5-DAY CATEGORICAL FORECAST

# COMPARING SINGLE CONTROL & ENSEMBLE FCSTS

IF ENSEMBLE IS MORE USEFUL THAN CONTROL FCST,  
WHAT EXPLAINS THE DIFFERENCE?

- 1) **Expected value: Ensemble mean better than control? YES**
- 2) Case dependent variations in spread: Ensemble has skill?
- 3) More detailed pdf from ensemble (m vs. 1 members)?
- 4) Is it only 2nd moment (spread), or further details in ensemble?

1) **Ensemble mean** has lower RMS error than control fcsts =>  
Smaller random error/ Better estimate of 1st moment / Higher resolution



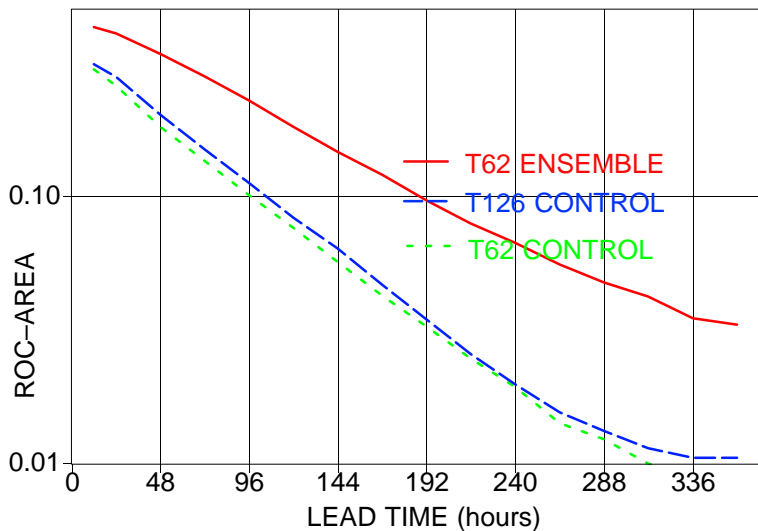
1000 hPa height RMS error for the T170/T126 (for the first 60 hrs, then T62) control forecast (short dash blue/long dash green) and the T126 (for first 60 hrs, T62 afterwards) ensemble mean (solid red), for June 28 – July 22 2000, over the NH extratropics.



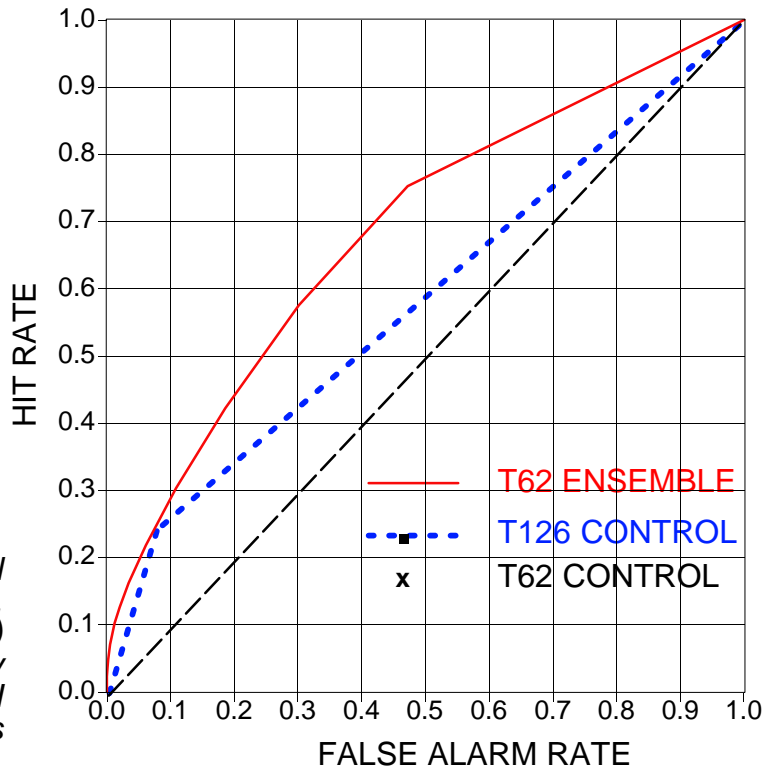
# COMPARING SINGLE CONTROL & ENSEMBLE FCSTS

- 1) Expected value: Ensemble mean better than control?
- 2) *More detailed pdf from ensemble (m vs. 1 members)? YES*
- 3) *Case dependent variations in spread: Ensemble has skill?*
- 4) Is it only 2nd moment (spread), or further details in ensemble?

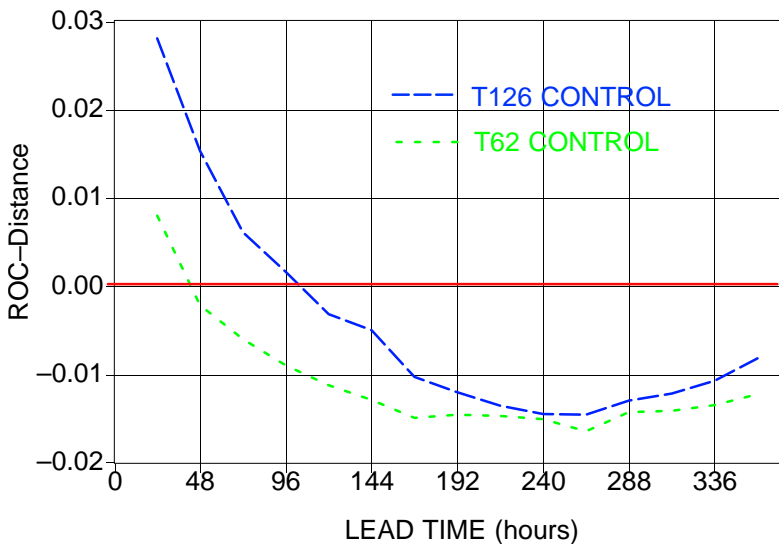
## RELATIVE OPERATING CHARACTERISTICS



ROC (Relative Operating Characteristics) area skill score for the T126 (dashed) and T62 (dotted) control, and the 14-member T62 ensemble forecasts (solid) predicting events defined in terms of 10 climatologically equally likely bins for the 500 hPa height, NH extratropics, for April–June 1999. Scale on vertical axis is logarithmic.



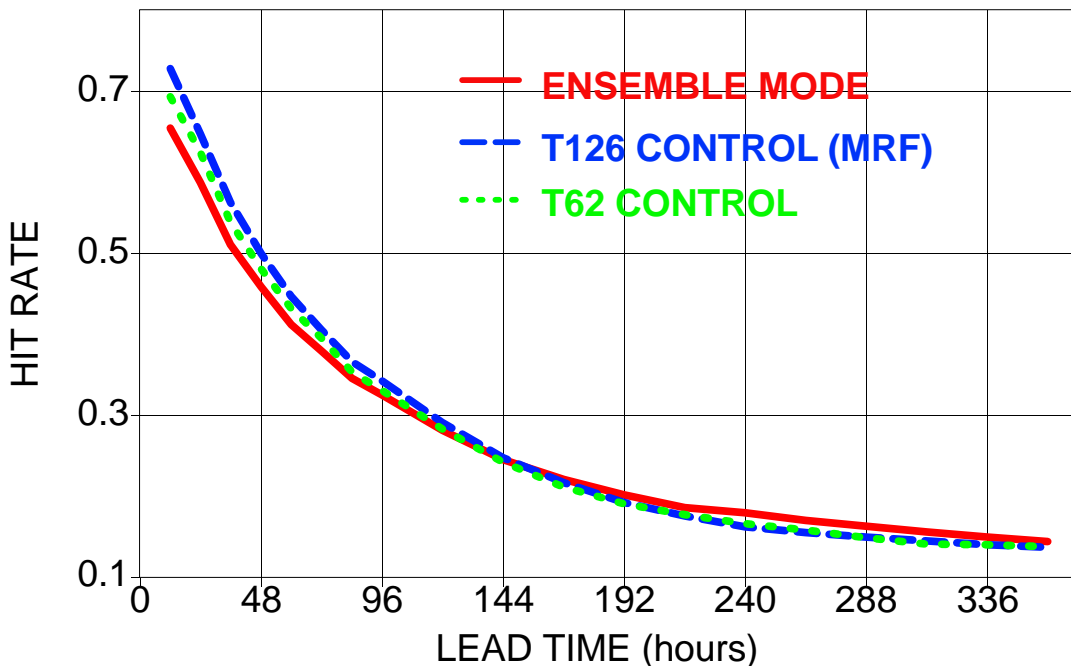
ROC (Relative Operating Characteristics) curve for a 5-day lead time 14-member T62 ensemble of forecasts and for the T126 and T62 control forecasts predicting events defined in terms of 10 climatologically equally likely bins for the 500 hPa height, NH extratropics, April–June 1999.



Same as figure above except for ROC-distance, defined (on linear vertical axis) as the distance between a control point and the closest point on the ensemble polygon. Positive (negative) values indicate the control point is above (below) the ensemble curve.

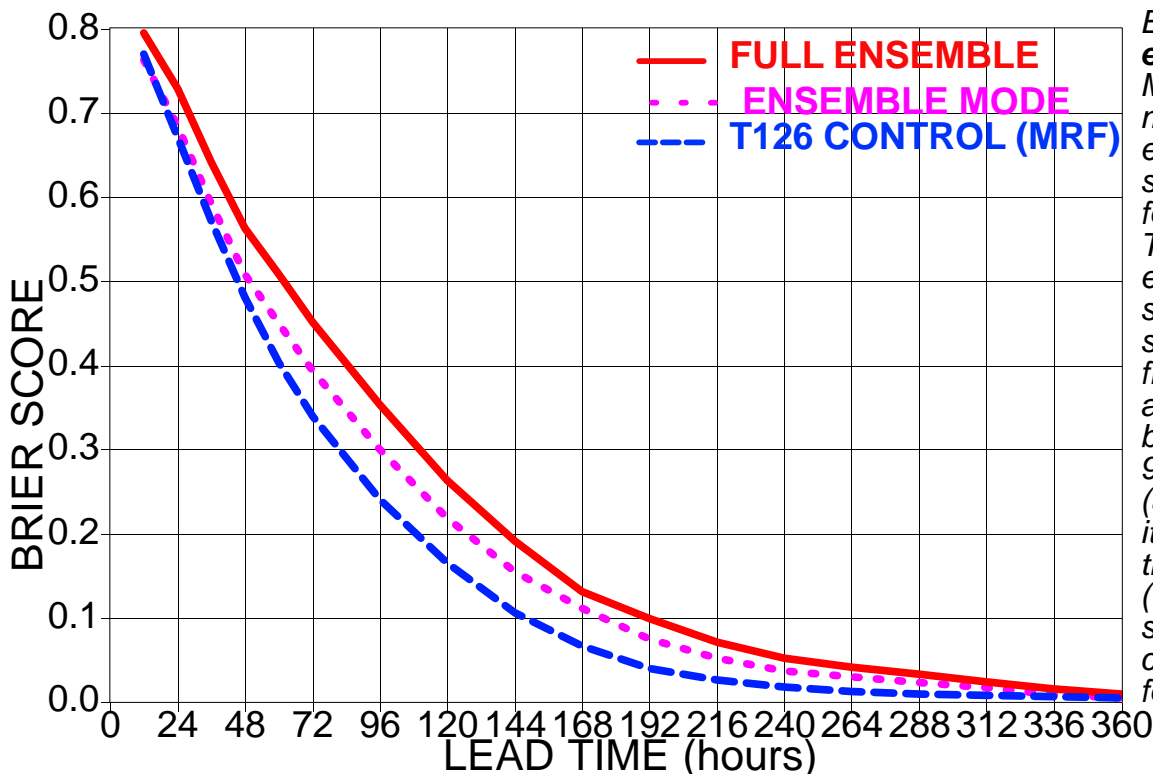
# COMPARING SINGLE CONTROL & ENSEMBLE FCSTS

- 1) Expected value: Ensemble mean better than control?
- 2) *More detailed pdf from ensemble (m vs. 1 members)?*
- 3) **Case dependent variations in spread: Ensemble has skill? YES**
- 4) Is it only 2nd moment (spread), or further details in ensemble?



**Ensemble mode not much better than control**

Overall reliability of T62 and T126 controls and 10-member ensemble mode (most frequent value) forecasts for the NH extratropics, for March–May 1997.



**Brier Skill Score for the NH extratropics, for March–May 1997.** Forecasts are made for 10 climatologically equally likely bins; results shown here are the average for the two extreme bins. The bin where the control or ensemble mode falls is assigned a probability corresponding to the observed frequency of the verifying analysis falling into the same bin ( $P$ ), while the remaining 9 bins are assigned  $(1-P)/9$  (assuming perfect reliability). Note that depending on the value of the mode ( $1 \leq M \leq 10$ ), the corresponding observed frequency for the ensemble (but not for the control) varies widely.

# 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

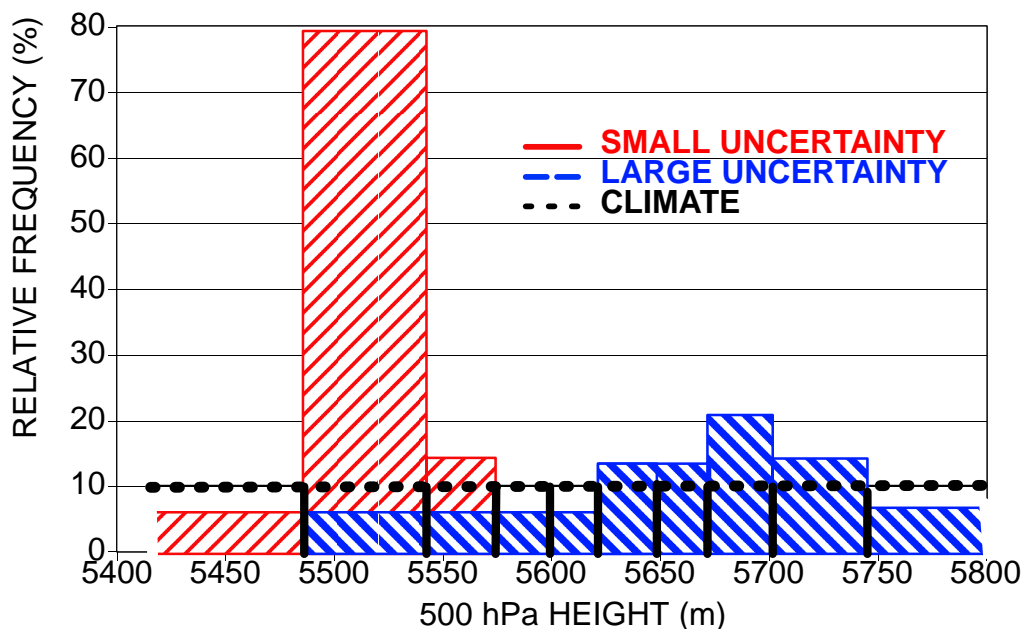
## DATA:

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

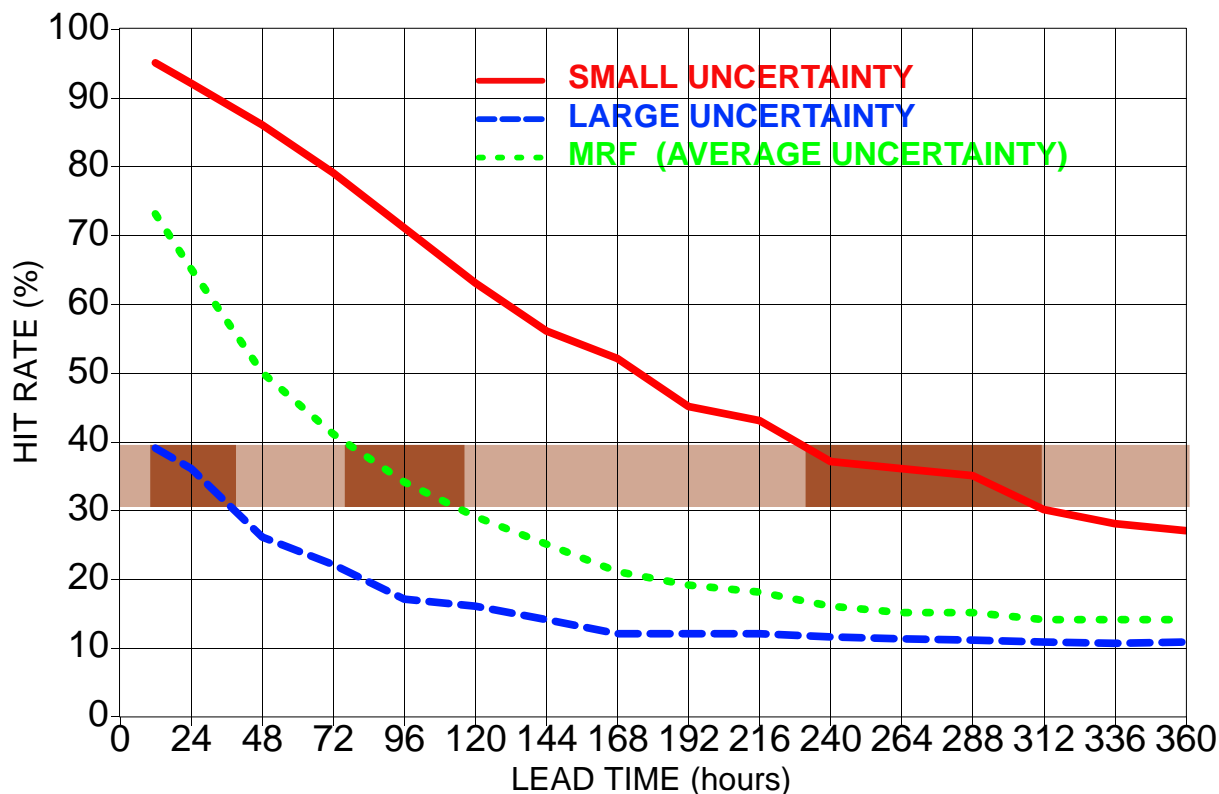
14 perturbed fcsts and high resolution control

## VERIFICATION:

**Hit rate** for ensemble mode and hires control fcst



# SEPARATING HIGH VS. LOW UNCERTAINTY FCSTS



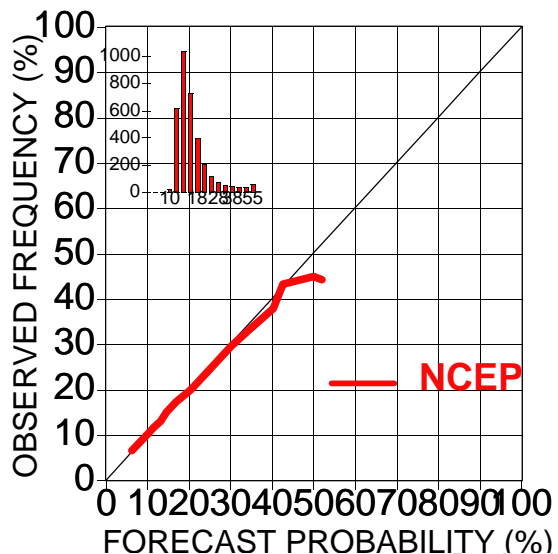
**THE UNCERTAINTY OF FCSTS CAN BE QUANTIFIED IN ADVANCE**  
**HIT RATES FOR 1-DAY FCSTS**

**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**

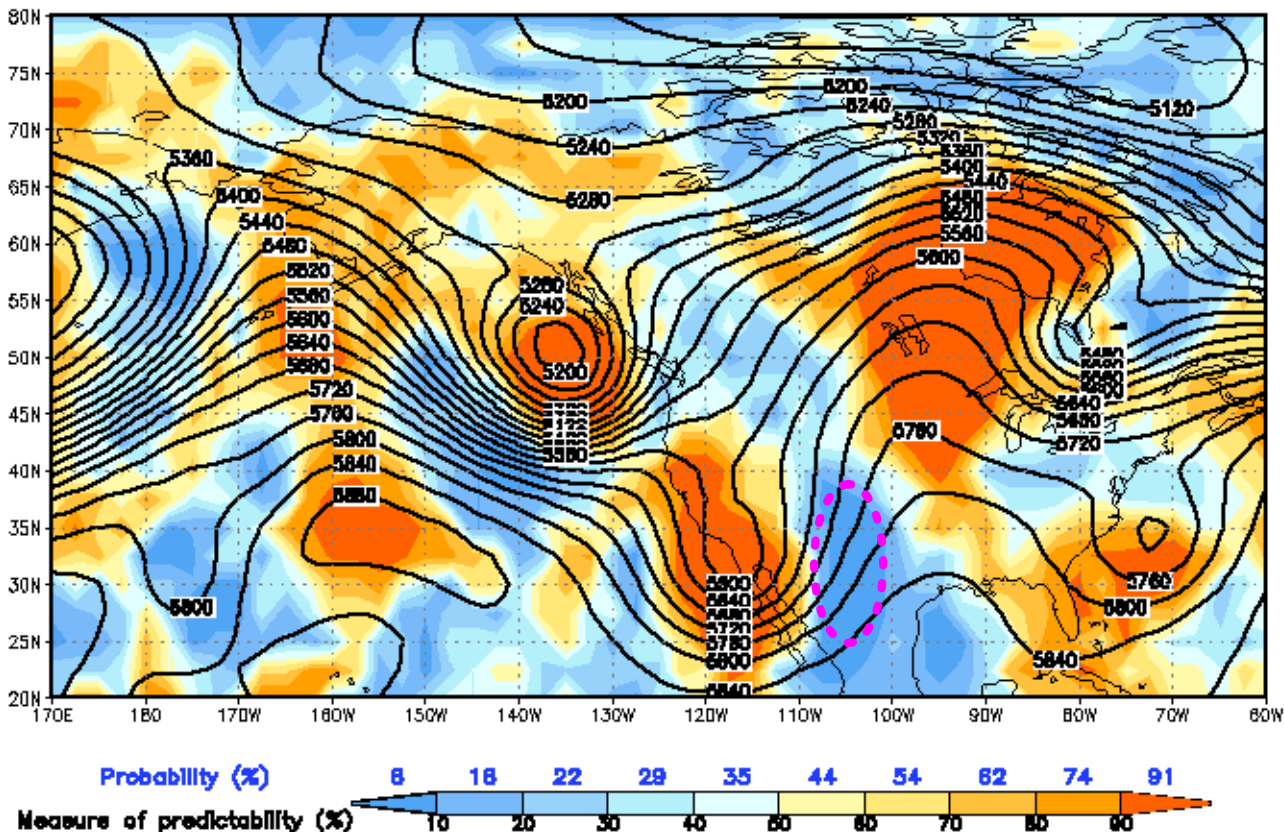
**1-2% OF ALL DAYS THE 12-DAY FCST CAN BE MADE WITH MORE CONFIDENCE THAN THE 1-DAY FCST**

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

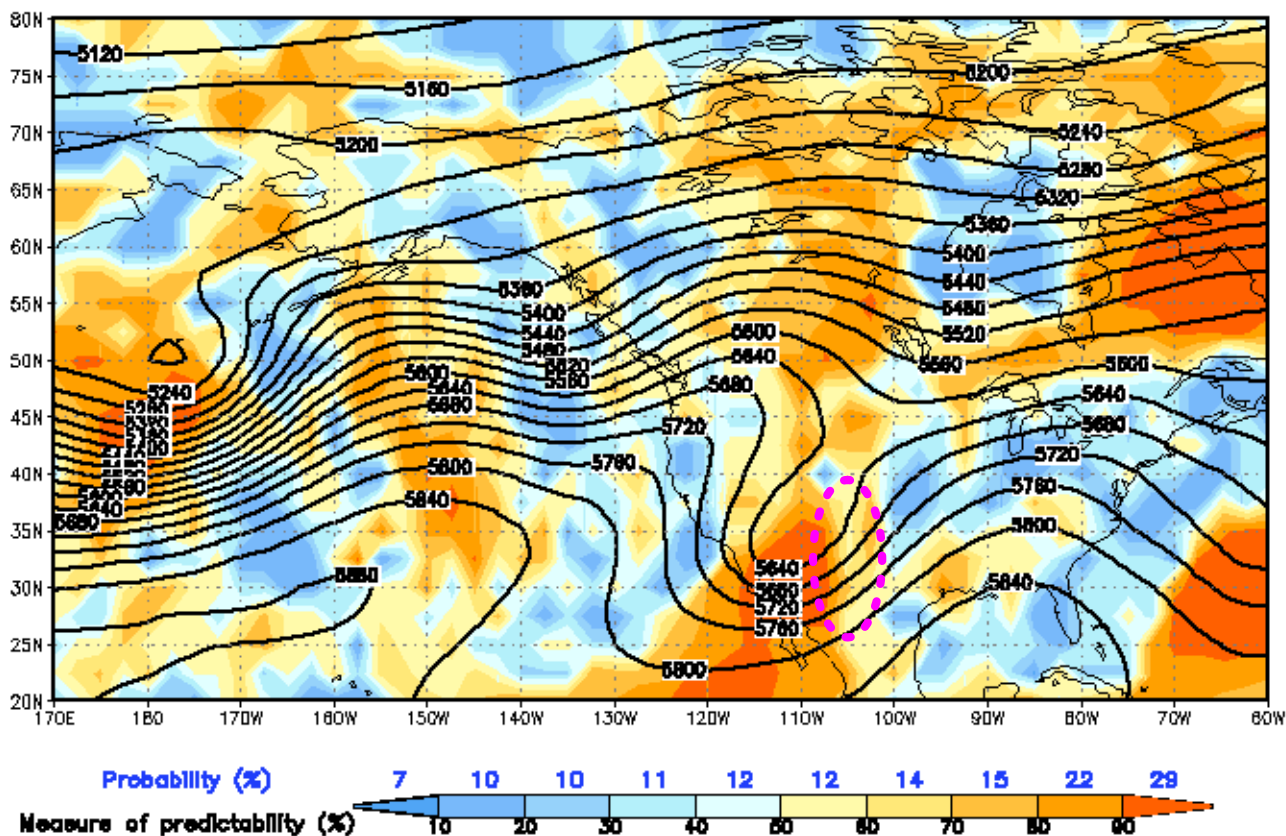


Reliability diagram for 240-hour lead time 500 hPa height NH extratropics forecasts between March and May 1997. Forecast probabilities are based on how many ensemble members fell in any of 10 climatologically equally likely bins at each gridpoint, and are calibrated using verification statistics from the winter of 1995-96. Insert in upper left corner shows in how many events a particular forecast probability was used for the most likely bin (ensemble mode).

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

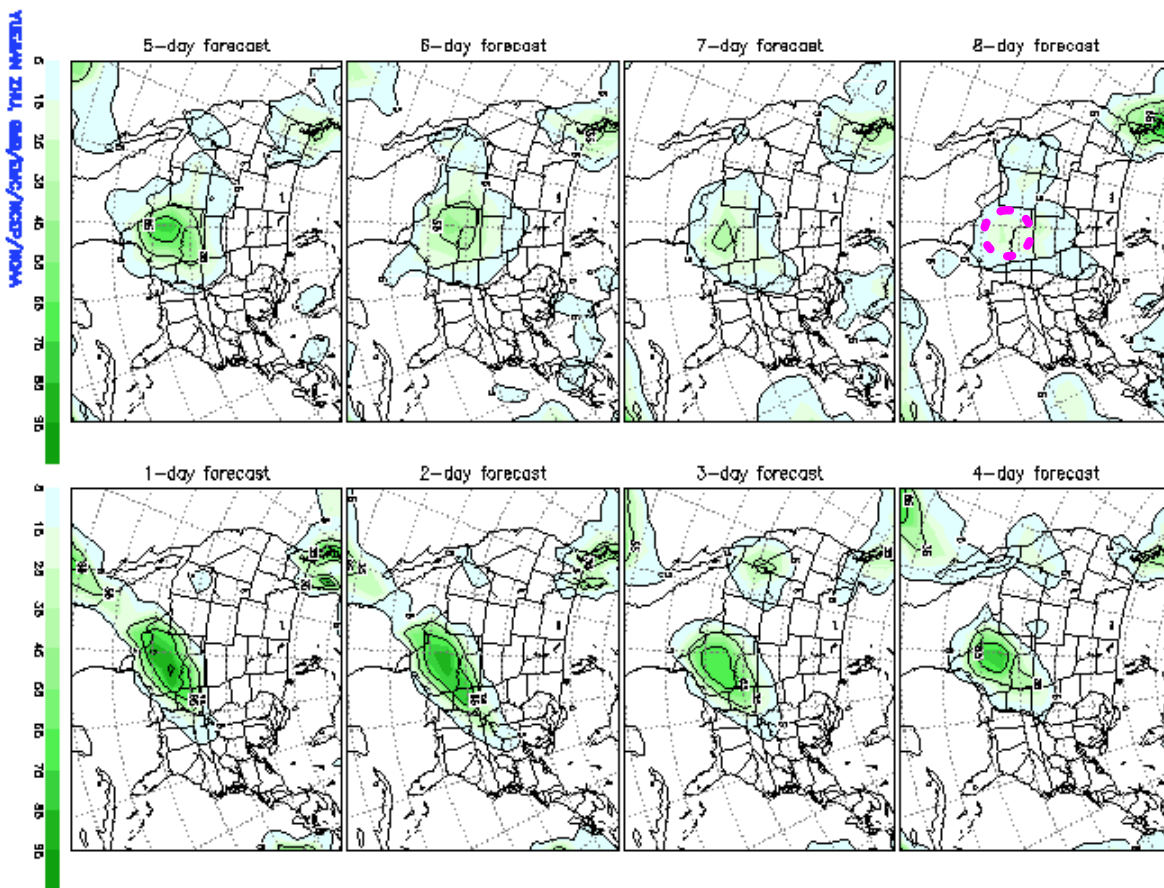
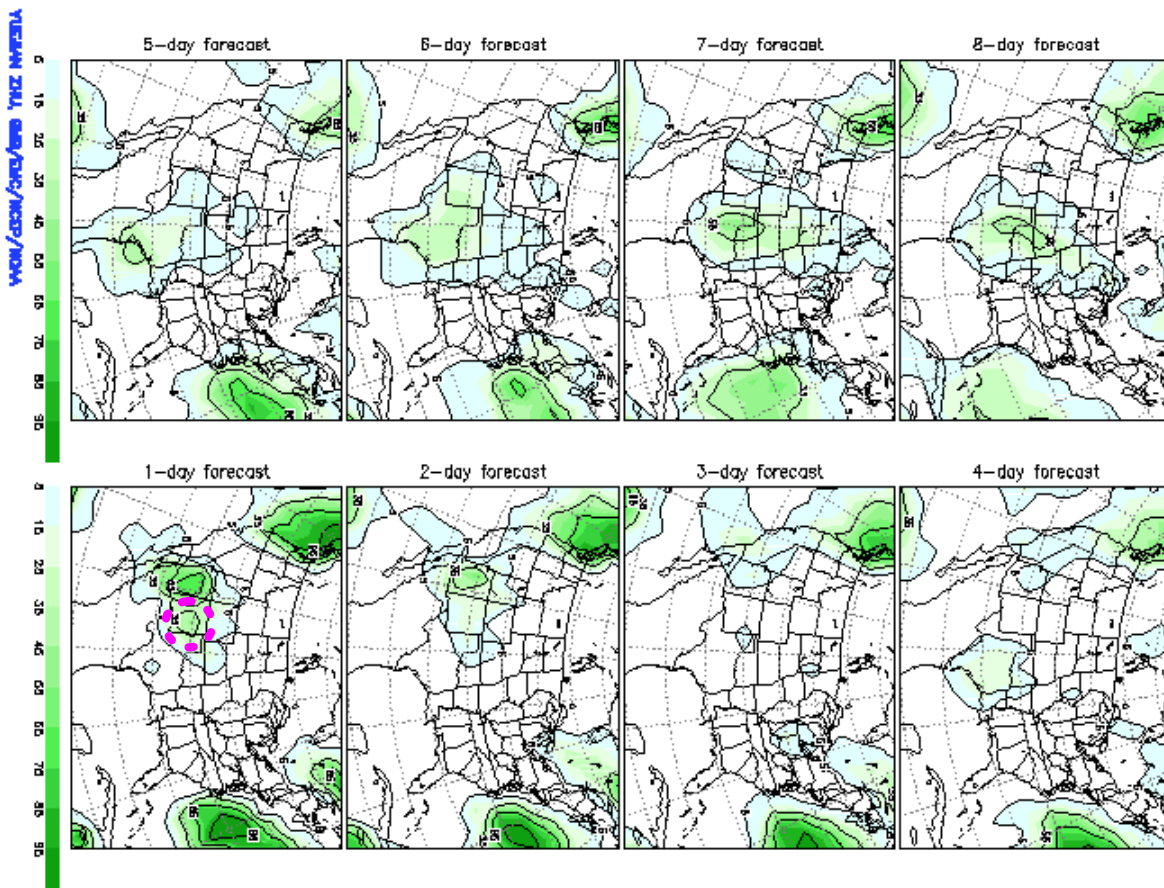


Relative measure of predictability (colors)  
for ensemble mean forecast (contours) of 500 hPa height  
ini: 2000102700 valid: 2000110400 feat: 192 hours



Ens Prob of Precip Amount Exceeding 0.5 Inch (12.7 mm/day) Valid Period: 2000102712-2000102812

Ens Prob of Precip Amount Exceeding 0.5 Inch (12.7 mm/day) Valid Period: 2000110312-2000110412

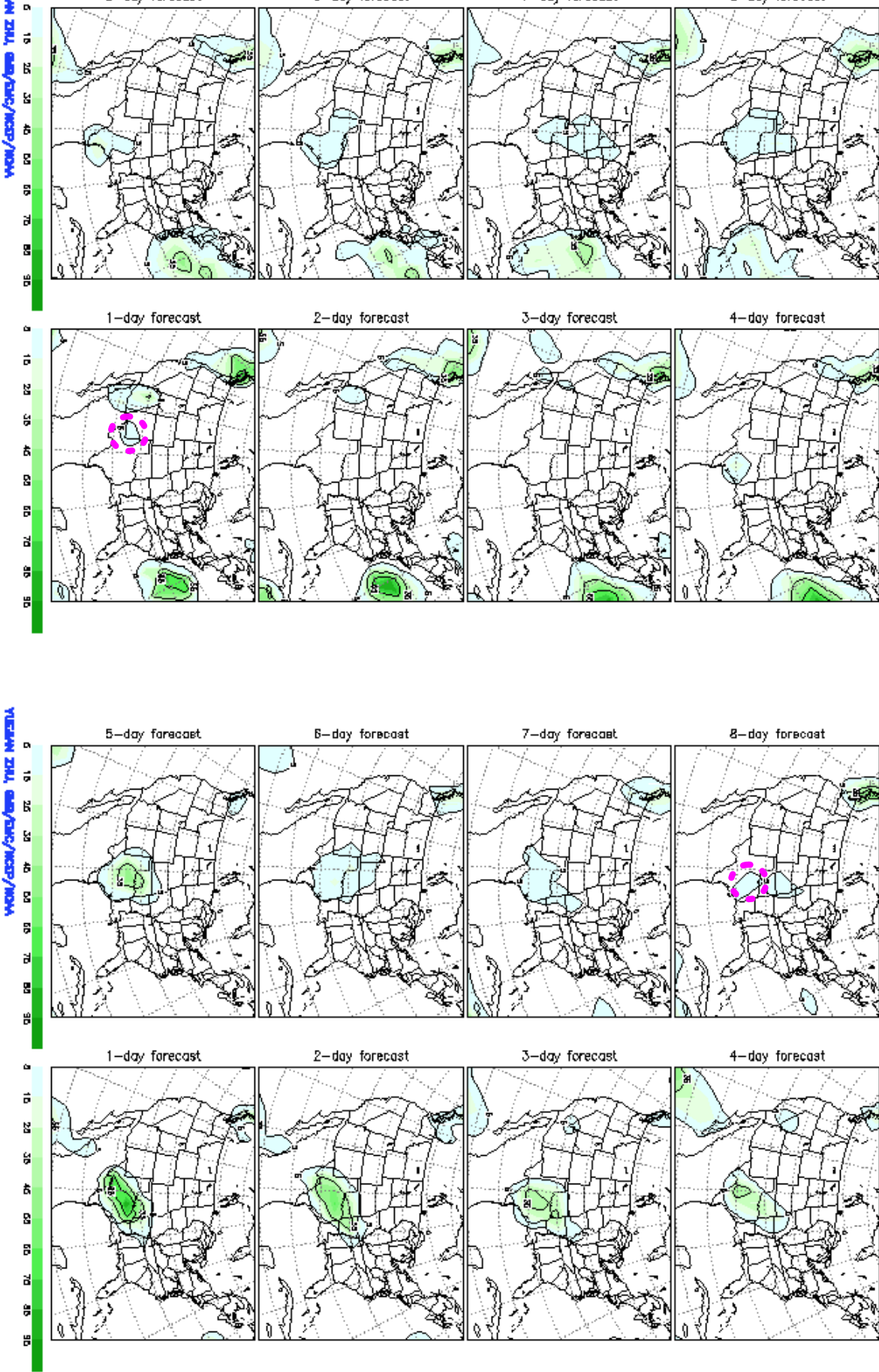


TOZH, Z.: VALIDATION OF PROBABILISTIC FORECASTS

0 10 20 30 40 50 60 70 80 90  
 PROBABILITY (%)

Ens 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

0 10 20 30 40 50 60 70 80 90  
 PROBABILITY (%)



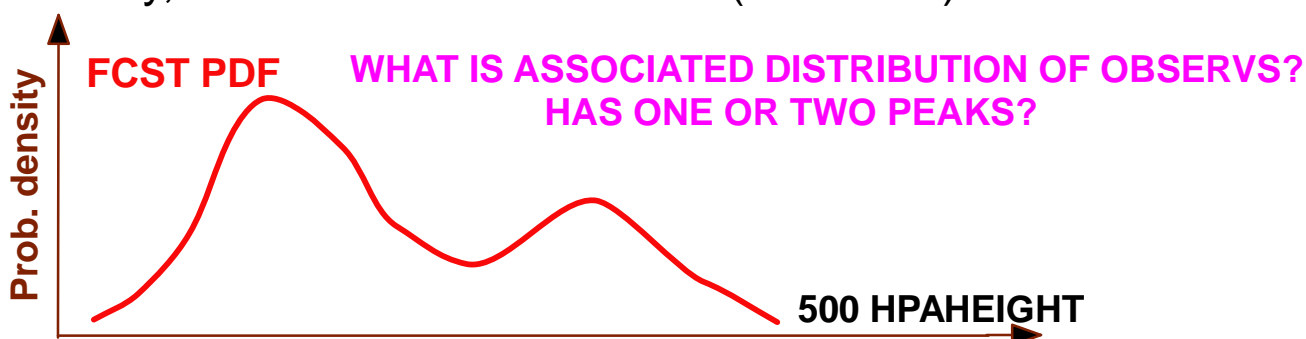
## COMPARING SINGLE CONTROL & ENSEMBLE FCSTS

- 1) Expected value: Ensemble mean better than control?
- 2) *More detailed pdf from ensemble (m vs. 1 members)?*
- 3) *Case dependent variations in spread: Ensemble has skill?*
- 4) **Is it only 2nd moment (spread), or further details in ensemble?**

## CAN ENSEMBLES SKILLFULLY PREDICT BIMODALITY?

WORK IN PROGRESS

Difficult to verify, **NEEDS LOTS OF DATA** (too much)

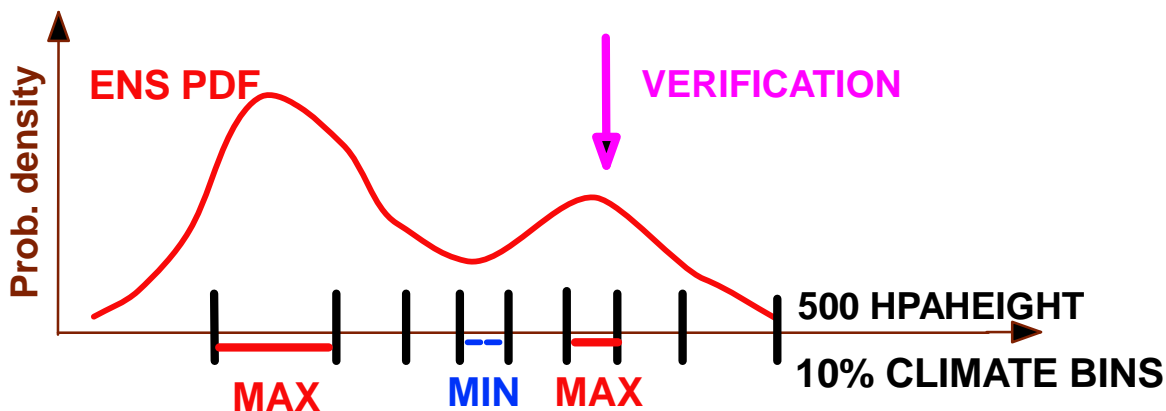


Each fcst pdf pattern needs large number of realizations to establish associated distribution of observations

### APPROACH:

**Use climate pdf as reference** (10 climatologically equally likely bins)  
**Drastically reduce dof** by compositing pdf according to location of max

- 1) Identify bimodal distributions wrt climate pdf
- 2) Locate local maxima & minima in terms of 10 climate bins
- 3) Establish frequency of verifying analysis falling in max/min bins





# CAN ENSEMBLES SKILLFULLY PREDICT BIMODALITY?

1) Given overall ensemble fcst distribution –

*Does bimodality occur more frequently than expected by chance?*

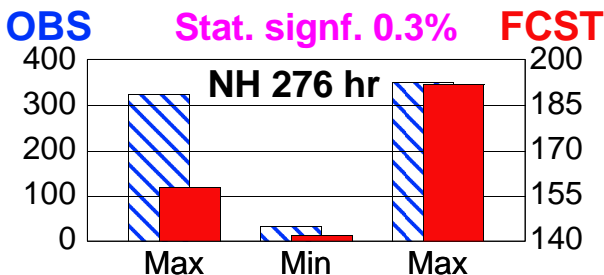
Ratio between multi/unimodal fcst pdfs	12	168	288	360h
NH	0.12	1.1	12	23
SH	0.93	5.3	14	17

Many bimodal pdfs must be due to sampling; have not tested stat. signif

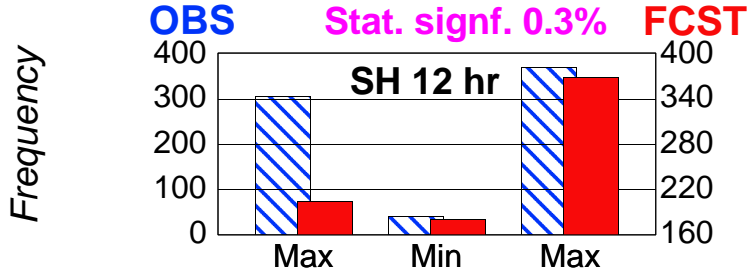
2) *In bimodal fcst cases, do obs confirm bimodality?*

**COMPOSITE RESULTS** for NH & SH extratr. for Nov 2000–Feb 2001

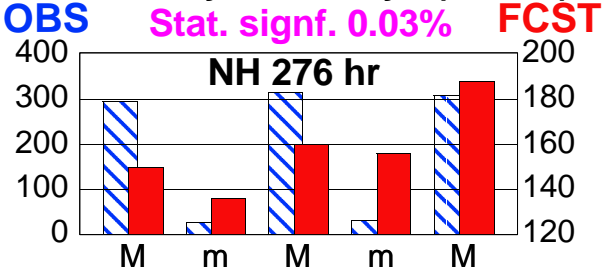
NH Bimodality: 6–16 days



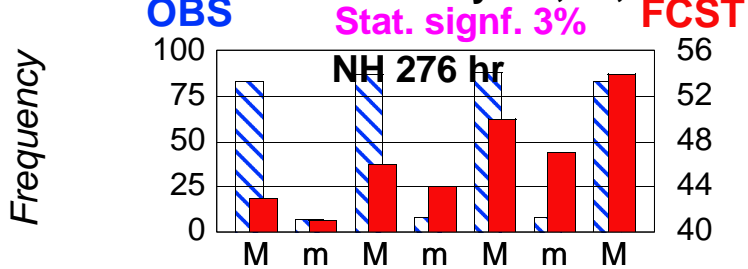
SH Bimodality: 0.5–7 days



NH Trimodality: 9–13 days (exc 12)

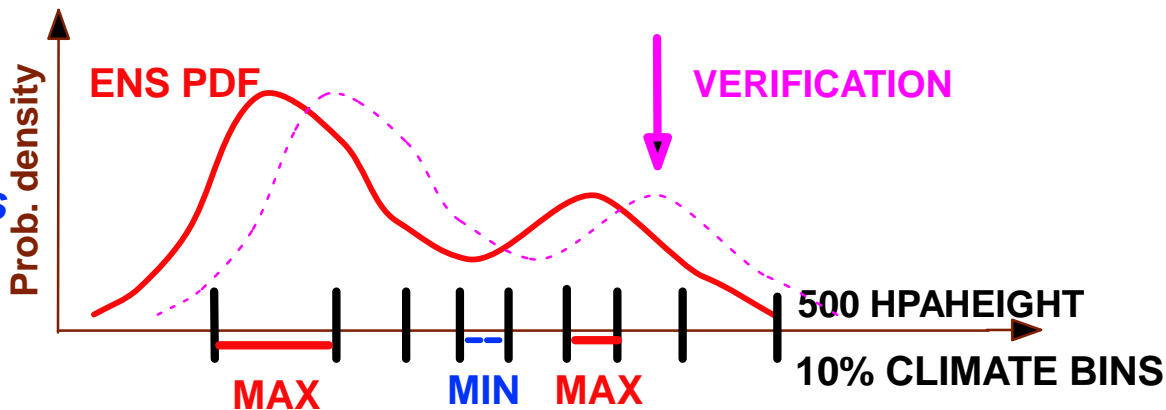


NH Quatremodality: 10, 11, 11.5 d



**PROBLEMS:**  
Bias in fcst model seriously hinders analysis

Bin-resolution (10) too coarse at short lead

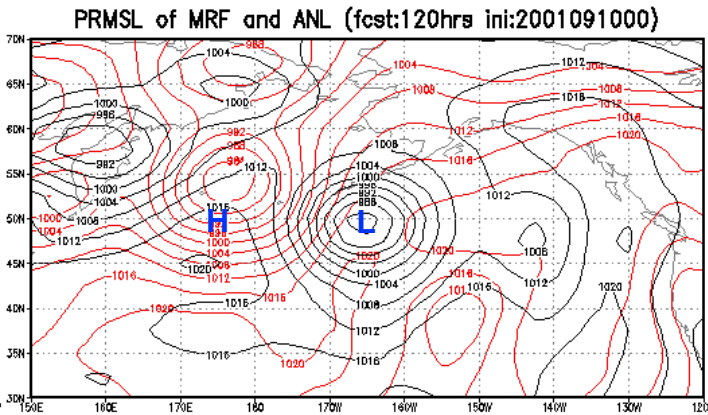
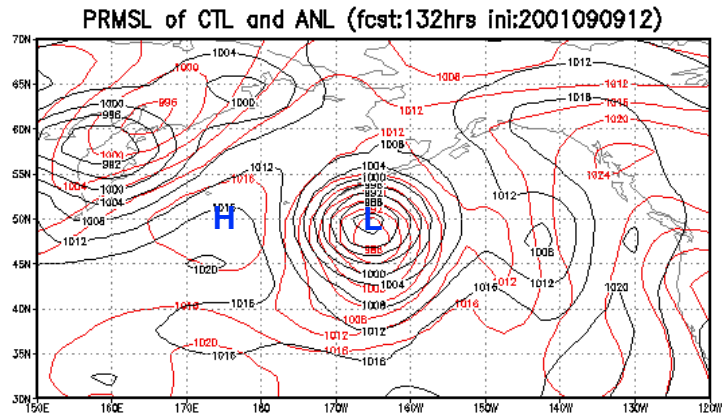


**EXPECTATION:** Verification of bias-reduced fcsts will show stronger multimodal behavior

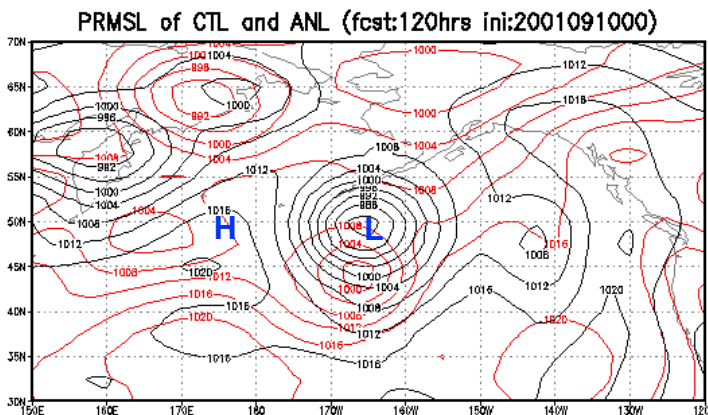
# CAN ENSEMBLES SKILLFULLY PREDICT BIMODALITY?

4) Does multimodality as described here have fcst implications?

## CASE STUDY OF LARGE VARIATIONS IN CONSECUTIVE CONTROL FCSTS

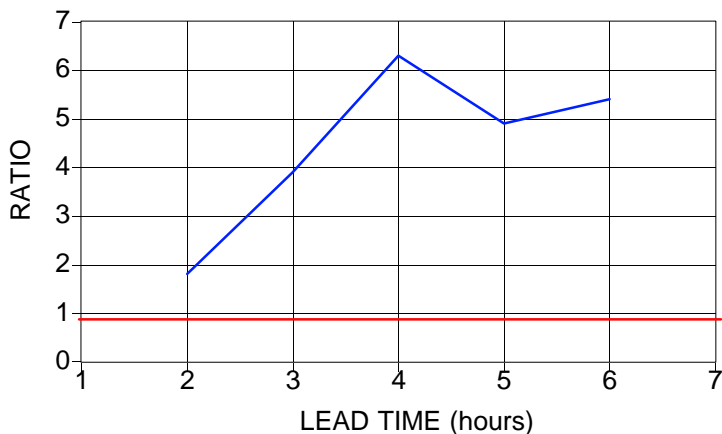


IS THIS PERHAPS RELATED TO MULTIMODALITY?



### USE 50-MEMBER TIME-LAGGED ENSEMBLE

initialized 0909 & 0910 00 & 12Z, 0911 00Z

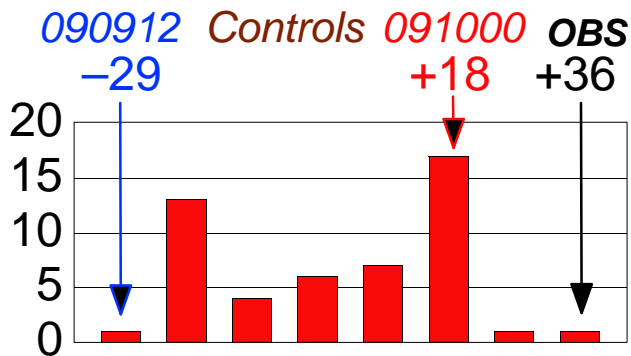


a) # bimodal gridpoints vs average # for Sept 2001 (Ratio)

NUMBER OF MULTIMODAL GRIDPOINTS MUCH HIGHER THAN USUAL

Difference in ratio significant? Probably yes (have not checked)

# CASE STUDY OF LARGE VARIATIONS IN CONSECUTIVE CONTROL FORECASTS



Distribution of High-Low MSLP difference

**STRONGLY BIMODAL**

Statistically significant? Have not tested

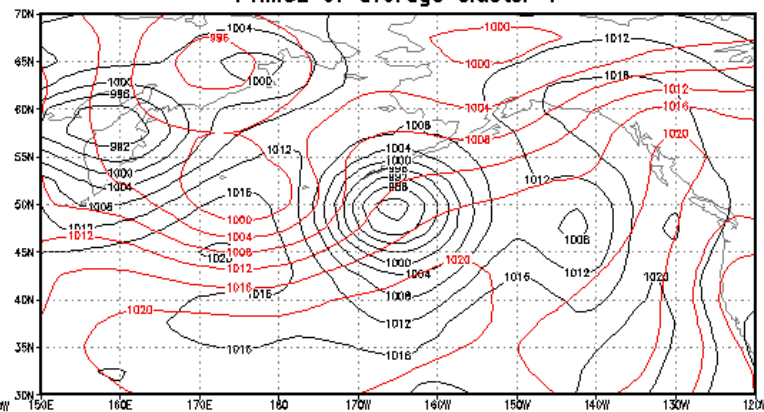
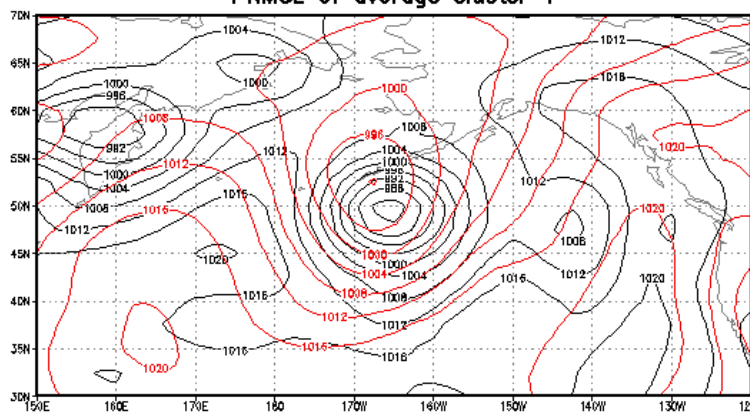
## CLUSTER ANALYSIS – Two dominant patterns

**GOOD CLUSTER (19 members)**

**BAD CLUSTER (20 members)**

PRMSL of average cluster 4

PRMSL of average cluster 1

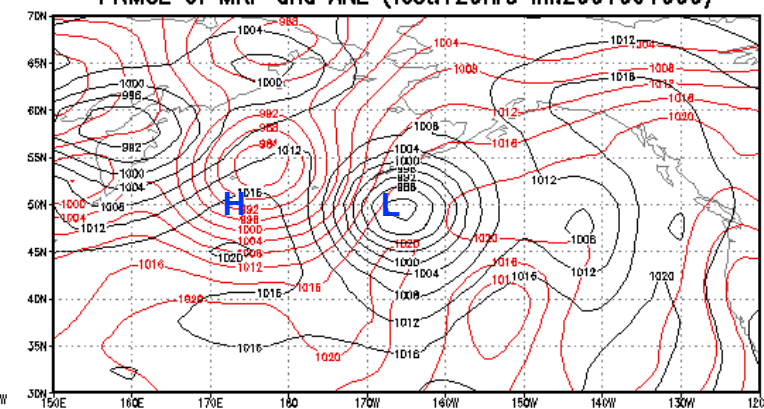
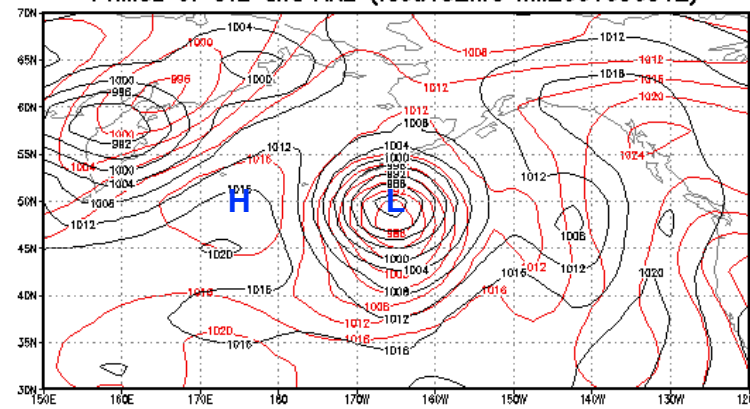


**GOOD CONTROL FCST**

**BAD CONTROL FCST**

PRMSL of CTL and ANL (fcst:132hrs ini:2001090912)

PRMSL of MRF and ANL (fcst:120hrs ini:2001091000)



**CAN CASES LIKE THIS**

**BE IDENTIFIED BY STAT METHODS AS LIKELY REAL?**

## **CONCLUSIONS**

### **1) Types of weather forecasts wrt uncertainty**

***DICHOTOMOUS (CATEGORICAL) VS. PROBABILISTIC***

### **2) Types of weather forecasts wrt generation**

***STATSTICAL; DYNAMICAL – SINGLE VS. ENSMEBLE  
IN ANY CASE, PDF IS DESIRED, GENERAL FORMAT  
ONLY LIOUVILLE EQS PROVIDE THAT – NEED FOR POSTPROC.***

### **3) Attributes of (probabilistic) forecasts**

***RELIABILITY (NO BIAS) &  
RESOLUTION (SMALL RANDOM ERROR)***

### **4) Probabilistic verification measures**

***BRIER, RANKED PROB., ROC, INFO CONTENT, ETC.***

### **5) Value of ensemble vs. control fcst**

***1ST MOMENT BETTER;  
TEMOPRAL VARIATIONS IN 2ND MOMENT  
BETTER DEFINED PDF***

### **6) How much detail can ensembles faithfully define in pdf?**

***BIMODALITY CAPTURED –  
JUMPS IN CONSECUTIVE CONTROL FCSTS ASSOCIATED WITH IT***