A Graphical Technique for Diagnosing Significant Error Tendencies in Meteorological Forecast Models

Matthew C. Sittel

Lockheed Martin Information Technology, Offutt AFB, Nebraska, USA

Robert J. Craig

Air Force Weather Agency, Offutt AFB, Nebraska, USA

Who We Are / What We Do

Model Quality Control Team (QC)
Air Force Weather Agency (AFWA)
Offutt AFB, Nebraska, USA

Mission Statement:

"To research, compile, and provide the best model verification information for Air Force Weather personnel to make better decisions."

AFWA QC Verification Efforts

- Penn State/NCAR MM5 Model run at AFWA

- ~30 different domains (“theaters”)
  45, 15 or 5 km grid point spacing

- Theaters run 2 or 4 times daily
  Forecasts out to as far as 72 hours

AFWA QC Verification Efforts

- Verification includes air temperature, wind, pressure, precipitation, present weather

- Surface and upper air data

- Point verification for military bases and other locations of operational interest

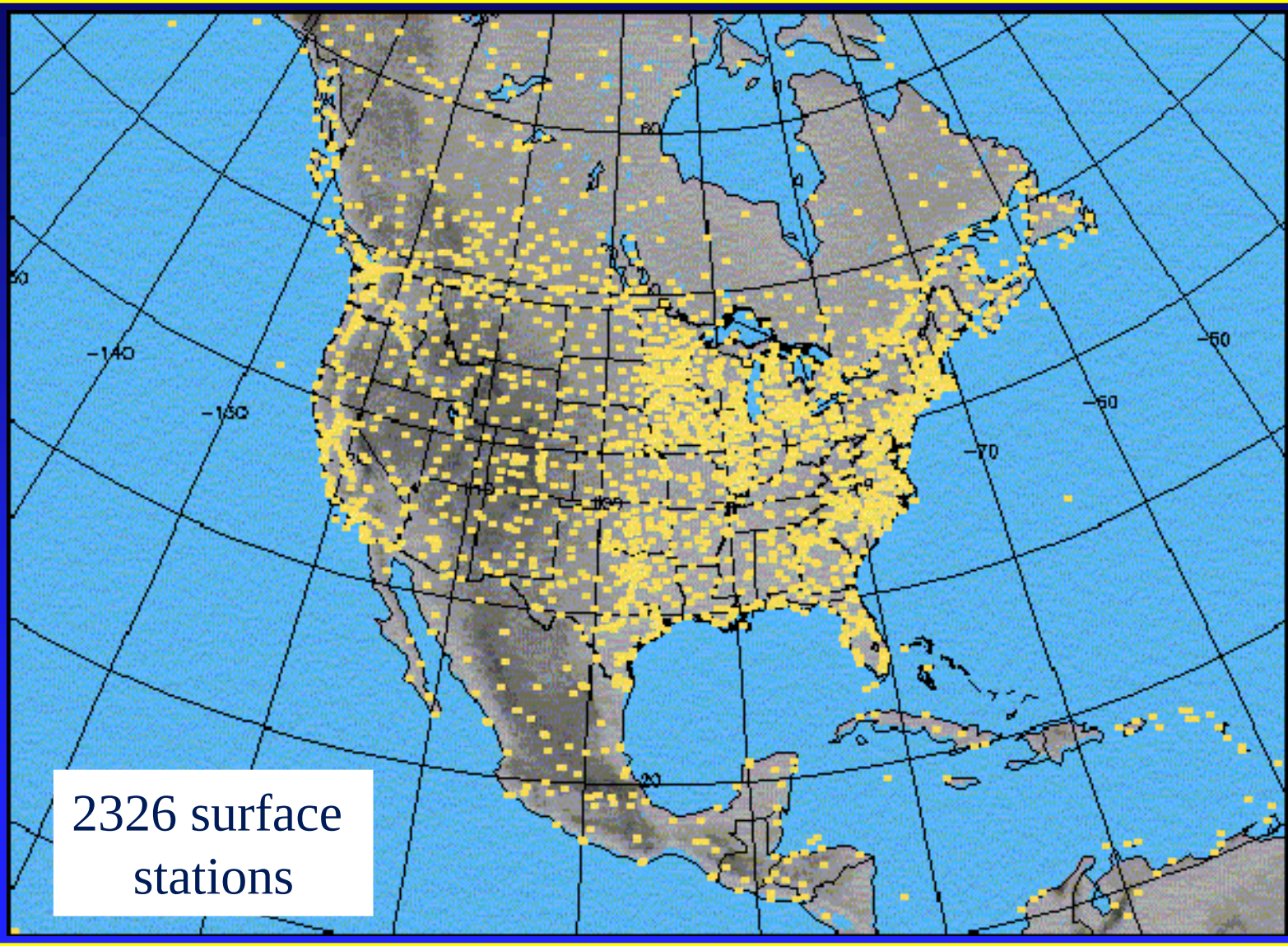
Point Verification

- Model forecast values determined at each valid observation point

- 4-point bilinear interpolation used for continuous fields

- Closest grid point (“nearest neighbor”) used for all other fields

Sample Station Coverage Map



Global Data Counts

- ~ 10,000 surface observations hourly
- ~ 700 upper air sounding sites (11 mandatory levels-1000 to 100 mb)
- Theaters can overlap; observation sites can be in multiple theaters
- ~ 30 theaters, up to 4 runs/day
- How much data IS that?

How much? Too much!

~ 7.5 MILLION model/observation pairs…DAILY!

(and ~ 700 MB disk space/day)

The Problem

- Field users do not have the time or training to examine that much data.

- We must provide useful data for all locations in the world where operations may take place.

- How can we reduce the data amount yet still provide useful information?

Forecasts aren't perfect!

- Field users want perfect forecasts, since an incorrect forecast may be very costly.

- But…models do make errors!

- Description of errors: How much? How often? Can we summarize without losing too much detail?

Model Forecast Errors

- Forecast Error (E):

  E = Model forecast minus Observed value

- E=4°C is a worse forecast than E=2°C

- But a 4°C error near a critical value (e.g., 0°C for aircraft) can be more costly than a 4°C error in warm weather

User-Defined Errors

- Users have different definitions of a "bad" forecast

- Define an "extreme error" based on the field user's criterion for a bad forecast (e.g., "model off by more than 4°C")

- Forecast Error Magnitude Threshold (FEMT) comes from the extreme error definition (± 4°C in above example).

Categorizing Errors

- Long-term model performance requires examination of results from many model runs to determine forecast error distribution

- Categorization of the errors based on FEMT simplifies calculations and depictions of error distribution at each observation site

Categorizing Errors

- One FEMT (e.g., 4 °C) defines four possible categories for a forecast error:

E	E	E	E
less than -4 °C	greater than or equal to -4 °C but less than 0 °C	greater than or equal to 0 °C but less than or equal to +4 °C	greater than +4 °C

Categorization Example

Forecast	7 °C			
Observed	12 °C			
Error (E)				

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
0	0	0	0

Categorization Example

Forecast	7 °C			
Observed	12 °C			
Error (E)	- 5 °C			

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
0	0	0	0

Categorization Example

Forecast	7 °C			
Observed	12 °C			
Error (E)	- 5 °C			

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
1	0	0	0

Categorization Example

Forecast	7 °C	15 °C		
Observed	12 °C	14 °C		
Error (E)	- 5 °C	+1 °C		

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
1	0	0	0

Categorization Example

Forecast	7 °C	15 °C		
Observed	12 °C	14 °C		
Error (E)	- 5 °C	+1 °C		

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
1	0	1	0

Categorization Example

Forecast	7 °C	15 °C	19 °C	19 °C
Observed	12 °C	14 °C	13 °C	12 °C
Error (E)	- 5 °C	+1 °C	+6 °C	+7 °C

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
1	0	1	0

Categorization Example

Forecast	7 °C	15 °C	19 °C	19 °C
Observed	12 °C	14 °C	13 °C	12 °C
Error (E)	- 5 °C	+1 °C	+6 °C	+7 °C

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
1	0	1	2

Categorization Example

After 60 forecasts, the error distribution might look like this:

E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
6	13	24	17

Categorization Example

23 of the 60 forecast errors were extreme.

EXTREME			EXTREME
E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
6	13	24	17

Categorization Example

23 of the 60 forecast errors were extreme.
Of those 23 errors, 17 were positive in sign.

EXTREME			EXTREME
E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
6	13	24	17

Categorization Example

23 of the 60 forecast errors were extreme.
Of those 23 errors, 17 were positive in sign.

38% extreme errors (23 / 60)
74% positive extremes (17 / 23)

EXTREME			EXTREME
E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
6	13	24	17

Extreme Errors

38% extreme errors
74% positive extremes

The frequency and tendency of extreme errors is of particular interest to the operational user of model forecasts.

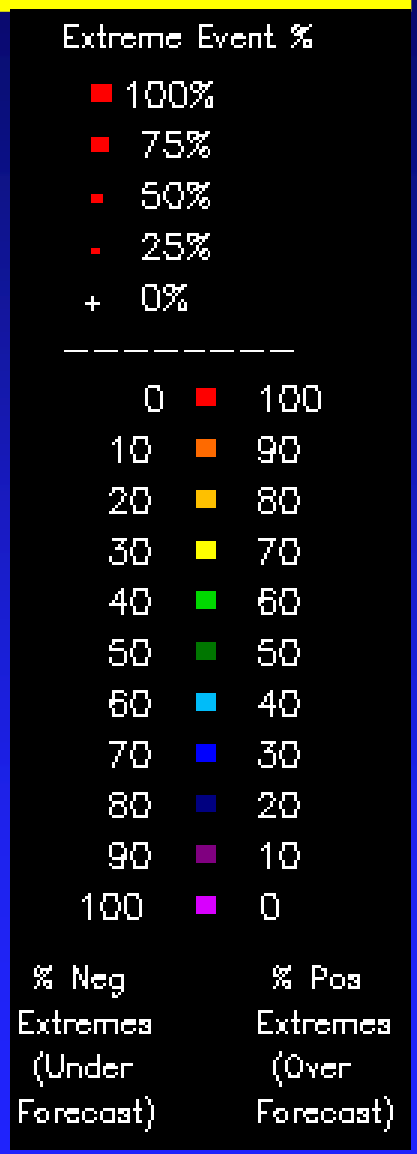
We use these two values as attributes of one map symbol.

Extreme Errors

Percentage of extreme errors

Size: larger box, more errors

(plus symbol for NO errors)



Extreme Errors

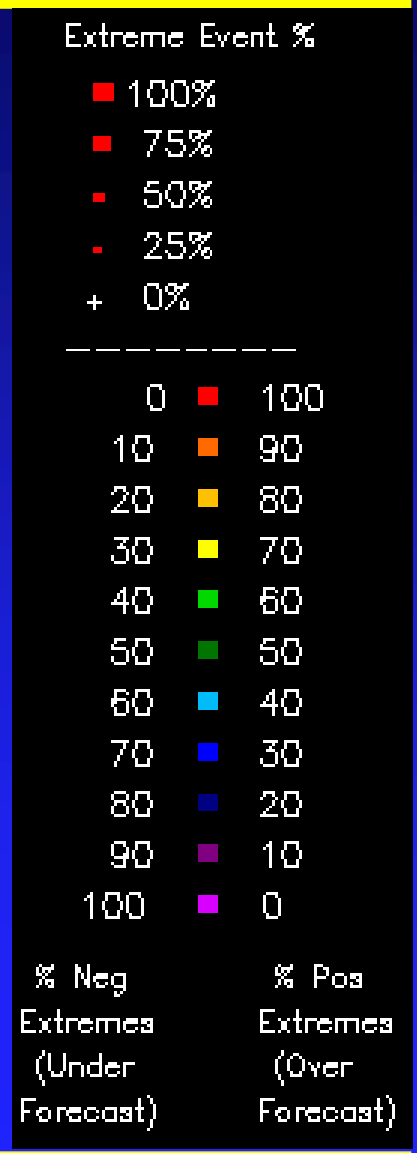
Percentage of extreme errors

Size: larger box, more errors (plus symbol for NO errors)

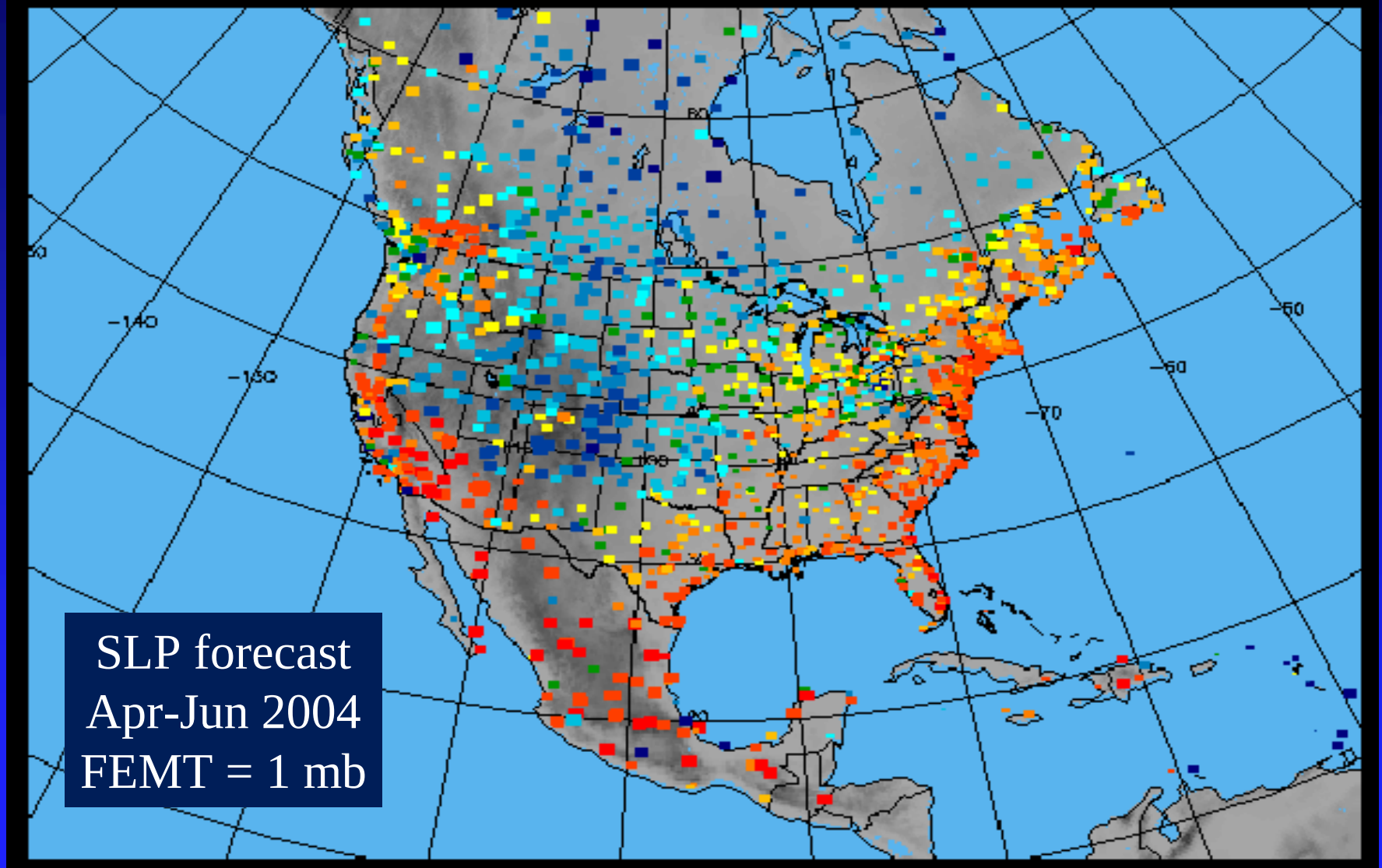
Tendency of extreme errors

Color: purples represent more negatives, reds more positives

Plot results for each location



Graphical Accuracy Product (GAP)

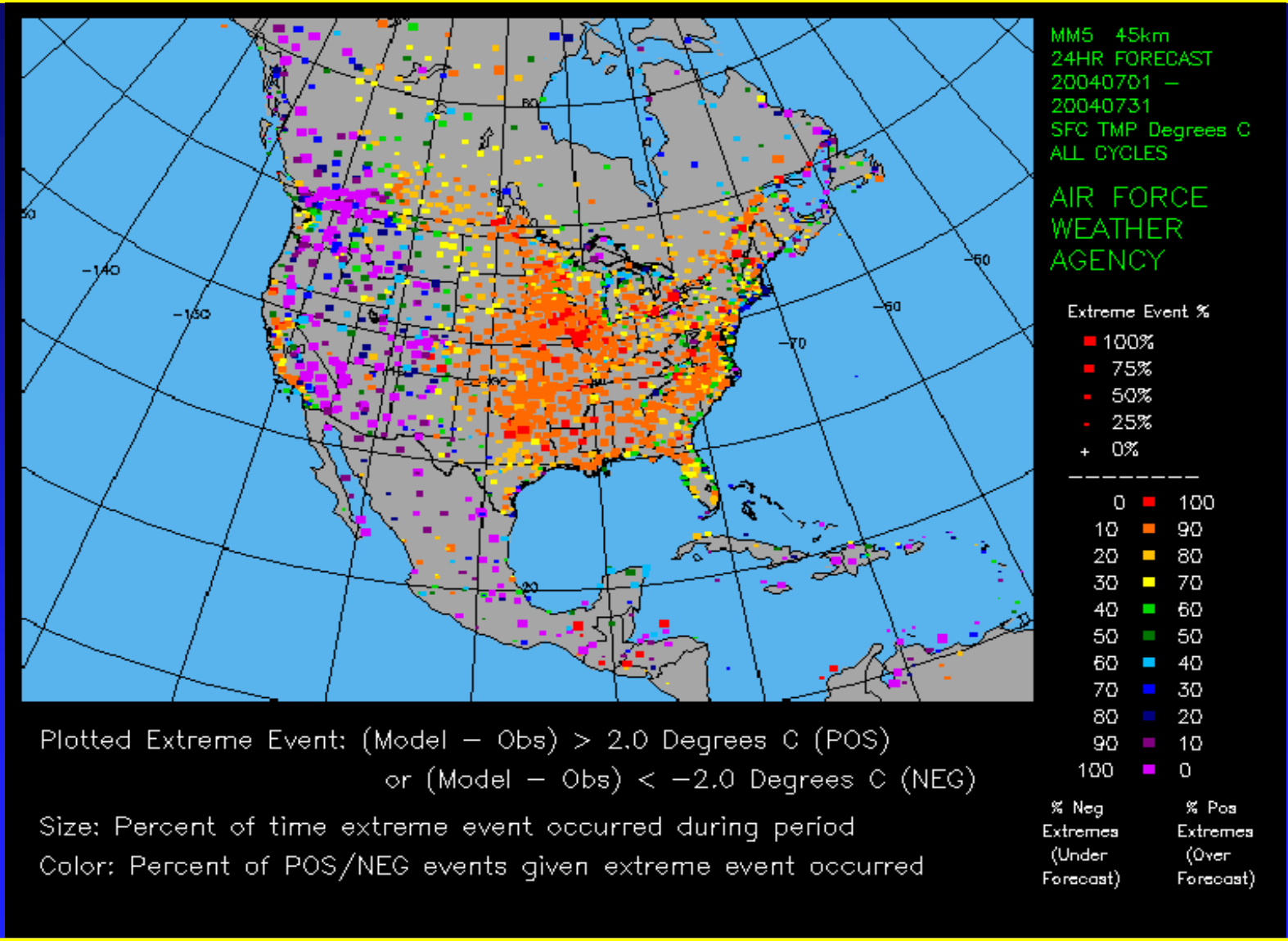


Sample GAPs

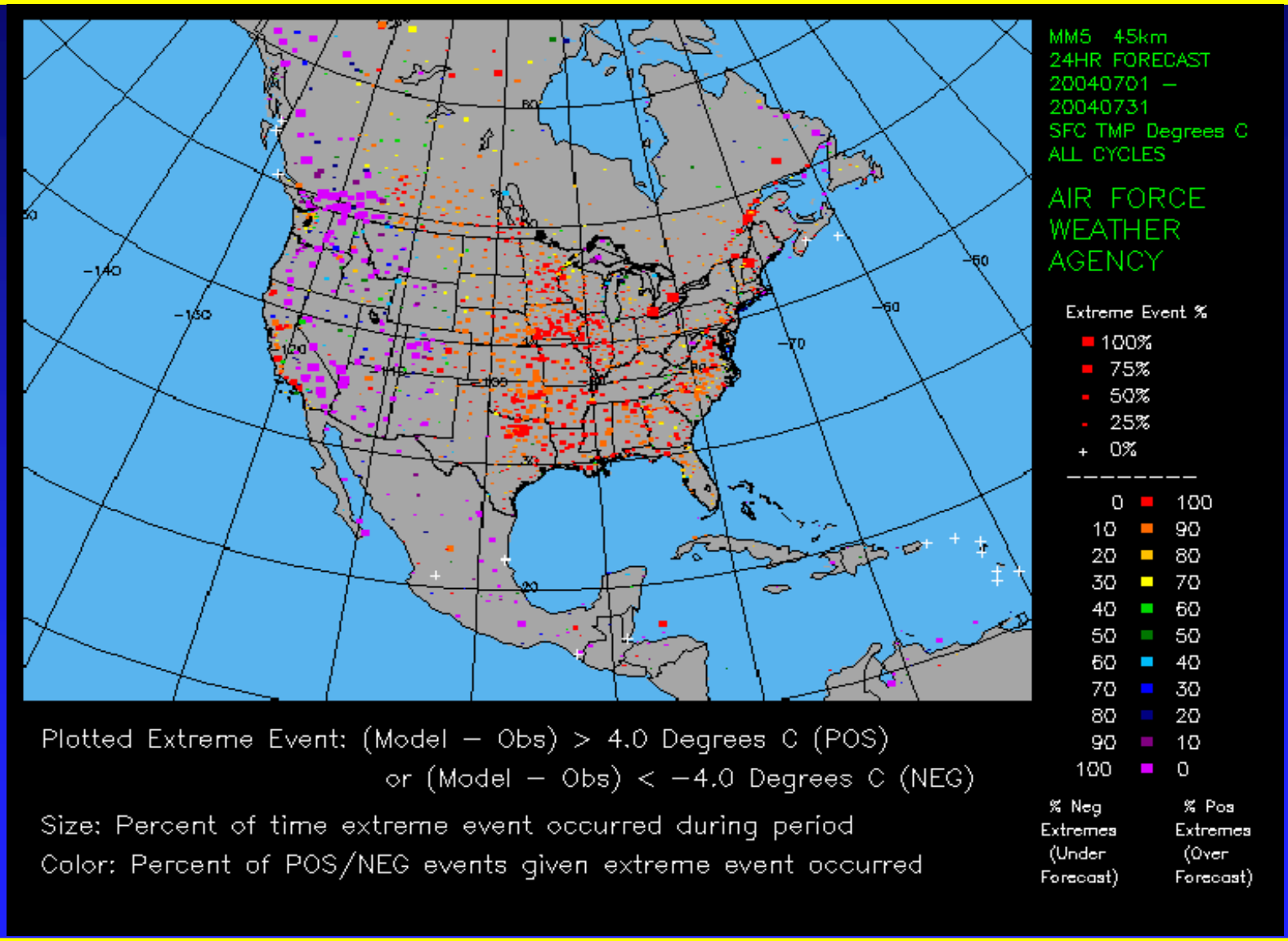
Examples of MM5 24-hour surface forecast performance during the period July 1-31, 2004*

* Note: inclusion on map is limited to those stations with at least 40% observation availability during the month.

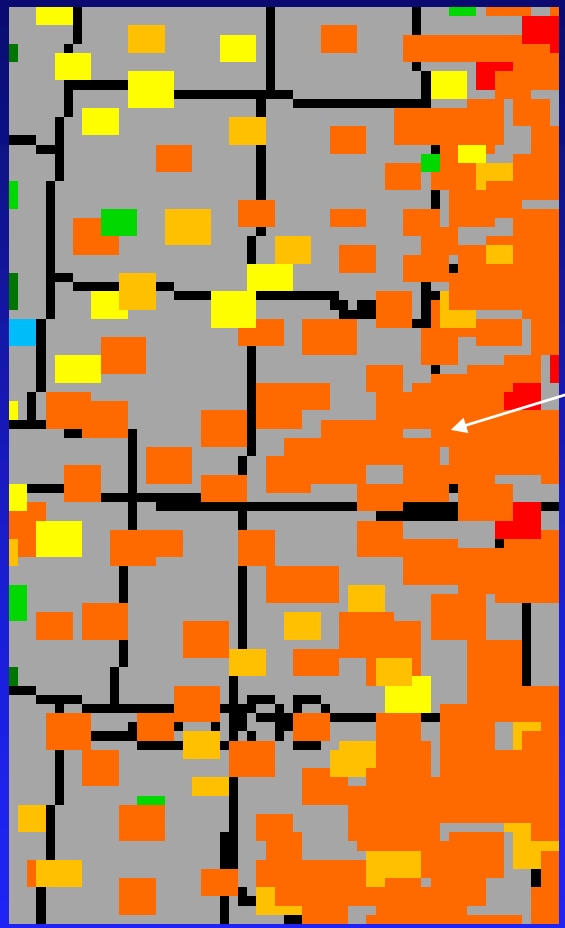
Temperature, FEMT=2 °C



Temperature, FEMT=4 °C

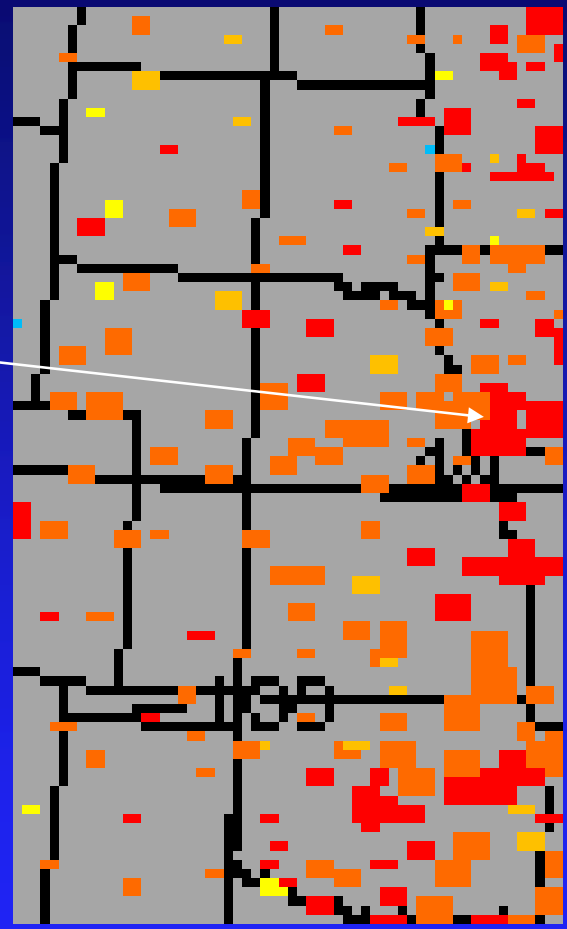


Offutt Distribution Example



Different FEMTs

Different Symbols



FEMT=2 °C

FEMT=4 °C

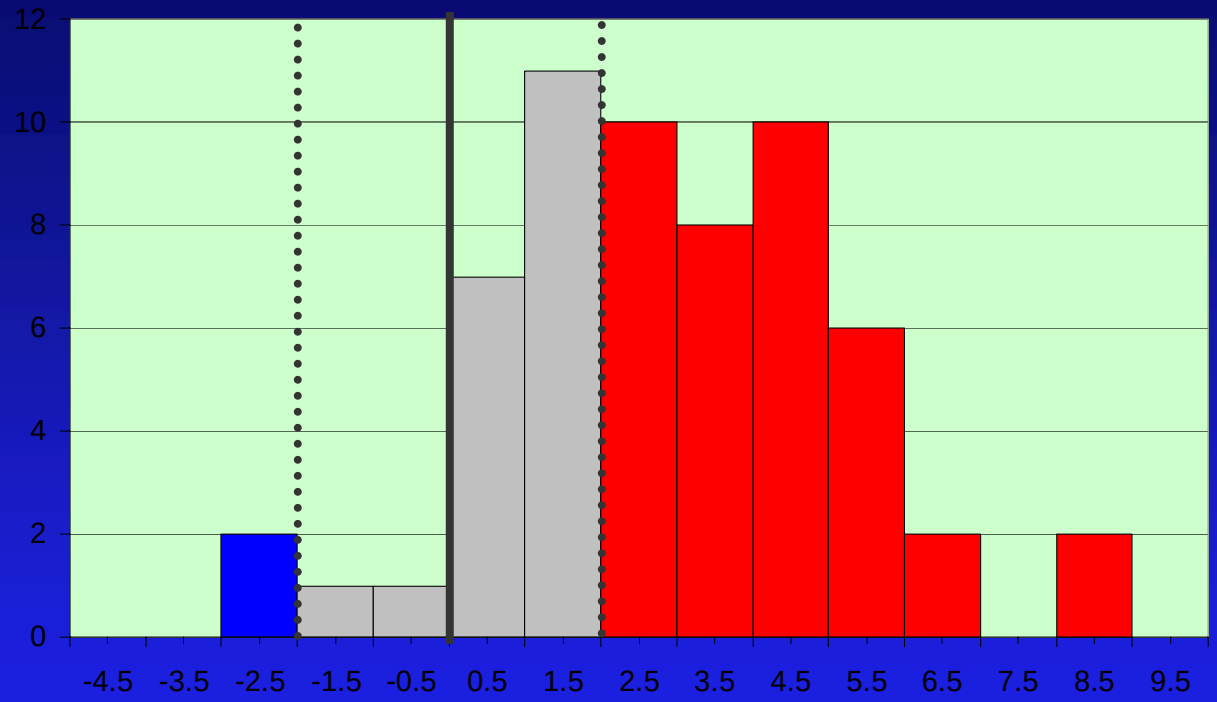
Offutt Distribution Example

FEMT = 2 °C

60 errors

67% Extremes

95% Positive Extremes



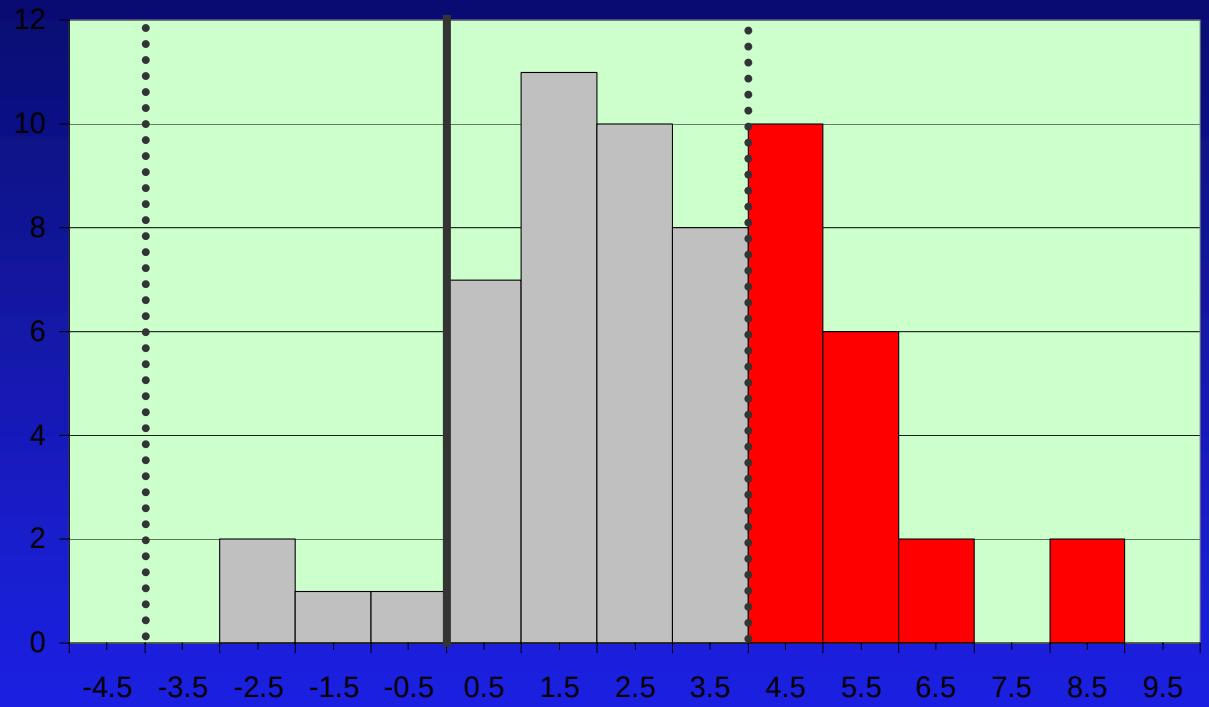
E < -2 °C	-2 °C <= E < 0 °C	0 °C <= E <= +2 °C	E > +2 °C
2	2	18	38

Offutt Distribution Example

FEMT = 4 °C
60 errors

33% Extremes

100% Positive Extremes

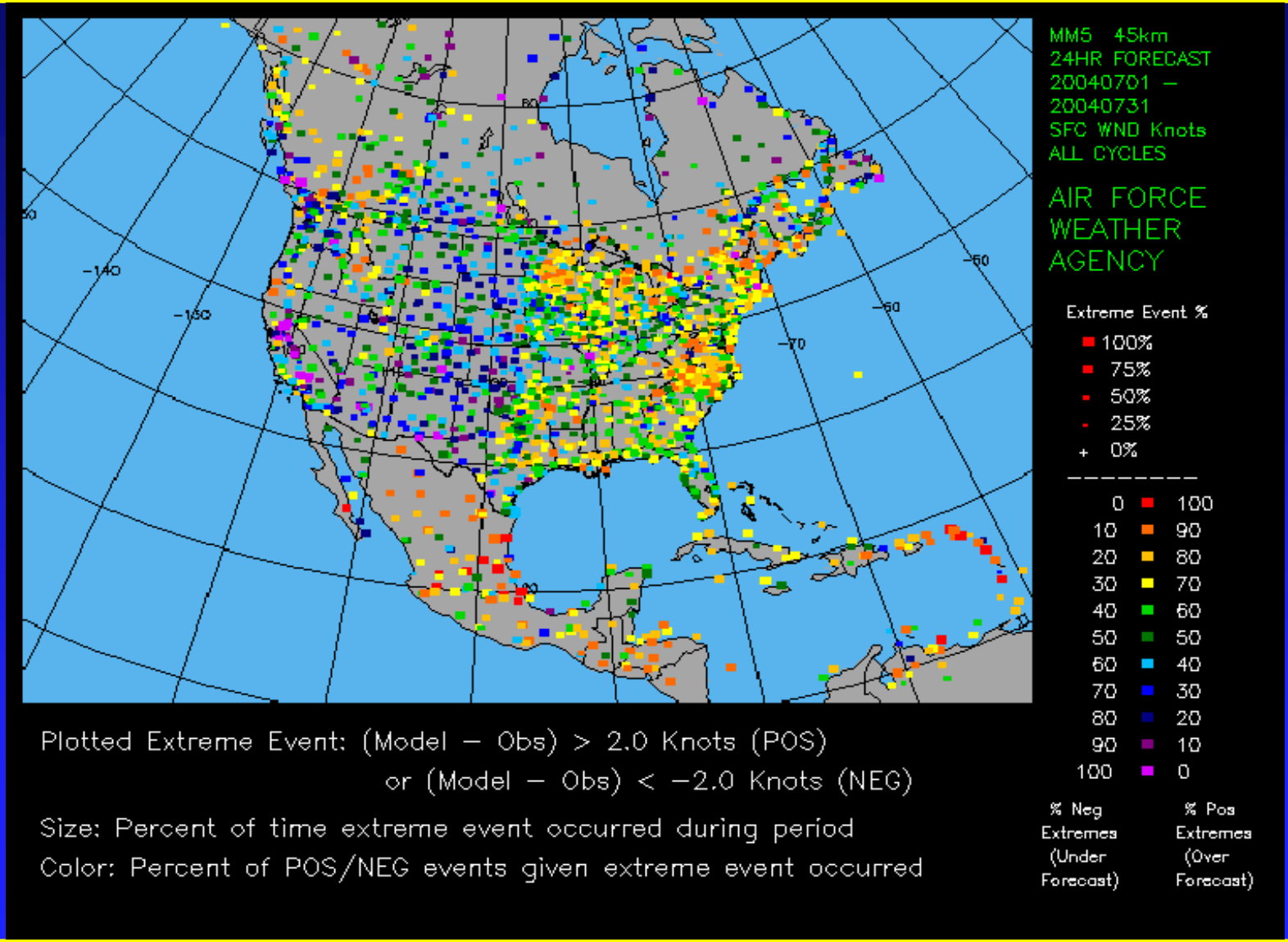


E < -4 °C	-4 °C <= E < 0 °C	0 °C <= E <= +4 °C	E > +4 °C
0	4	36	20

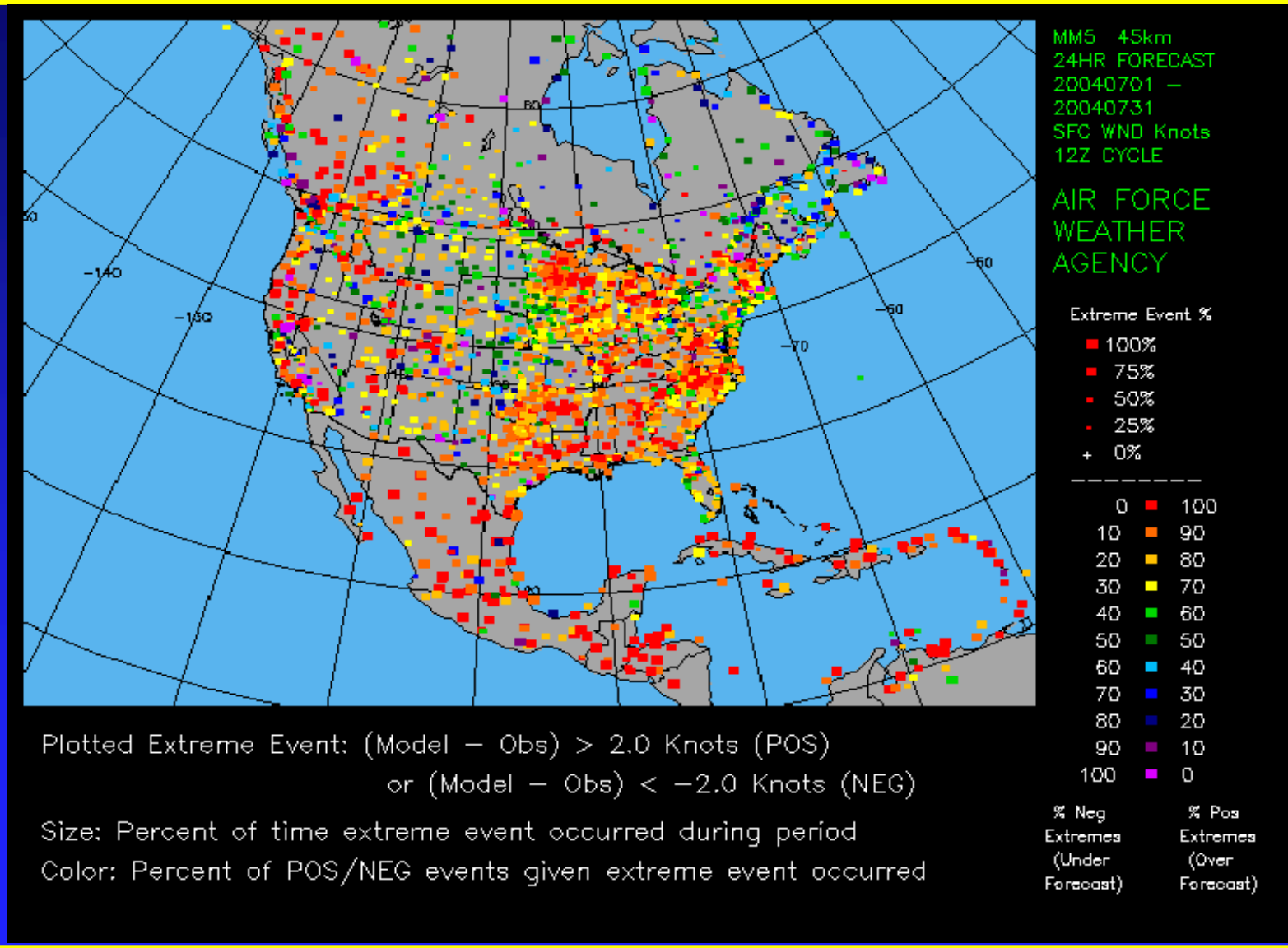
The Versatile GAP

- Adjustment of FEMT changes the categorization
- Length of time for the categorization can be varied
- Valid for any geographic domain
- FEMT can be defined as conditions for one or more variables (e.g., precipitation and temperatures < 0 °C)

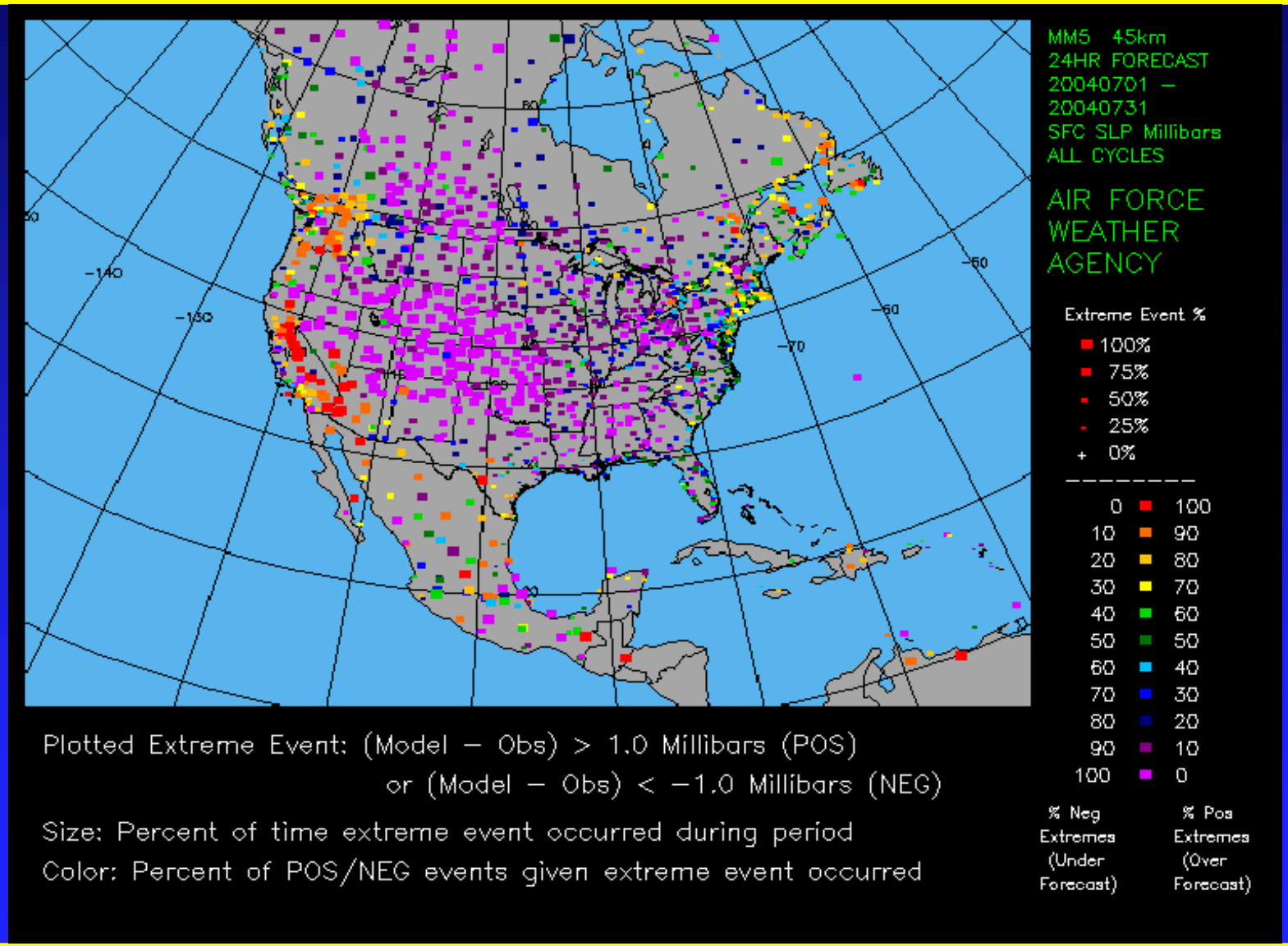
Wind Speed, FEMT=2 kts, all cycles



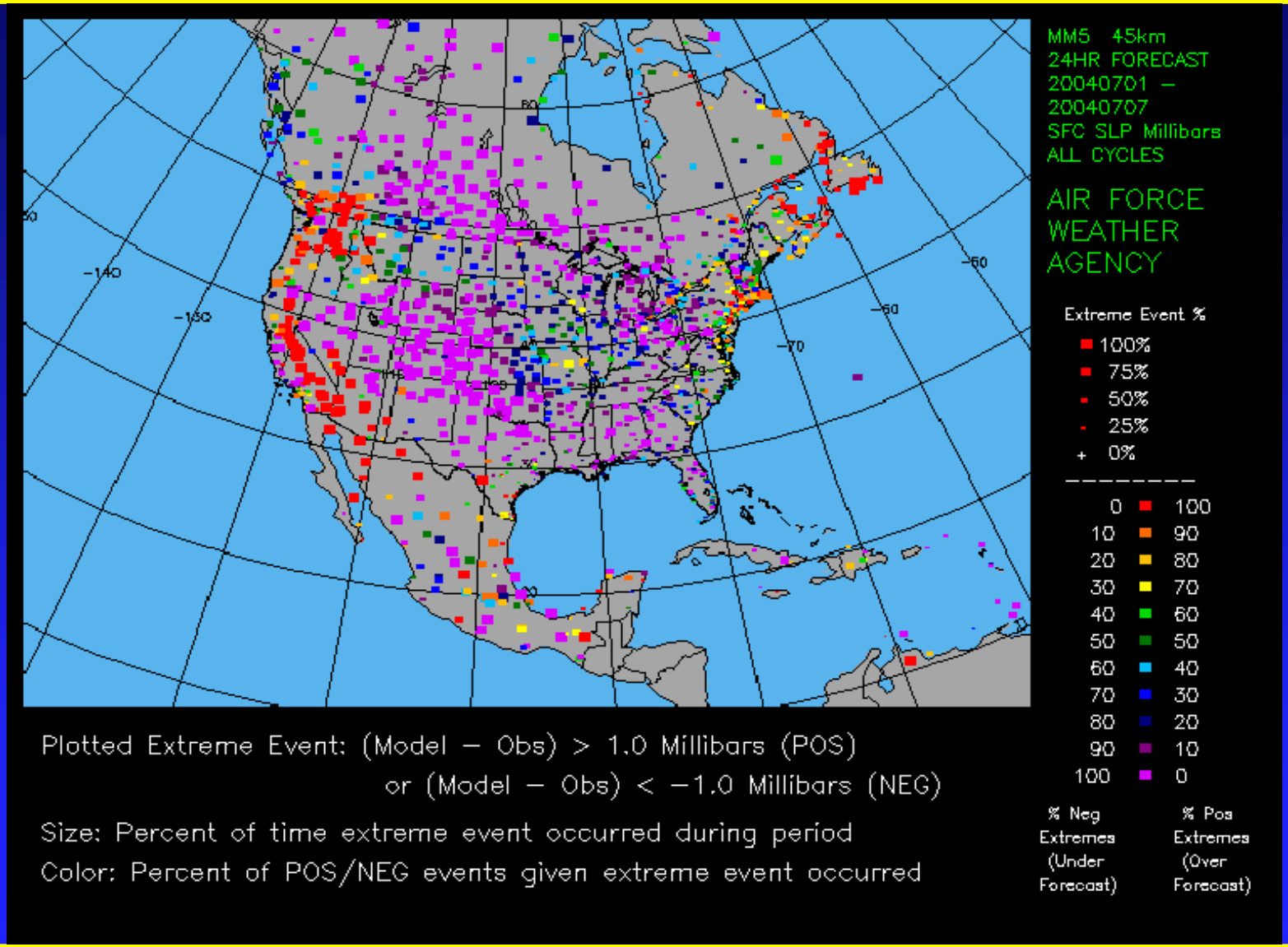
Wind Speed, FEMT=2 kts, 12Z only



SLP, FEMT=1 mb, 31 days



SLP, FEMT=1 mb, 7 days



GAP Summary

- Illustrates frequency and tendency of extreme errors in model forecasts
- Versatile in choice of variable, location and error thresholds
- Adjustable to highlight event of interest
- Greatly reduces amount of data for end user, yet still provides useful information

Questions/comments?

E-mail:

Matthew.Sittel@afwa.af.mil