

Practically-perfect Hindcasts



Mike Kay
CIRES/FSL
kay@fsl.noaa.gov

So what's the goal?

Provide meaningful feedback to people making areal forecasts

Sounds easy enough, right?

If it's so easy then why are we having a whole workshop where the theme is about making verification more meaningful?!

SOUTH PARK



The underpants gnomes



They had a business model all planned out that was going to make them rich:

Step 1: Collect underpants

Step 2: ???

Step 3: **PROFIT!!!**

*We're kinda like the underpants
gnomes*



We in the verification community have a basic business model as well:

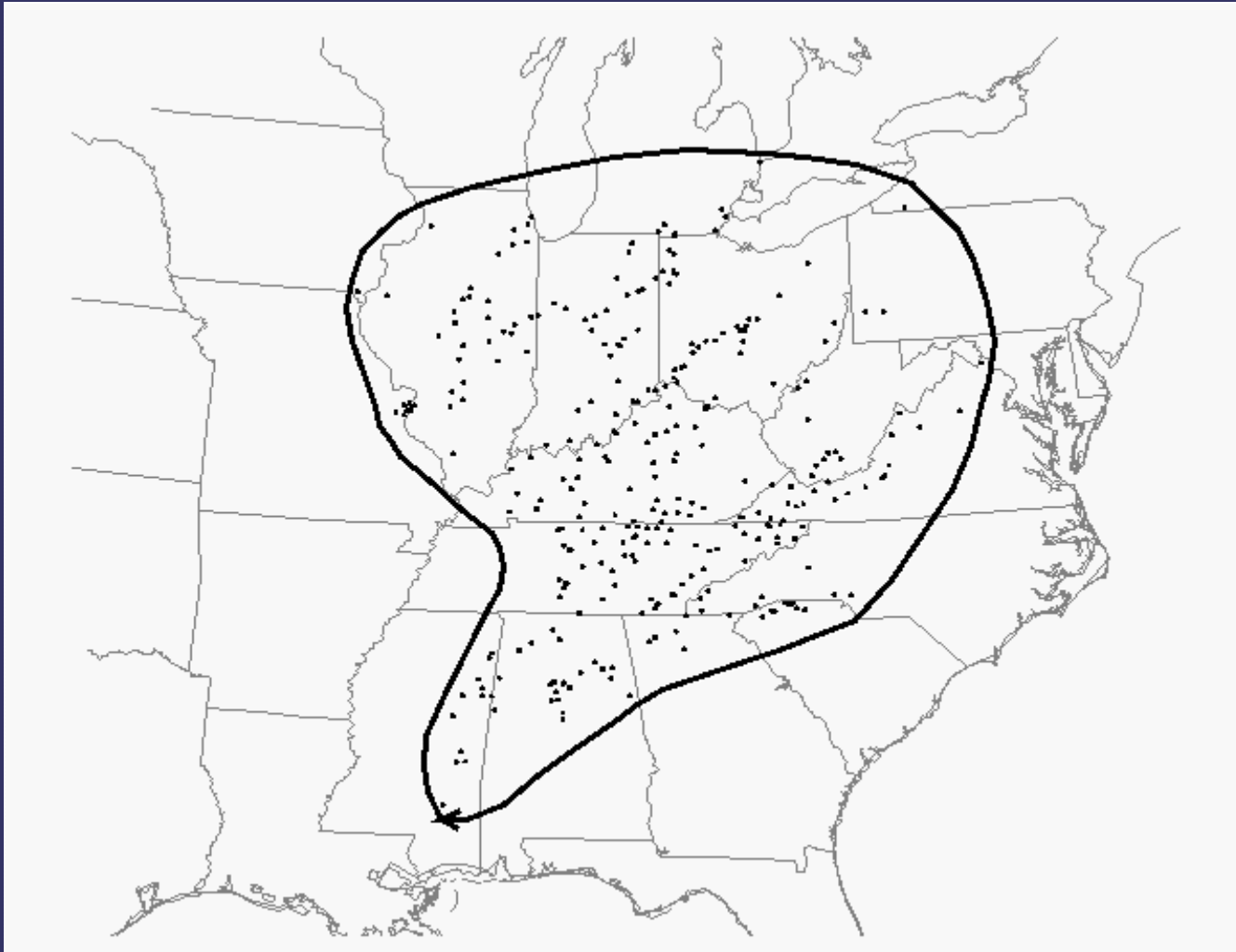
Step 1: Gather forecasts and observations and perform some type of verification

Step 2: ???

Step 3: **Forecast improvement!!!**

Let's focus on Step 2...

What's the problem?



Hypothetical 'perfect' Convective Outlook for 3 April 1974
adapted from Weiss et al. (1980)

Just how bad (or good) is it?

For this outlook, the POD = 1.0, the FAR = 0.71 and the CSI = 0.29 (verified on a 40km grid similar to the MDR grid traditionally used)

On the largest tornado day in history with what is considered a perfect forecast, the CSI is only 0.29!

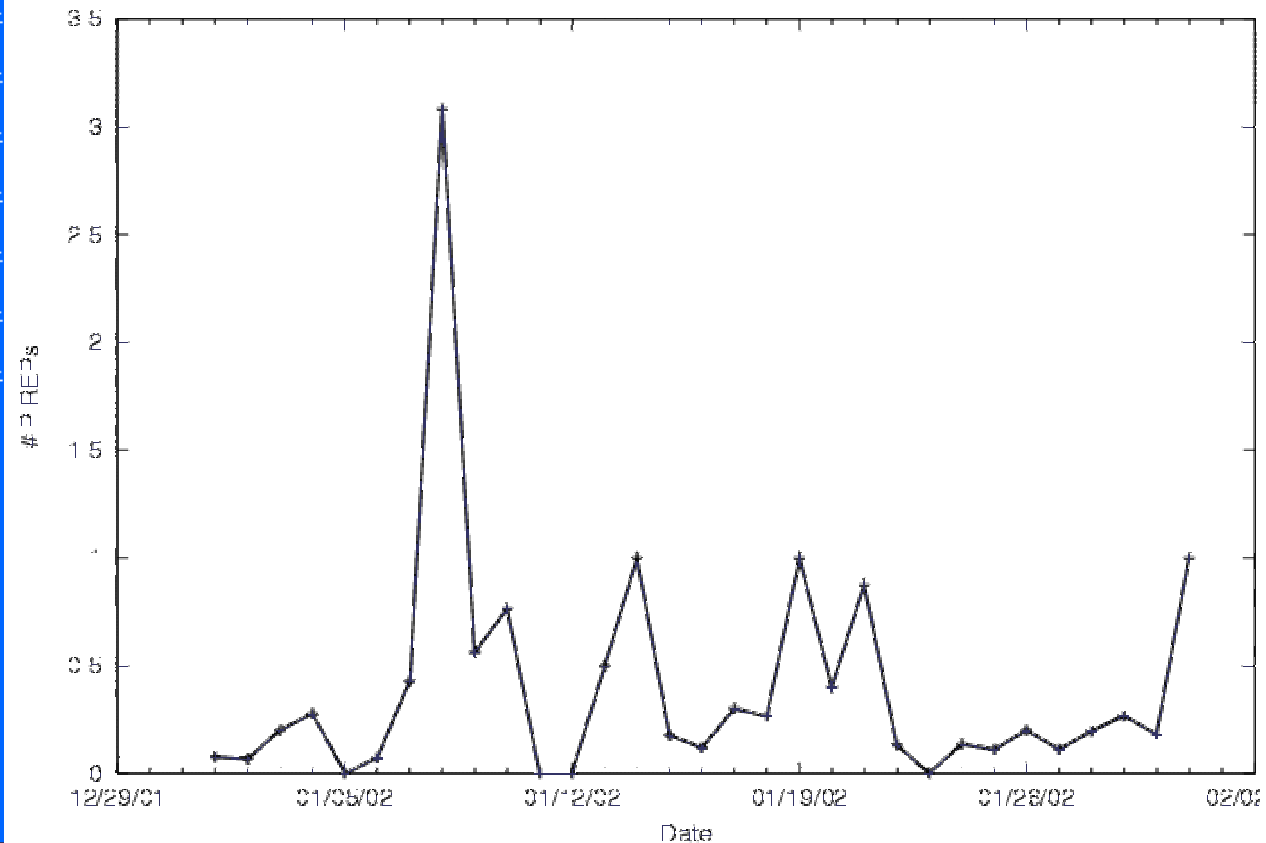
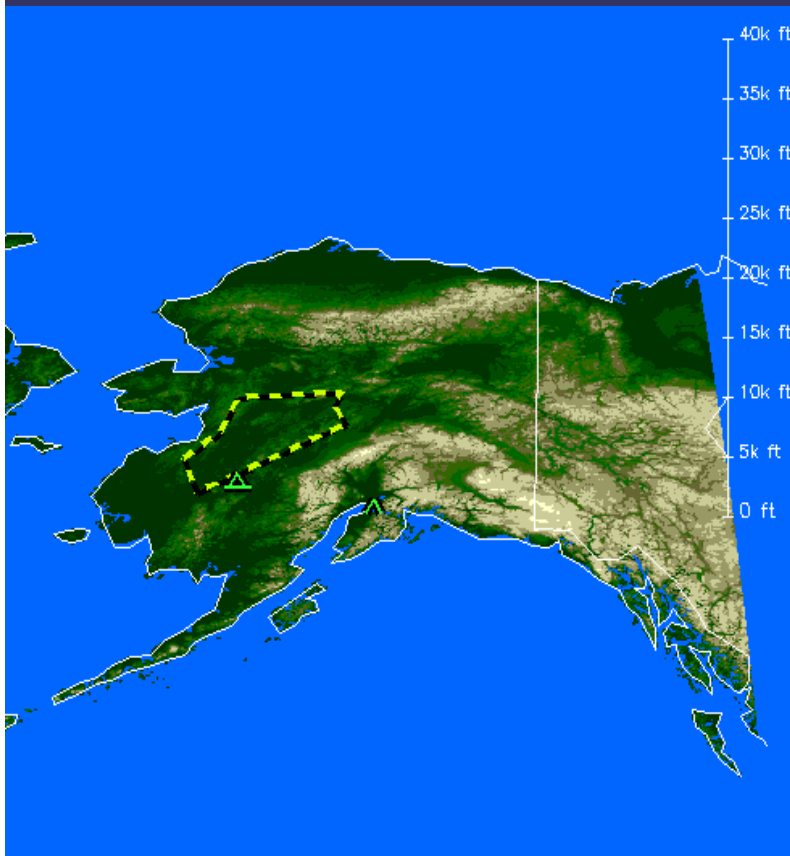
What's killing us? The FAR

Dead space is a fact of life and we've got to deal with that somehow

Surely you can't be serious...

AIRMETs for Alaska

Avg. PIREPs per AIRMET -Jan. '02



Average of 25 AIRMETs/day during this period

I am serious and don't call me Shirley

What we need is a method of evaluation that shows the practical performance of forecasts

In rare-event forecasting, absolute ranges of performance are essentially irrelevant and are prone to misinterpretation

Telling a forecaster that their CSI is 0.04 is not going to help them improve their forecasts!

Areal forecasts

SPC (severe storms), AWC (icing, turbulence)

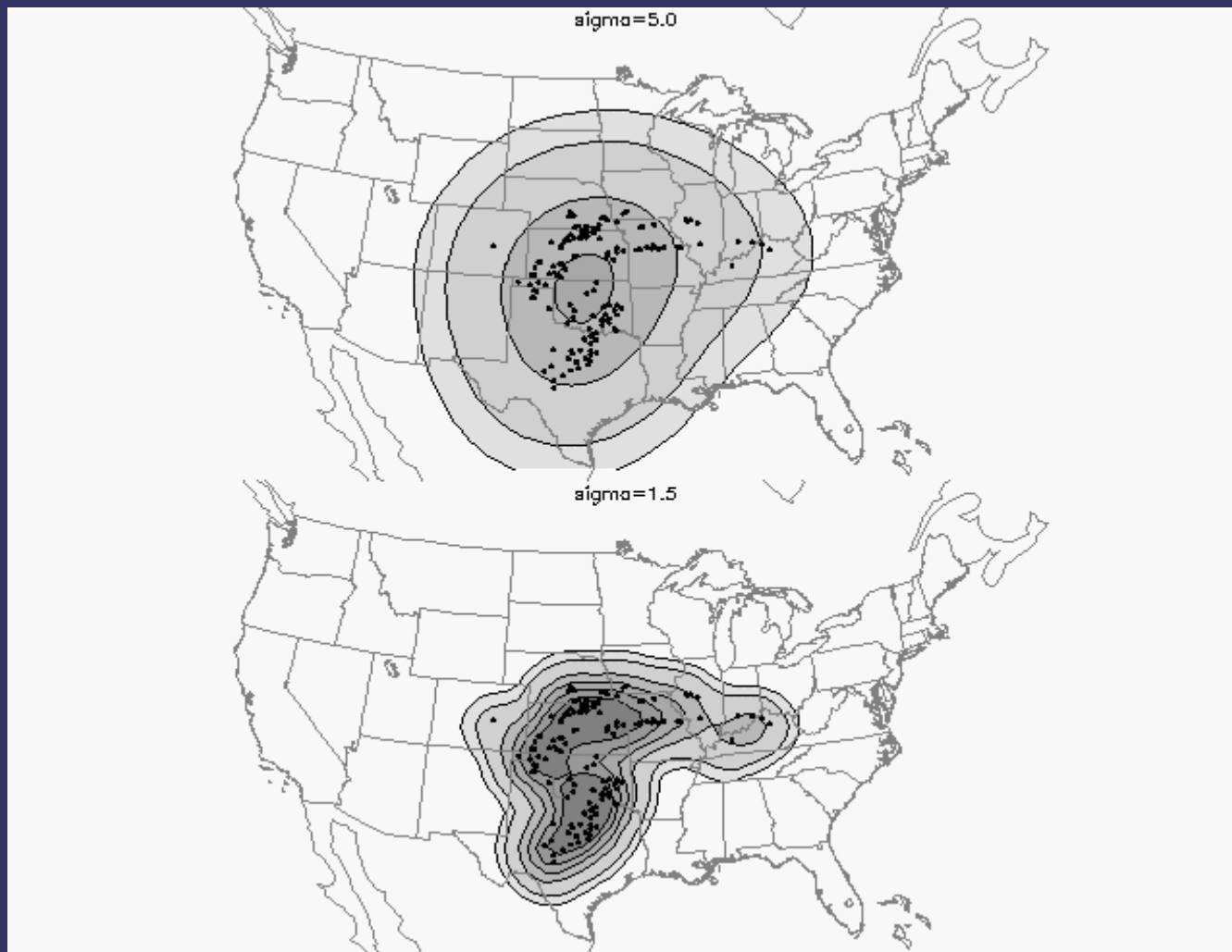
The practically-perfect method

Let's generate an artificial forecast (hindcast) using the “events” that represents the best-possible forecast that a person could make if they had perfect knowledge of the event locations ahead of time (yet still maintain the scale of the forecast!)

Spatially smooth “events” using non-parametric density estimation (similar to objective analysis)
events are either 1 (occur) or 0 (did not occur)

Smoothness/scale of the forecast is controlled by a confidence factor that represents forecaster uncertainty

The smoothing parameter



Larger smoother = greater uncertainty

What can we do with them?

These forecasts can be used for a number of interesting purposes:

1. Provide objective limits to forecast skill from situation to situation so that relative performance can be assessed
 - This works for both dichotomous and probabilistic forecasts
2. Provide a visual aid for **subjective** forecast evaluation!!!! (a BIG hit with the forecasters who work **visually** and not **statistically**)

Let's simulate reality and look at forecast skill

The practical range of forecast skill for rare events is not $[0, 1]$ (e.g., Superoutbreak example presented earlier).

Let's consider a set of simulations which looks at:

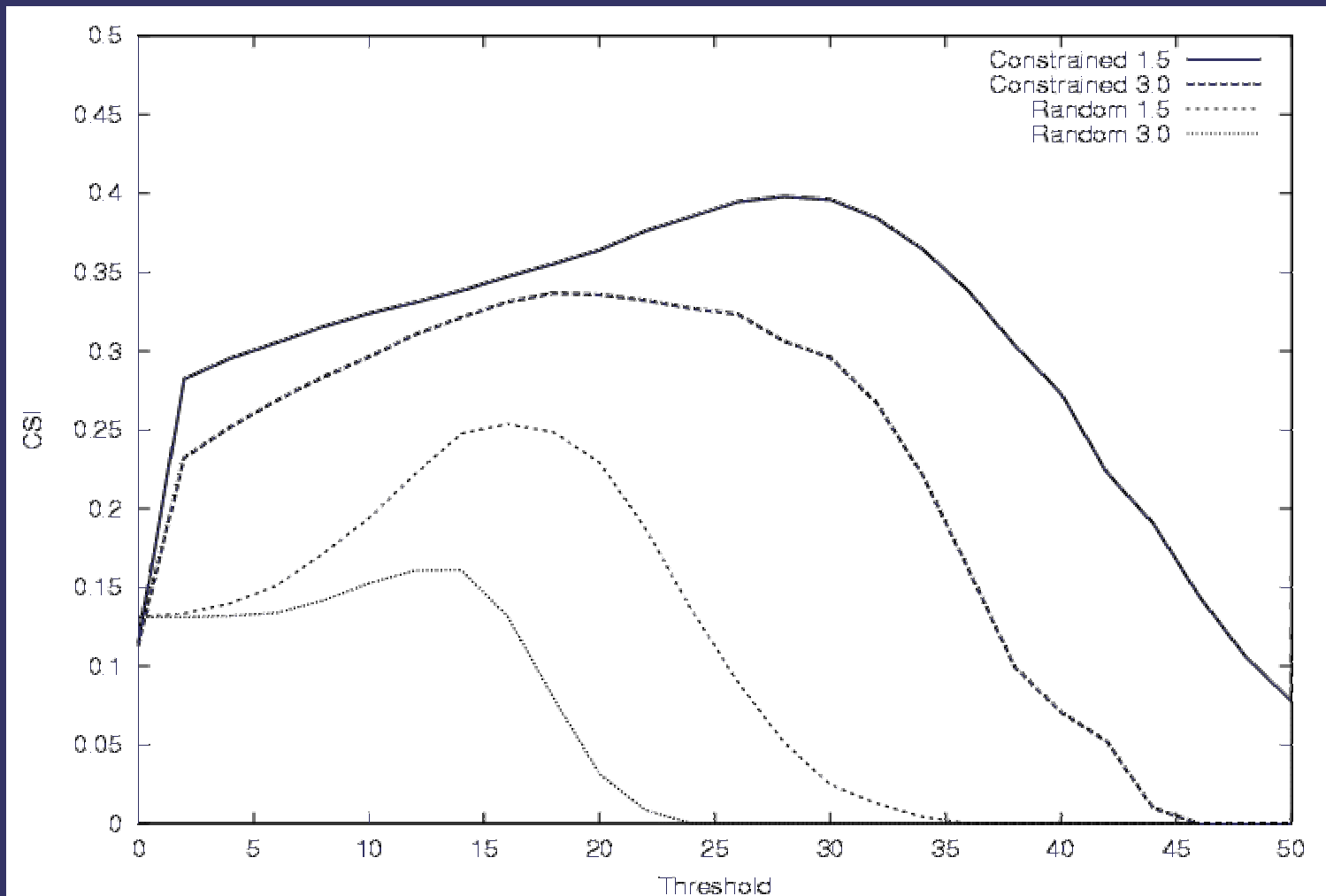
Event density: dense or sparse coverage

Event location: random or constrained

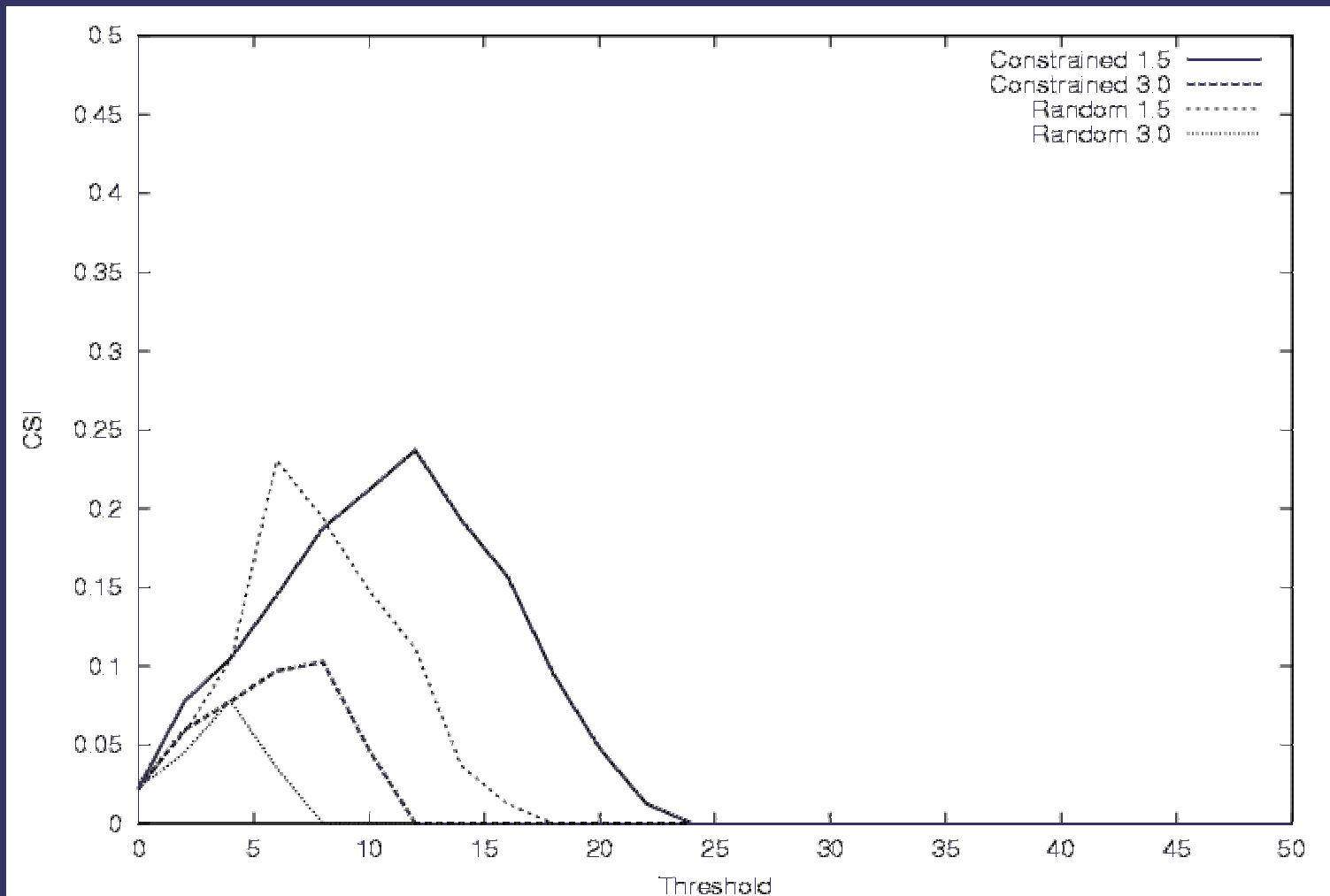
250 simulations of each situation

How does forecast skill vary in this set of situations?

Range of CSI for dense coverage



Range of CSI for sparse coverage



It follows your intuition

When a lot of reports are expected

Absolute performance measures will increase

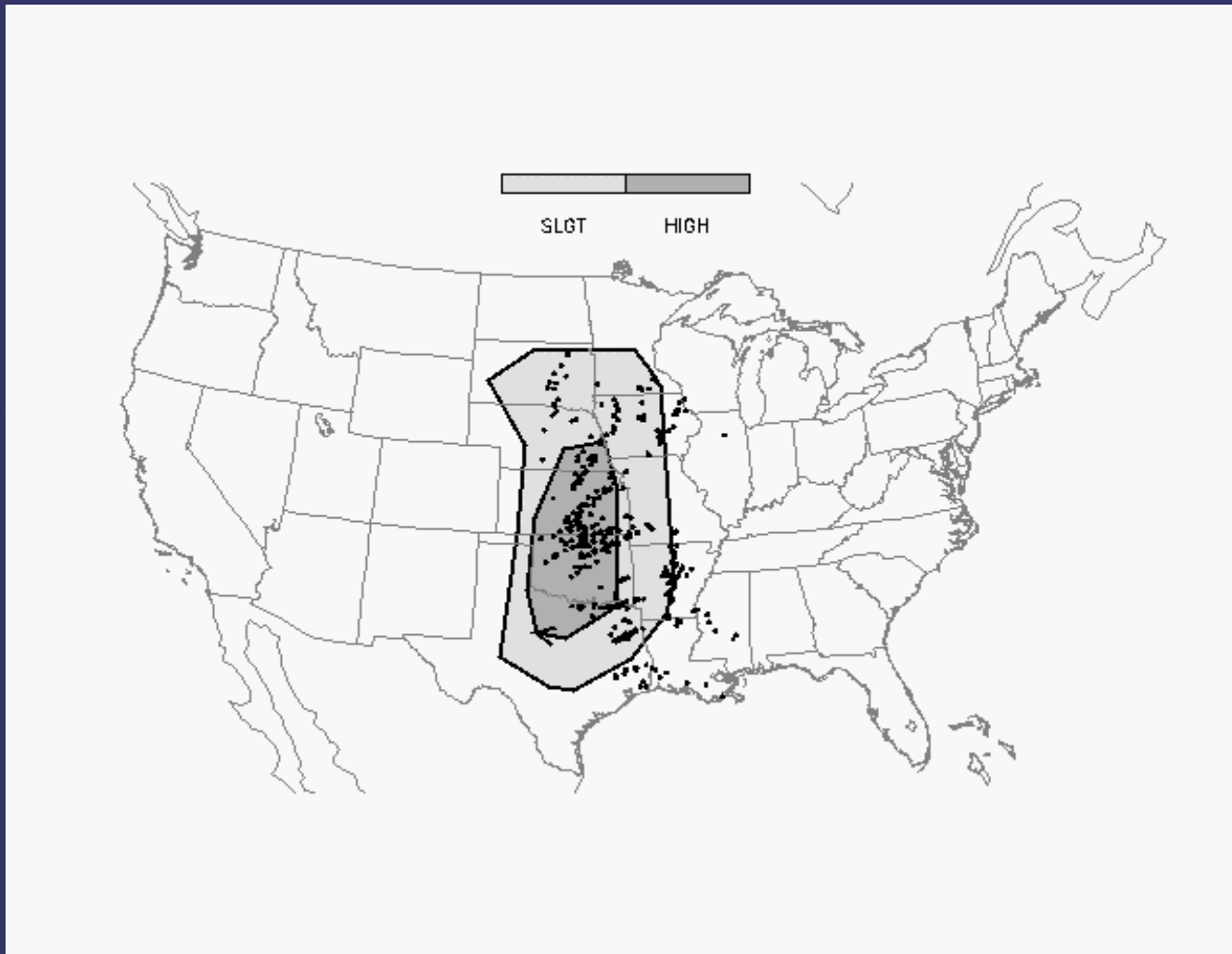
Accurate forecasts are possible at long lead times if you have an idea of event location

When sparse coverage is expected

Scores are lower

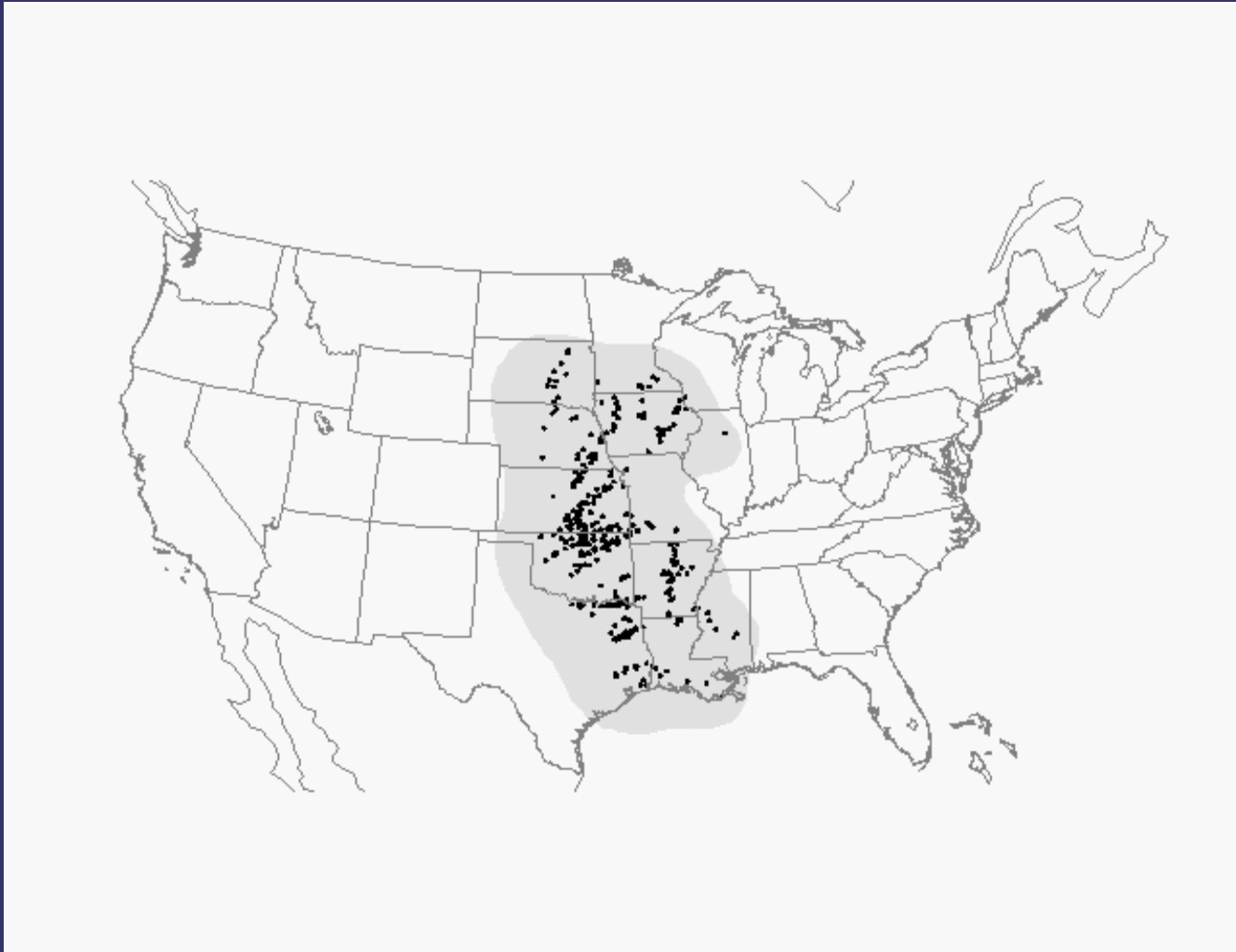
Shorter lead time/higher confidence doesn't help the scores very much

Dichotomous Forecast



CSI=0.34 on an 80km grid

Practically-perfect Forecast



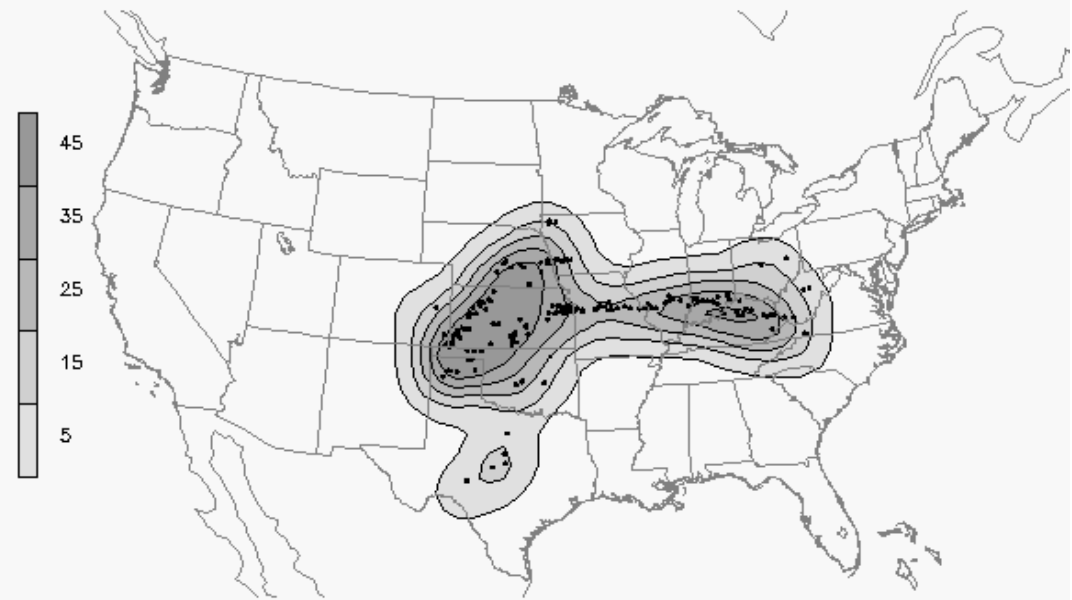
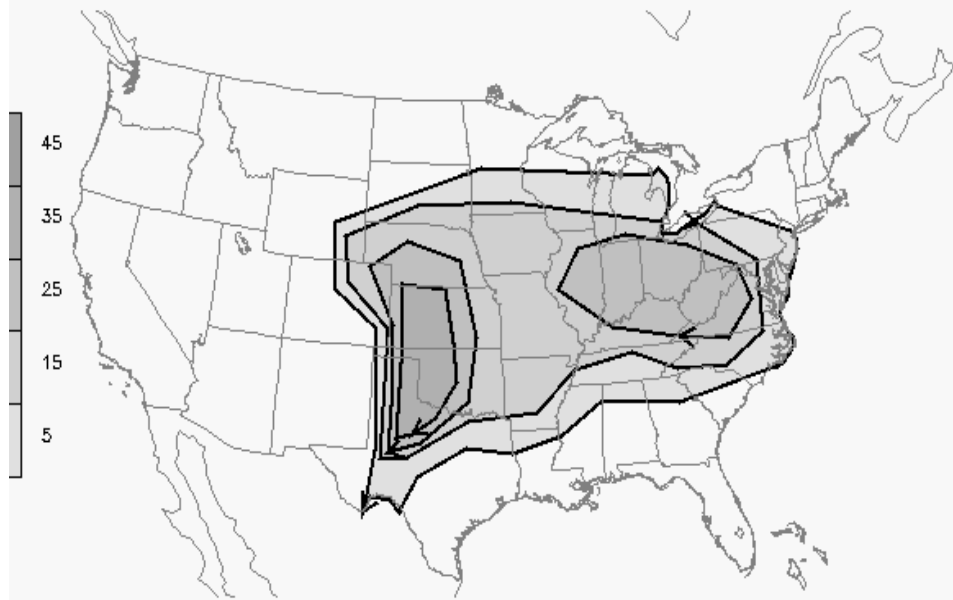
Min. CSI=0.05 Max. CSI=0.48

The Conv. Outlook can be considered 78% of the way to perfect

Probabilistic Forecasts

Actual Forecast

Practically-perfect Forecast



Probabilistic P. P. Forecasts

Since our practically-perfect forecast can be viewed as probabilities of events occurring, we can treat it as a probabilistic forecast and compute anything we'd want (Brier Score, etc.)

We can use this as a “better” reference forecast than climatology (esp. for rare events!)

Often you will have negative skill scores because the practically-perfect forecast is so “unbeatable”.

Conclusions

Forecasts will improve when we can bring verification methods to the forecasters that they can relate to.

Practically-perfect forecasts represent a step in that direction.

They provide a tool for assessing relative performance

They also represent a visual tool for subjective verification. This is more important than you might think