

# Verification of Ensemble Forecasts - A Survey

Laurence J. Wilson  
Meteorological Service of Canada  
Montreal, Quebec

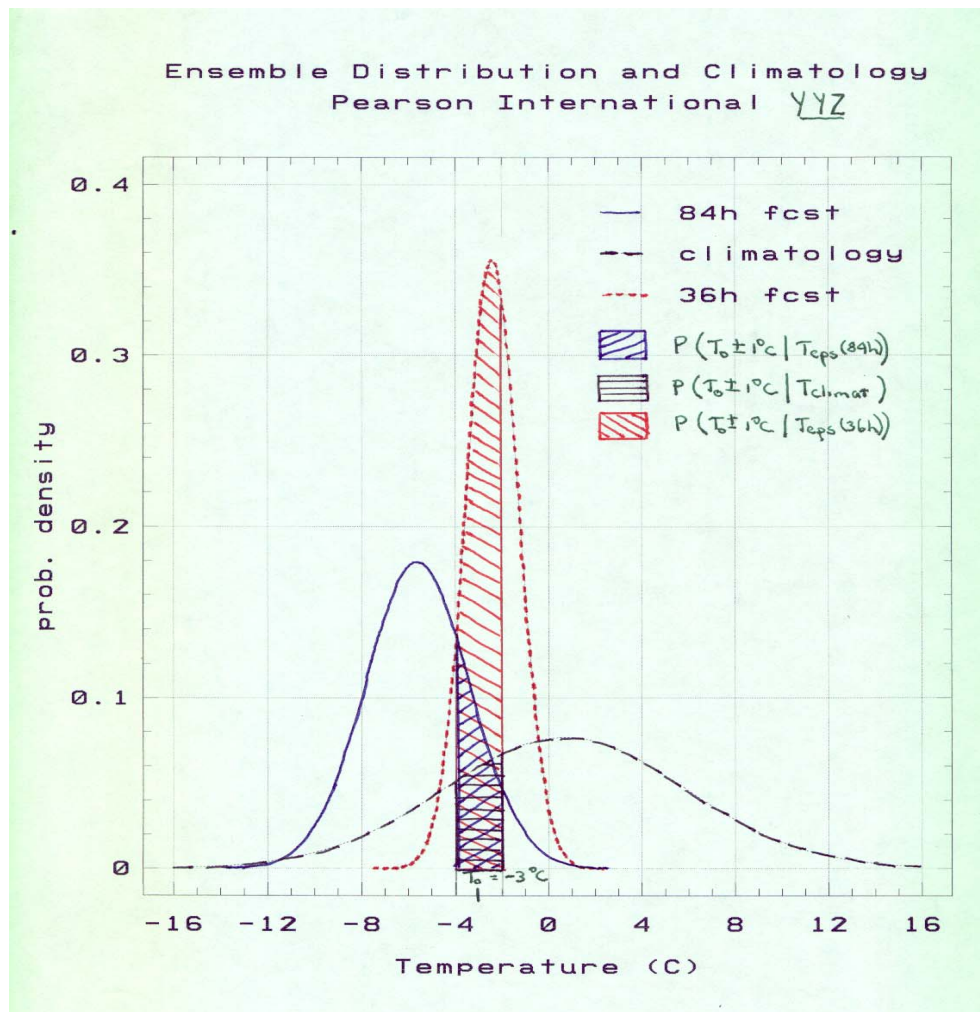
# Outline

- **The ensemble verification problem**
  - Attributes applied to the ensemble distribution
- **Verification of the ensemble distribution**
  - Wilson 1999
  - RPS and CRPS
  - Rank Histogram
- **Verification of individual ensemble members**
- **Verification of probability forecasts from the ensemble**
  - Reliability tables
  - The ROC

## Verification of the ensemble

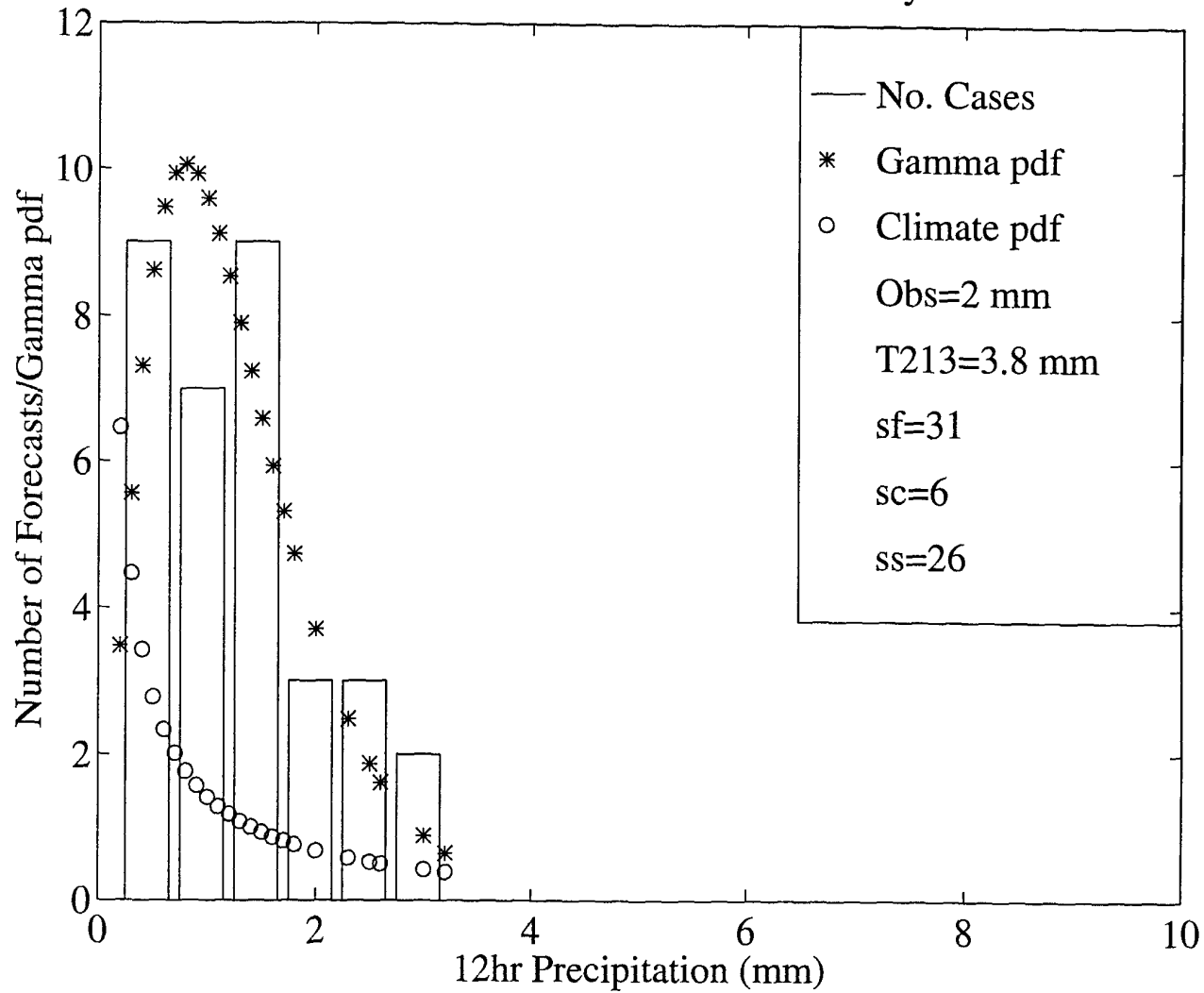
- Problem:
  - how to compare a distribution with an observation
- The concept of “consistency”:
  - For each possible probability distribution  $f$ , the a posteriori verifying observations are distributed according to  $f$  in those circumstances when the system predicts the distribution  $f$ . (Talagrand)
  - similar to reliability
- The concept of “non-triviality”
  - the eps must predict different distributions at different times

# Strategy for ensemble verification



# Ensemble verification - distribution

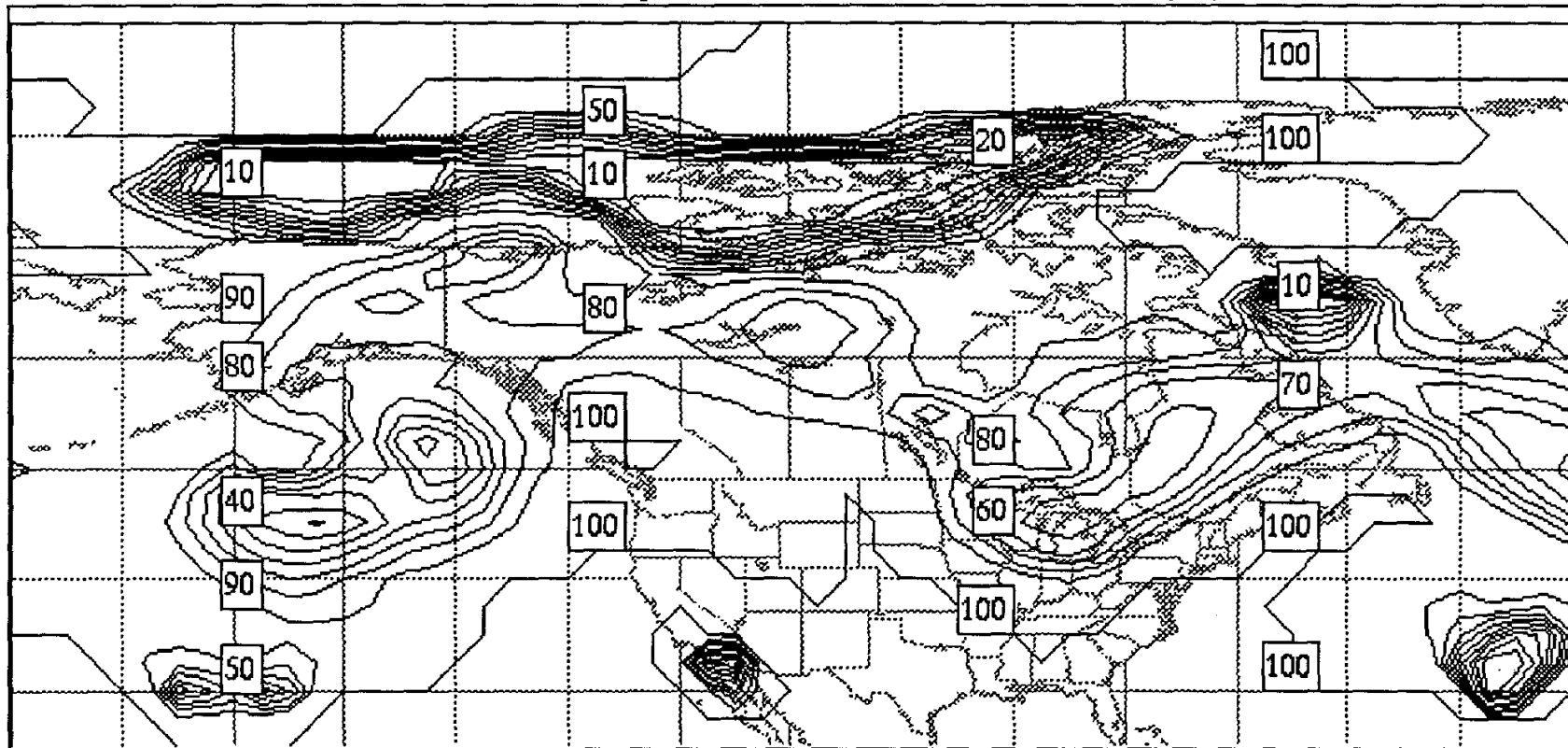
YVR: 72-hr Ensemble Forecasts Valid May 23 1996



# Ensemble verification - 500 mb

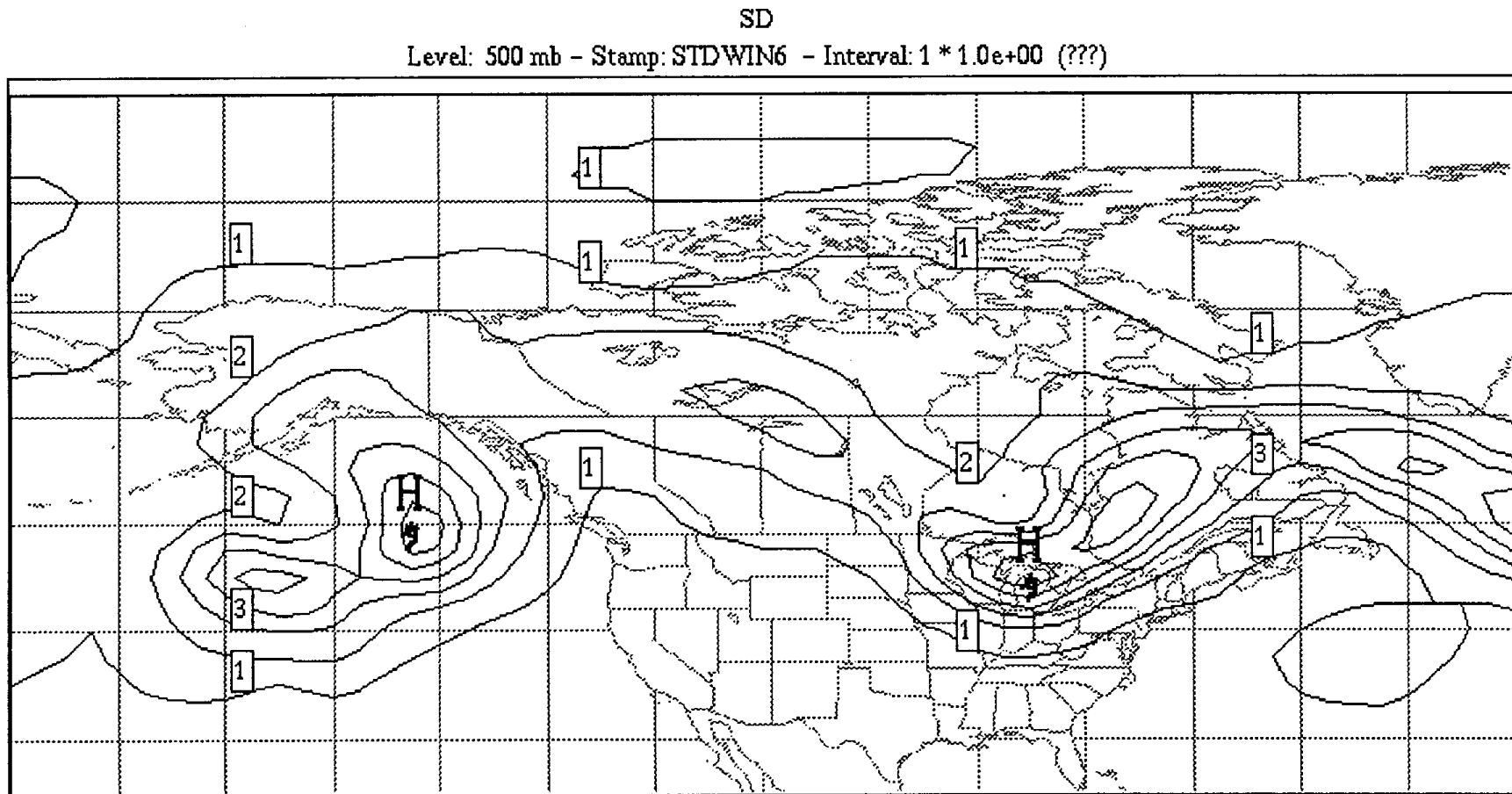
SK

Level: 500 mb - Stamp: SCORWIN4 - Interval:  $10 * 1.0e-02$  (???)



36 hour fest valid 00:00Z October 03 1996

# Ensemble verification - 500 mb



36 hour fcst valid 00:00Z October 03 1996

## Comments on “Wilson” score

- Sensitive both to “nearness” of the ensemble mean and to ensemble spread
- Verifies the distribution only in the vicinity of the observation; variations outside the window have no impact
- Believed to be strictly proper - shown empirically
- Related to Brier Score for a single forecast

$$Sc = 1 - \sqrt{BS}$$

- Can account for forecast “difficulty” by choosing window based on climatological variance

# Verification of approximations to the eps distribution

- The Rank probability score (RPS)

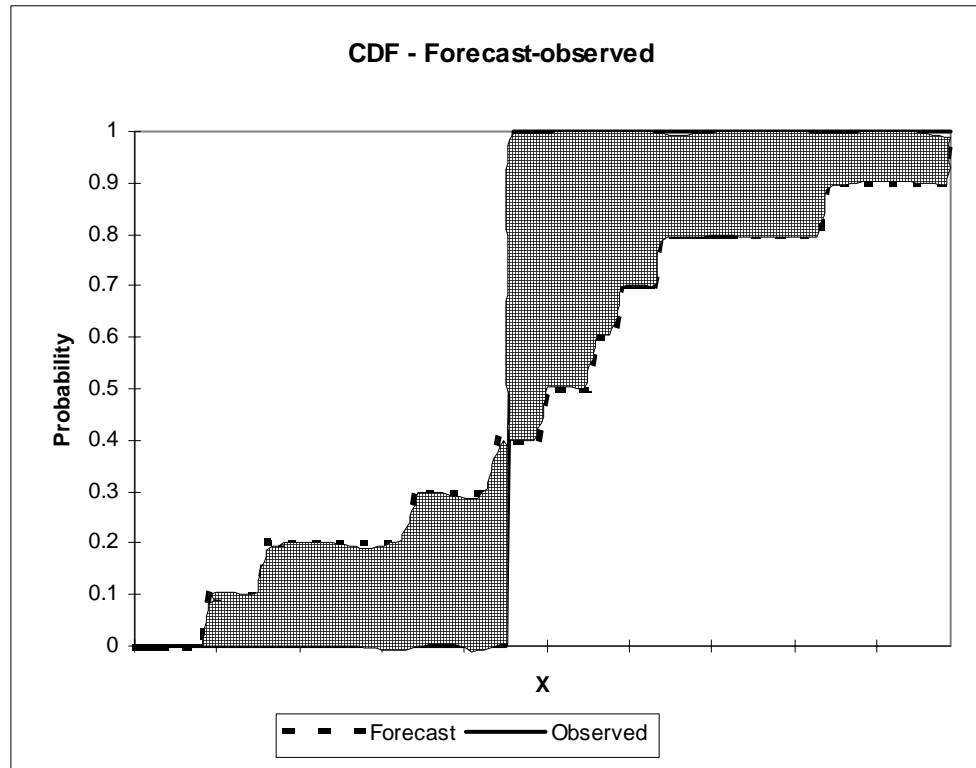
$$RPS = \frac{1}{K-1} \left[ \sum_{i=1}^K \left( \sum_{n=1}^i P_n - \sum_{n=1}^i O_n \right)^2 \right]$$

– discrete form, choose categories; samples distribution according to categories

- Continuous RPS

$$CRPS (P, x_a) = \int_{-\infty}^{\infty} [P(x) - P_a(x)]^2 dx$$

# CRPS example



# Rank Histogram (Talagrand Diagram)

- Preparation

- order the members of the ensemble from lowest to highest - identifies  $n+1$  ranges including the two extremes
- identify the location of the observation, tally over a large number of cases

- Interpretation

- Flat indicates ensemble spread about right to represent uncertainty
- U-shaped - ensemble spread too small
- dome-shaped - ensemble spread too large
- asymmetric - over- or under-forecasting bias
- **This is NOT a true verification measure**

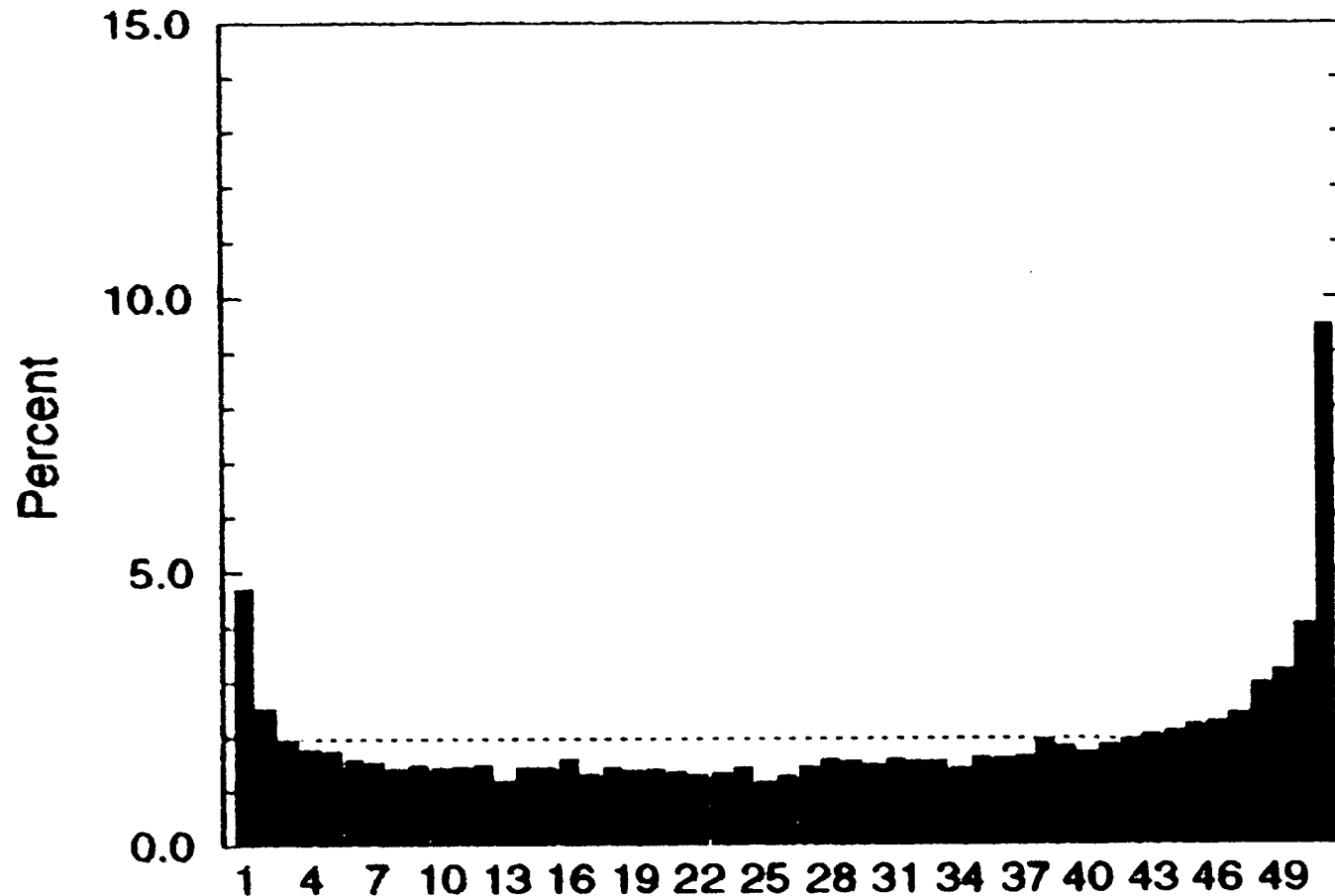


# Rank Histogram

50 ens. members, sample size 22032

Real ensemble, T 850, D+6

Delta: 3952146 Expected: 21600 (182.97)



## Verification of individual members

- Preferred for comparison with operational model than verification of ensemble mean
- Unperturbed control
  - compare with full resolution model
- Best and worst member
  - a “posteriori” verification - less use to forecasters
  - select over a forecast range or individually at each range
- Methods
  - all that apply to continuous fields: RMSE, MAE, bias, anomaly correlation etc.
  - preferable to verify against data than analysis.

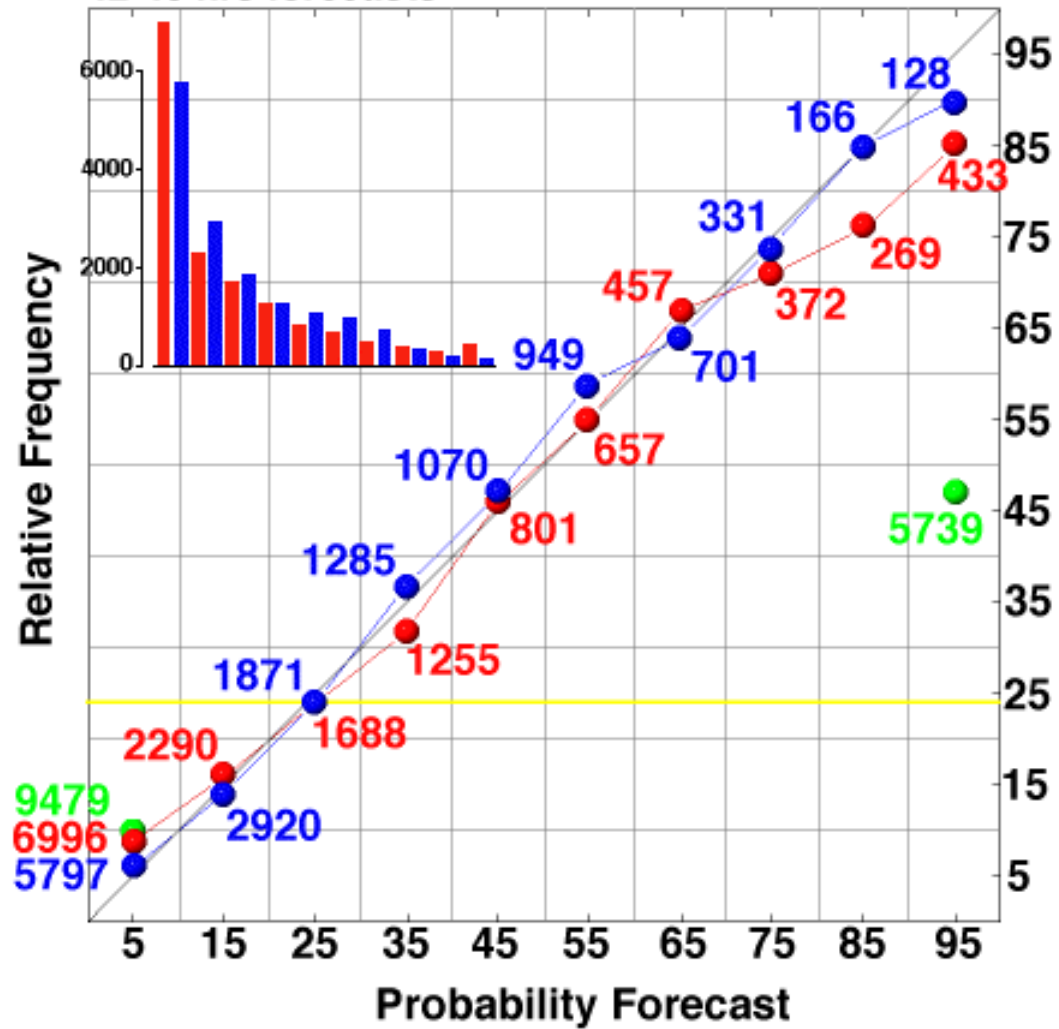
# The Ensemble mean

- Popular, because scores well with quadratic rules
- Should NOT be compared to individual outcomes:
  - different sampling distribution
  - not a trajectory of the model

# Verification of probability forecasts from the Ensemble

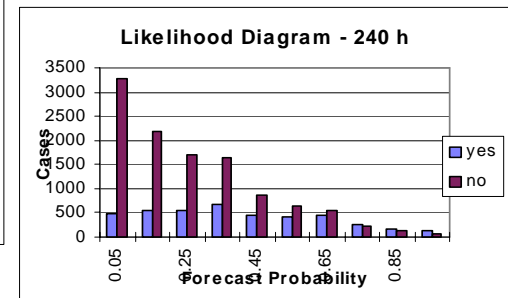
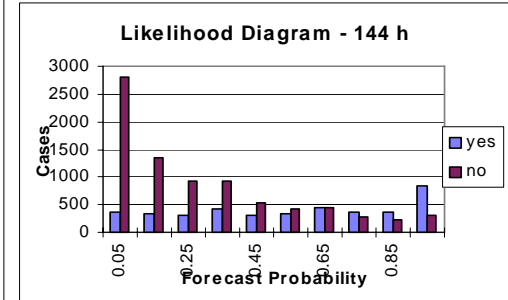
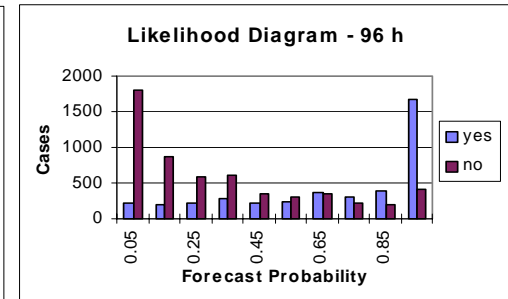
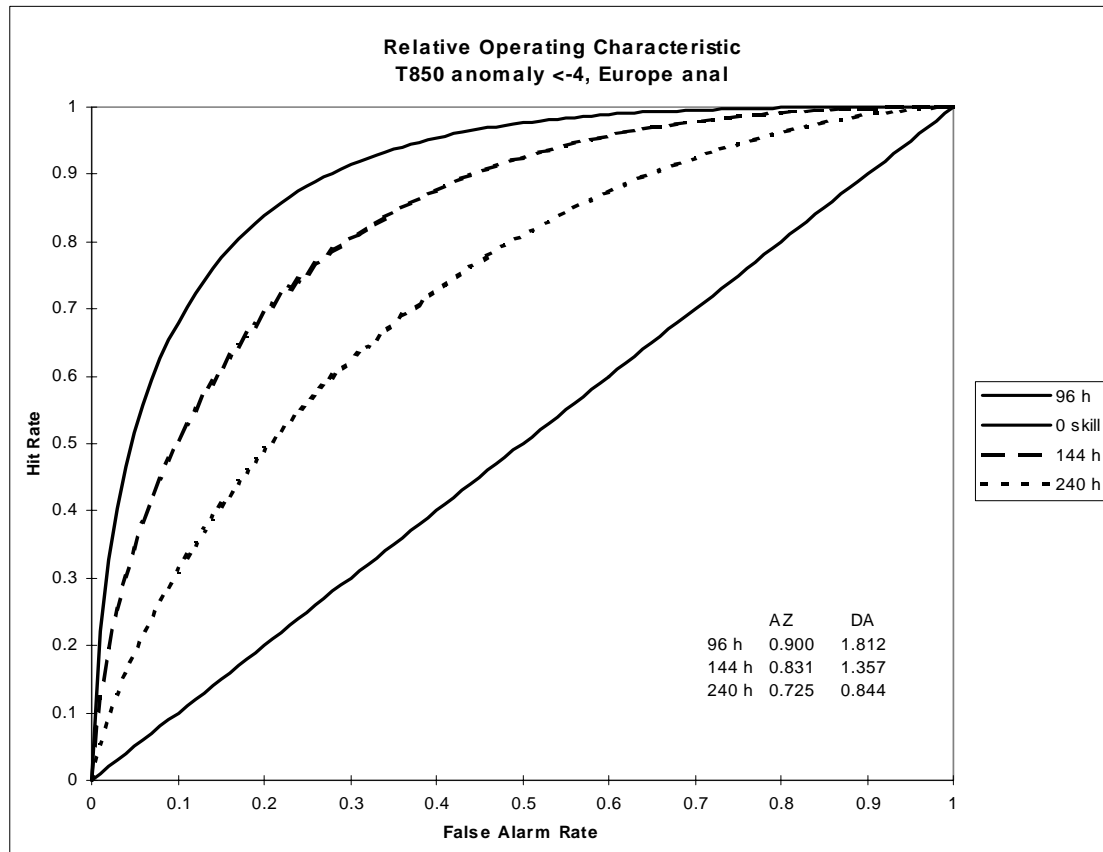
- Same as verification of any probability forecasts
- **Reliability Table (with unconditional distribution of forecasts) + ROC (with likelihood diagram) sufficient for complete diagnostic verification**
- Reliability table: Distribution conditioned by fcst
- ROC: Distribution conditioned by obs.
- Attributes:
  - reliability
  - sharpness
  - resolution
  - discrimination

**Reliability Table**  
42-48 hrs forecasts



# ROC - ECMWF Ensemble Forecasts

## Temperature 850 mb anomaly <-4C (vs. analysis)

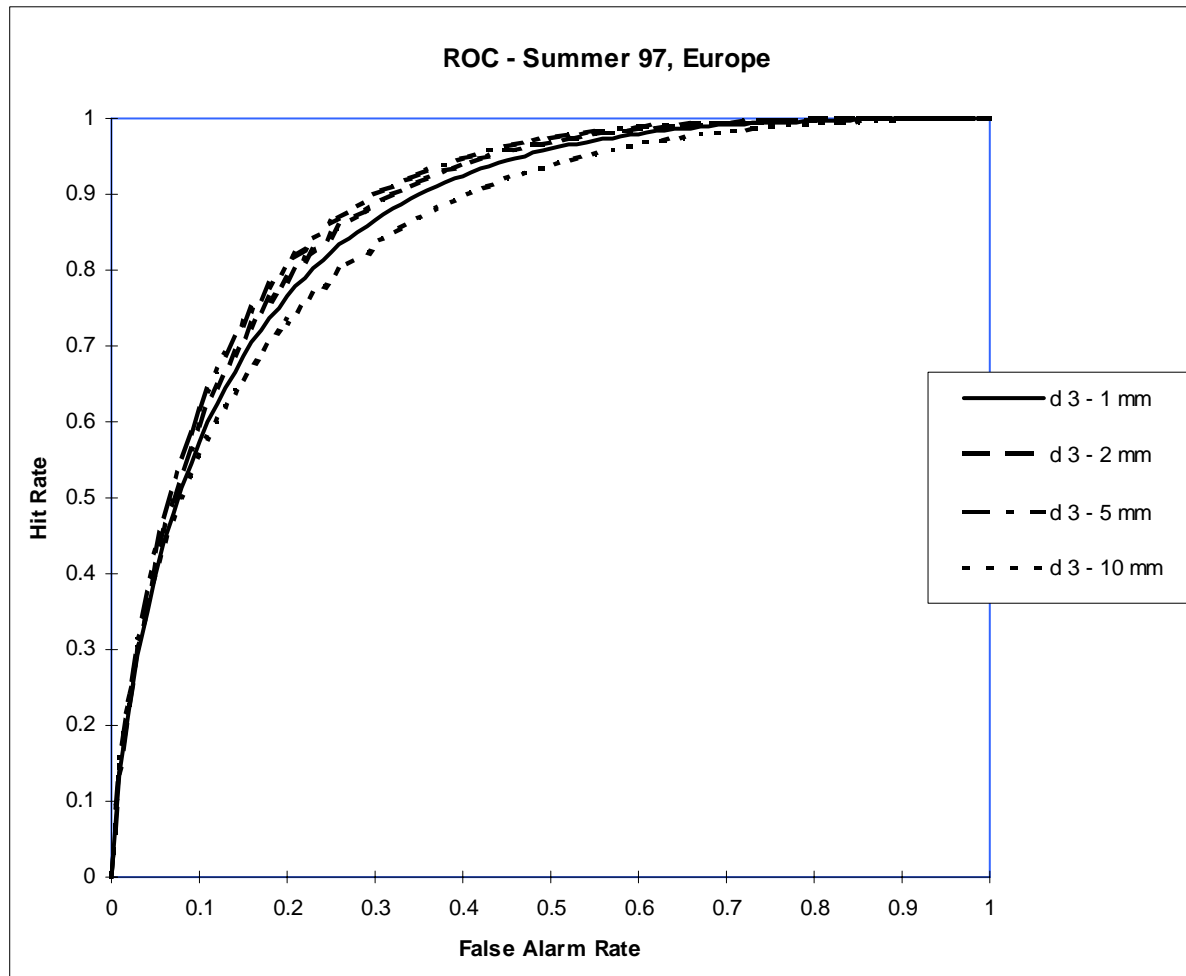


# ROC Issues

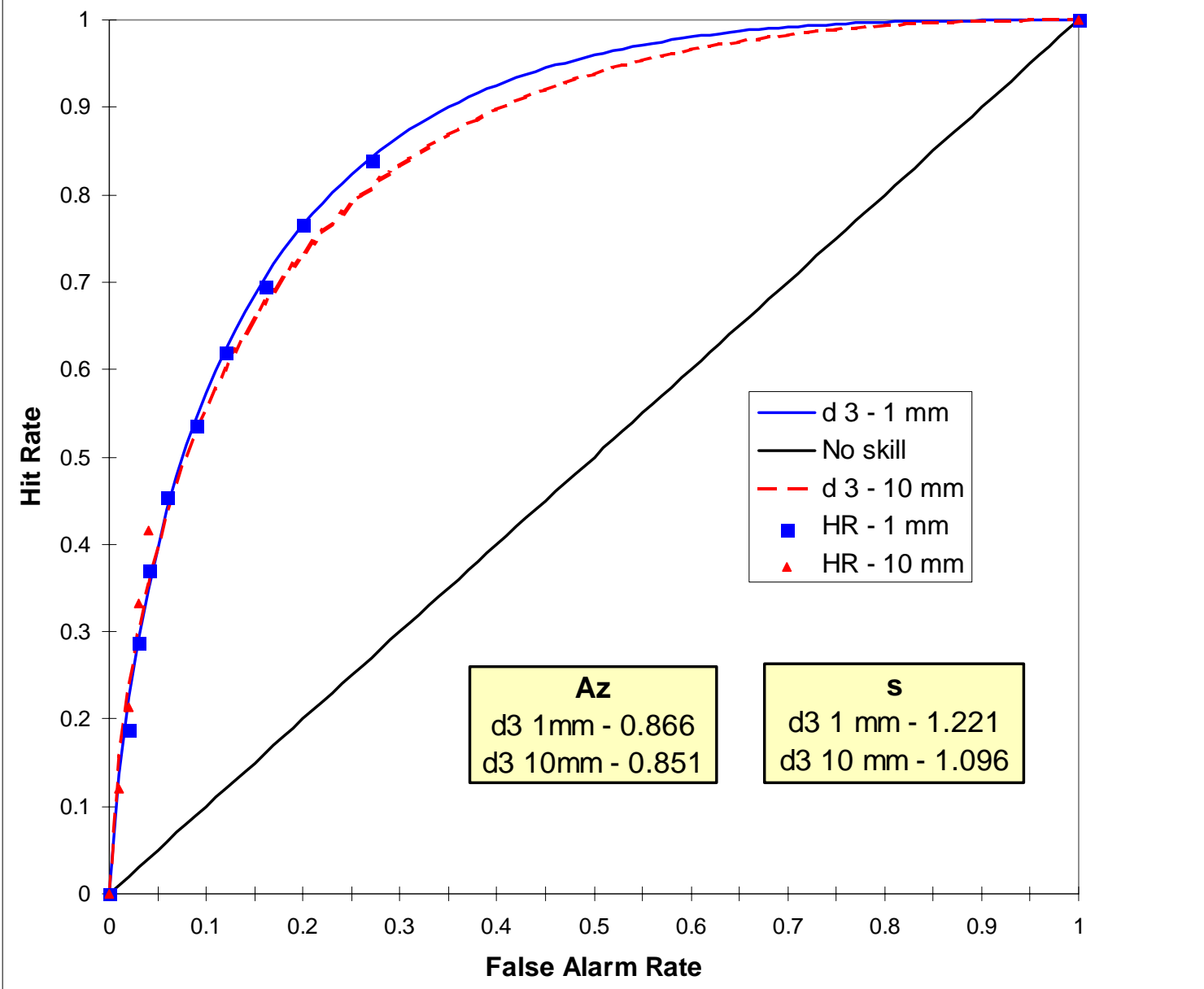
- Empirical vs. fitted
- No. points needed to define the ROC
- ROC and value (“potential value”)

# ROC - threshold variation

(Wilson, 2000)



### ROC - Summer 97 -Europe



## Summary

- Verification of the ensemble distribution - depends on how it is to be used by forecaster
- Two aspects: verification of distribution vs. verification of probabilities from the distribution
- Several measures shown, characteristics identified
- Sufficiency of Reliability table and ROC graph for diagnostic verification of probability forecasts