Probabilistic verification of ECMWF precipitation forecasts.

Anna Ghelli, ECMWF

Thanks to Ervin Zsoter and Francois Lalaurette

Summary

- **Probabilistic scores**
  - **Reliability diagram**
  - **Brier Skill Score**
  - **Relative Operating Curve (ROC)**
  - **ROC area**
- **EPS performance**
- **Conclusions**

Probabilistic Scores

Reliability diagram Plot the observed frequency against the forecast probability, where the range of forecast probabilities is divided into K bins (for example, 0-5%, 5-15%, 15-25%, etc.). The sample size in each bin is often included as a histogram or values beside the data points.

Answers the question: How well do the predicted probabilities of an event correspond to their observed frequencies?

Brier skill score

Answers the question: What is the relative skill of the probabilistic forecast over that of climatology, in terms of predicting whether or not an event occurred?

Range: minus infinity to 1, 0 indicates no skill when compared to the reference forecast. Perfect score: 1.

ROC curve Plot hit rate vs false alarm rate, using a set of increasing probability thresholds to make the yes/no decision.

Answers the question: What is the ability of the forecast to discriminate between events and non-events?

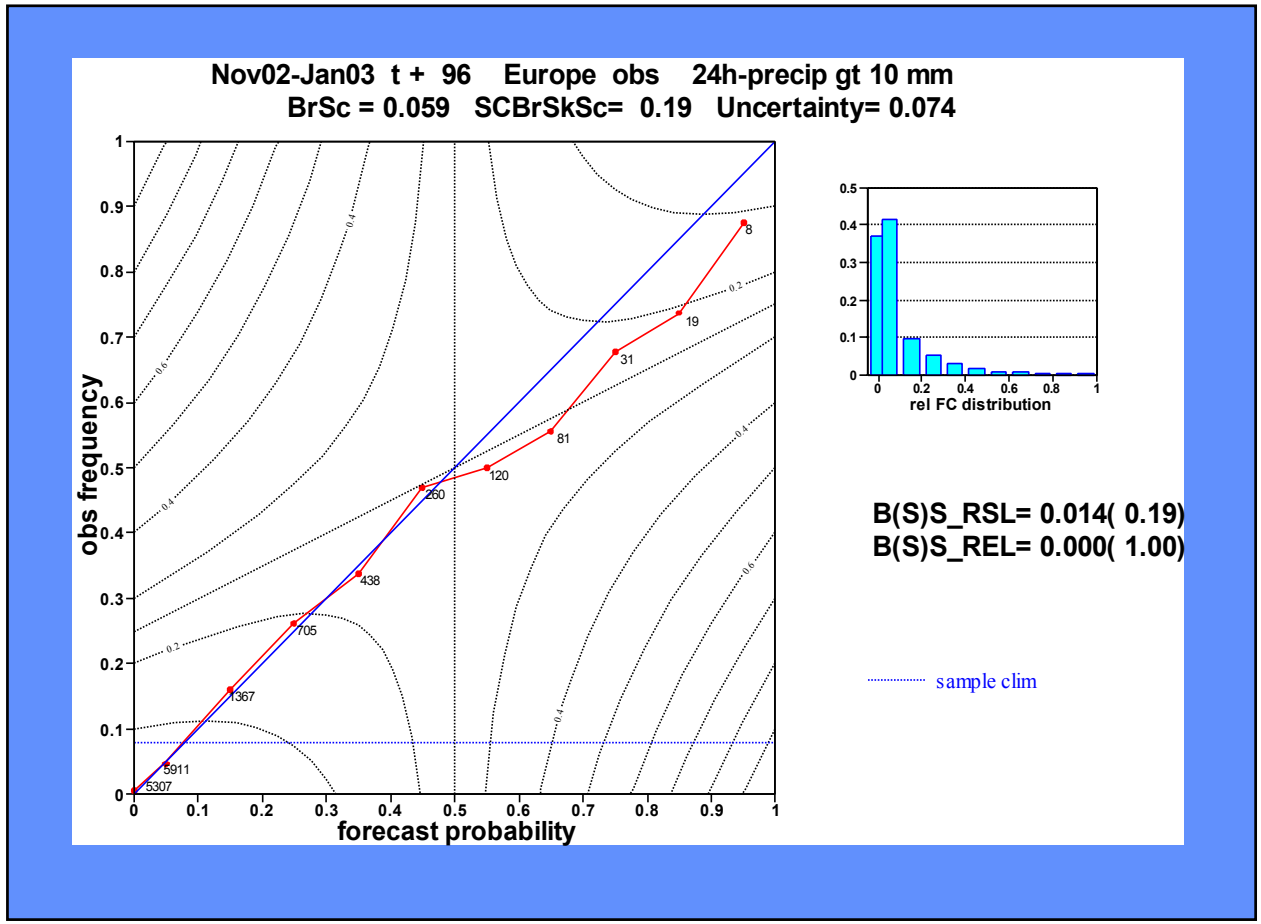
ROC area: Range: 0 to 1, 0.5 indicates no skill. Perfect score: 1

Reliability diagram for November 2002 to January 2003.

Area: Europe
Threshold: 10mm/24h

The low and medium observed frequency are well forecast.

The higher the observed frequency the more over confident the model is.

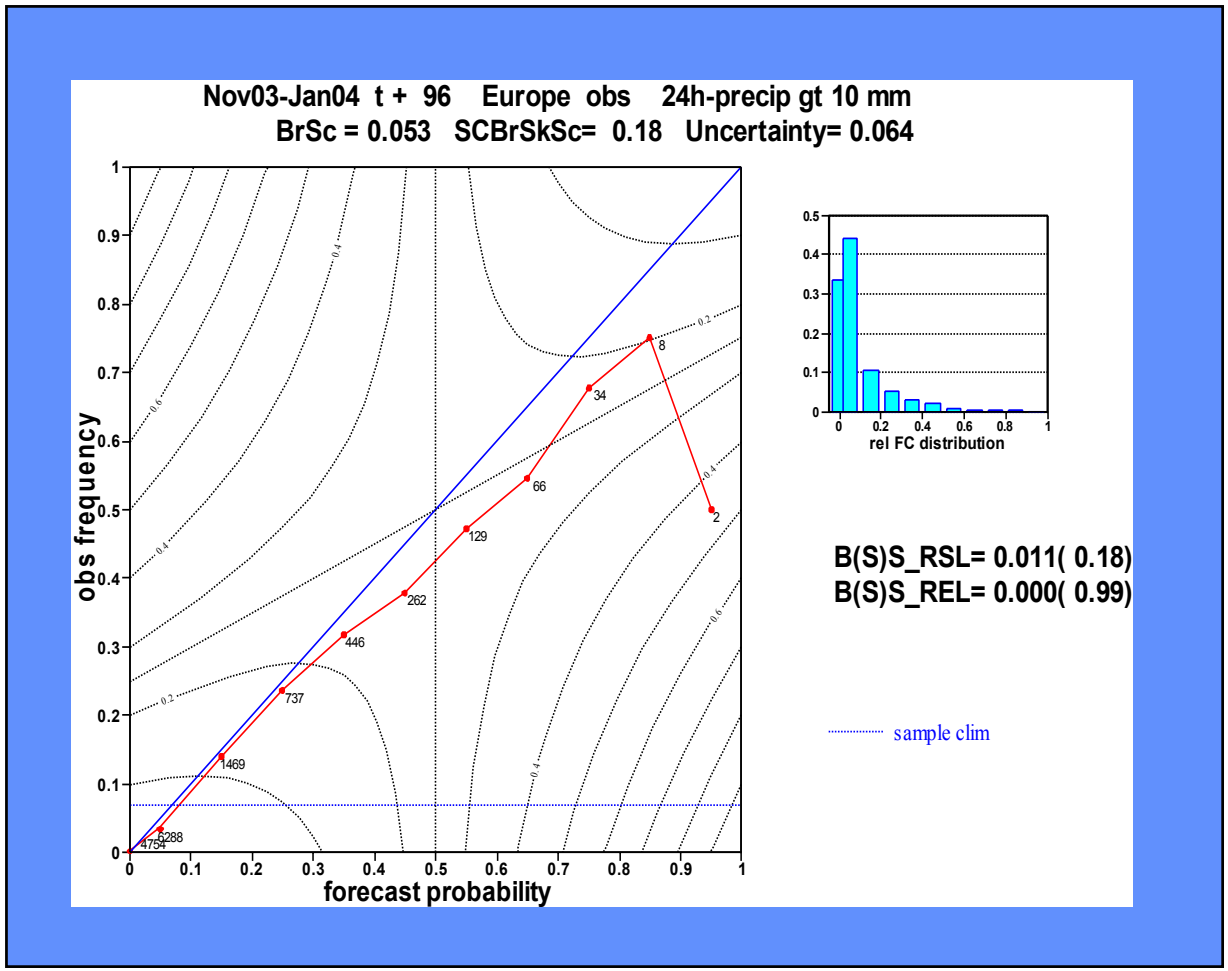


Reliability diagram for November 2003 to January 2004.

Area: Europe
Threshold: 10mm/24h

The low observed frequency is well forecast.

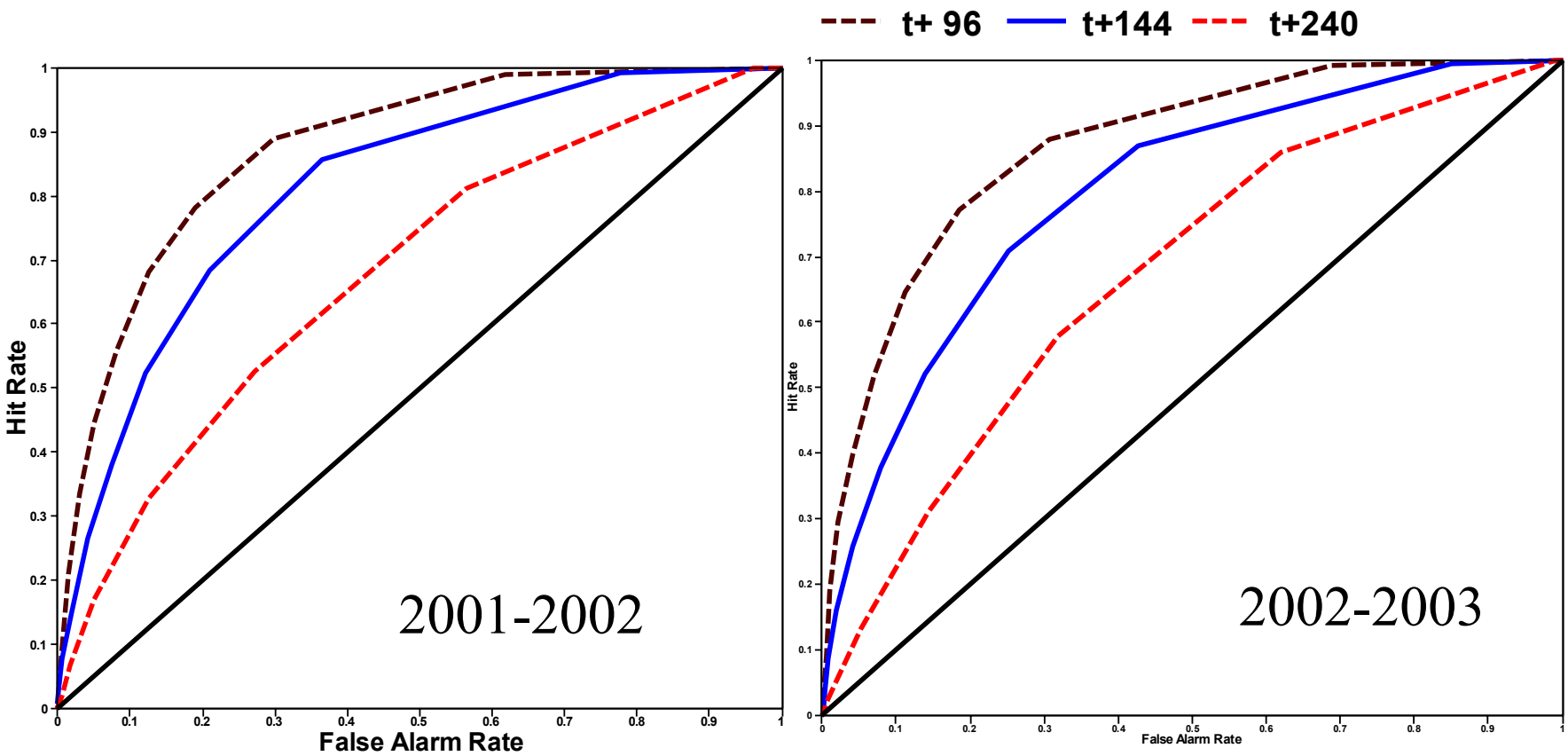
The higher the observed frequency the more over confident the model is.



ROC curves for winter periods

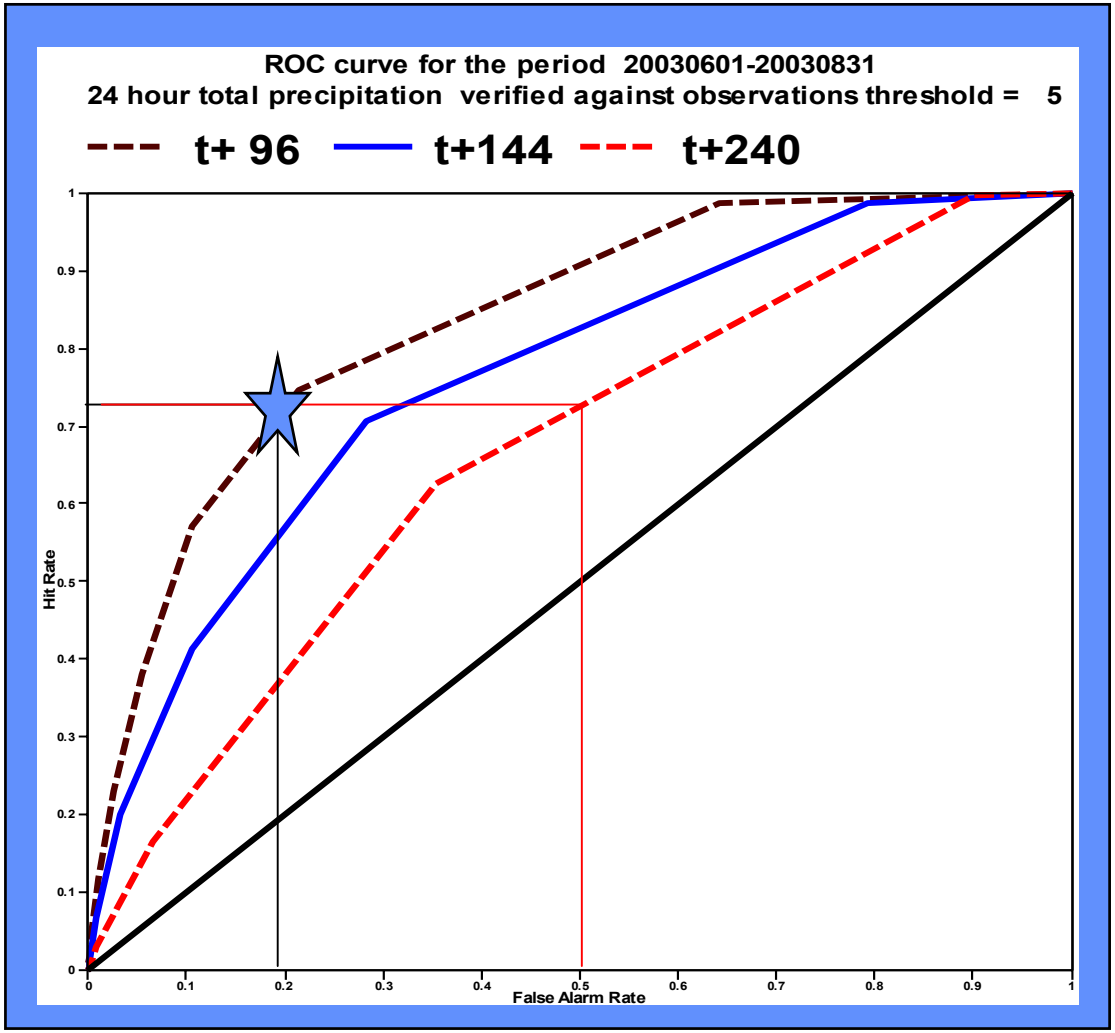
Area: Europe
Threshold: 5mm/24h

ROC measures the ability of the forecast to discriminate between two alternative outcomes
Diagonal → no skill



Area: Europe
Threshold: 5mm/24h

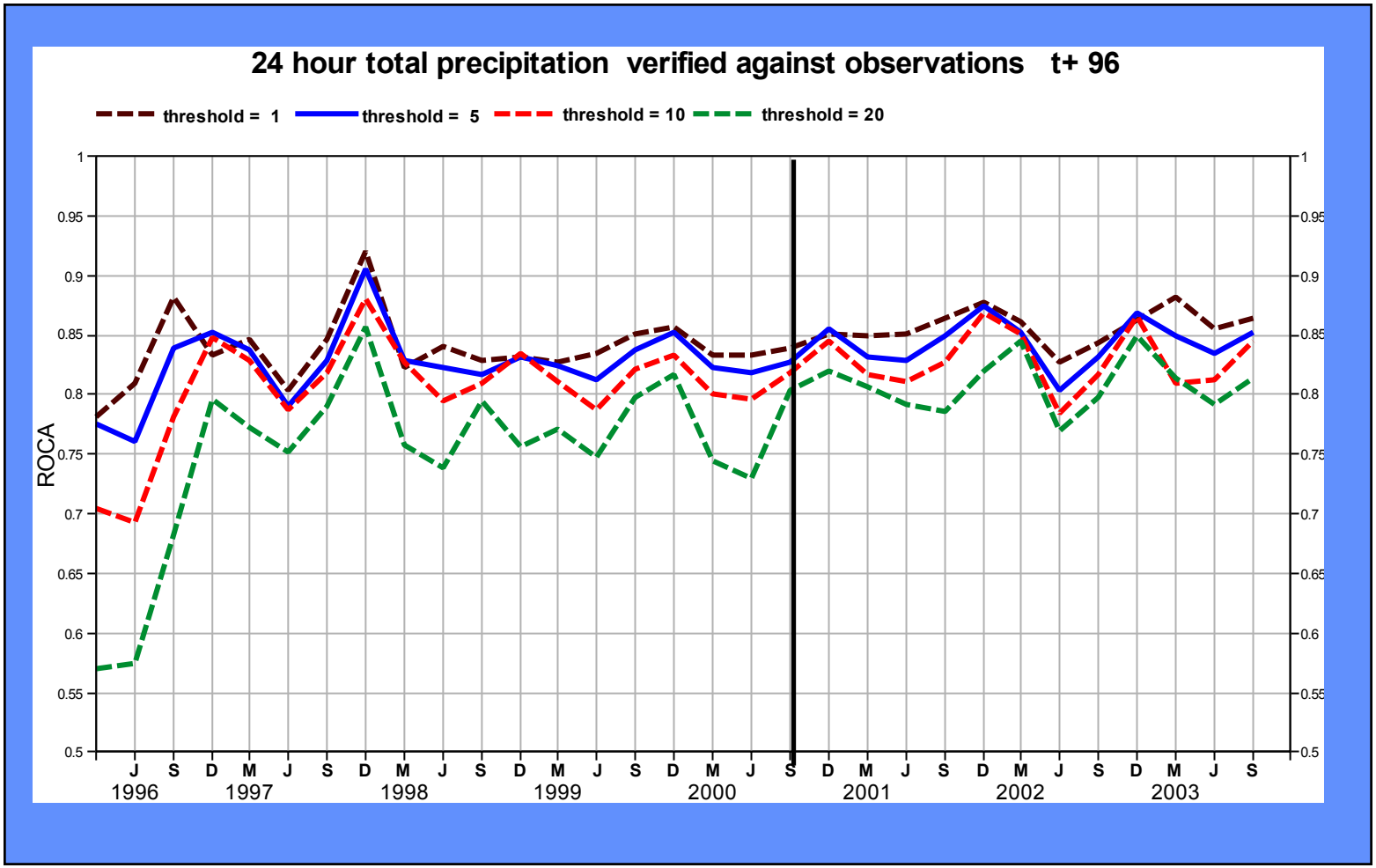
The star represents one deterministic forecast: ROC changes as lead time changes



ROC measures the ability of the forecast to discriminate between two alternative outcomes

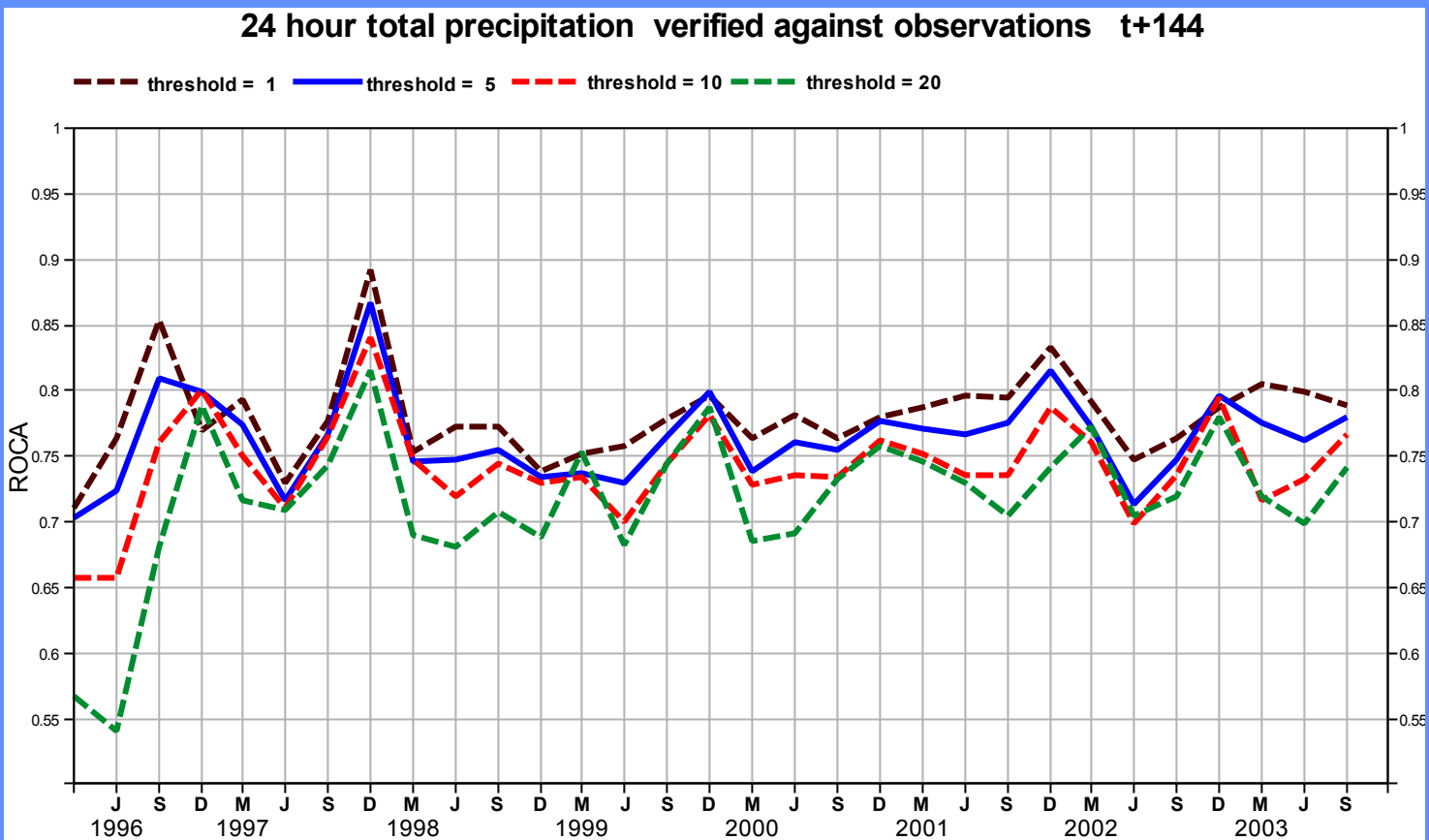
-- ROC curves for different lead times -- summer 2003

-- ROC Area for different thresholds – Forecast range D+4

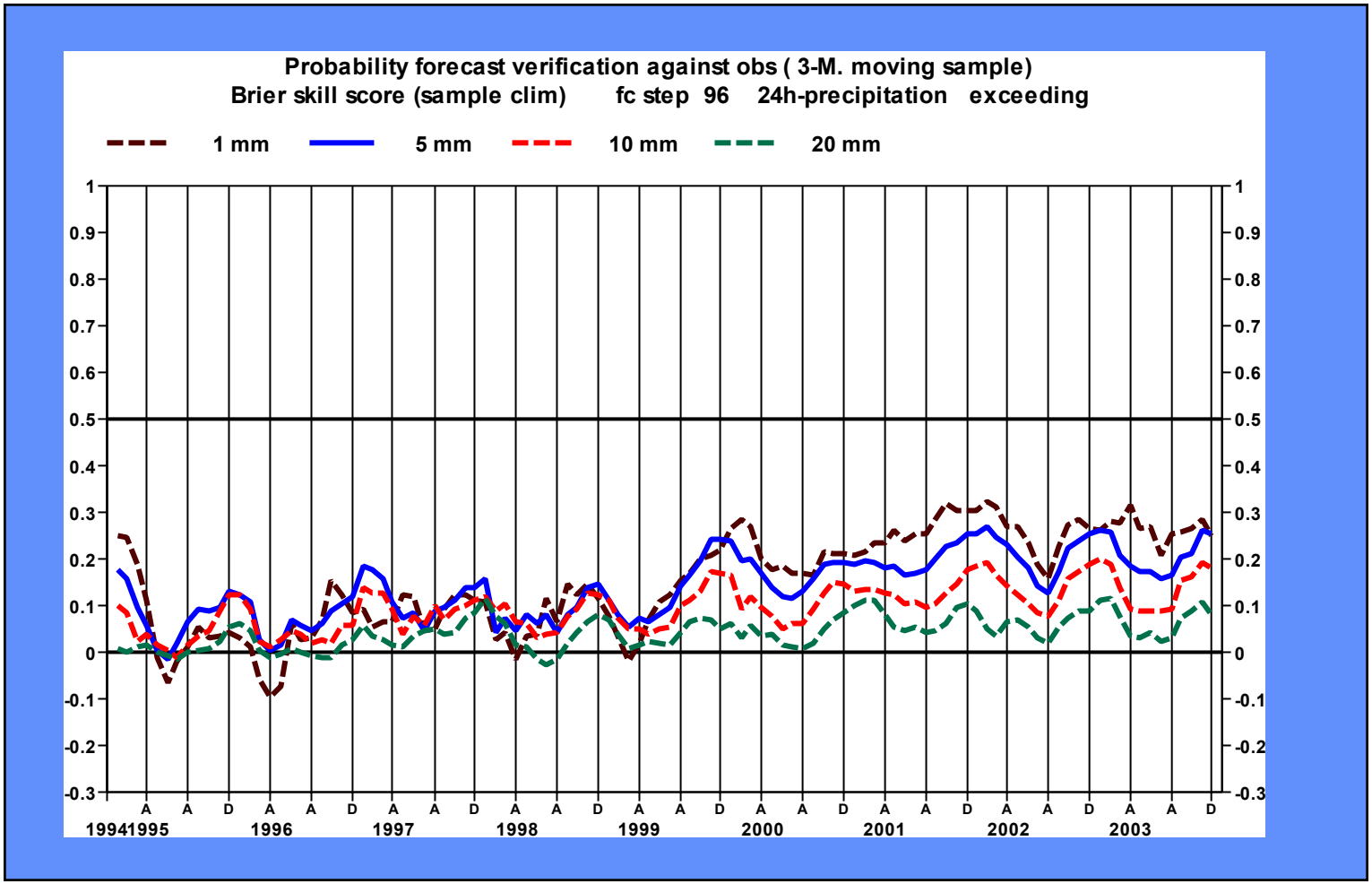


Sensible improvements of the EPS since Autumn 2000

-- ROC Area for different thresholds – Forecast range D+7



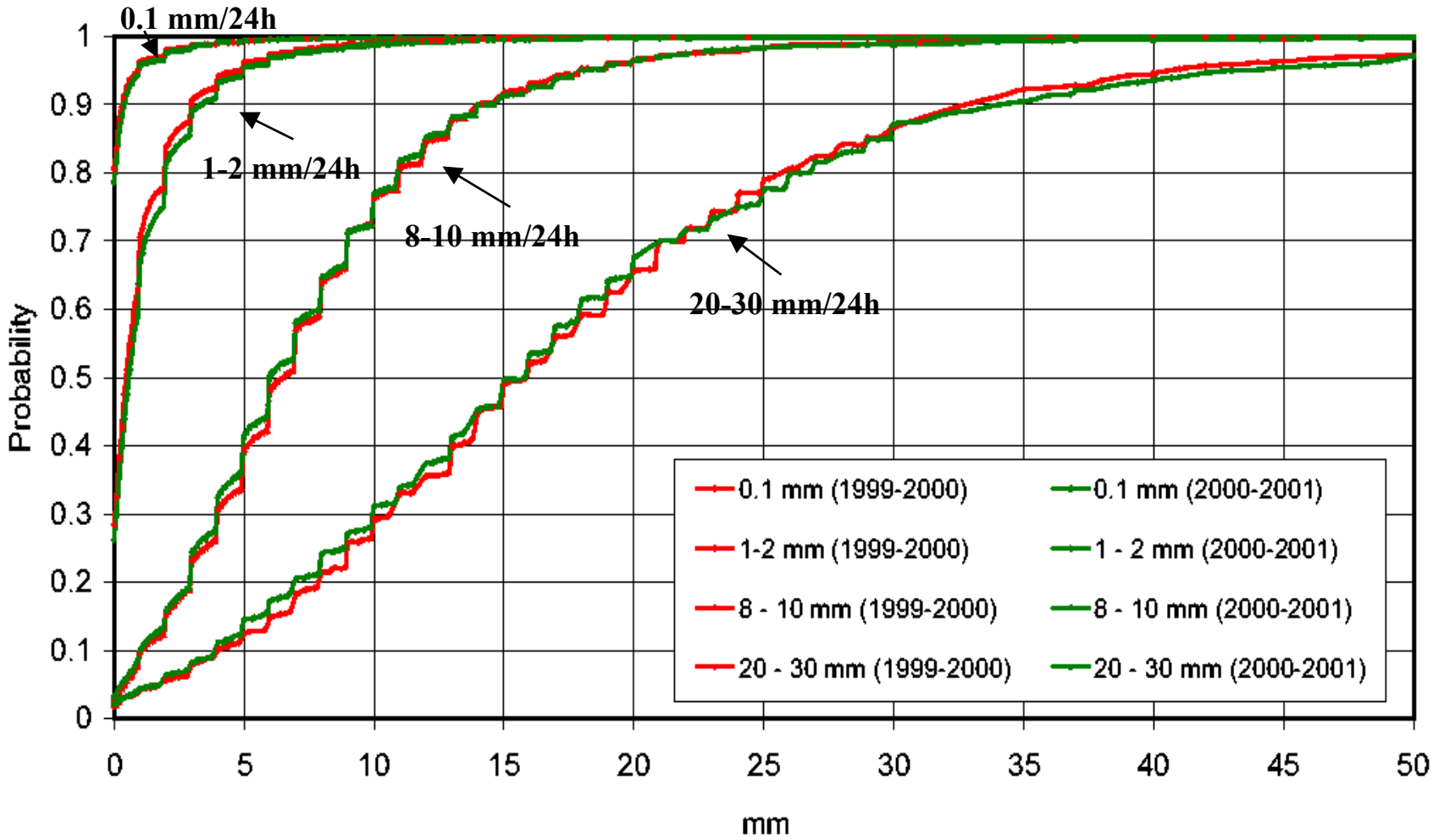
Brier Skill Score (BSS) for different thresholds – Forecast range D+4



Improvements of the EPS in 1999 (increase of vertical resolution and change in cloud scheme) and in Autumn 2000 (change in horizontal resolution)

Downscaling probabilities

Short range verification of the deterministic model against observations provides conditional sub grid scale distributions



Downscaling probabilities

"Downscaled" probabilities are then constructed as the sum of the "subgrid-scale" PDFs from each EPS member (a_i):

$$P(X_{loc} > \alpha) = (1/\mathrm{N}) \sum_{i=1}^{\mathrm{N}} p(X_{loc} > \alpha \mid X_f = \alpha_i)$$

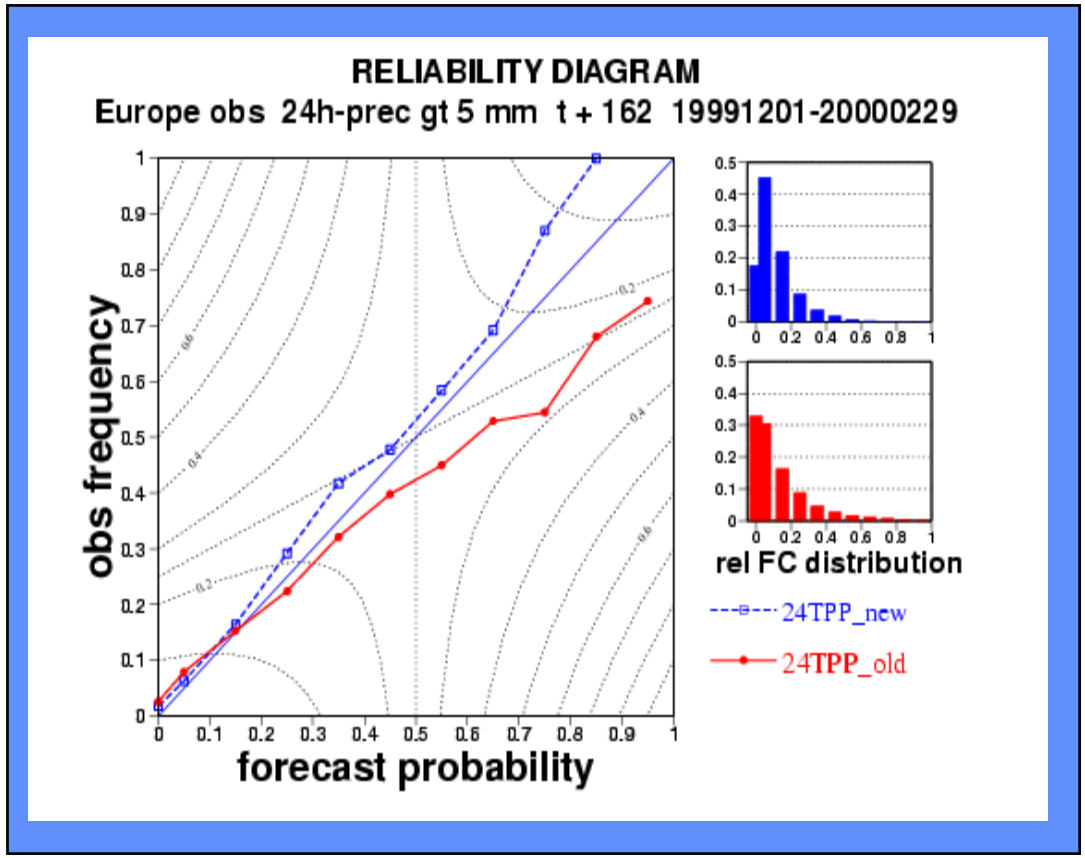
In contrast with the usual EPS probabilities that are relative proportions:

$$P(X_{grid} > \alpha) = (1/\mathrm{N}) \sum_{i=1}^{\mathrm{N}} H(\alpha_i - \alpha)$$

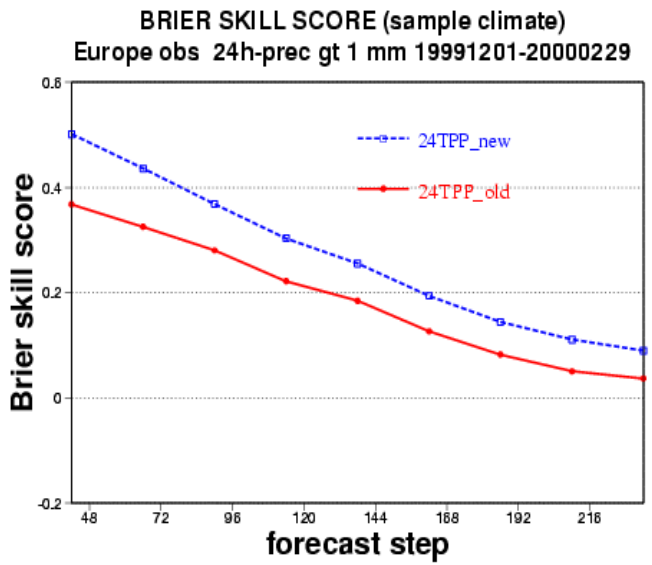
Downscaled vs Grid point Reliability Diagrams

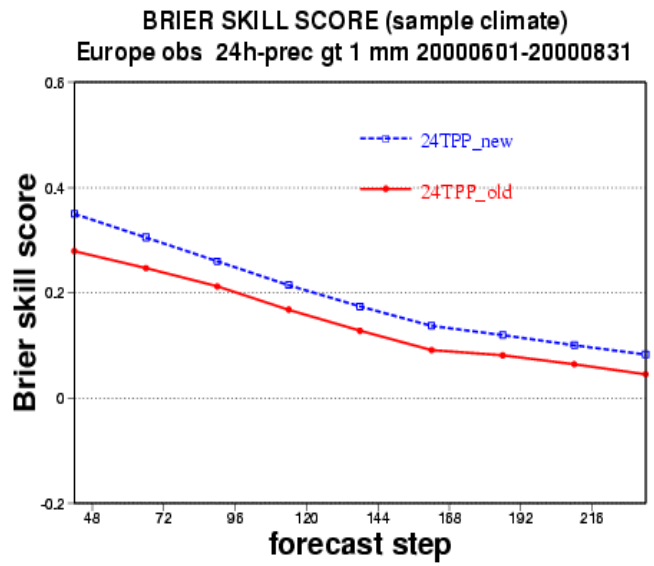
The blue curve shows the new reliability curve.

While the grid point diagram shows slight over-forecast (probabilities are too high), the downscaled approach is closer to the diagonal and slightly under-confident at higher observed frequency.

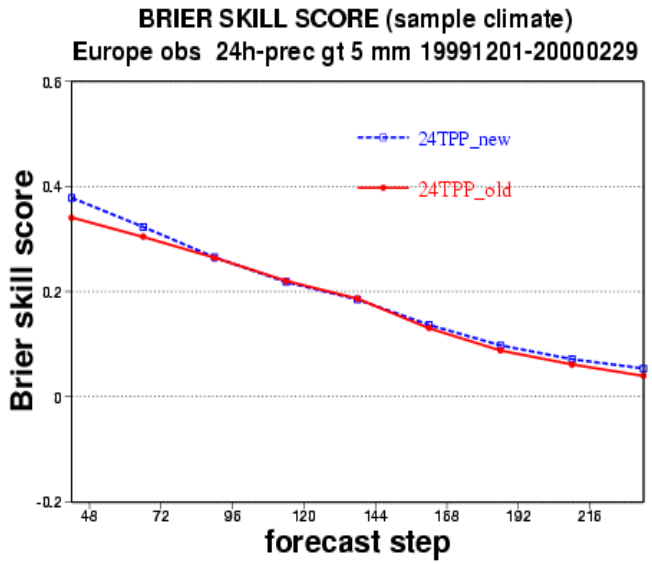


Downscaled vs Grid point Skill Scores

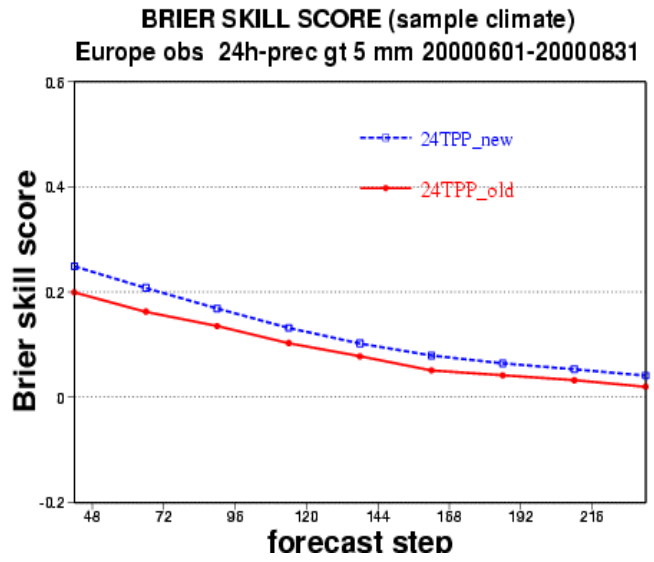




Winter



Summer



Conclusions

- Increase in resolution has had positive impact on the precipitation forecast. Largest impact on the higher precipitation thresholds
- August 2002 summer floods in Europe have not been forecast quite as accurately as other summers. The kink shows up in both ROCA and BSS.
- Example of downscaling of probabilities