Forecast verification Laboratory

T-NOTE

Juan Ruiz & Celeste Saulo

A 3 hourly accumulated precipitation forecast for South Eastern South America obtained with a 4-km numerical model is verified in this laboratory.

The forecast database consist of 11 runs, 48 hours long starting at 12 UTC.

A satellite based precipitation estimate (CMORPH) is used as verifying truth. The estimated precipitation resolution is 8 km.

Verification is performed using the 3 hourly accumulated forecasted and observed precipitation.

Guide lines:

Marginal PDF verification:

-Think about the main idea behind these methods.

-Which information is provided by the PDF of forecasts and observations. Relate this information with the frequency bias score.

-Can you estimate the length of the sipin-up period in the model using figures 2-4?

-Can you recognize the main characteristics of the forecast deduced from figures 2-4 in the forecast example of figure 1?

-How does the model represent the diurnal cycle of precipitation?

-Is there any trend in the forecast error?

Spatial matching verification scores:

- Think about the main idea behind spatial matching verification scores.

-If we use data interpolated to the model grid. Are we performing a model oriented or user oriented verification?

-Think about the shortcomings associated with an aggregation of data sets corresponding to different hours / seasons.

-Which one is the reference forecast for the ETS?

- Analize ETS, POD and FAR. What information can we get from these scores?

-Which are the main limitations of these kind of scores?

-Would these scores detect “near hits”? Using figure 5, where do you think the ETS will be penalizing the forecast too much?

Filter based verification scores:
-Are all the scales present in a forecast equaly predictable?
-Can we asses forecast skill at different spatial and / or temporal scales?
-Think about how the fraction skill score evaluates model performance (use figures 8 to 11).
-Analize the FSS values as a function of precipitation threshold, forecast lead time and spatial scale.
-Can we use this score to compare the skill of low resolution models with high resolution convective allowing models? Why?

Object oriented verification:
-Think about the main idea behind object oriented verification methods.
-Using figures 13 to 16 analyze one possible method to identify precipitating systems in the forecast and in the observations.
-Think about different possible matching criteria and analyze the results obtained using the distance criteria from Davies et al. 2006.
-Can we use object oriented verification without using a matching criteria?
-What kind of information (not provided by the scores used so far) can be obtained with object oriented methods? How can this technique help to asses forecast uncertainty regarding location?
-Using figures 19 to 21 analyze the sensitivity of the method to the precipitation threshold used to identify the objects.
-Which are the main limitations in object oriented verification methods?

Verification of probabilistic forecast:

-Discuss the difference between probabilistic and deterministic forecasts and their verification.
-Which are the most appropiate kind of forecast for limited predictability phenomena?
-Which is the information provided by the rank histogram? Which properties of the forecast can be inferred from the rank histogram in figure 26?
- What information is provided by reliability diagrams? Analize the reliability diagrams provided in figure 27. Do this forecast need to be calibrated? Why?
-Is this result relevant for decission making based on the probabilistic forecast?
-What can we say about the probabilistic forecast of precipitation over 20 mm for the 24 hours forecast lead time?
-What information can we get from the Brier Skill Score? Identify the periods where the forecast seems to provide useful information. Does the Brier Skill Score can be improved using forecast calibration?
-Analize the results obtained using a probabilistic precipitation forecast generated using larger horizontal displacements. Are the improve in the scores consistent with is observed in the forecast examples (Figures 29 to 32).
-Does this analysis provide information about the magnitude of displacement errors?

Example: forecasted precipitation (left) and observed precipitation (right) for a 18 hours forecast.

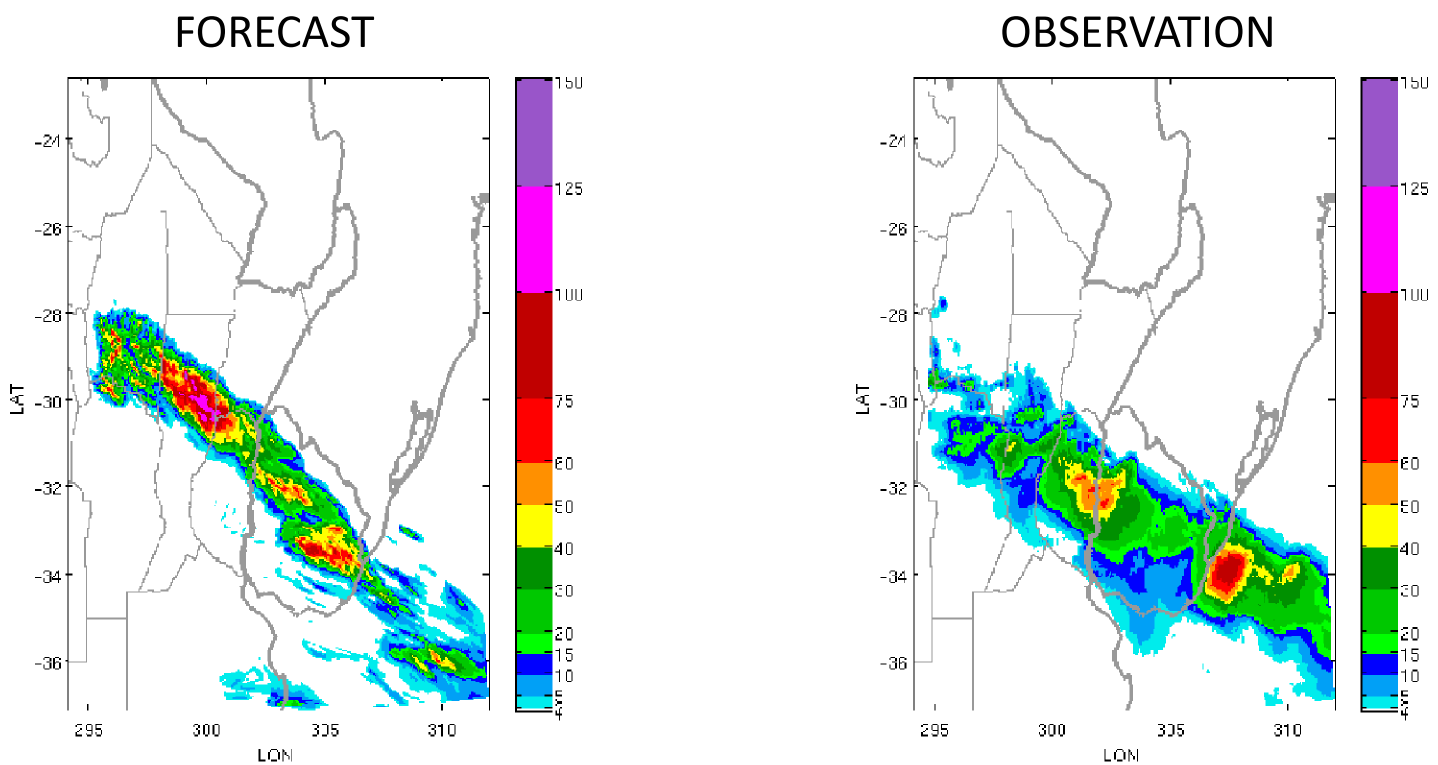
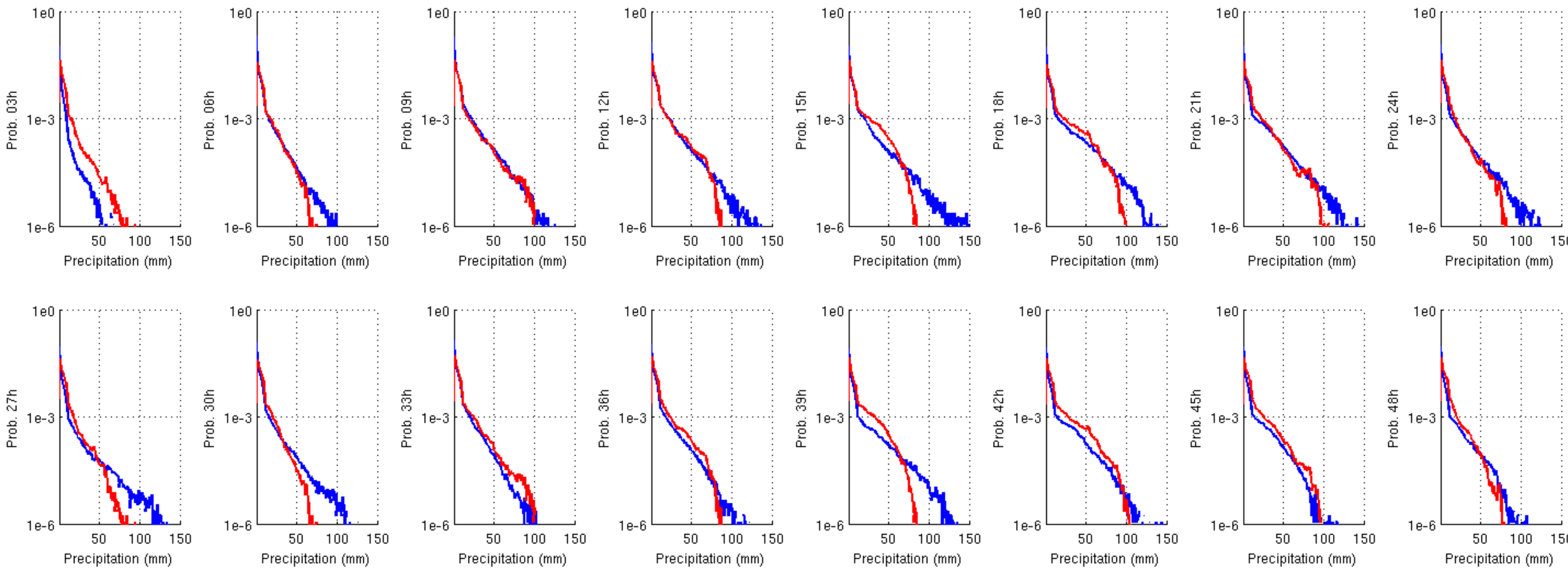


Figure 1: Example for one particular day (18 hour forecast)

Probability density function of the observed 3 hour accumulated precipitation (red) and forecasted precipitation (blue) for forecast lengths from 3 to 48 hours.



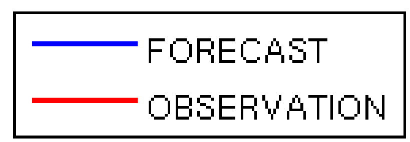


Figure 2: Computed over the entire sample

Frequency bias as a function of precipitation threshold and for forecast length between 3 and 48 hours. Values higher than one indicates that forecast is overestimating the frequency of the event with respect to the observations.

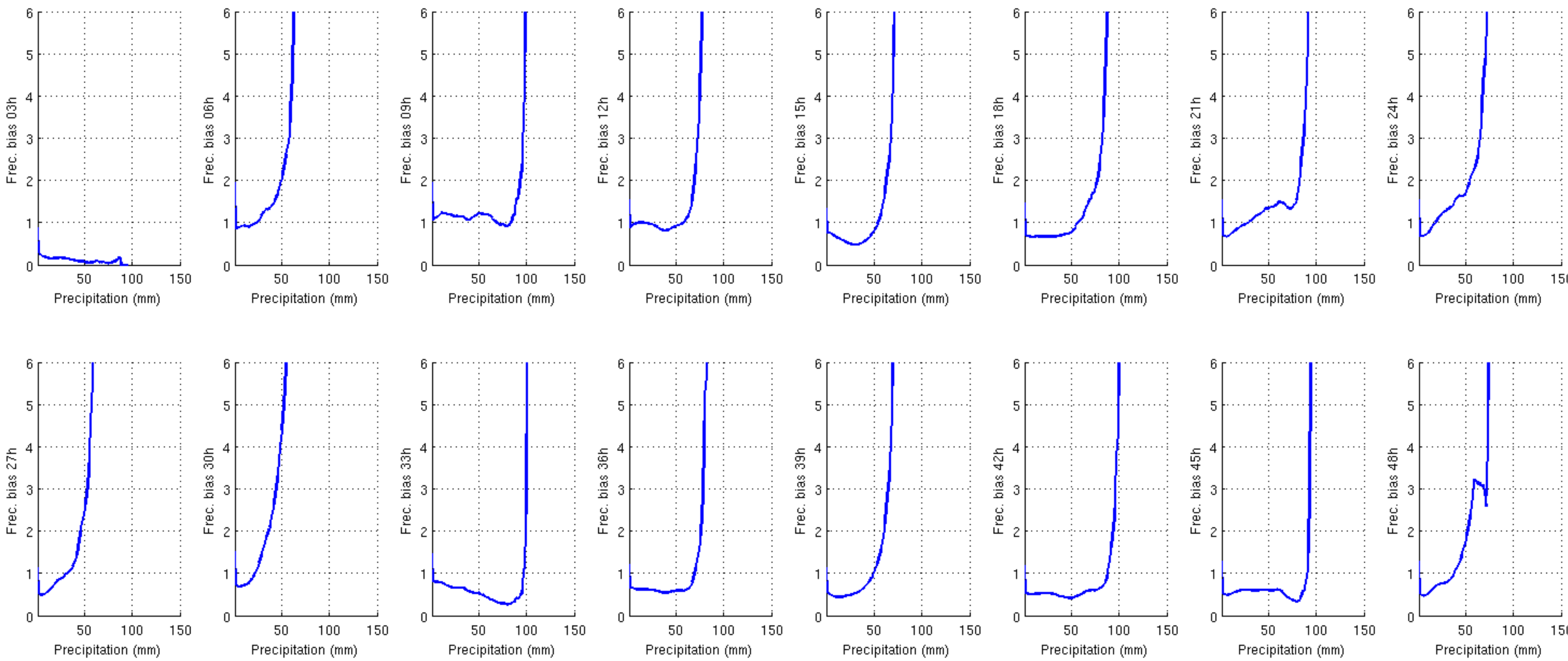


Figure 3: Computed over the entire sample.

Domain and event averaged forecasted and observed precipitation as a function of forecast lead time.

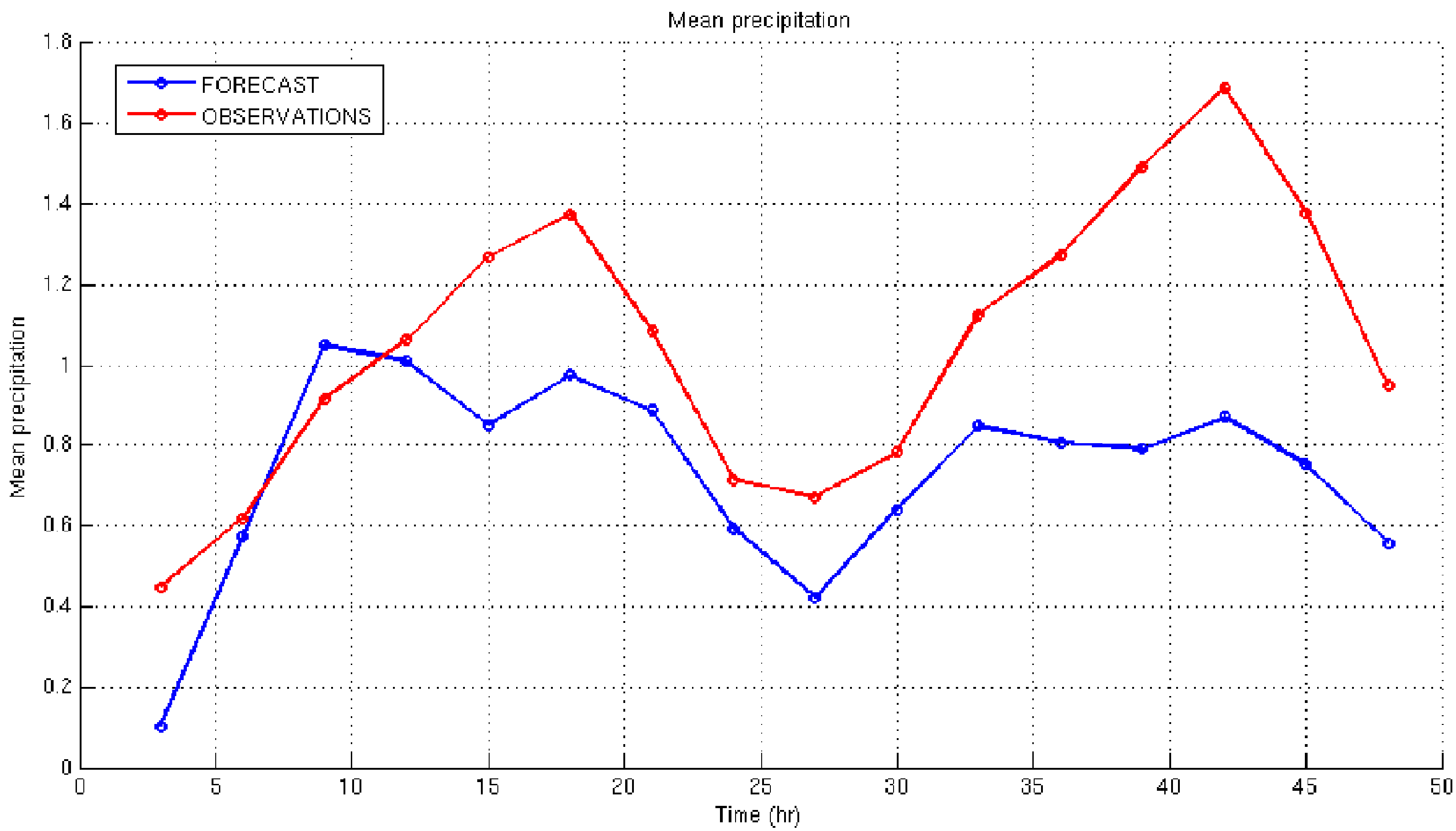


Figure 4: Computed over the entire sample.

Example: Areas that contributes to different elements of the contingency table. White= correct negative, yellow= hit, green = false alarm and purple = miss. For the same time

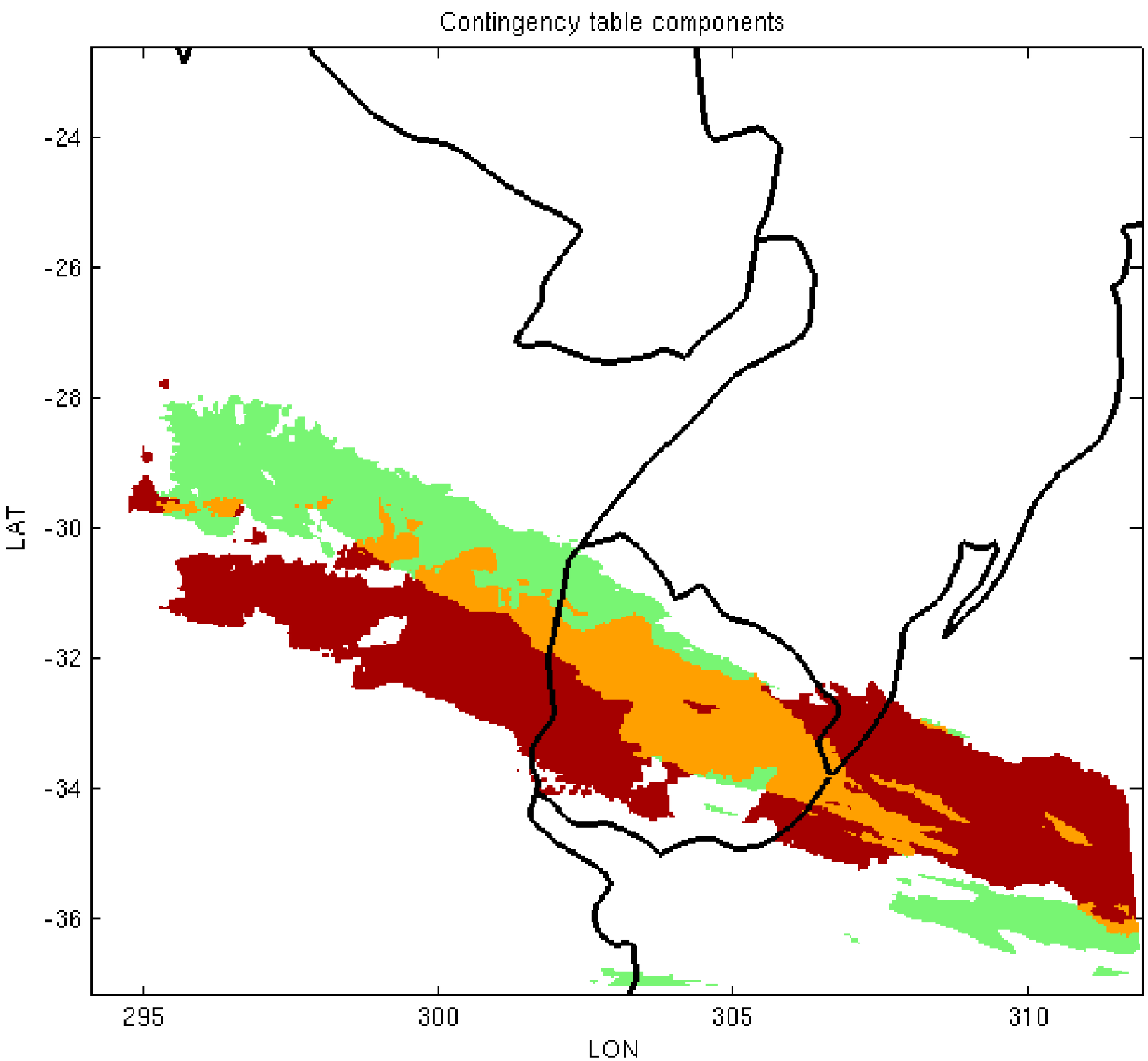


Figure 5: Example for one particular day (18 hour forecast)

ET score (blue), probability of detection (red) and false alarm ratio (green) as a function of the precipitation threshold and for forecast lengths between 3 and 48 hours.

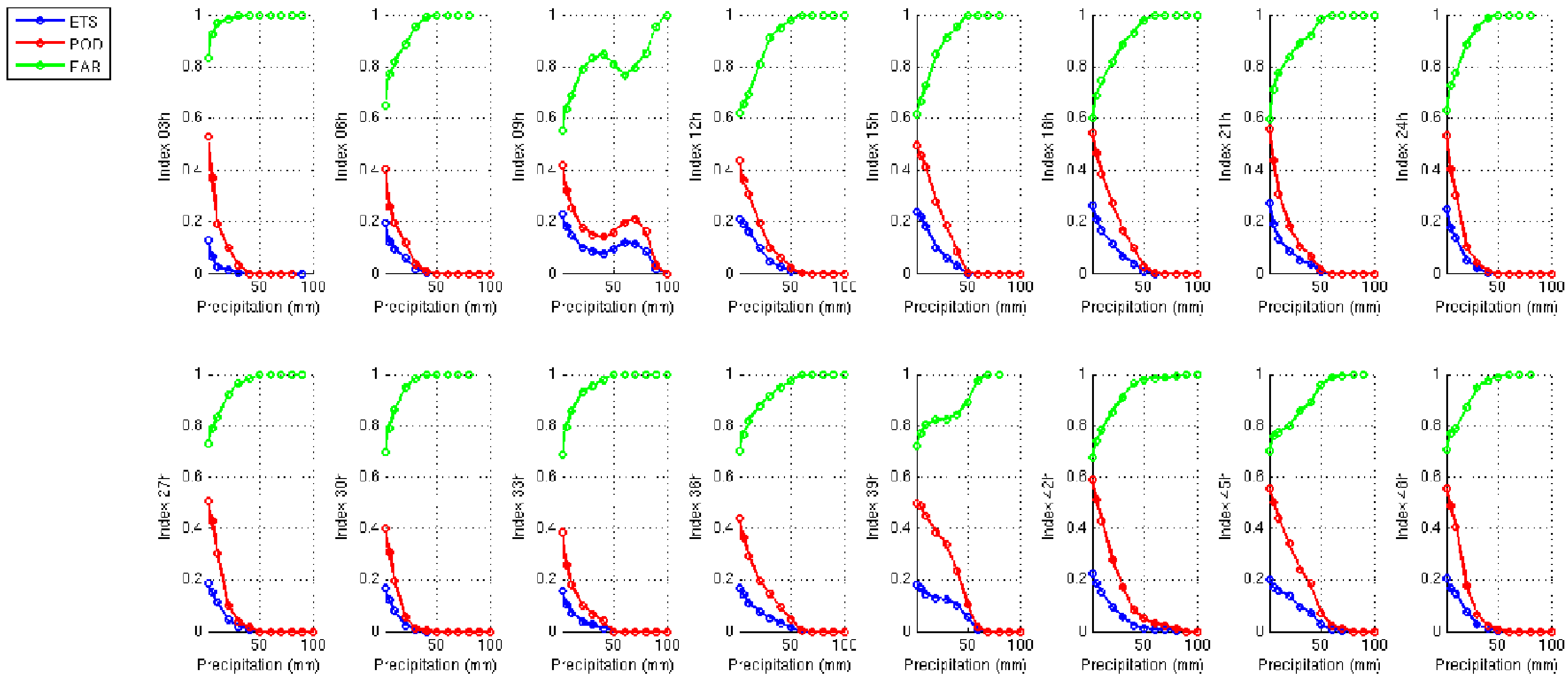


Figure 6: Computed over the entire sample.

Comparison between original precipitation (upper panels) and frequency of precipitation greater than 10 mm (lower panels) for forecasted (left panels) and observed (right panels) precipitation. A 160 km box radius has been used to compute precipitation frequencies.

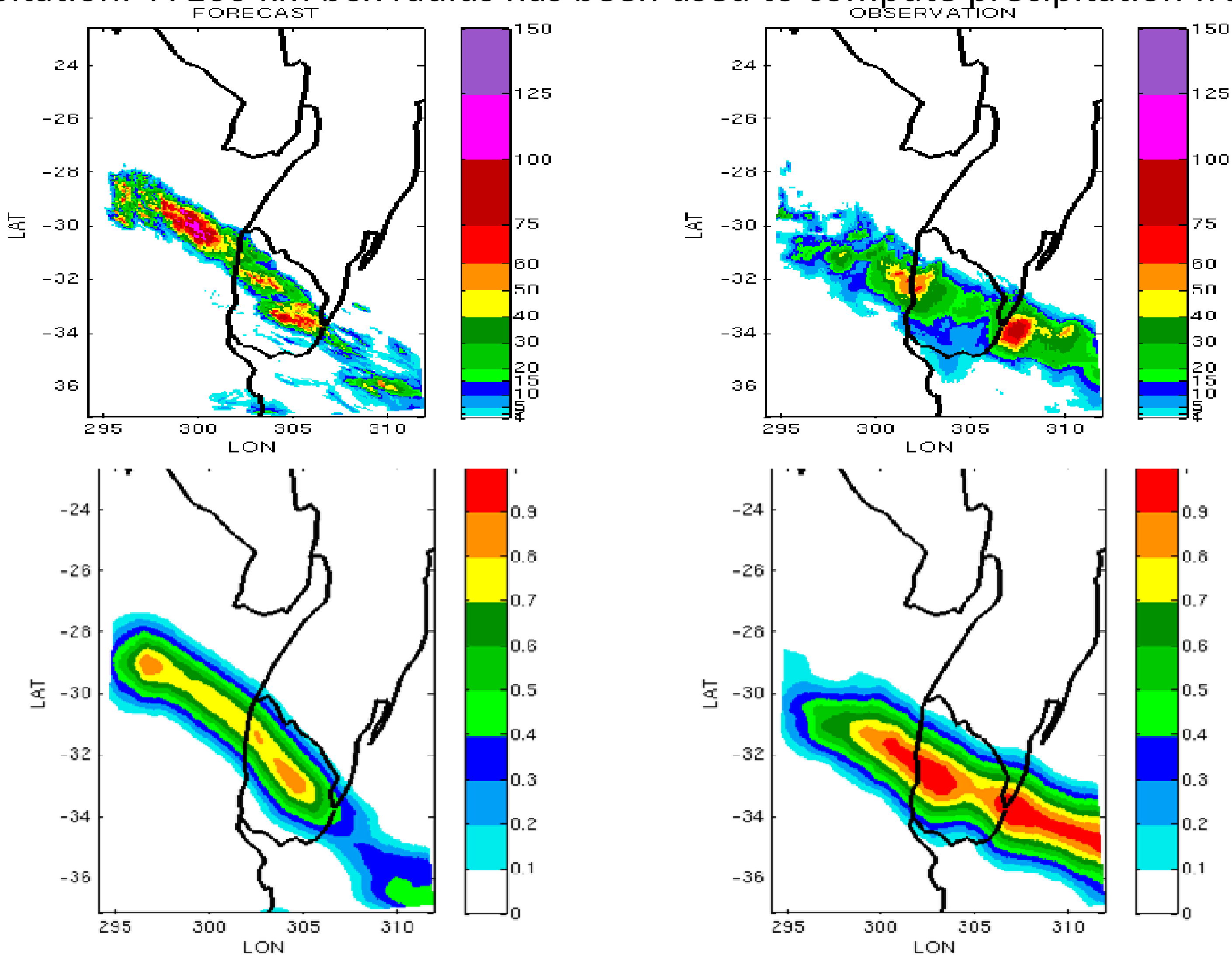


Figure 8: Example for one particular day (18 hour forecast)

Difference between forecasted frequency and observed frequency using a 160 km box radius.

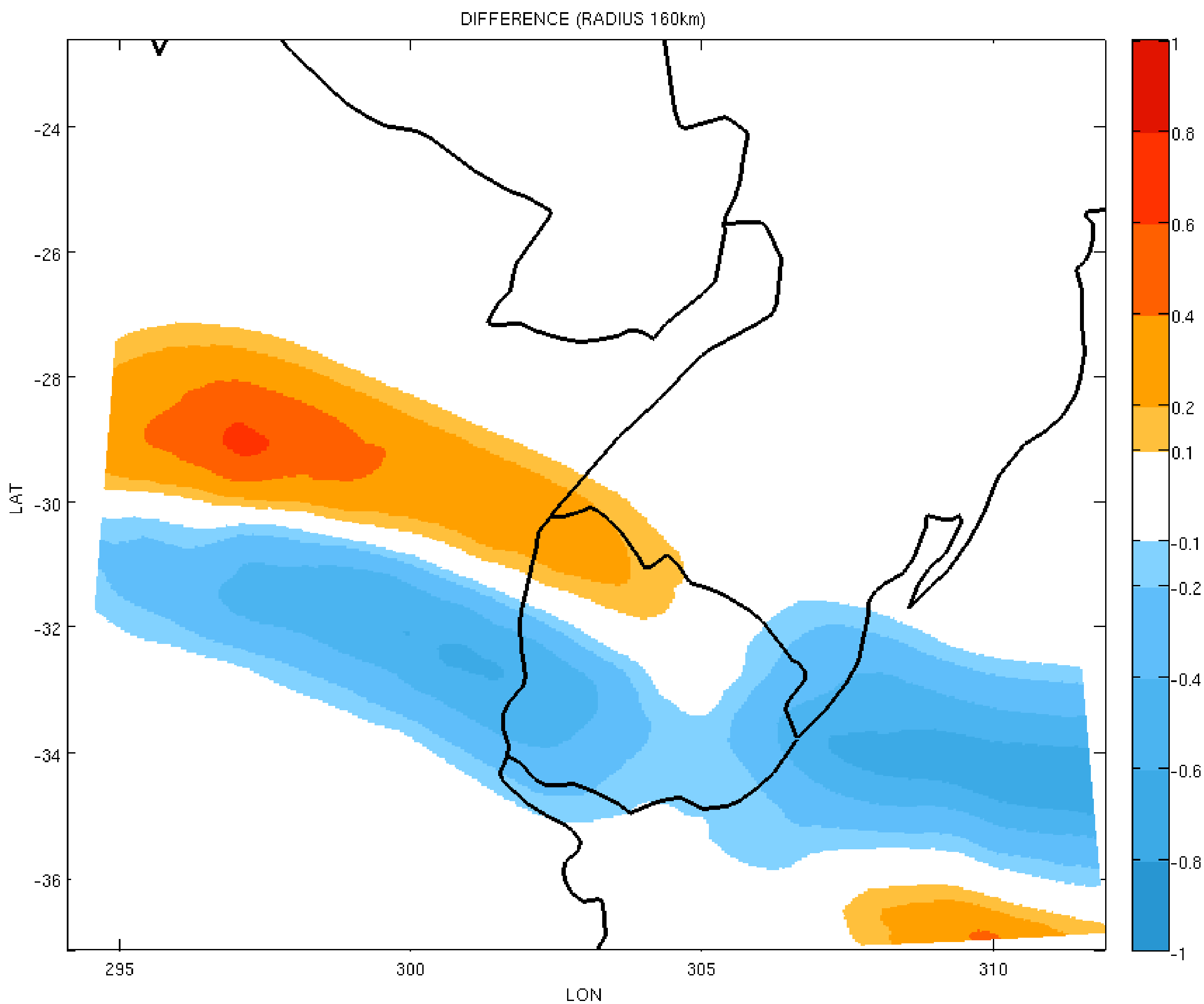


Figure 9: Example for one particular day (18 hour forecast)

Comparison between original precipitation (upper panels) and frequency of precipitation greather than 10 mm (lower panels) for forecasted (left panels) and observed (right panels) precipitation. A 20 km box radius has been used.

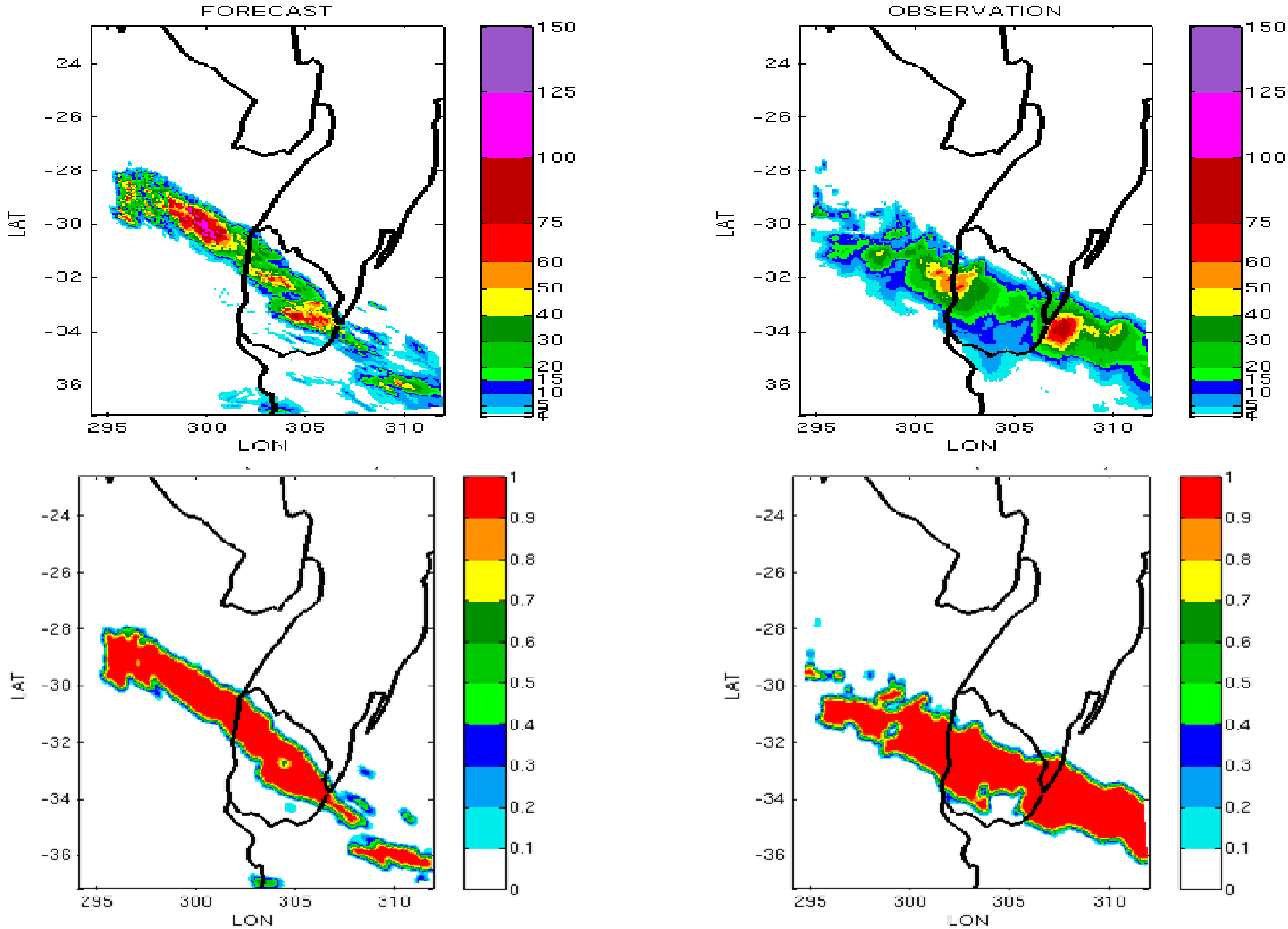


Figure 10: Example for one particular day (18 hour forecast)

Difference between forecasted frequency and observed frequency using a 20 km box radius.

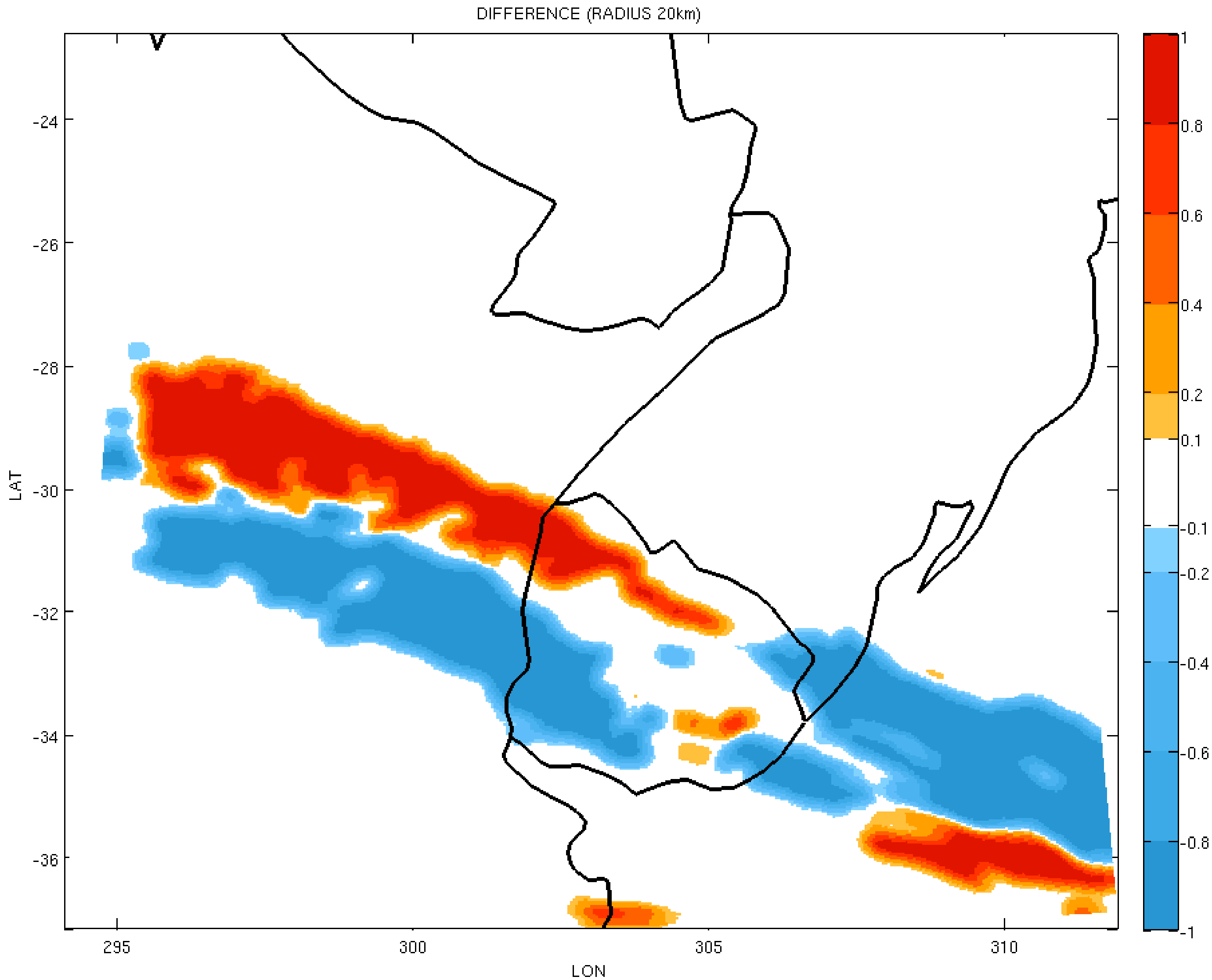


Figure 11: Example for one particular day (18 hour forecast)

Fraction skill score as a function as area size for the 1 mm and 20 mm thresholds.

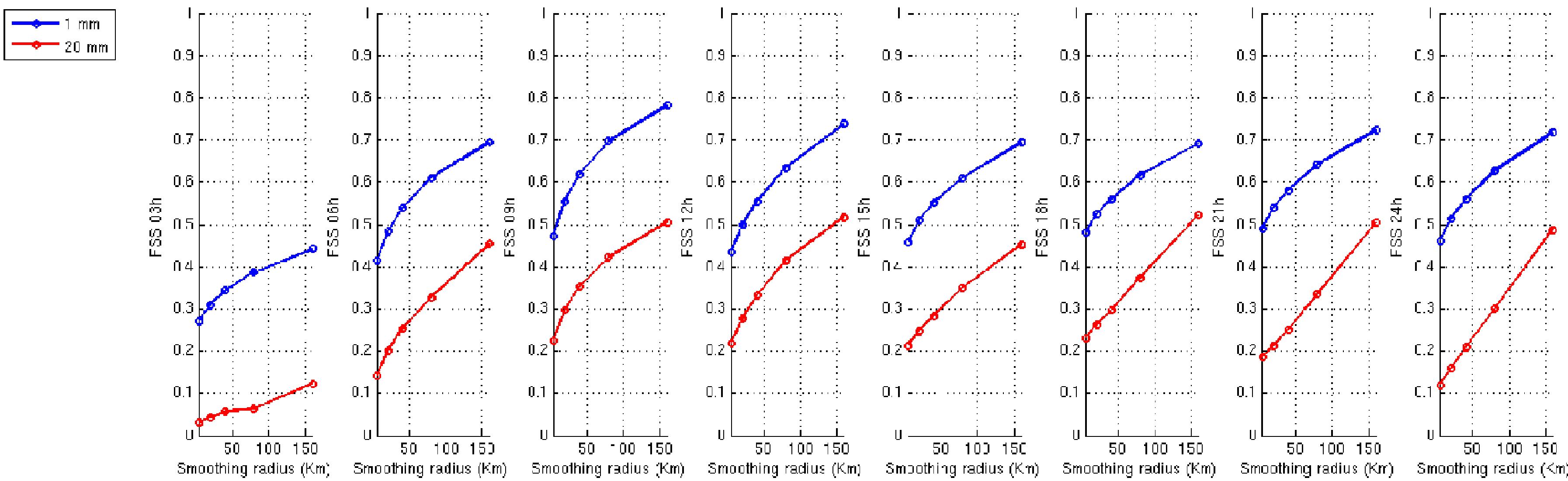


Figure 12: Computed over the entire sample

Smoothed forecasted (left) and observed (right) precipitation fields using 240 km radius boxes.

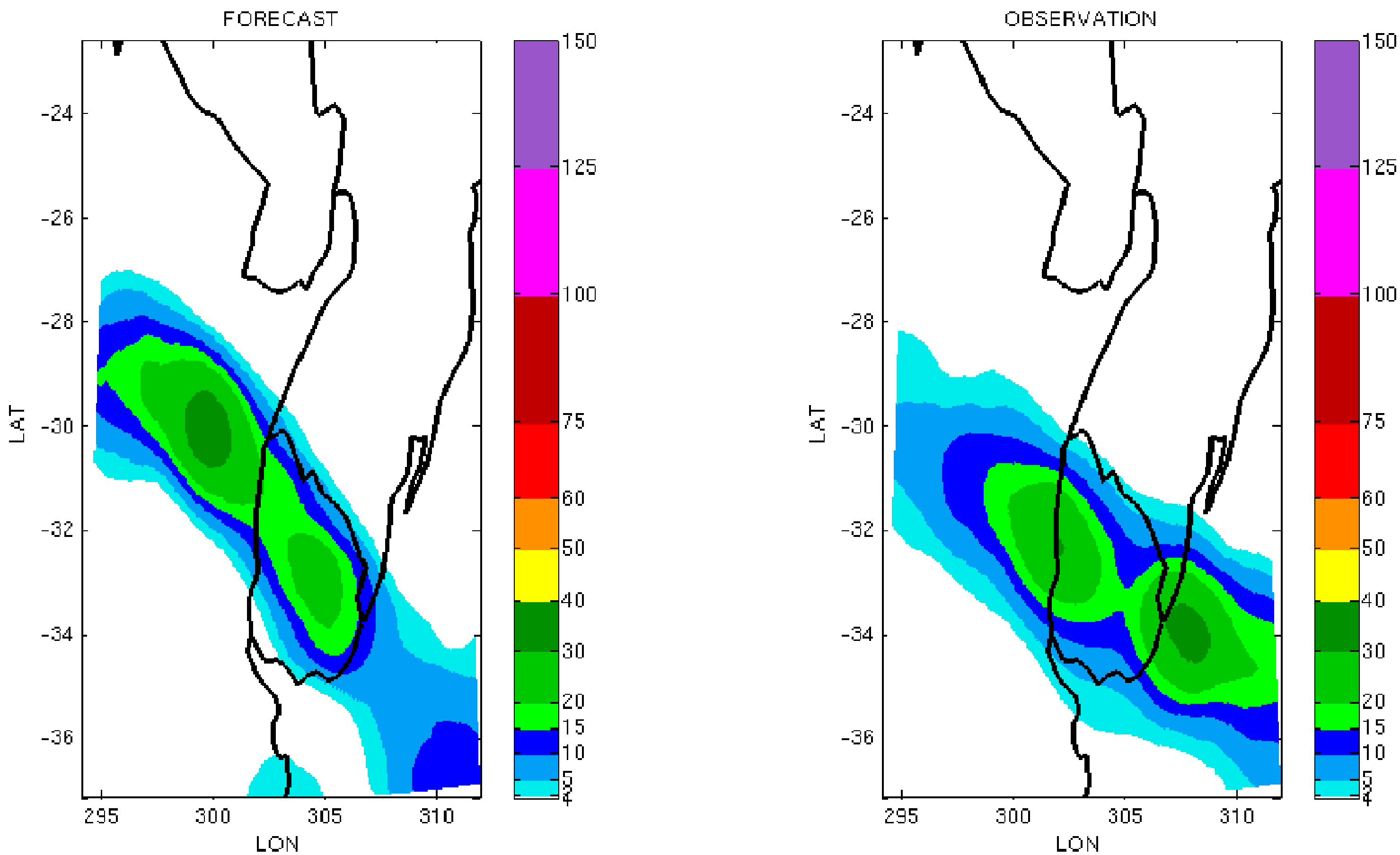
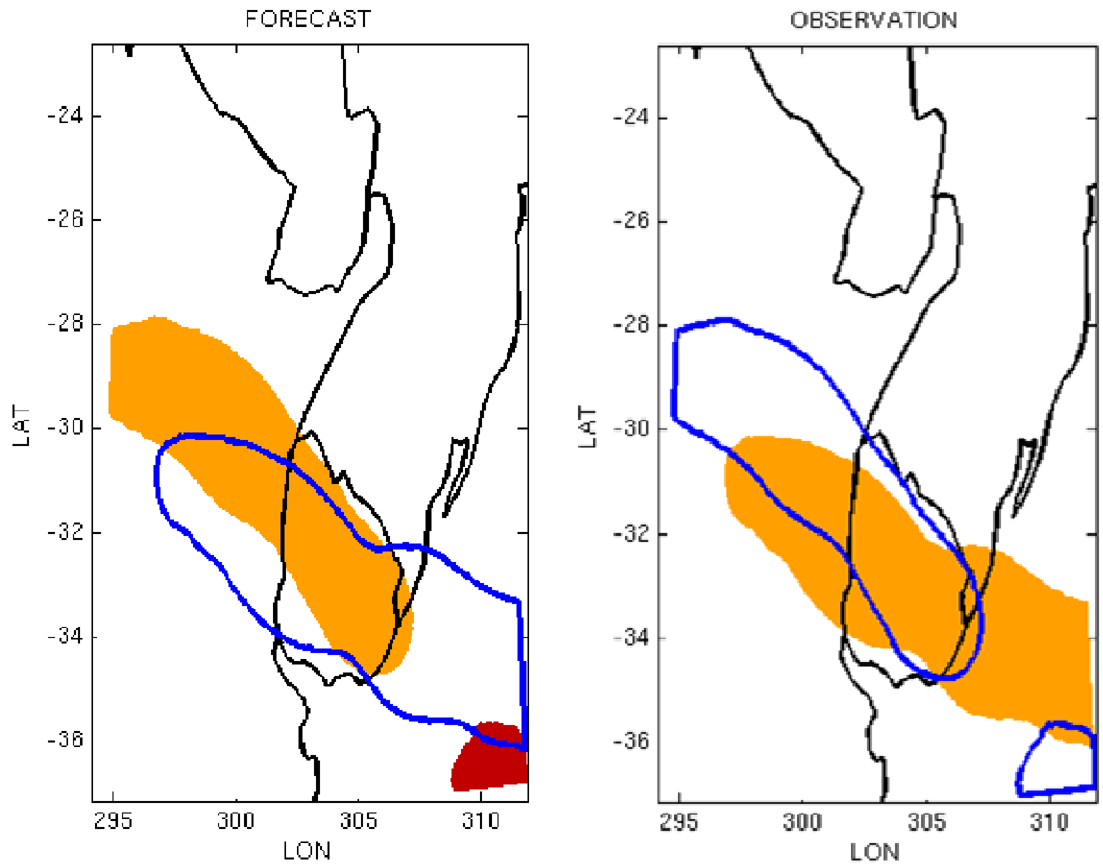


Figure 13: Example for one particular day (18 hour forecast)

Objects identified in the forecast and in the observations (shaded). Lines indicate objects in the corresponding observed or forecast data set. A 10 mm threshold has been used to identify the objects.



Forecast (shaded), observation (contour)

Forecast (contour), observation (shaded)

Figure 14: Example for one particular day (18 hour forecast)

Characteristics of the object identified in the forecast.

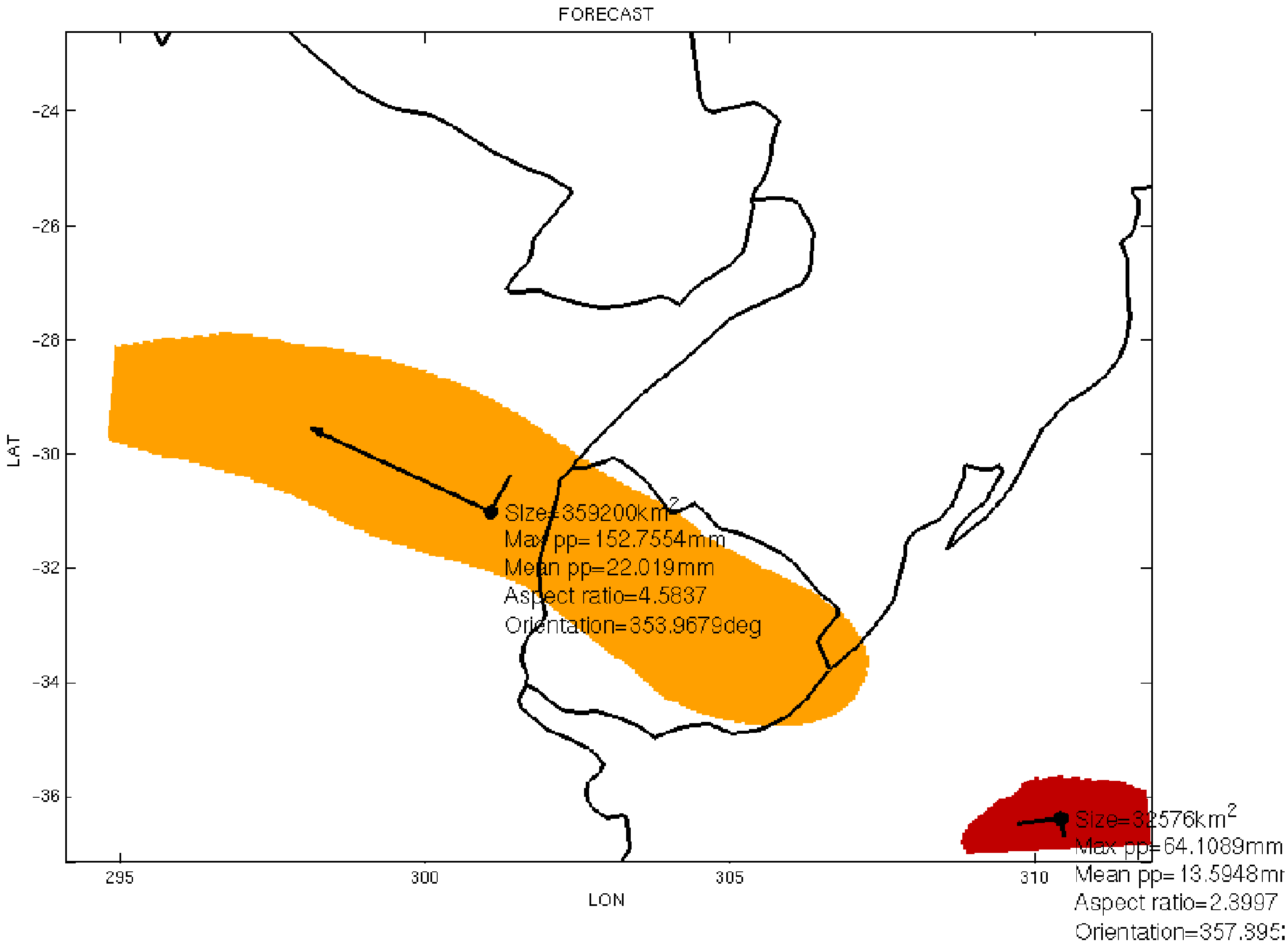


Figure 15: Example for one particular day (18 hour forecast)

Characteristics of the object identified in the observations.

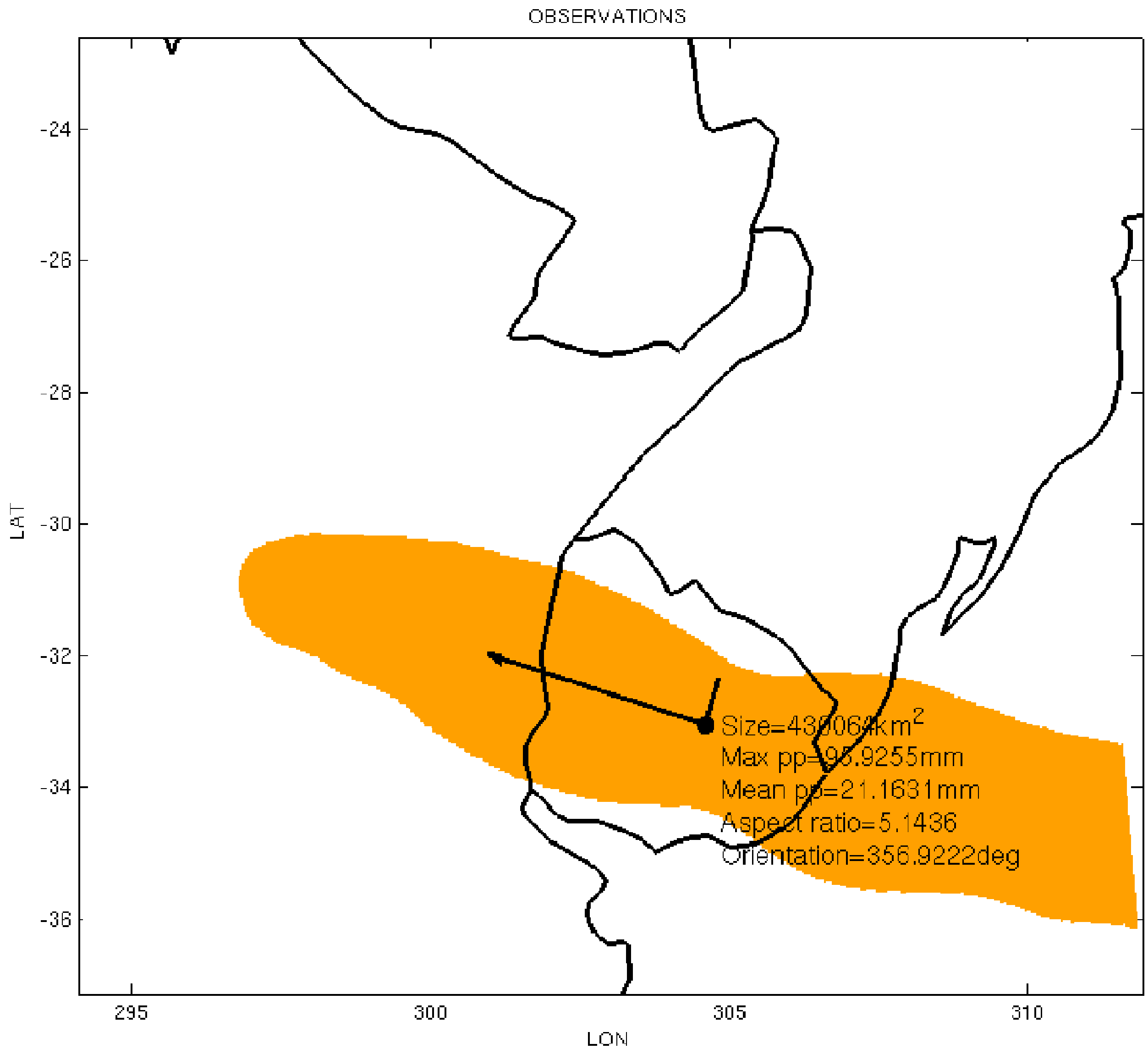


Figure 16: Example for one particular day (18 hour forecast)

Matching between forecasted (shaded) object and the observed object (contour). The Distance factor of this matching is indicated in the figure. The distance between the centroid of the two objects is also depicted.

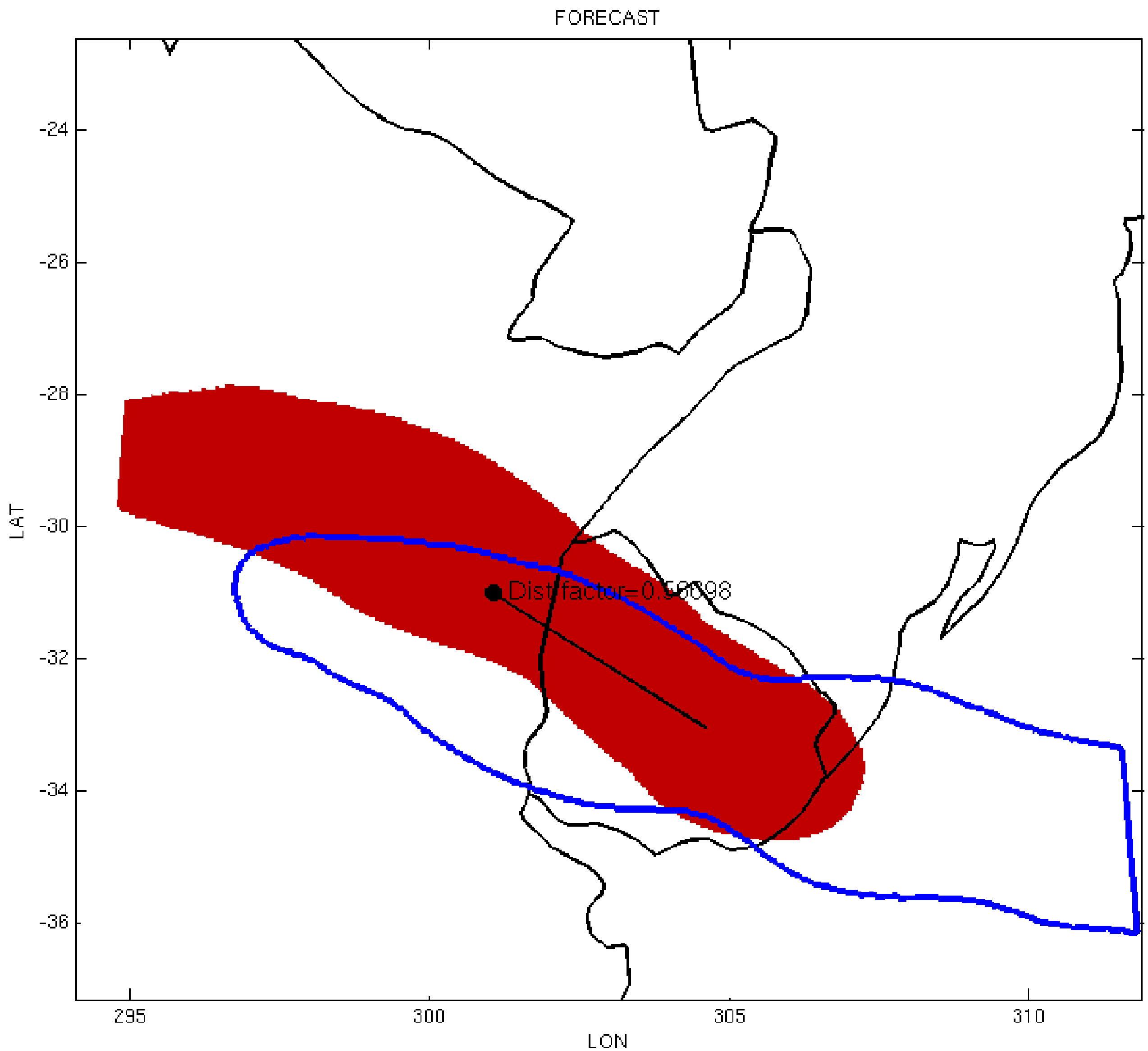


Figure 17: Example for one particular day (18 hour forecast)

Matching between forecasted (shaded) object and the observed object (contour). The Distance factor of this matching is indicated in the figure. The distance between the centroid of the two objects is also depicted.

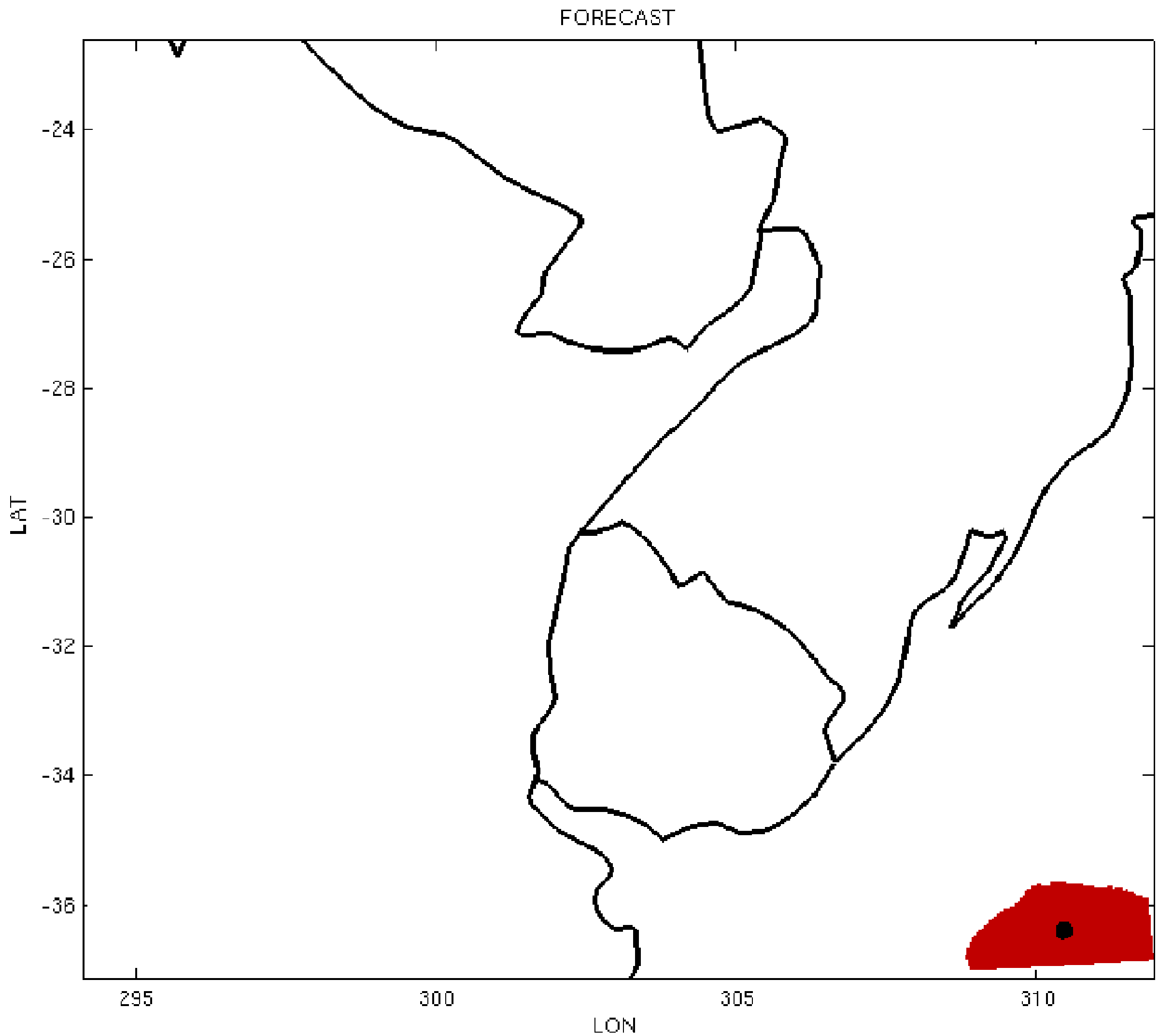
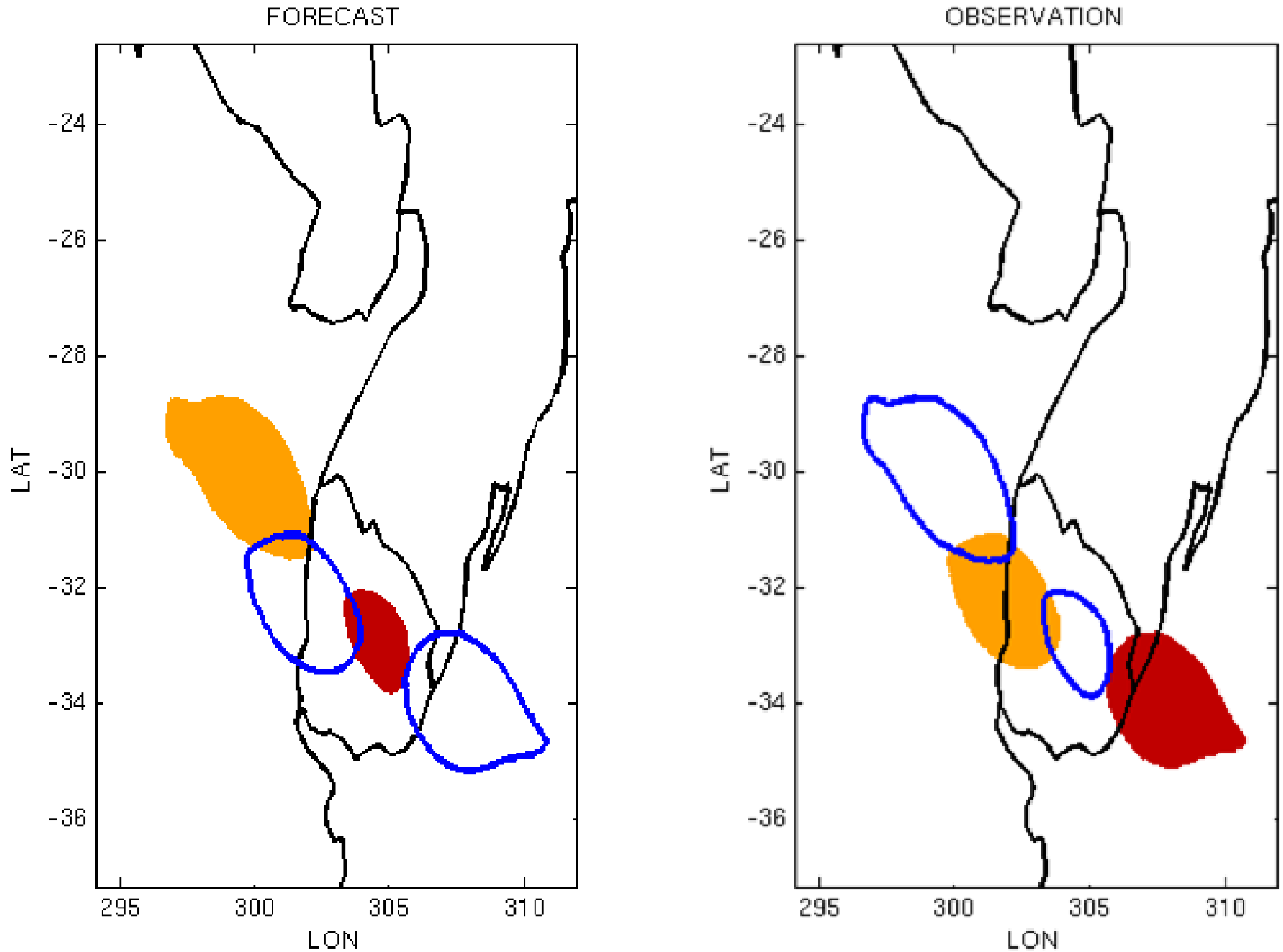


Figure 18: Example for one particular day (18 hour forecast)

Objects identified in the forecast and in the observations (shaded). Lines indicate objects in the corresponding observed or forecast data set. A 20 mm threshold has been used to identify the objects.



Forecast (shaded), observation (contour)

Forecast (contour), observation (shaded)

Figure 19: Example for one particular day (18 hour forecast)

Characteristics of the object identified in the forecast.

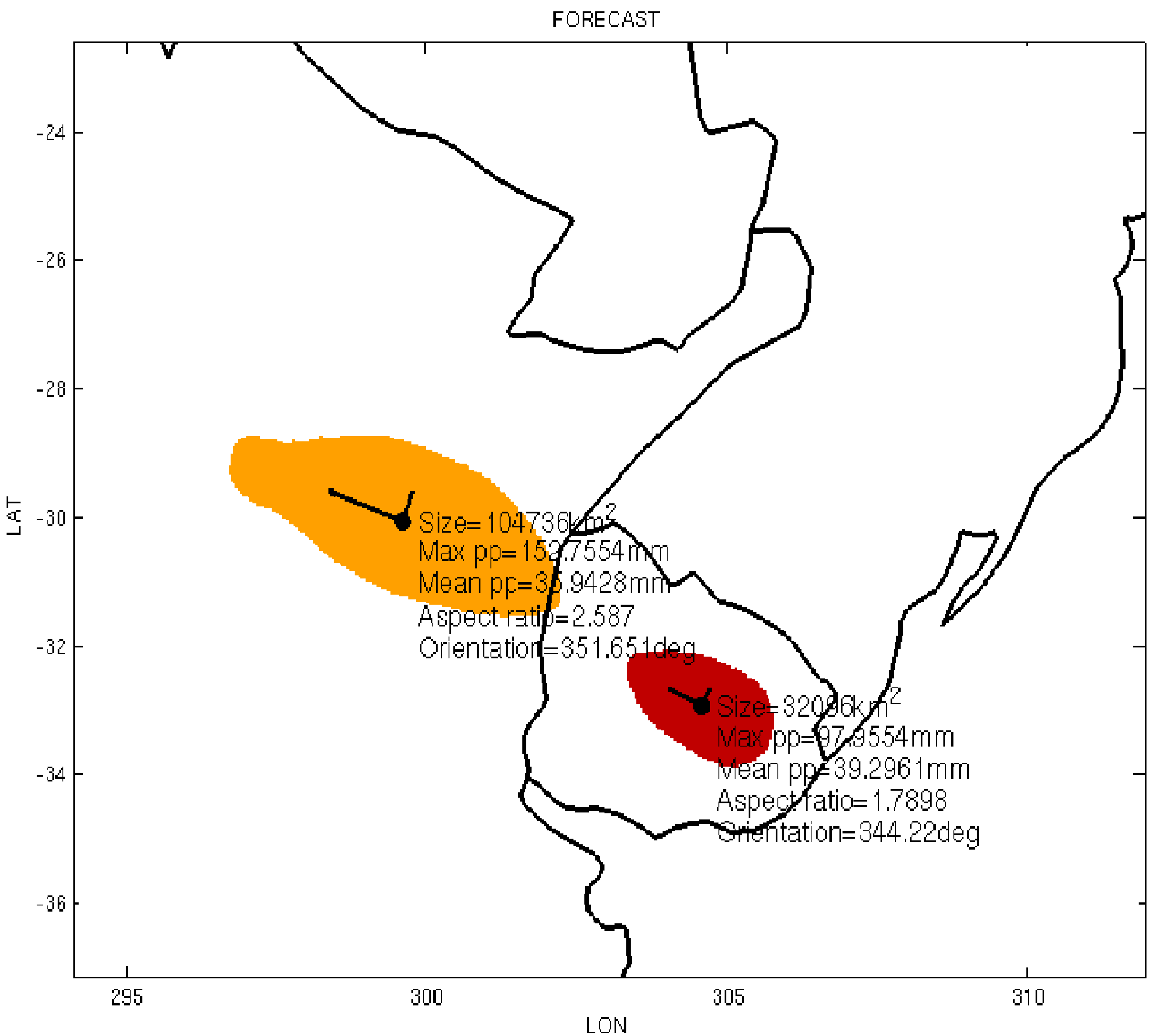


Figure 20: Example for one particular day (18 hour forecast)

Characteristics of the object identified in the observations.

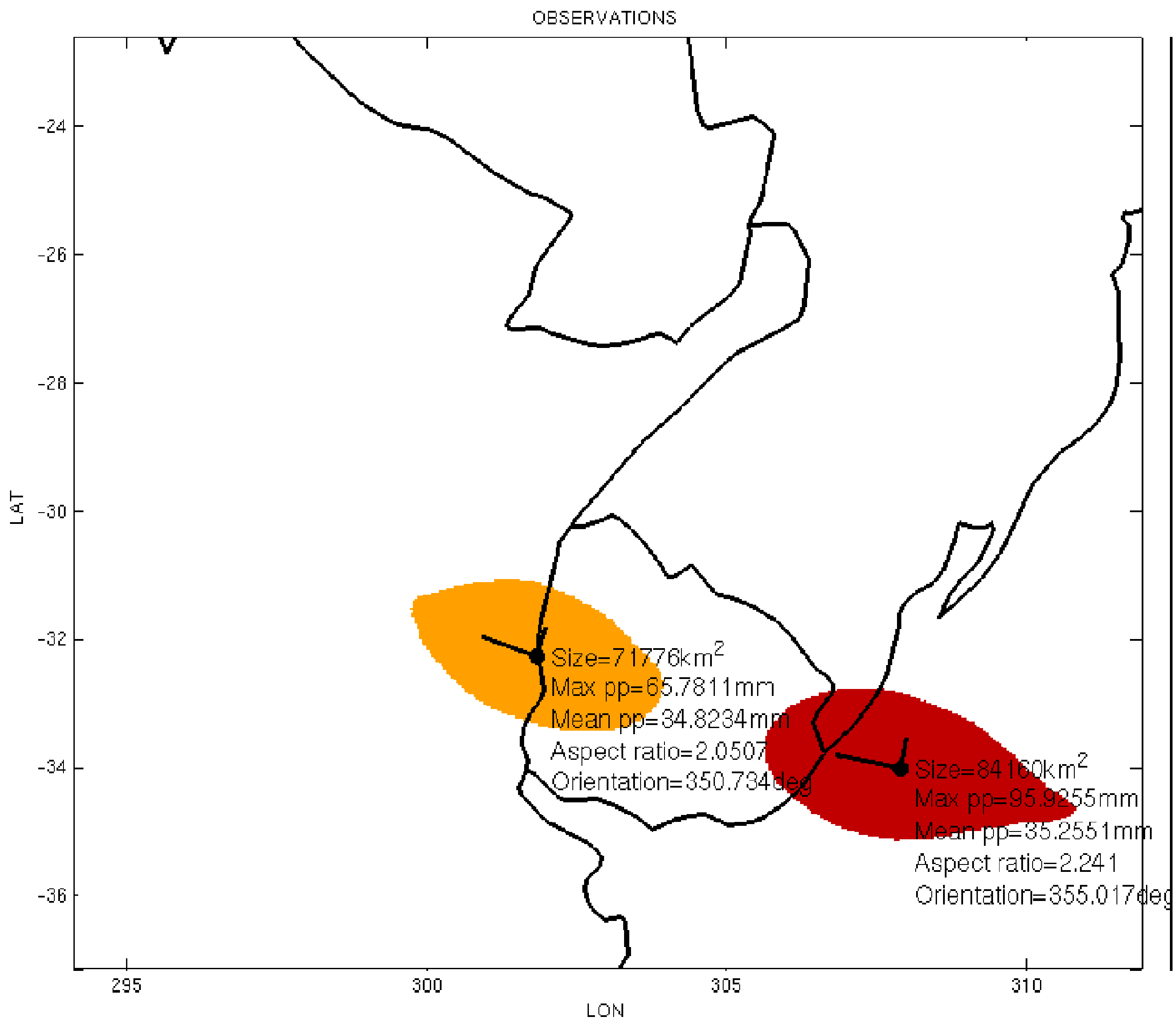


Figure 21: Example for one particular day (18 hour forecast)

Matching between forecasted (shaded) object and the observed object (contour). The Distance factor of this matching is indicated in the figure. The distance between the centroid of the two objects is also depicted.

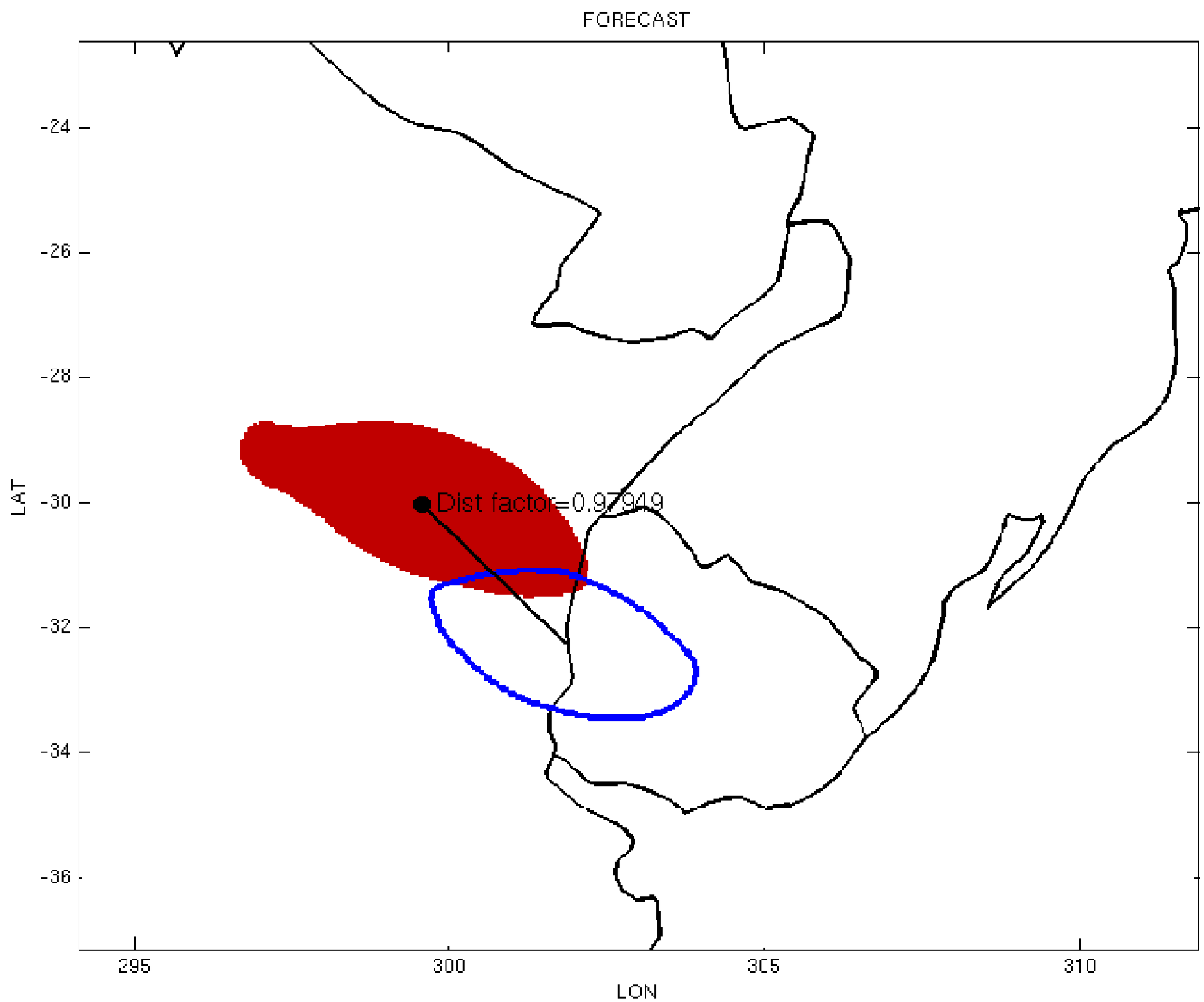


Figure 22: Example for one particular day (18 hour forecast)

Matching between forecasted (shaded) object and the observed object (contour). The Distance factor of this matching is indicated in the figure. The distance between the centroid of the two objects is also depicted.

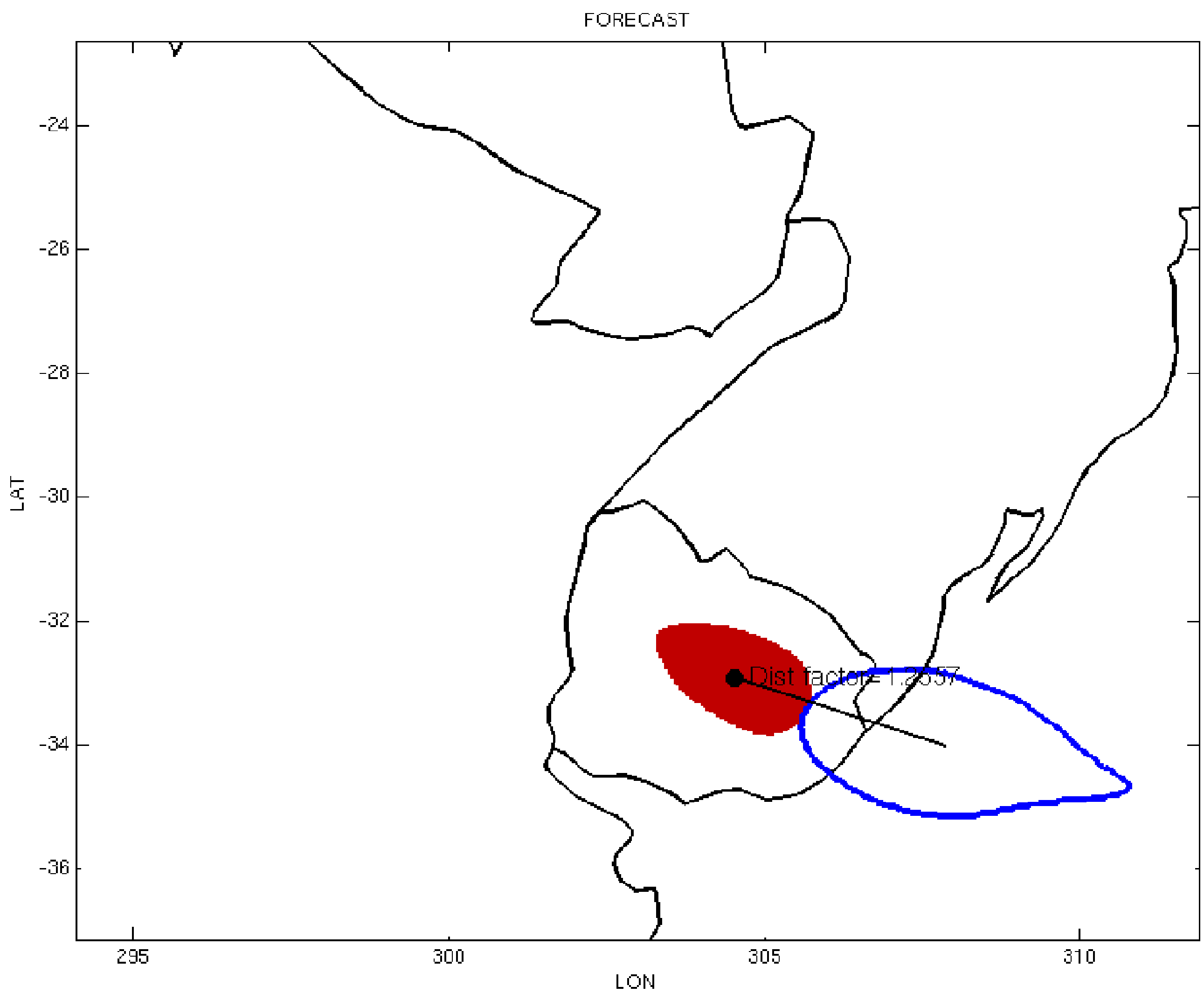


Figure 23: Example for one particular day (18 hour forecast)

Probability of having 3 hourly accumulated precipitation over 10 mm. Ensemble of forecast was generated by a pragmatic ensemble approach (Theis, 2005) horizontal displacement of deterministic forecast (max displacement= 160 km.

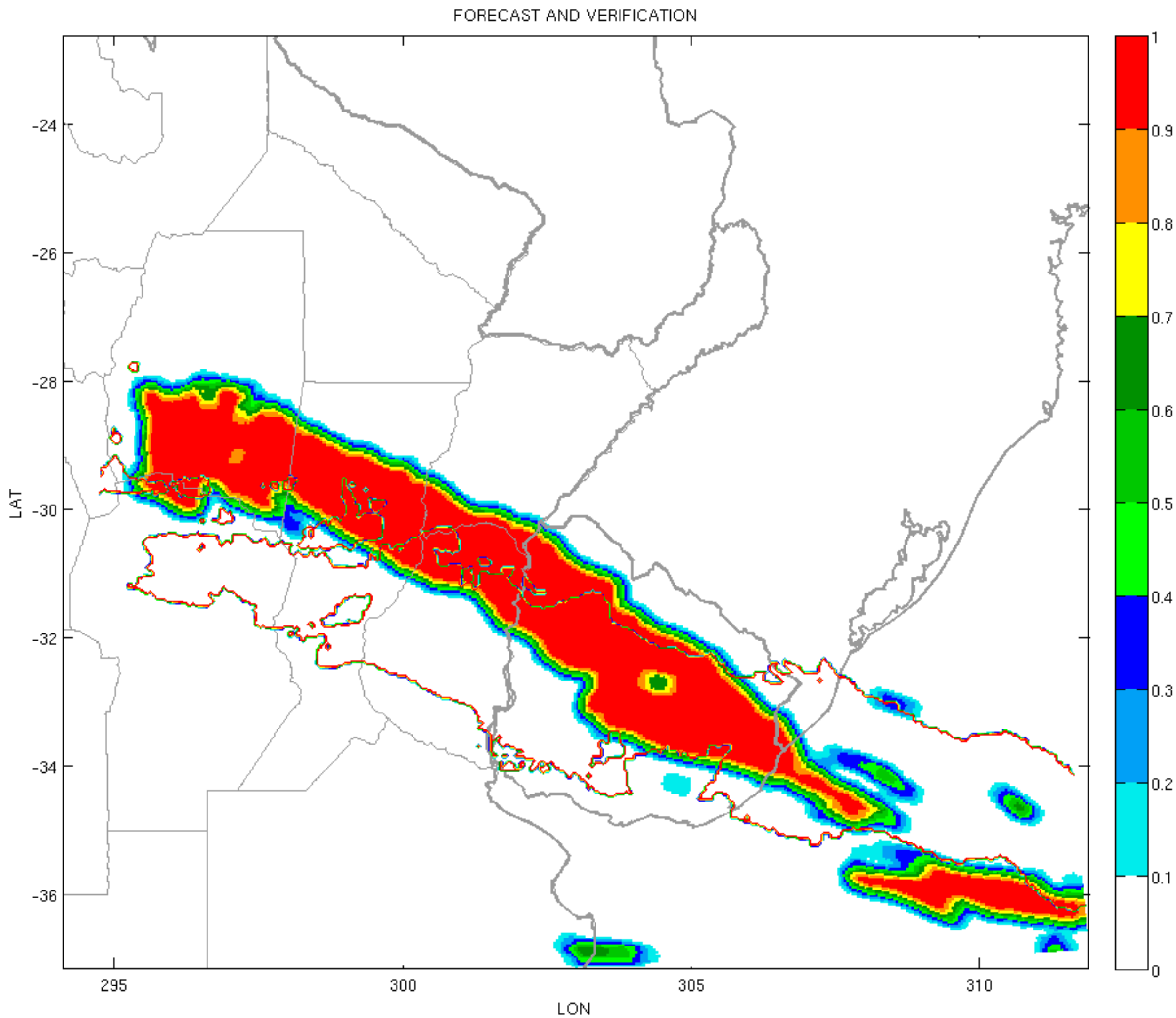


Figure 24: Example for one particular day (18 hour forecast)

Difference between forecasted and “observed” probability.

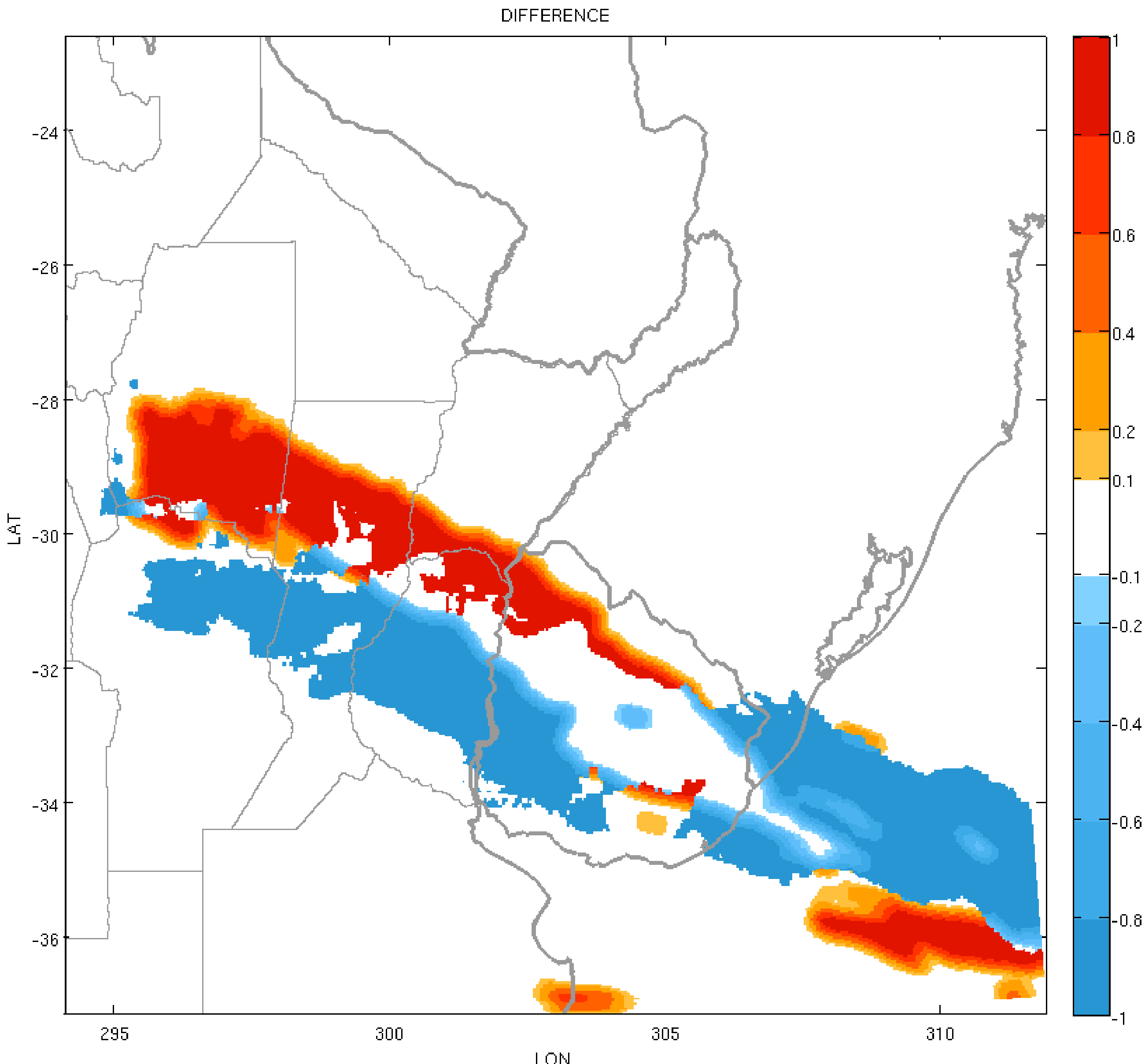


Figure 25: Example for one particular day (18 hour forecast)

Rank histogram as a function of forecast lead time for the probabilistic forecast generated using horizontal displacements of deterministic forecast (maximum displacement: 40 km)

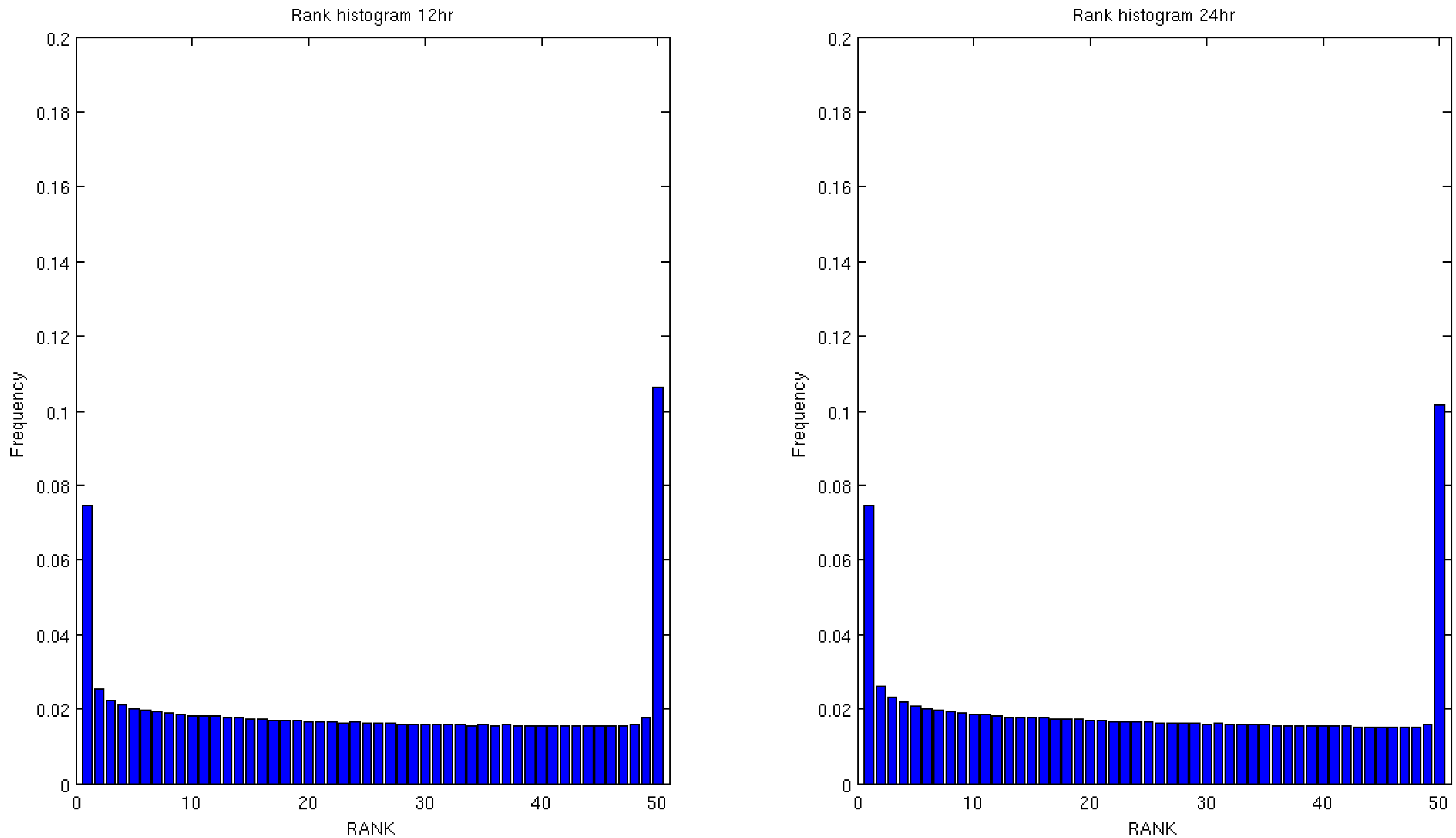


Figure 26: Computed over the entire sample

Reliability diagrams for 1 mm and 20 mm thresholds as a function of forecast lead time. Probabilistic forecast was generated using horizontal displacement of deterministic forecasts (maximum displacement 40 km).

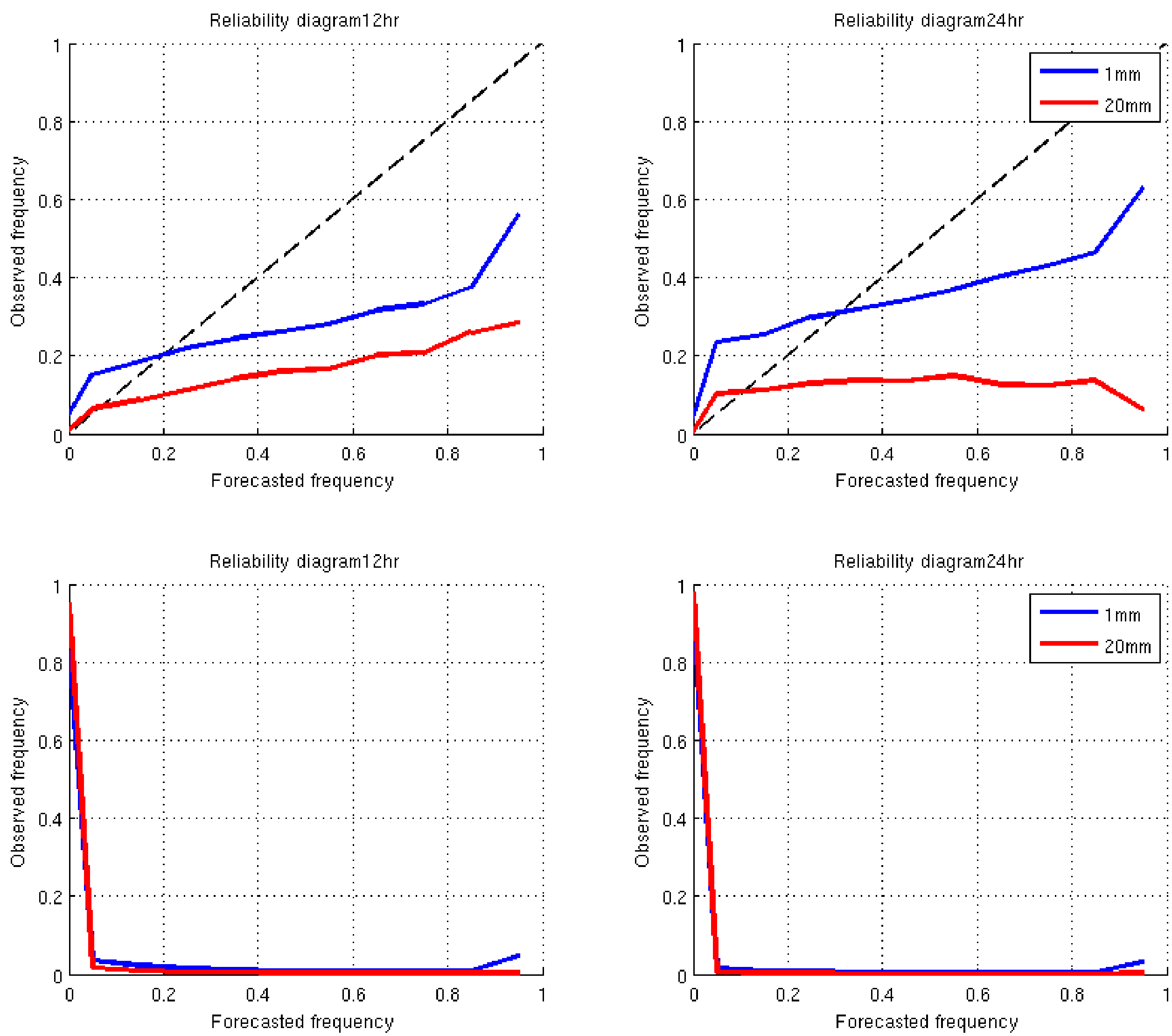


Figure 27: Computed over the entire sample

Brier skill score as a function of the forecast lead time for the 1 mm and 20 mm thresholds.

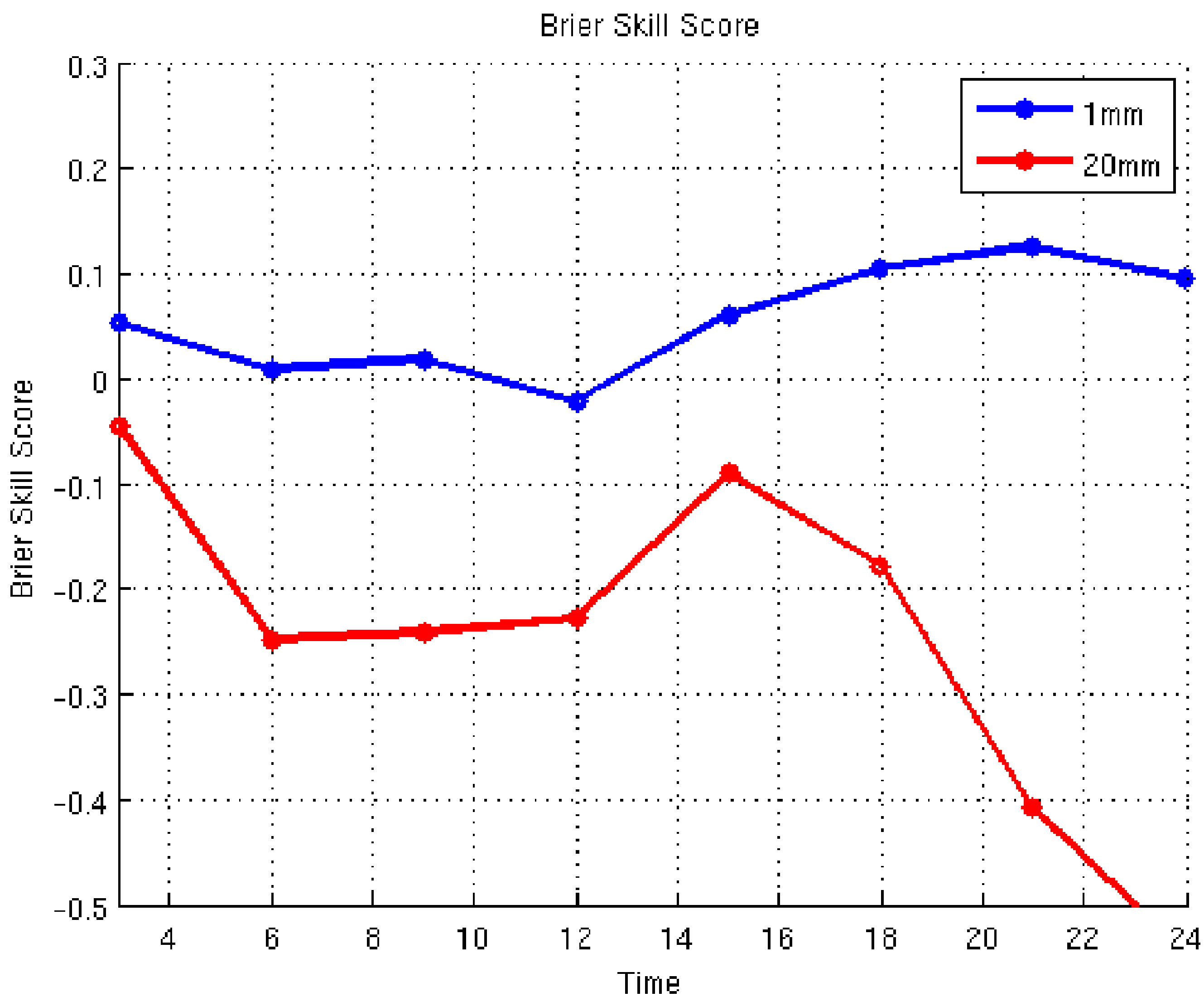


Figure 28: Computed over the entire sample.

Probability of having 3 hourly accumulated precipitation over 10 mm. Ensemble of forecast was generated using a pragmatic ensemble approach (Theis, 2005) of deterministic forecast (max displacement= 140 km.

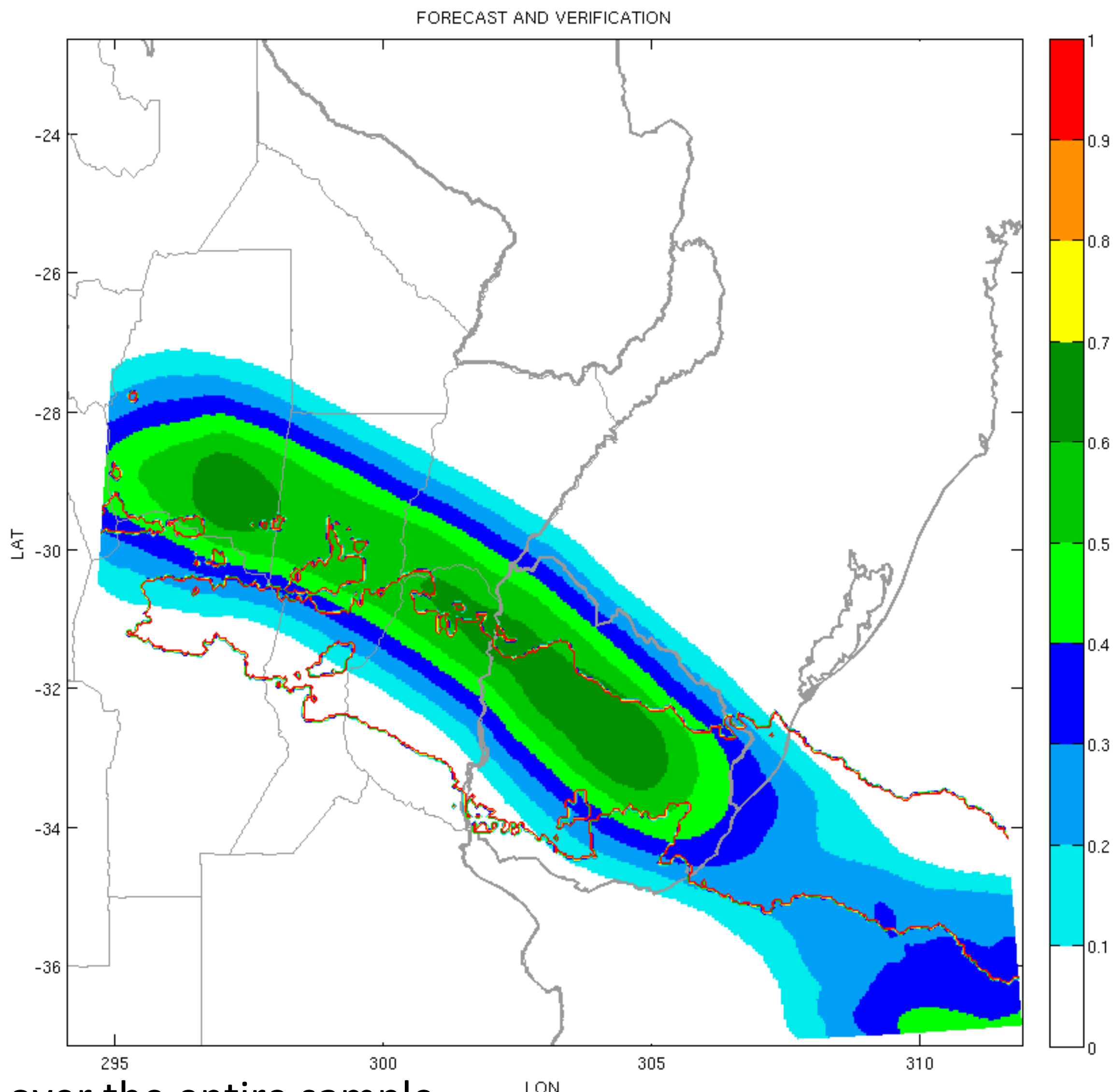


Figure 29: Computed over the entire sample

Difference between forecasted and "observed" probability.

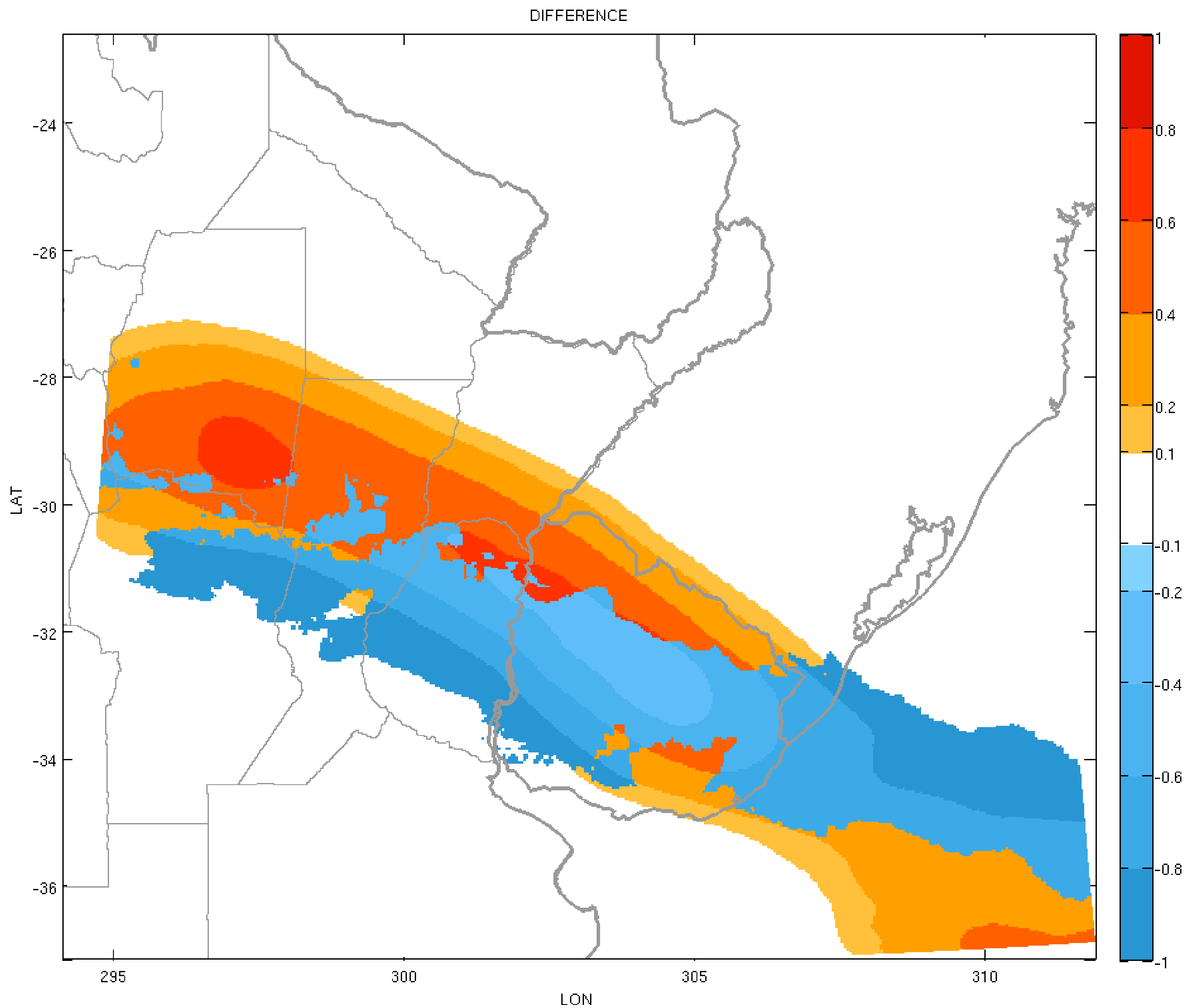


Figure 30: Computed over the entire sample.

Reliability diagrams for 1 mm and 20 mm thresholds as a function of forecast lead time. Probabilistic forecast was generated using horizontal displacement of deterministic forecasts (maximum displacement 240 km).

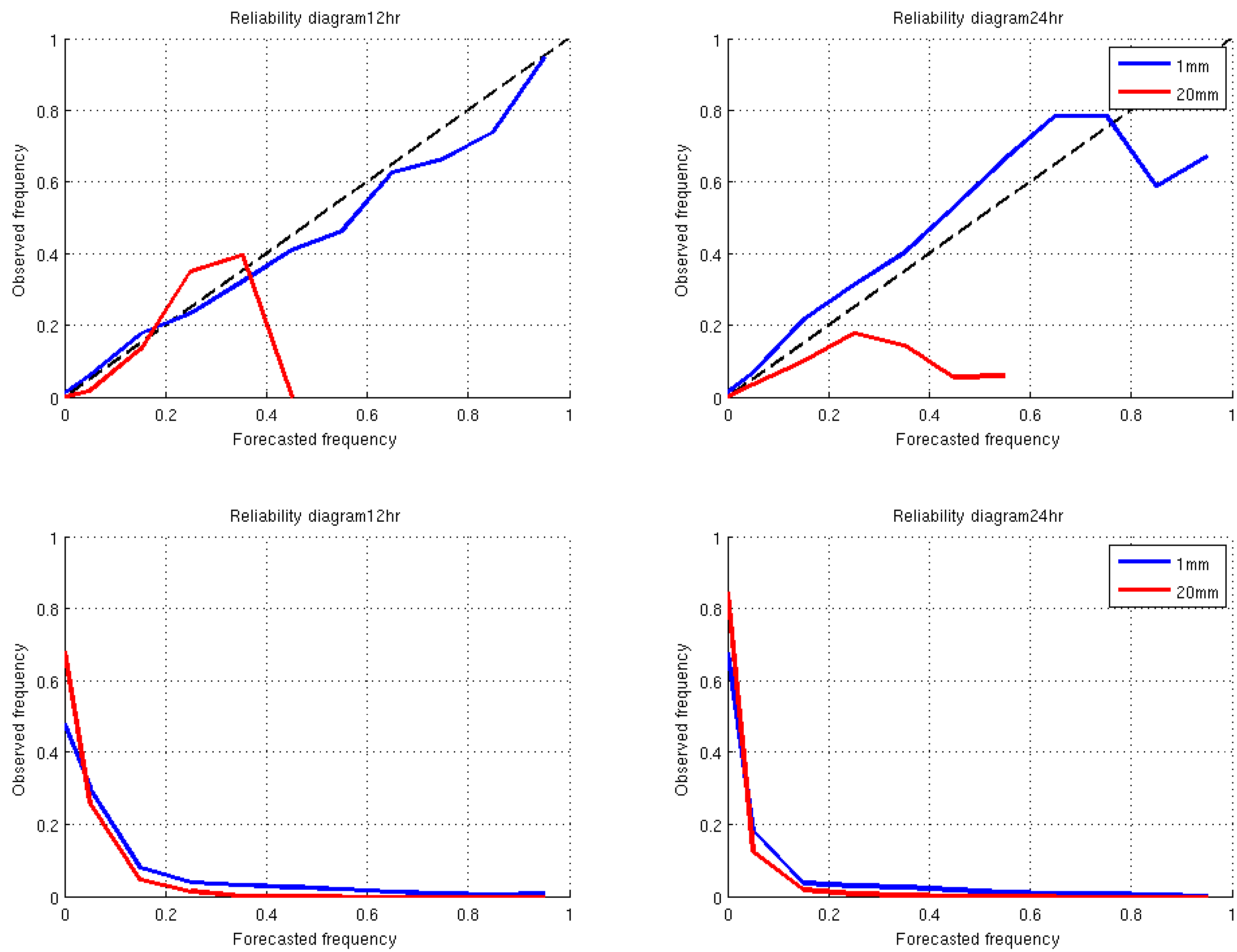


Figure 31: Computed over the entire sample.

Brier skill score as a function of the forecast lead time for the 1 mm and 20 mm thresholds.

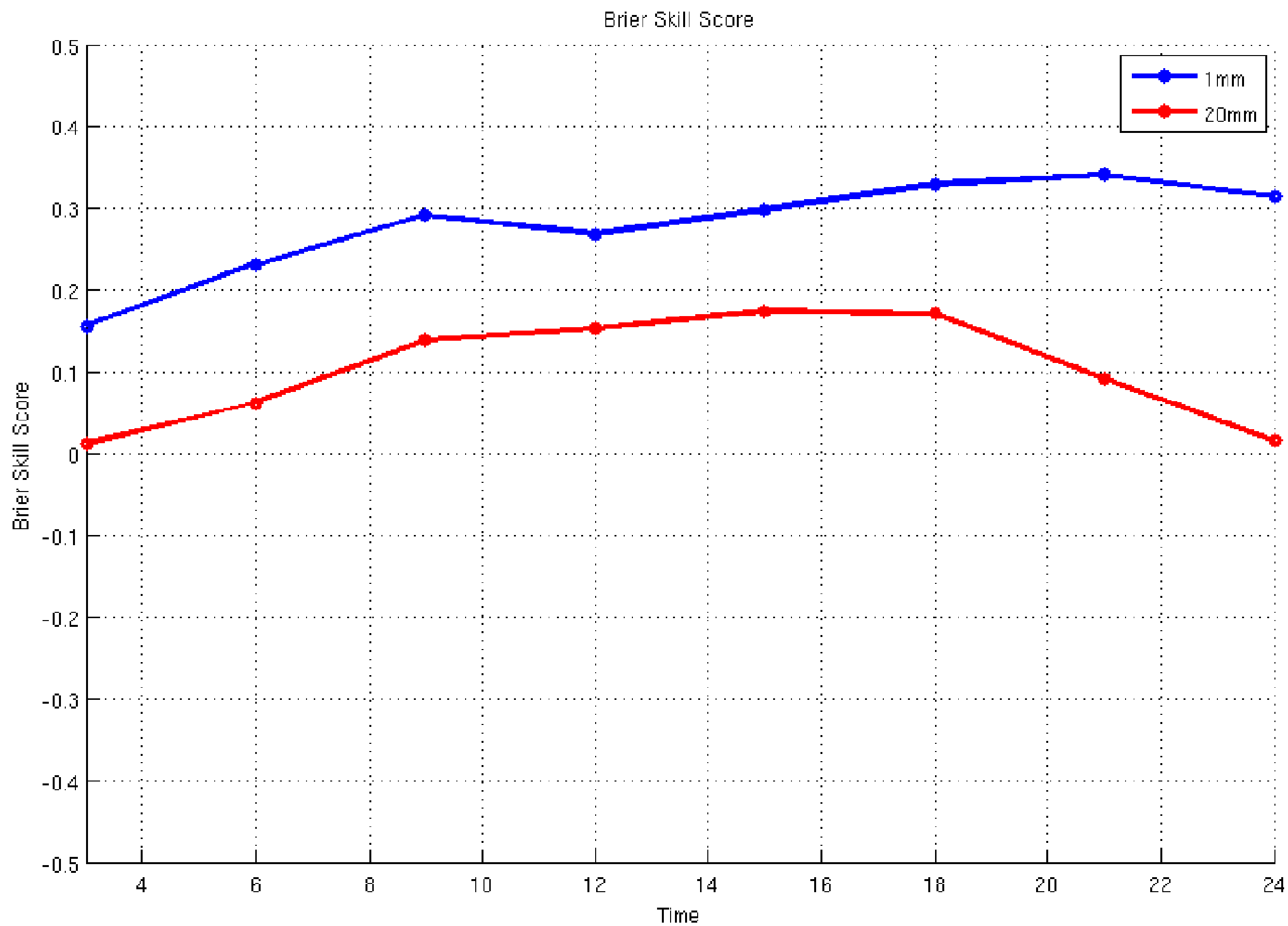


Figure 32: Computed over the entire sample