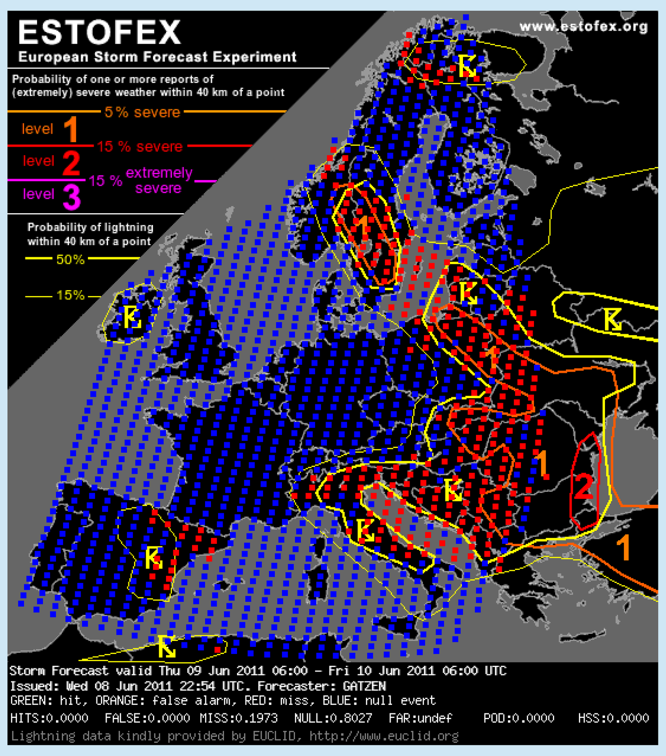
European Storm Forecast Experiment

Verification of multi-categorical thunderstorm forecasts over the period 2009-2013

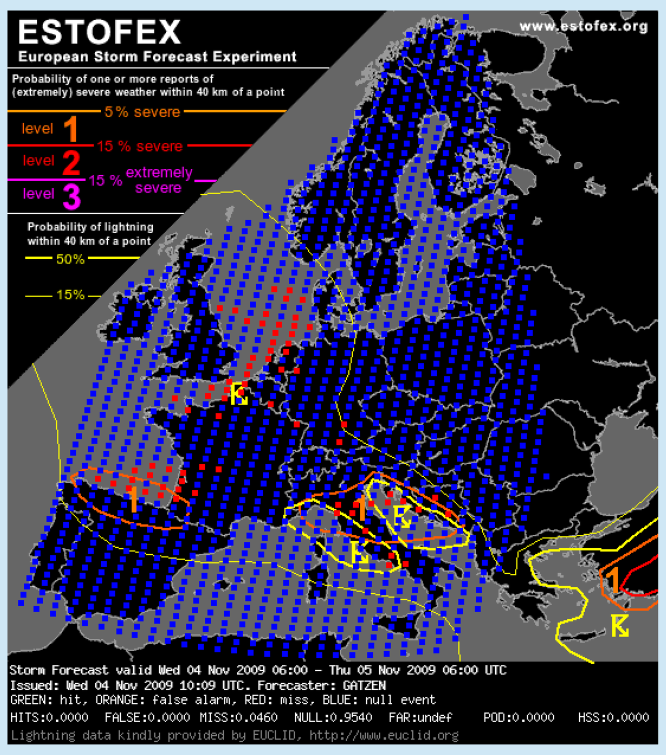
Oscar van der Velde, Pieter Groenemeijer, Helge Tuschy, Christoph Gatzen, Johannes Dahl, Oliver Schlenczek, Tomas Pucik, Marko Korosec, Georg Pistotnik



ESTOFEX has been issuing severe storm forecasts since 2002. The past yes/no thunder forecasts have already been subjected to verification [1,2]. Since September 2009 we started publishing two thunder probability lines. This allows a forecaster to convey more accurately the level of activity expected. We have tentatively marked the two lines with 15% and 50% probability. These separate three areas of (very) low, medium and high probability.

To the left, two example forecasts are presented. The squares represent the 1566 verification grid points (same as in [1]), corresponding to areas equal to circles of radius 40 km. Red indicates locations where lightning was detected by the EUCLID network.

To verify probability forecasts, one needs to slowly gather the relative observed frequency over many forecasts. We do this with histograms summing points across the map with a certain observed frequency of the times it was in each forecast category. The frequency of points in each probability category can also be plotted as map.



Individual forecasts can only be verified in the spatial domain: probability is approximated by the regional coverage of storms. In this sense, it is clear that the top left forecast was practically almost perfect, while the bottom forecast, although not missing any event, does not reflect the occurred clustering of storms so well. Here we will numerically verify whether grid points were in the right forecast area by looking at the regional density around that point (9 points).

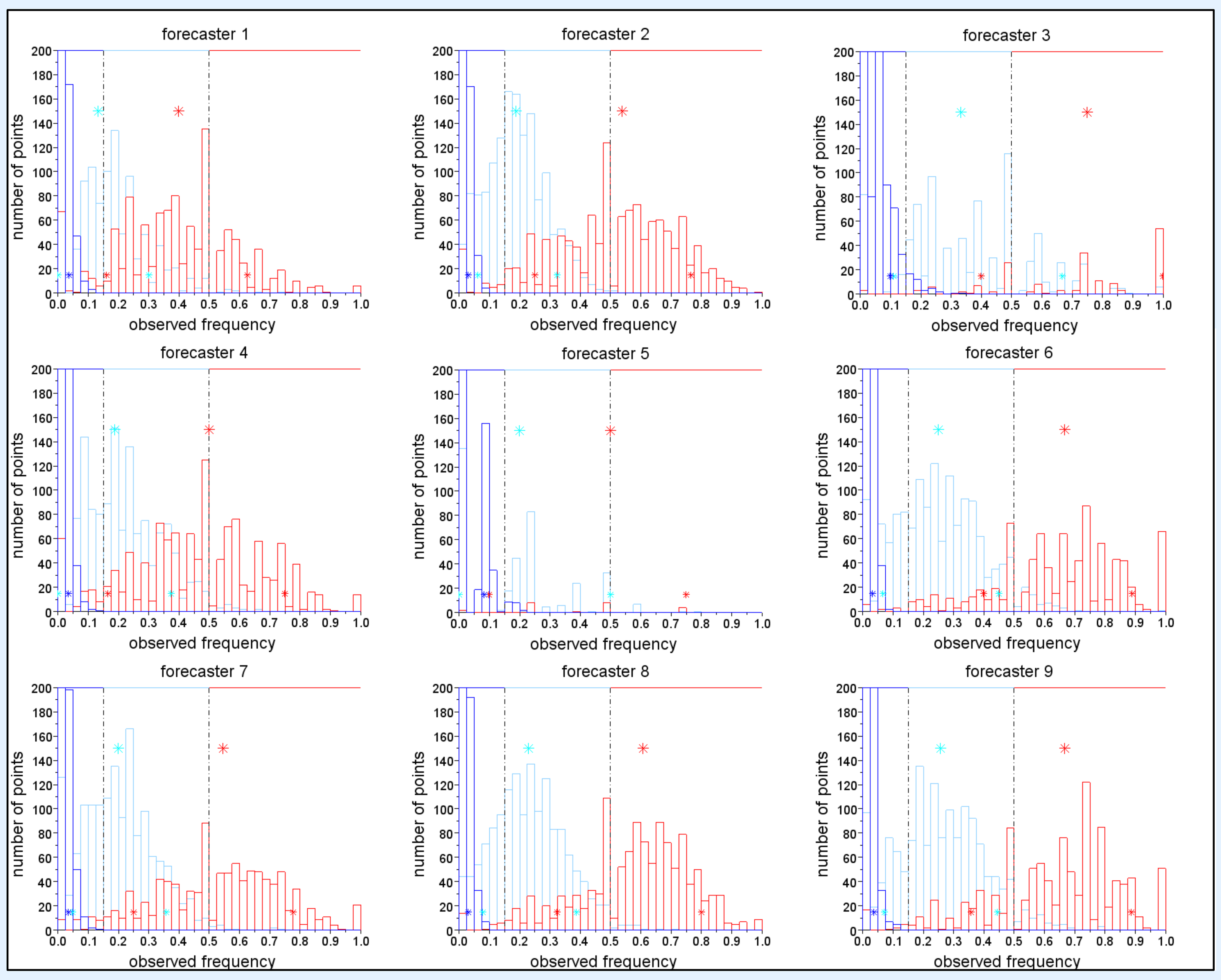
References

[1] Groenemeijer, P. H. *et al.* (2007), Verification of Dichotomous Lightning Forecasts at the European Storm Forecast Experiment (ESTOFEX), ECSS 2007, Trieste. www.estofex.org/files/estofex_lightning_verification.ppt

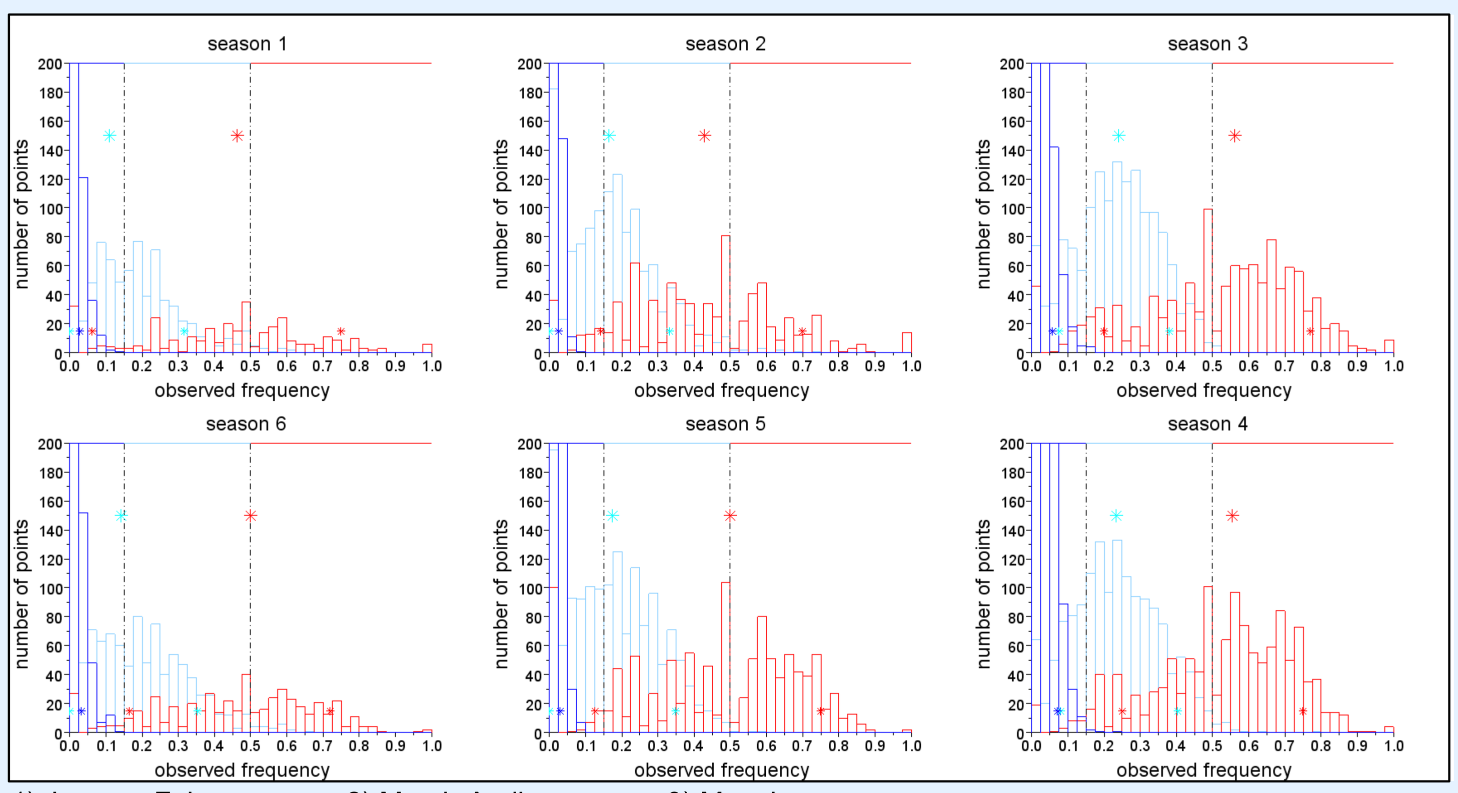
[2] Brooks, H. E., *et al.* (2011), Evaluation of European Storm Forecast Experiment (ESTOFEX) forecasts. Atmospheric Research, Volume 100, Issue 4, June 2011, Pages 538-546

Contact: oscarvdvelde@yahoo.com or inflow@estofex.org

Forecast performance by forecaster and season



Each plot shows three distributions: outside any thunder line (0-15% probability); between thin and thick line (15-50% probability); within thick line (50-100% probability). The star symbols indicate median and 10/90% percentile of each distribution. Number of forecasts: 70, 184, 22; 114, 15, 149; 154, 295, 157.
The graph answers the question: of all the times a point was included in a certain forecast area (e.g. 15-50%), how often was lightning observed? To avoid issues with lightning detection efficiency and histogram spikes, included are all points on the map with >3 thunder days detected per year, included at least 4 times in the corresponding category. Note that forecasters 3 and 5 did not make enough forecasts for solid statistics.



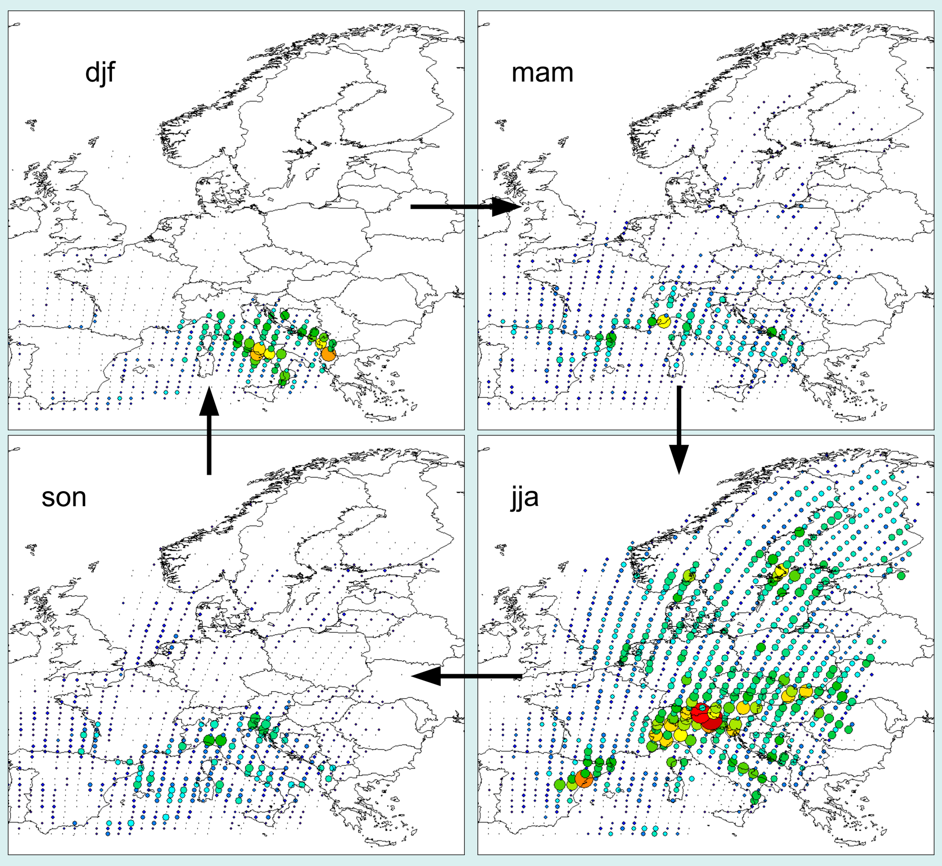
1) January-February 2) March-April 3) May-June
6) November-December 5) September-October 4) July-August

Conclusions

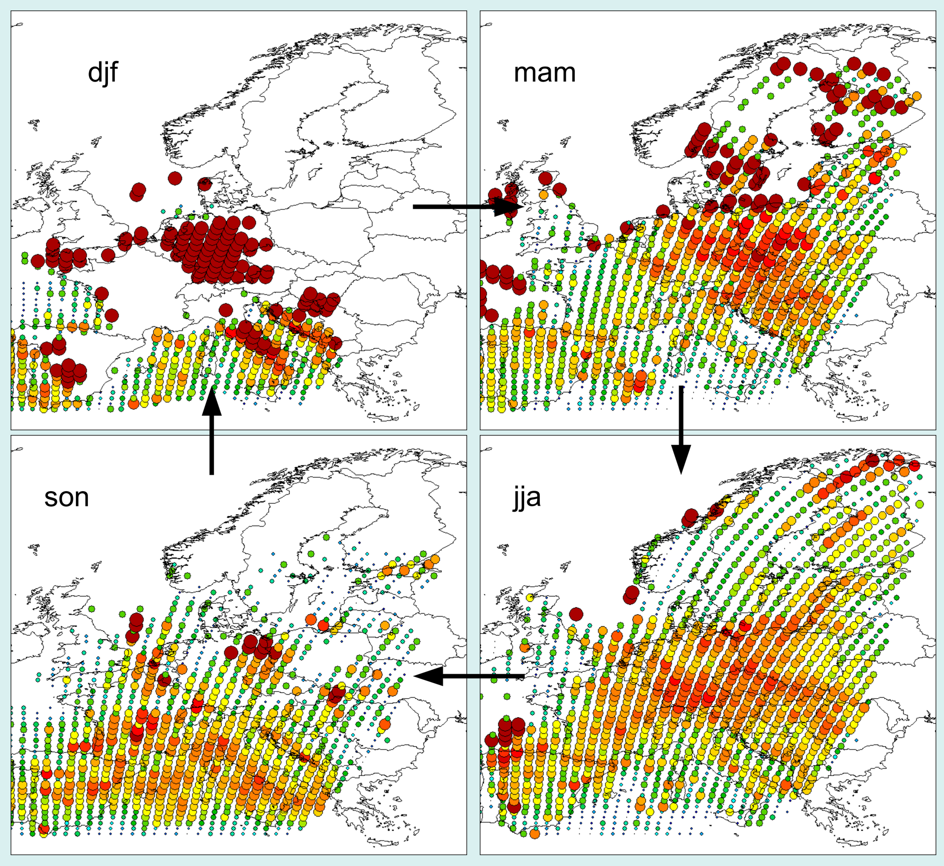
1. Low probability threshold is 8% in summer, only 2% in winter (overforecasting medium probability)
2. High probability threshold is closer to 30-40% especially if we consider the long tail (overforecasting high probability)
3. Forecaster calibration can be grouped: forecasters 6-8-9 (best), 2-7 and 1-4.
3. We should adapt better to seasonal differences.
4. Mapping of biases can identify problem regions and can help future forecasting decisions.
5. No long-term improvement detected.

We would like to thank Dr. Gerhard Diendorfer and EUCLID for their lightning data,which was indispensable for this study and our daily forecast quality monitoring.

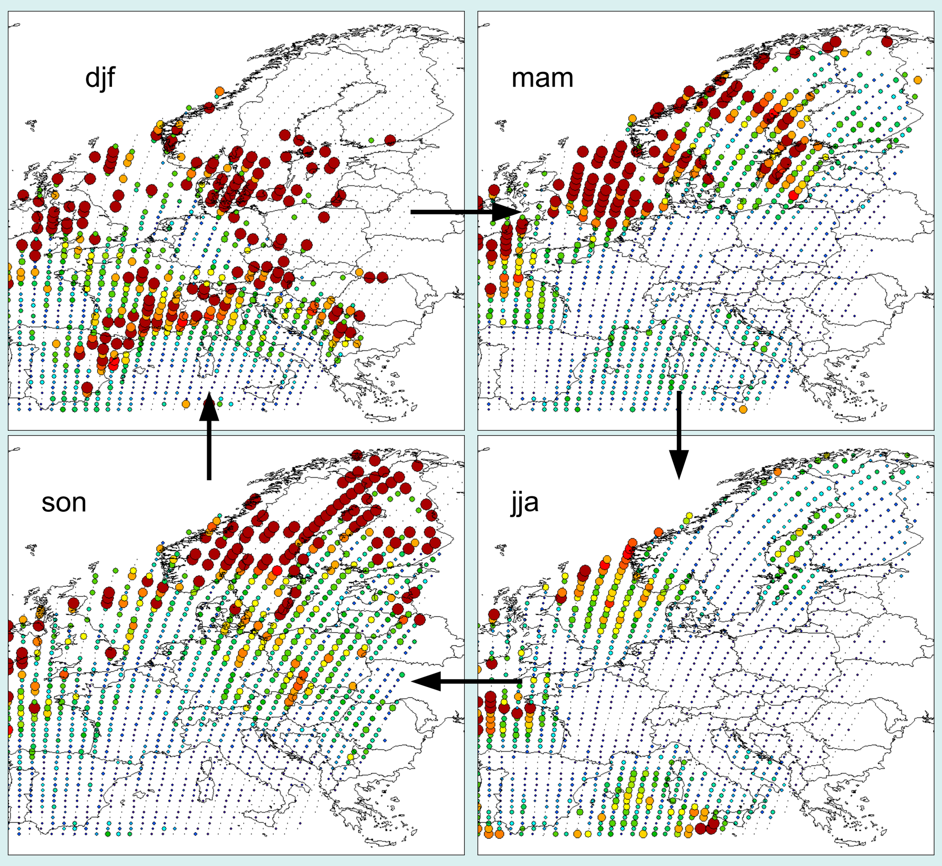
Seasonal forecast error distribution



Observed frequency (0-20%, e.g. green 10%) over all forecasts per season when the point fell outside the thunder areas (<15% probability). This includes days with and without storms. Shows "systematic misses" in areas with frequent storms. In winter misses are concentrated along the upwind slopes of Italy and Balkan. In spring it's the coastal region from Catalonia to northern Italy. In summer, northeastern Spain, northern Italy but interestingly also Netherlands, southern Norway, the east coast of Sweden and Slovakia yield concentrations of misses. In autumn, a band from Biscay to Balearic Sea and Sardinia, and the north coasts of Ligurian Sea and Adriatic trigger unforeseen storms.



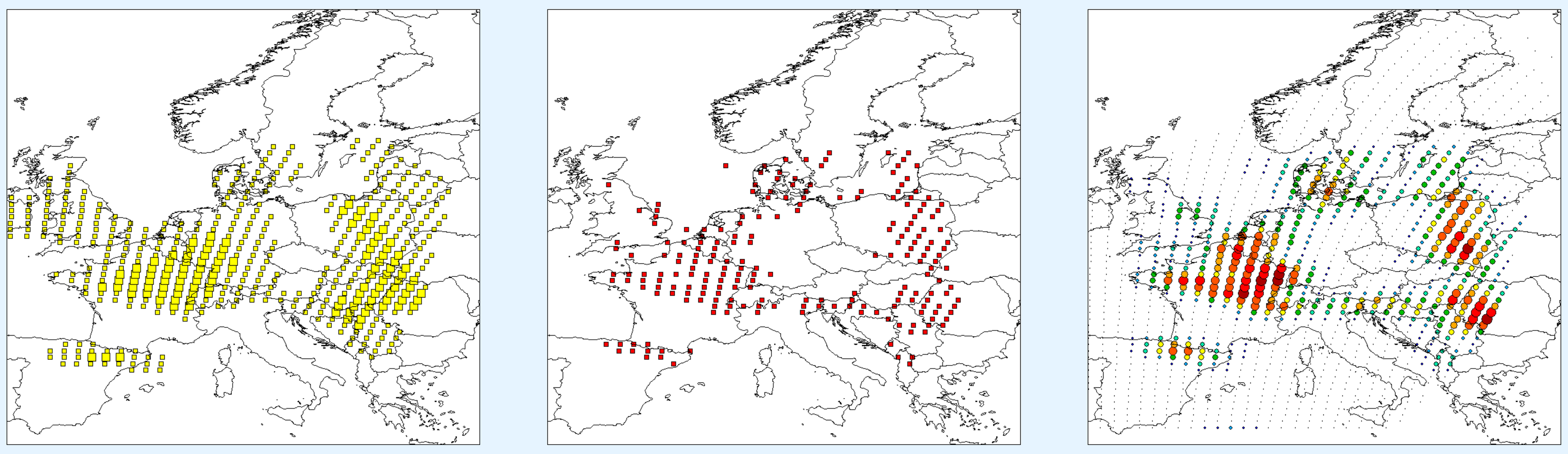
Observed frequency (range 0-100%, cyan 25%, yellow 50%) over all forecasts per season when the point was included in the highest probability area (50-100%). Ignore brown (only 1 forecast, occurred). Red are points with occurrence matching the forecast well. Green and blue are problematic regions where high probability forecasts systematically failed. For example, France in spring or Sweden in summer. The French underoccurrence could perhaps be related to a moisture/CAPE bias in the GFS model which the forecasters use (among others).Another explanation could be that the majority of forecasters live in central Europe and are optimally tuned for their specific continental conditions. Similar patterns appear for the medium probability area observed frequency distribution.



Percentage of occurred thunderstorms (range 0-100%, cyan 25%, yellow 50%) which fell outside the thunder areas (<15% probability). Shows the "surprise misses" in areas where few thunderstorms occur. The few storms that do occur are often missed correctly, because expected probability based on our judgment of forecasting data and climatology would not exceed 15%.
The smallest dots are locations without storms in that season. The majority of brown circles represent just single thunderstorm appearances.

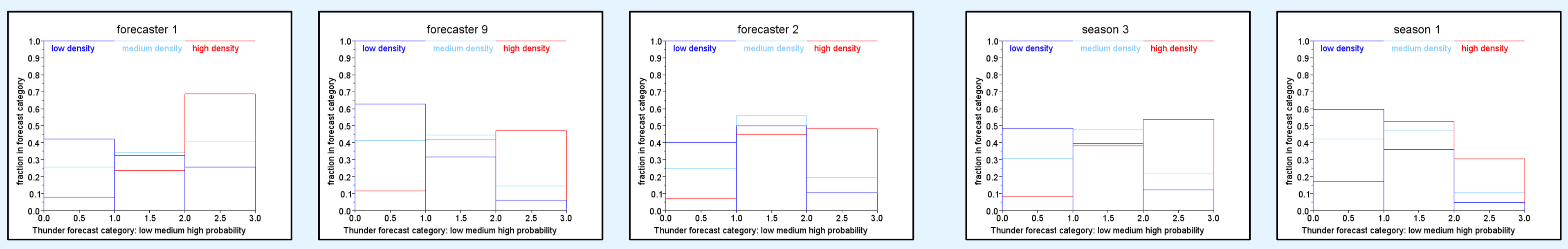
Several features can be noticed:
In winter, surprise misses concentrate along the north coast of the western Mediterranean Sea. Also downstream of the Adriatic storms can percolate deeper inland than we expected. Note that we do not see this in the autumn months, which is active there.
In spring, surprise misses concentrate west of France, north of Denmark and over the Baltic Sea and Sweden.
In summer, the northern Baltic Sea, northern North Sea, Celtic Sea and the western Mediterranean Sea show most surprises.
In autumn it can be noted that the entire northeastern 1/3rd of the European land shows a significant amount of surprise misses, suggesting a role for large active low pressure areas when the land has not cooled down yet (in winter it stops).

Performance of individual forecasts (storm coverage)



Forecast Observed 9-point Observed Density

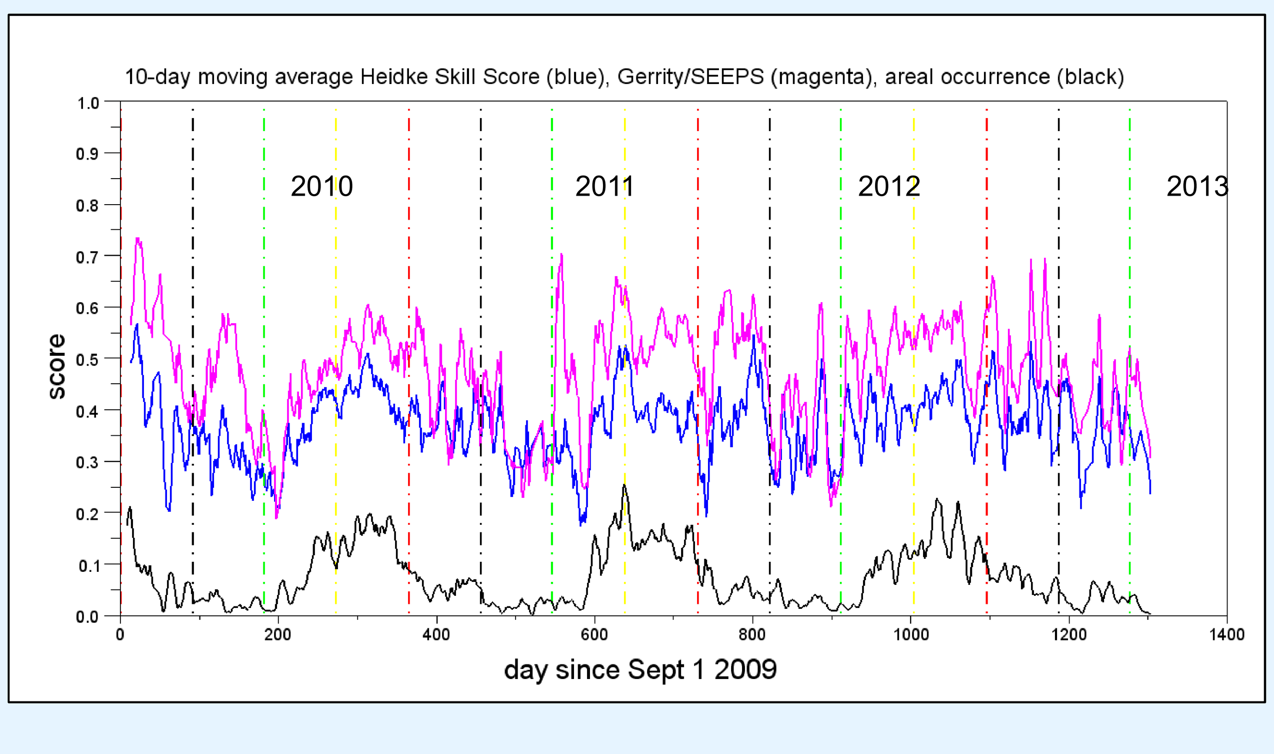
In a forecast of multiple probability categories, high probability should ideally also correspond to a high storm coverage. The medium probability area is usually used to express uncertainty about the expected coverage, but at times one can be certain about a relatively low coverage of storms. The low probability area (outside the thunder areas) may ideally catch any isolated storms not easily accounted for by forecast areas. A way to verify single forecasts is to convert the observed storm points into "regional density" using 8 surrounding points. These may then be compared to the forecast probability, e.g. 2 out of 9 = 22% (should fall in 15-50% forecast region).
In this example, the British Isles medium thunder probability was not warranted, except in the eastern UK. The Spanish maximum was offset.



Question: If a thunderstorm occurred, which forecast probability area was it included in? Was the forecast for that day appropriate for this location? The graphs show the fraction of times a storm point surrounded by a 9-point regional coverage (low 1/9, medium 2-4/9,or high 5-9/9) was included in a forecast area category (low <15%, medium 15-50%, high 50-100%).
Ideal case: red at the right, blue at the left.

Three forecasters with different characteristics are highlighted:
Forecaster #1 made large areas of high probability. He caught indeed a large fraction of high density storms, but also a significant amount of very isolated and medium coverage inside those areas (blue bars in the righthand column). **Forecaster #9** is the opposite, making smaller areas with more misses but fewer false alarms. **Forecaster #2** is the most effective in catching medium density into medium probability areas, but in doing so, he makes the areas too large and catch isolated storms and false alarms.

Top right: in winter ("season 1" graph, DJF) high regional storm coverage occurs more often in the medium than in the high probability forecast areas. Also the misses (red and light blue bar in left column) are almost doubled compared to the summer ("season 3" graph, JJA)



The evolution of individual forecast scores over time (1 Sept 2009 – 27 March 2013) is displayed above. Vertical lines are placed at the first day of the months **December**, March, June and September.