A new scale decomposition of the Brier score for the verification of spatial probabilistic forecasts

B. Casati, L.J. Wilson

Meteorological Service of Canada

Environnement Canada Environment Canada

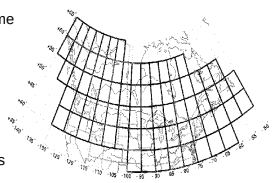
1. Aims and Motivations

Weather is characterized by features on different scales.
How accurate are the forecasts on different spatial scales ?

AIMS:
1. Assess the forecast **quality/skill** separately **on different scales**.
2. Determine the **no-skill/skill transition scale**.
3. Verify the forecast ability to reproduce the observed **scale structure**.

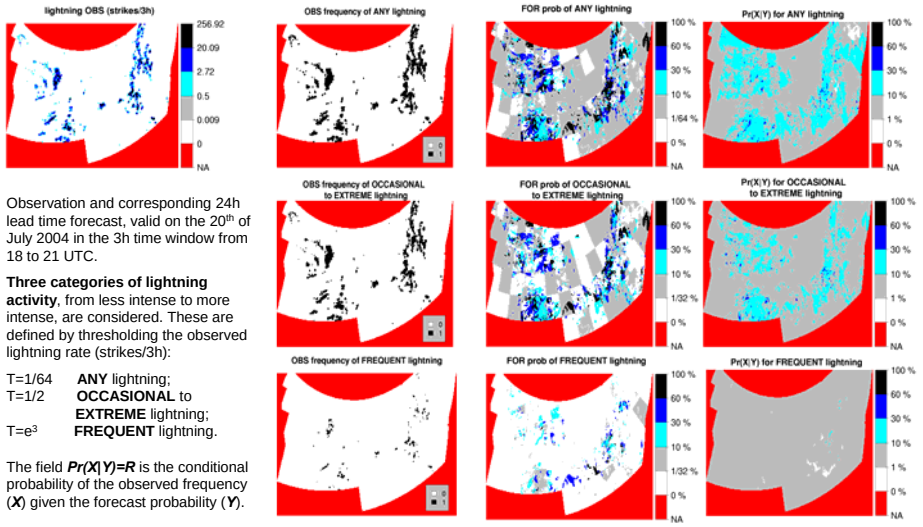
2. The CMC lightning probabilistic forecast

- **Probability of lightning occurrence** exceeding some set thresholds; 3 hour time windows, 24 km resolution.
- **Predictors** from the GEM 24 km resolution NWP model.
- **Statistical model**: tree structure regression models for each month and 5° x 5° lat-lon sector.
- **Training data set**: summers 2000 and 2001.
- **Observations** from the North America Lightning Detection Network (Vaisala Inc.).



W.R **Burrows**, C. Price, L.J.Wilson **(2005)**: "*Warm season lightning probability prediction for Canada and the northern United States*", **Wea & For, vol 20**, pp 971-988

3. Case study



Observation and corresponding 24h lead time forecast, valid on the 20th of July 2004 in the 3h time window from 18 to 21 UTC.

Three categories of lightning activity, from less intense to more intense, are considered. These are defined by thresholding the observed lightning rate (strikes/3h):

T=1/64	**ANY** lightning;
T=1/2	**OCCASIONAL** to **EXTREME** lightning;
T=e³	**FREQUENT** lightning.

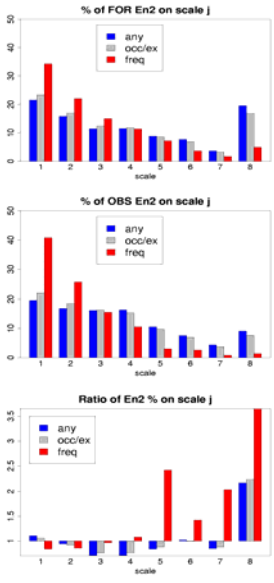
The field ***Pr(X\|Y)=R*** is the conditional probability of the observed frequency (***X***) given the forecast probability (***Y***).

Step 1. The forecast probability and observed frequency fields are decomposed into the sum of components on different scales by a 2D Haar wavelet filter:

$$X = \Sigma X_j$$

Each scale component X_j is a field with resolution of 24, 48, 96, 192, 384, 768, 1536, 3072 km ($j=1,\ldots,8$).

Step 2. The energy and energy bias are evaluated on each scale. These provide information on the scale structure and over/under-forecasting on different scales.



$$En2(X) = \overline{X^2}$$
$$En2(X_j) = \overline{X_j^2} \Rightarrow$$
$$\overline{X^2} = \Sigma \overline{X_j^2}$$
$$\%\ of\ En2(X_j) = \overline{X_j^2} / \overline{X^2}$$

The energy indicates the quantity of events present in the field at each different spatial scale.

The energy bias assesses over and under-forecasting on different scales.

The percentage of energy on different scales describes how the events are distributed across the scales.

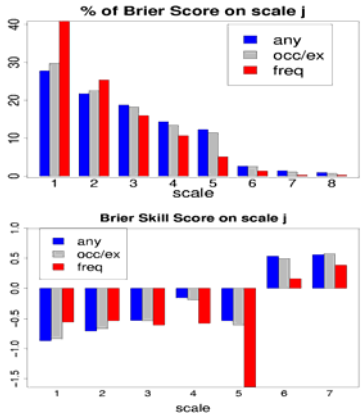
The ratio of forecast and observed energy percentages provides feedback on the differences between forecast and observed scale structure.

Case study: the quantity of events decreases as the scale increases; the frequent lightning category has more events on small scales and less on large scales; the largest scale shows the overall over-forecasting; 400 km features are over-forecast for the frequent lightning category.

4. Verification strategy

Step 3. Brier Score (BS) and Brier Skill Score (BSS) are evaluated on each scale. These provide feedback on the quality and skill on different scales and the no-skill/skill transition scale.

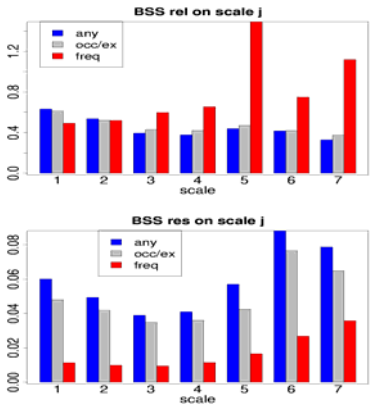
$$E = Y - X \;;\quad BS = \overline{E^2} \;;\; BS_j = \overline{E_j^2}$$
$$\%\ of\ BS_j = BS_j / BS$$
$$BSS_j = 1 - BS_j / \sigma_{X_j}$$



Case study: positive skill at large scales, negative skill at small scales, transition at 500 km; the frequent lightning forecast exhibits strong negative skill at the 400 km scale due to the over-forecasting of such features.

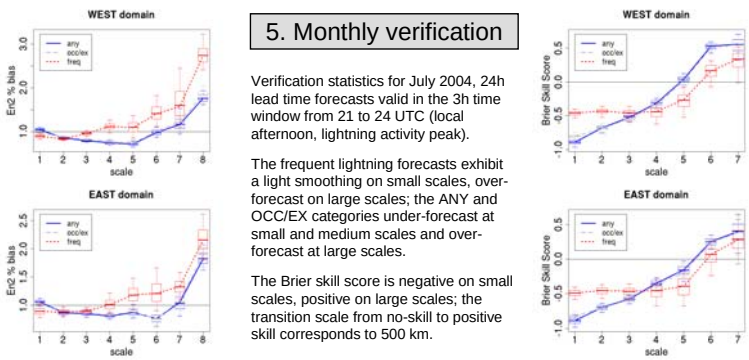
Step 4. The reliability and resolution components of the Brier score and skill score are evaluated on different scales. These diagnose which scales exhibit poor or good reliability and resolution and how these quantities affect the skill.

$$BS_{rel,j} = \overline{(Y_j - R_j)^2} \qquad BSS_{rel,j} = BS_{rel,j} / \sigma_{X_j}$$
$$BS_{res,j} = \overline{(R_j - \overline{X}_j)^2} \qquad BSS_{res,j} = BS_{res,j} / \sigma_{X_j}$$



Case study: the frequent lightning forecast lacks reliability on scale 5 due to the 400 km feature over-forecasting; large scales exhibit better resolution than small scales; the good resolution on scales 6 and 7 compensates the lack of reliability on the same scales, for all the three categories, leading to positive skill on these scales.

5. Monthly verification



Verification statistics for July 2004, 24h lead time forecasts valid in the 3h time window from 21 to 24 UTC (local afternoon, lightning activity peak).

The frequent lightning forecasts exhibit a light smoothing on small scales, over-forecast on large scales; the ANY and OCC/EX categories under-forecast at small and medium scales and over-forecast at large scales.

The Brier skill score is negative on small scales, positive on large scales; the transition scale from no-skill to positive skill corresponds to 500 km.

6. Conclusions

A new diagnostic verification technique for probabilistic forecasts defined over a spatial domain is presented. The technique is able to:

1. Quantify the forecast **bias on different scales** and provide information on the **scale structure** of forecast and observation fields.
2. Measure the **error and skill on different spatial scales** and determine the **no-skill skill transition scale**.
3. Evaluate **reliability and resolution on different scales**.

For the **CMC lightning probabilistic forecast**, the **transition scale** from no-skill to positive skill corresponds to the scale of the **5° x 5° lat-lon sectors** used to construct the tree structure regression model. The forecasts resolve the sector scale and larger scales but do not resolve the features within the sectors. The development of a new model, no longer based on the 5° x 5° lat-lon sector, has recently been started.

barbara.casati@ec.gc.ca