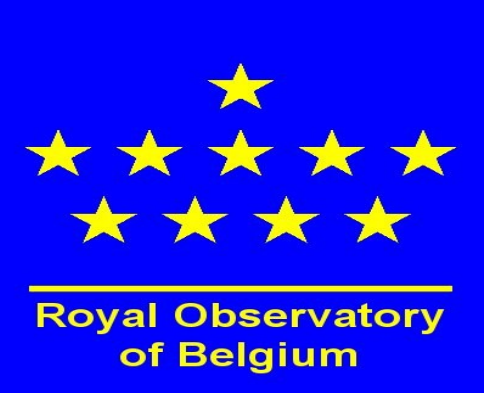

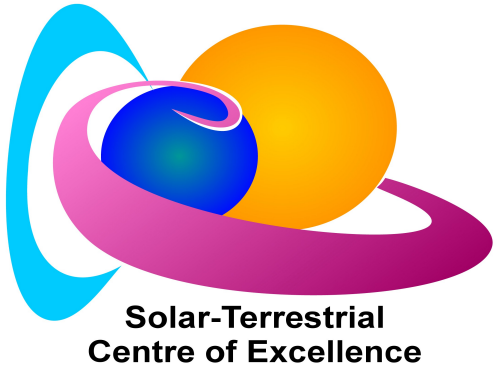

Discussion focus: Flare forecast verification

Jesse Andries and Andy Devos

Verification and validation

Forecast validation and verification:

- Statement
- How true was your statement?
- How useful was your statement?

Issues at each of those steps

Statement

- Is the statement well defined?
- Information content of the statement: Capture to the largest extent possible your knowledge about the future state of the system
- Also in the light of your capabilities:
  Admitting your weaknesses(uncertainty) is a strength!
- Whether the future state is uncertain because you are not capable enough or whether it is intrinsically uncertain is rather irrelevant.
  → probabilistic forecast

Concrete for Flare forecasts:

- Well defined: Prediction window?; C, M, X bin definitions? Which quantities: frequency of flares, fluence, max X-ray in absence of flare?
- Well defined: Standards: Rigid enough to allow unambiguous interpretation, but flexible enough to allow a variety and innovation in the statements made.
- ISES scales: quiet/C/M/X flares expected
  <> We expect C flares with also a significant probability for M flaring
- Human forecasting from experience <> statistical models
- Hit/miss statistics <> probabilistic verification

How true was the statement?

- To be measured against which “truth”, truth may be somewhat subjective in case of regional flare attribution. NOAA flare list or post analysis list on science data.
- To be evaluated on the statements with maximal information content: probabilistic forecast
- Probabilistic forecast: trade accuracy/precision <> reliability
- Reliability is a must!
- Statements must be verified to be maximally true

- Accuracy is wanted as high as possible but possibly limited by nature
- Long term probabilistic statistics reaches near complete reliability but with minimal accuracy
- Binary forecast forces precise statement and becoming therefore unreliable

How useful was the statement?

- Users often want binary stoplight forecast. The stoplight forecast should in principle be based on a convolution between the probability density and a cost-function.
- The cost function may/will depend on the customer
- User driven metrics may thus appropriately be hit/miss statistics on the stoplight forecast, but this makes only sense if the stoplight was very properly defined by means of a cost function:
  e.g. very limited value in hitt/miss statistics of X-flare occurrence based on 50% X-flare probability.
- The verification on the underlying probability density forecast is the more fundamental one.

How to advance/ coordinate/collaborate

- Reach agreement on verification techniques to allow consistency & best-practices in order to compare verification of local products.
  But taking into account similar considerations as in the format standardisation: Leave enough flexibility to vary and innovate.
- Inventory of techniques and used languages/libraries/codes (possibly to be shared)
- Share forecasts: in archive form! (WIS?)
- Common products & exact descriptions of each so that verification results are comparable: e.g. please explain exact meaning of errorbars on a probabilistic forecast?