Statistical adaptation for the prediction of rare meteorological events

Frédéric Atger
Météo-France

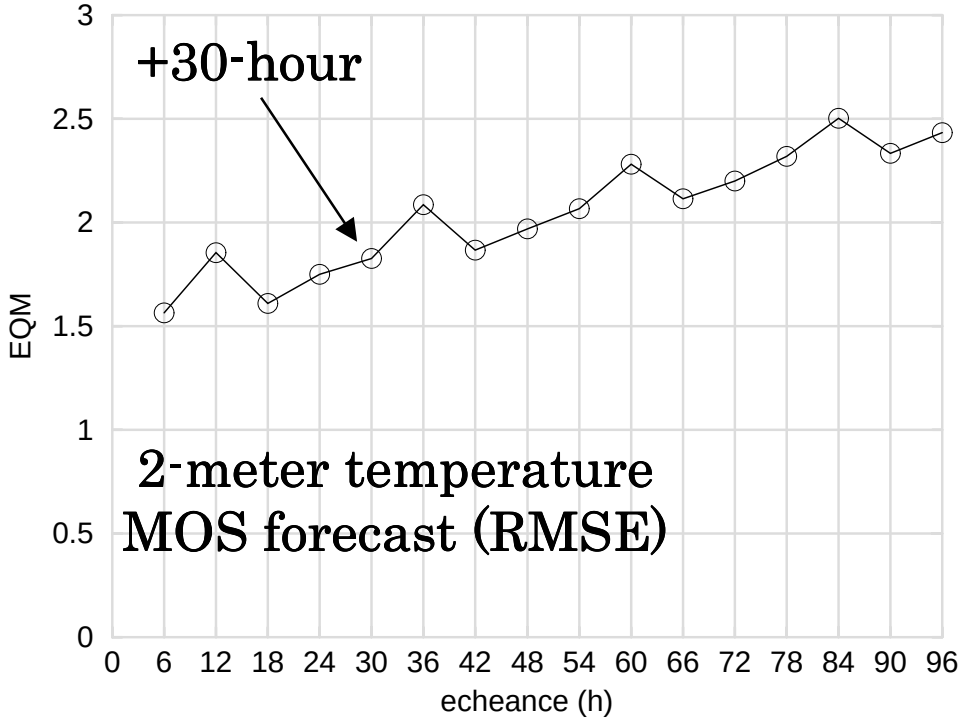
Forecasting rare events through statistical adaptation

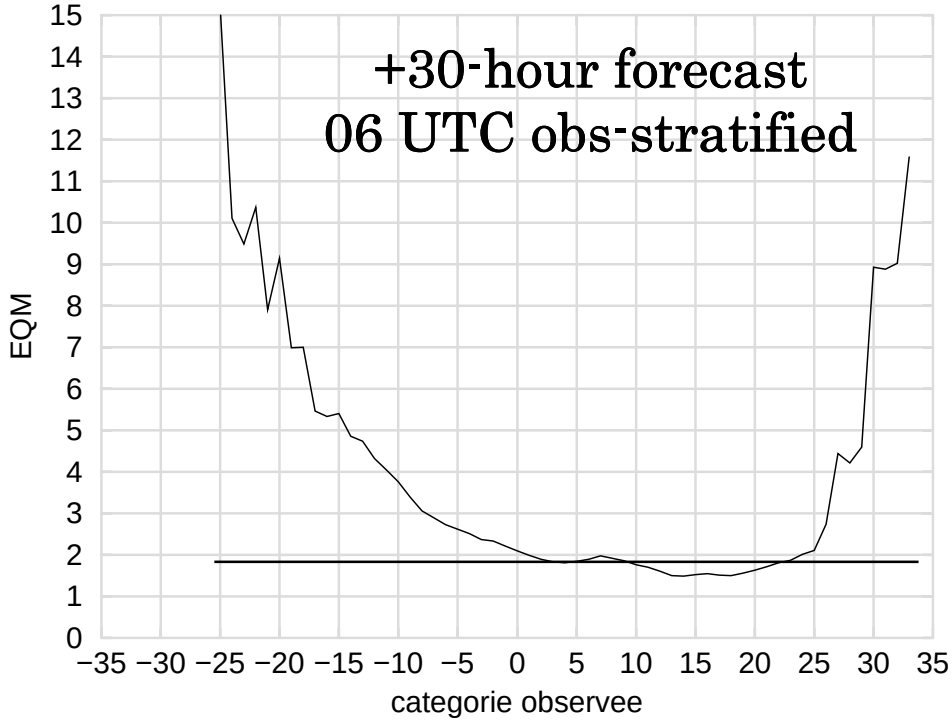
- There is no contradiction between the 2 terms…
- …but in practice emphasis is on frequent events
  - Forecasting well "most-of-the-time" 2-meter temperatures remains the main challenge…
  - …even though there is a definite requirement for predicting large departures from normal
- In the case of certain weather parameters, forecasting frequent occurrences is just useless: wind gusts (who cares for light gusts?), low visibility, low ceiling, etc
- In other cases, predicting extremes is the main requirement, e.g. humidity:
  - Higher values for protecting sensitive equipment
  - Lower values for forest fire prediction

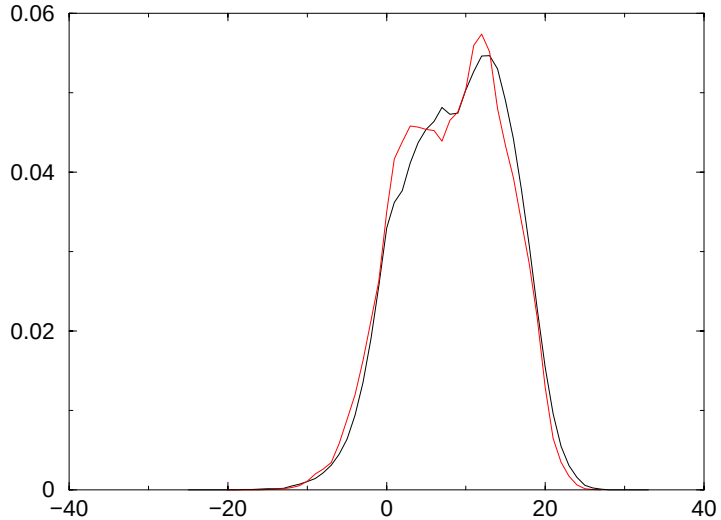
Outline

- Under-sampling
- MOS vs PP
- Verification issues
- Statistical model: linear vs non-linear
- Predictors
- Probabilistic vs deterministic

What's wrong with rare events ?

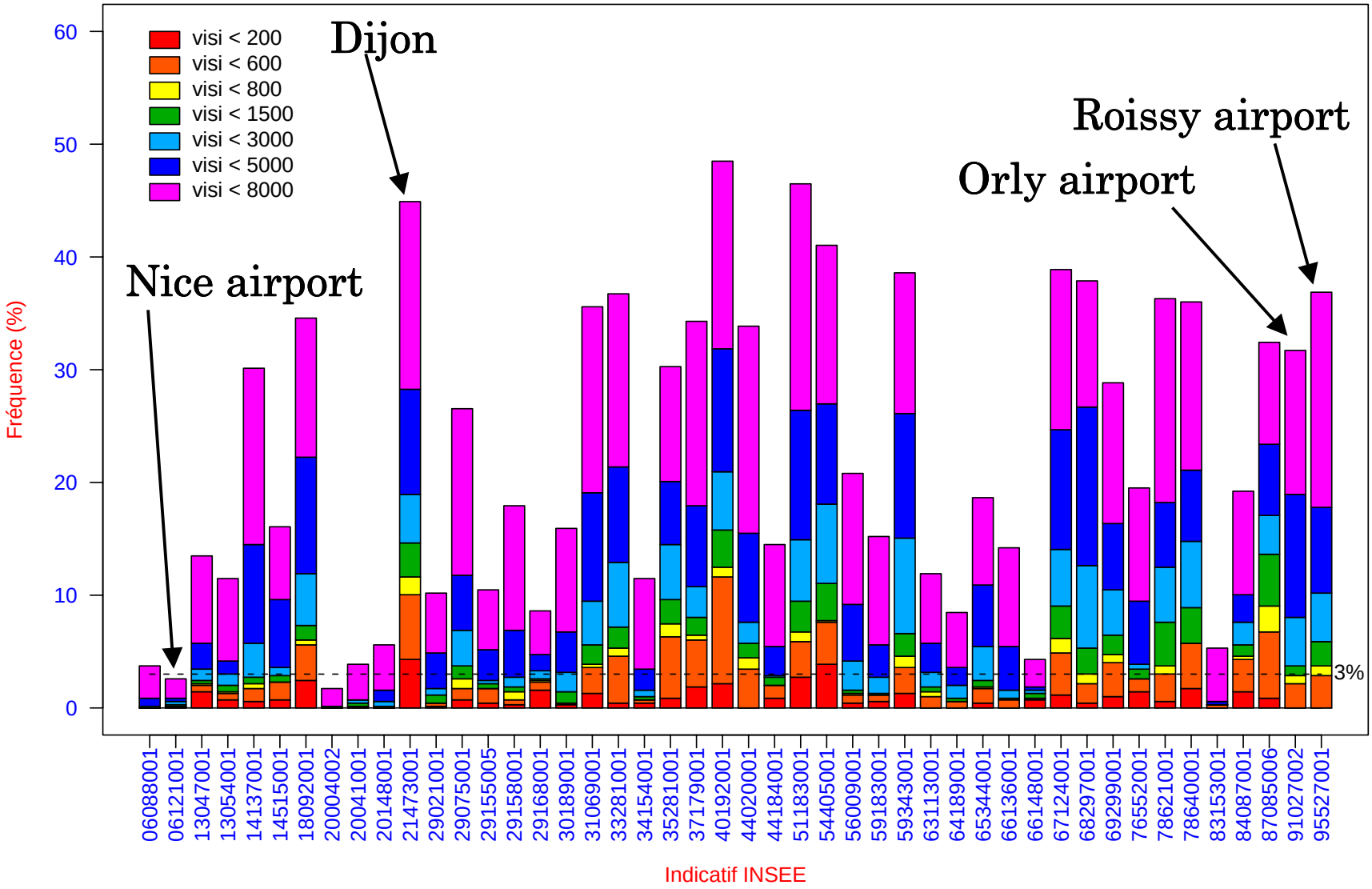






1st problem : under-sampling

Fréquences des différentes classes de visibilité
Hivers 97/98−98/99−99/00−00/01

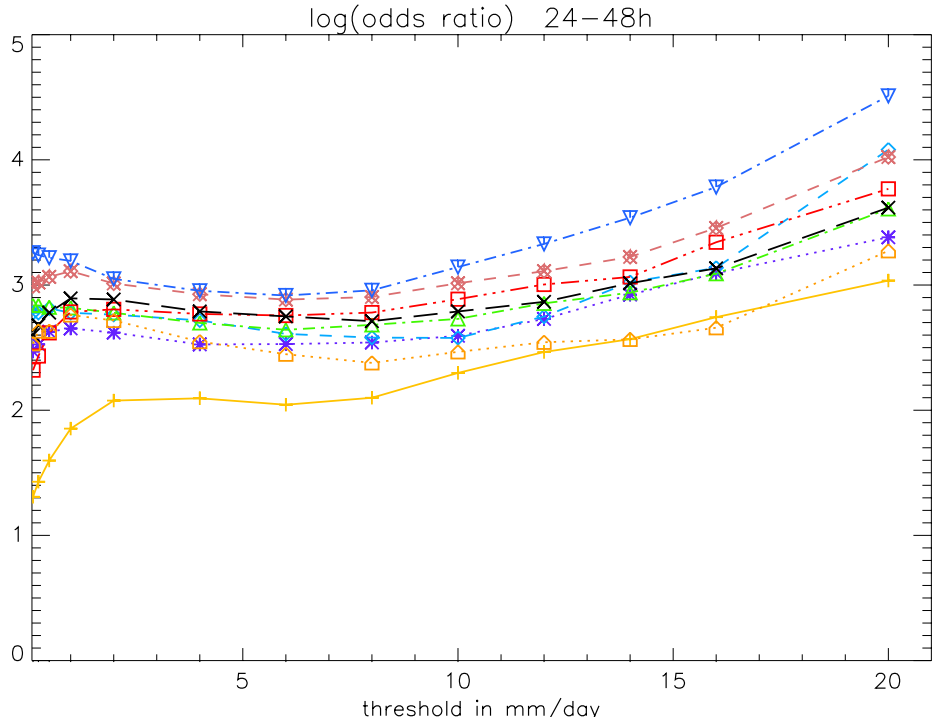


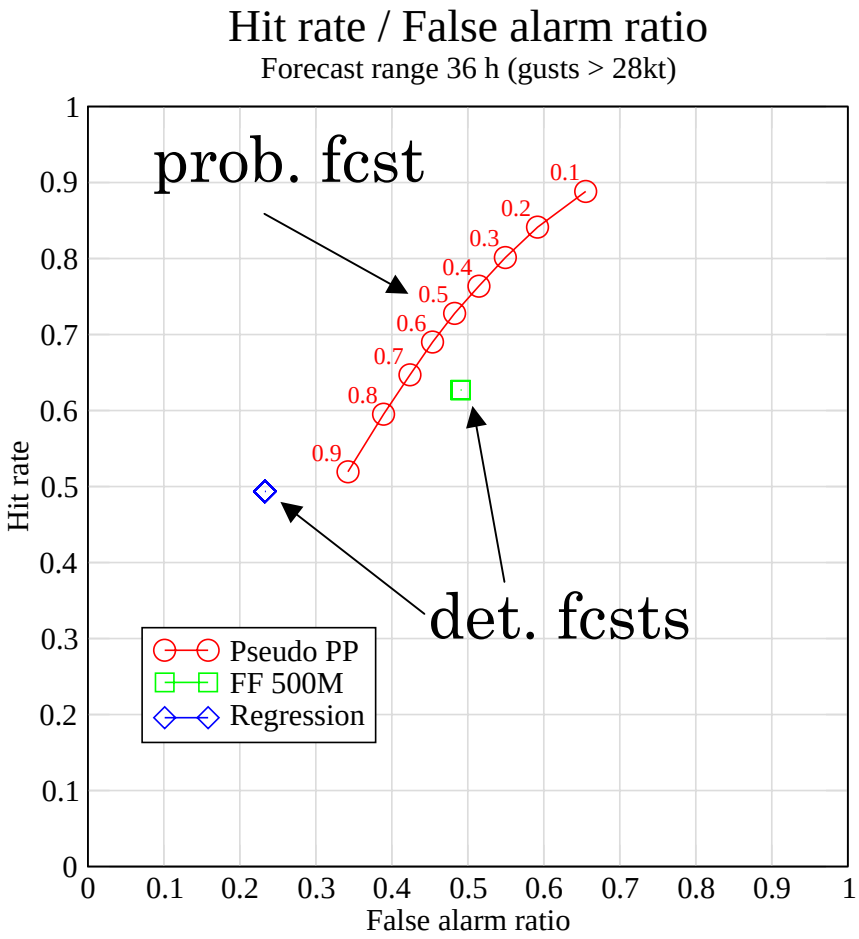
2nd problem : MOS vs PP

- The PP approach seems logical for rare events (larger samples ➔ stability) but:
  - Disregards model errors (analysis and forecast)
  - Not all potential predictors can be used
- The MOS approach takes into account model deficiencies (➔ better scores), but:
  - Samples may be small
  - Forecasts lose their dependency to the model forecast at longer lead-times (➔ uncomfortable interpretation)
- The pseudo-PP approach consists in using the forecasts at earlier lead-times as analyses
  - Samples are still small
  - Model errors are taken into account
  - All predictors can be used
  - And above all: "forecasts stick to the model"

3rd problem: validation

- Most verification criterions are unsuitable, because either too crude and meaningless (RMSE) or too sophisticated

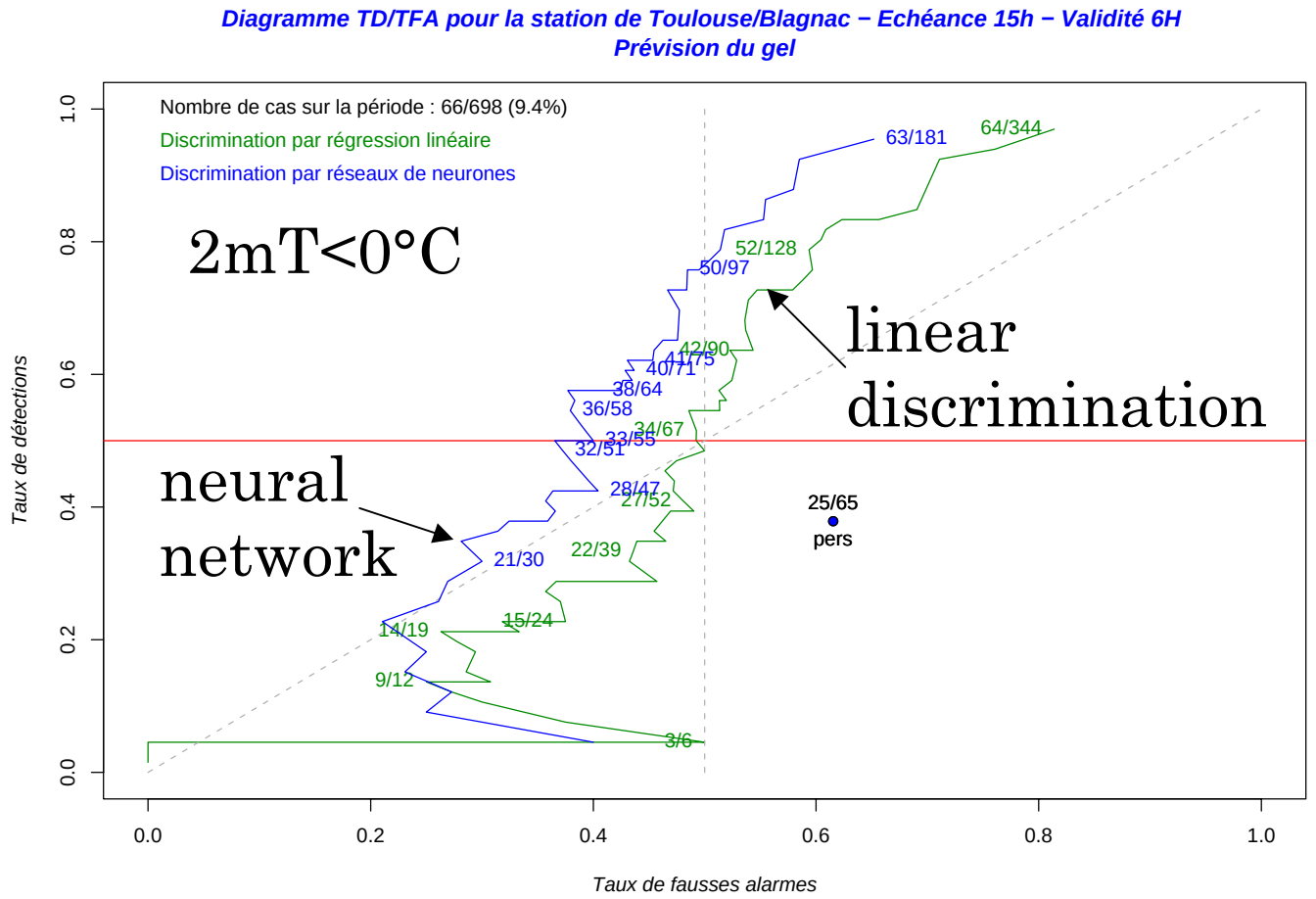




A simple contingency table tells a lot… and basic indicators such as the Hit Rate and False Alarm Rate (or ratio) are sufficient for picturing the main characteristics

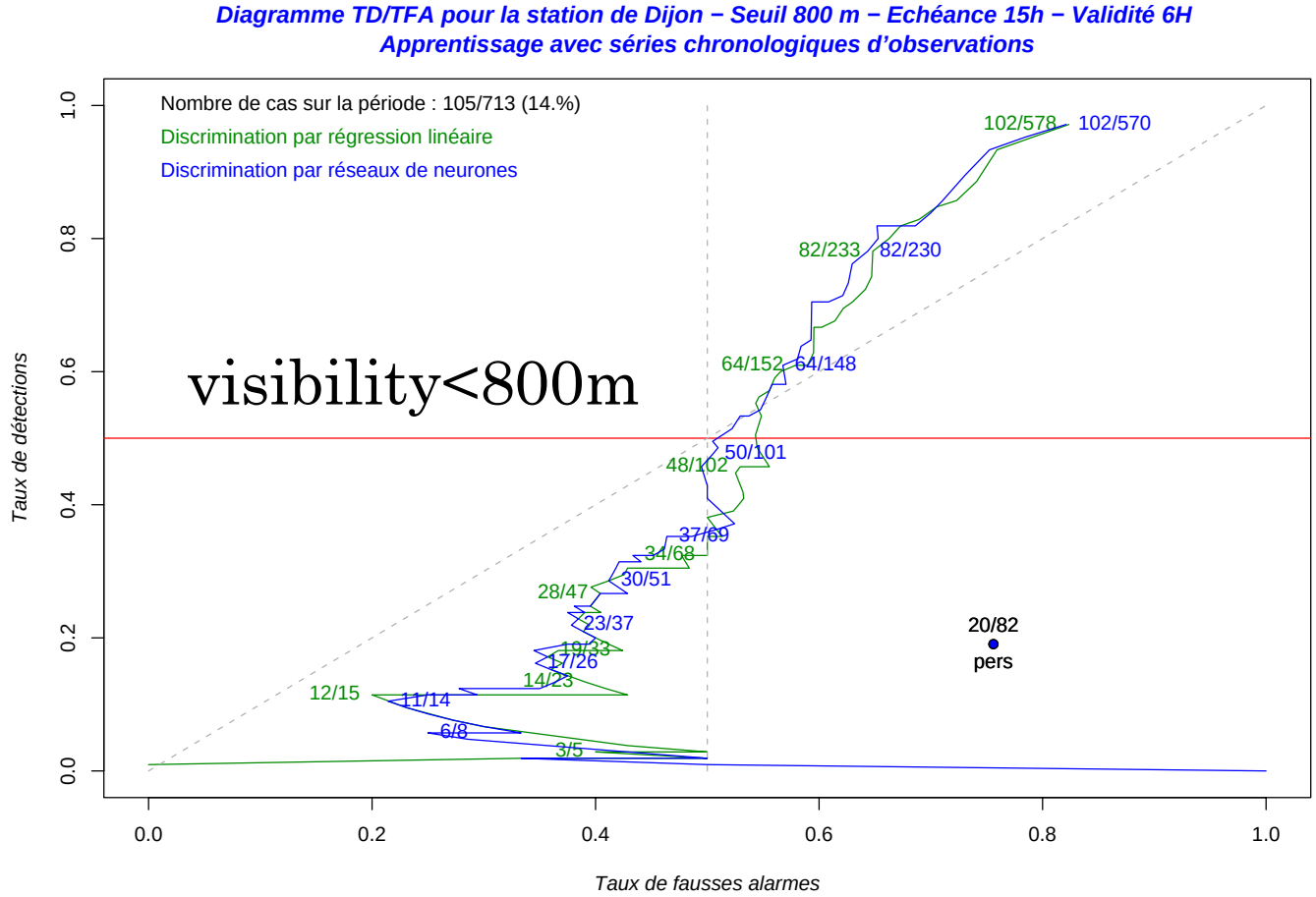
4th problem: linear or non linear?

- Linear models are unlikely to fit the complex relationship between predictors and predictand when considering the tails of the climate distribution ➔Non linear regressions, neural networks, etc, may help sometimes



5th problem: which predictors?

- Whatever the appropriateness of the statistical model, informative predictors are required, i.e. strong relationships between predictors and predictand



Probabilistic or deterministic?

- Statistical adaptation applied to the detection of rare events leads "naturally" to probabilistic products
- Forecasters often need some "first sight" deterministic guidance
- Both presentations should be provided
- Examples of products available on the web (for internal use by operational forecasters at Météo-France)
  - webVisibility.html
  - webGusts.html

Summary

- Forecasting rare events through statistical adaptation requires a specific approach
- Large sets of data are needed
- Validation requires suitable verification criterions (and large sets of data too)
- The choice MOS vs. PP is not straightforward
- Non linear statistical models may be required
- Informative predictors are not necessarily available
- Both deterministic and probabilistic products are useful