New scoring methods for weather forecasts

Isabelle Sanchez, Fabien Stoop, Michaël Zamo, Francis Pouponneau, Nicole Girardot

METEO FRANCE

METEO FRANCE

1 - Verification of sensible weather :

- Motivations

- Calculation

- Application

2 - Modified CRPS calculation

- Context

- Calculation

- First results

Conclusion

METEO FRANCE

Objectives :

- Score for final public production (illustration below)
- Synthetic score
- Values easy to understand
- Linked with user feeling
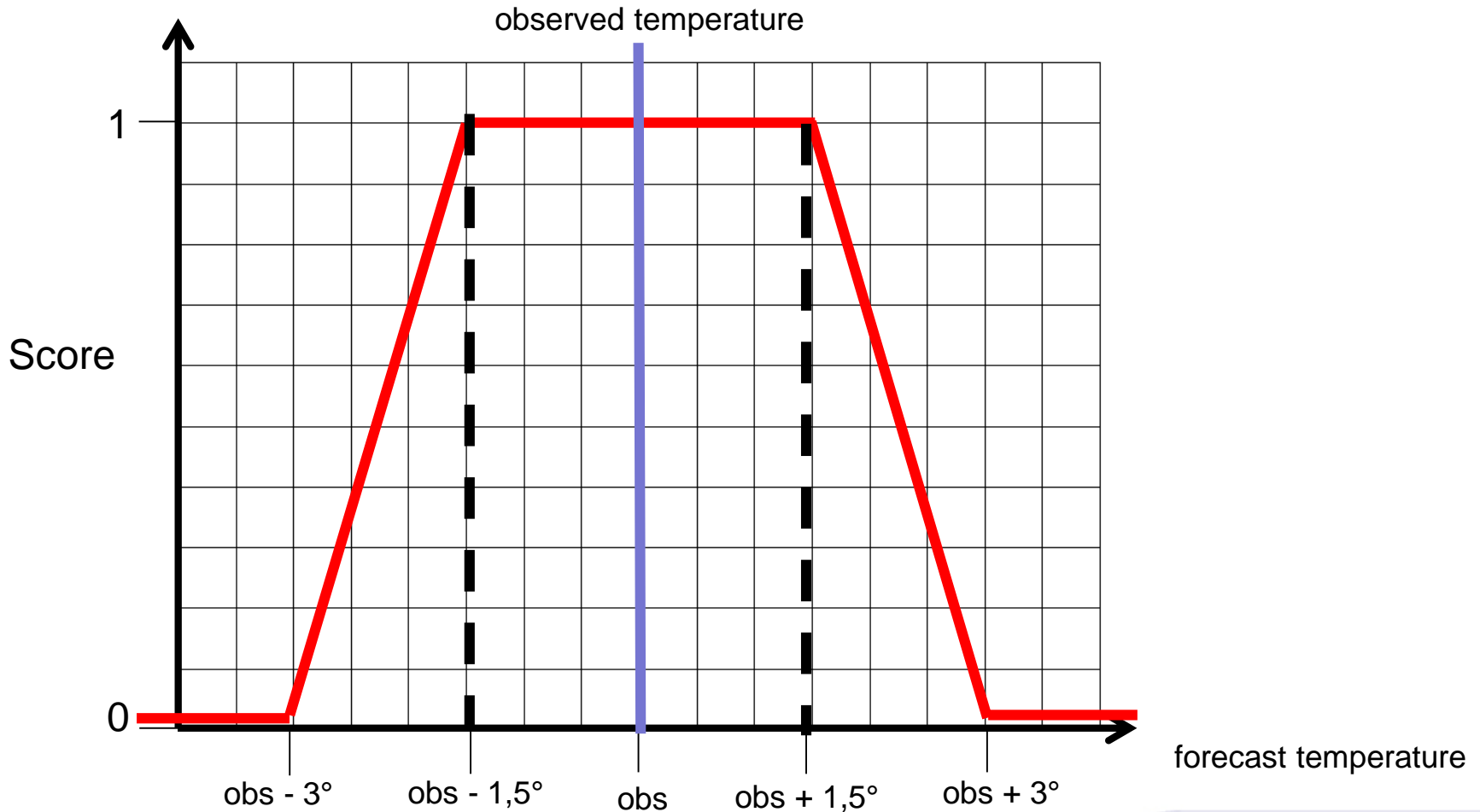- Same score from short to medium range (up to day 7)

The goals define the choices :

• Score for final public production → parameters : sensible weather (sun, rain, storm, snow…), total cloud cover, temperature, wind speed

• Synthetic score → aggregation of scores for these parameters

• Easy to understand → value from 0 (bad forecast) to 1 (good forecast)

• Linked with user feeling → put more weight where the error is more sensible (see next slides)

METEO FRANCE

Calculation for temperature :

(the score goes from 0 for bad forecast to 1 for good forecast)

─────── value of the score according to the forecast value

observed temperature

Score

1

0

forecast temperature

obs - 3°   obs - 1,5°   obs   obs + 1,5°   obs + 3°

METEO FRANCE

Calculation for windspeed :

The principle is the same as for temperature, except that the width of the tolerance interval enlarges when the windspeed becomes stronger.

Score

Figure 3: Note du taux de réussite pour les FF.
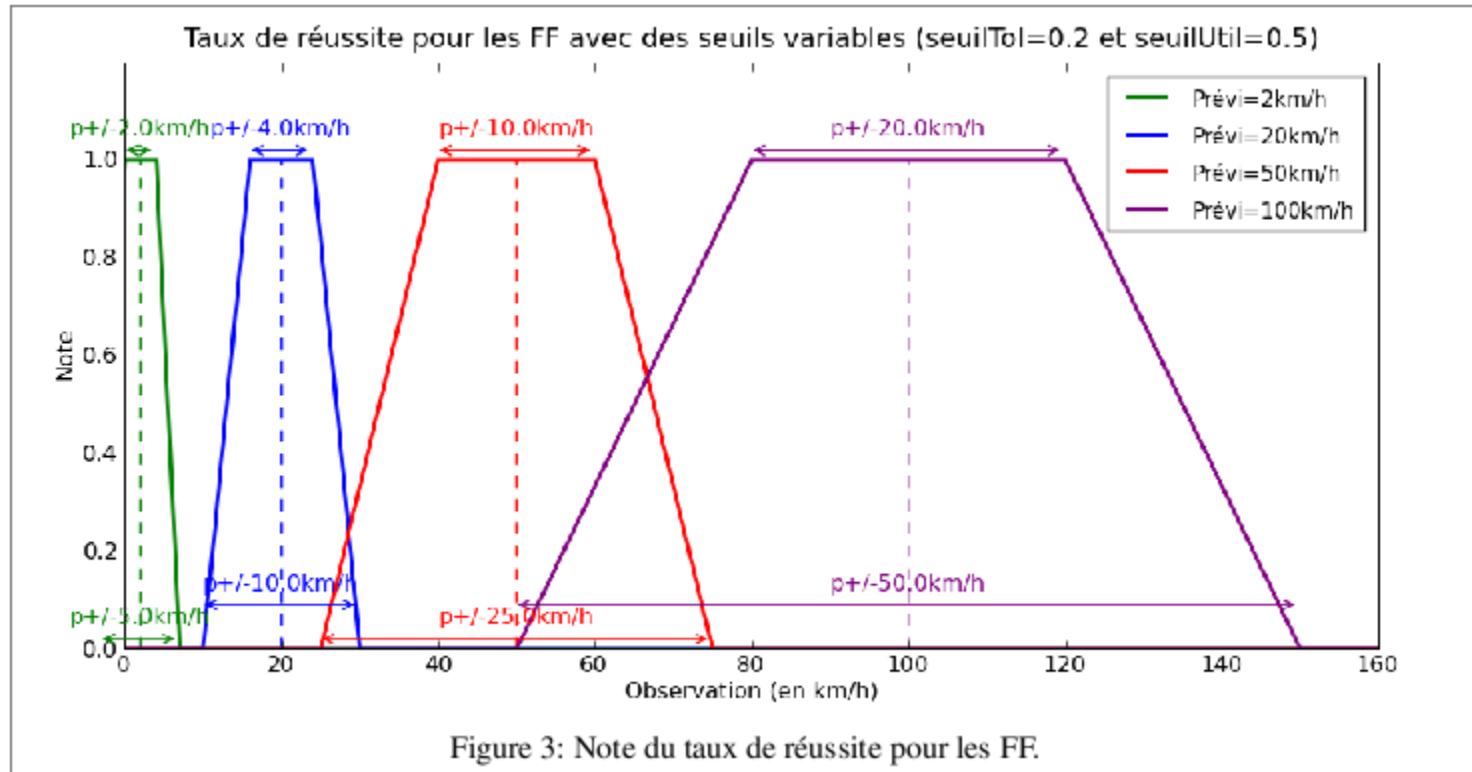
windspeed

Calculation for weather type :

Problem : the forecast classes of weather change with the time range.

In order to have a score valid for all time ranges, 6 weather types are considered :

- Clear sky
- Cloudy sky
- Rain
- Snow
- Thunderstorm
- Fog / mist

For each type, the forecast is expressed as a probability : Pf

the observation, Po, is 0 or 1.

The score for each type is :   **|1 – (Pf + Po)|**

The final score for sensible weather is the worst of the 6 values.

METEO FRANCE

Aggregation of temperature, windspeed and weather scores :

In order to get a synthetic value, the final score is computed as a weighted average of the three scores.

The weights are :

- 50% for weather type
- 30% for temperature
- 20% for windspeed

These weights aim to represent the importance of weather factors for the final users.

METEO FRANCE

Application :

The scores are computed every day.

Different kind of plots exist :

Score and tendancy for last 10 days   Score and tendancy for last 3 months

Indices moyens des 10 derniers jours						Indices moyens des 3 derniers mois entiers				
Région	**Indices CDPH**		**Indices CDPME**		**Région**	**Indices CDPH**		**Indices CDPME**	
	J1	**J3**	**J5**	**J7**		**J1**	**J3**	**J5**	**J7**
France	0.80 ↗	0.74 ↗	0.70 ↗	0.68 ~	**France**	0.82 ~	0.76 ~	0.69 ~	0.63 ↘
Nord	0.81 ~	0.75 ~	0.71 ↗	0.72 ~	**Nord**	0.84 ~	0.78 ~	0.71 ~	0.64 ↘
Ile-de-France-Centre	0.77 ↗	0.72 ↗	0.72 ↗	0.69 ↘	**Ile-de-France-Centre**	0.84 ~	0.79 ~	0.72 ~	0.65 ↘
Nord-Est	0.80 ↗	0.75 ↗	0.70 ~	0.70 ↘	**Nord-Est**	0.81 ↗	0.75 ↗	0.69 ~	0.63 ↘
Centre-Est	0.79 ↗	0.73 ↗	0.70 ~	0.69 ~	**Centre-Est**	0.80 ~	0.74 ~	0.66 ~	0.61 ~
Sud-Est	0.79 ↗	0.74 ↗	0.67 ~	0.65 ↗	**Sud-Est**	0.79 ↗	0.74 ~	0.66 ~	0.62 ~
Sud-Ouest	0.81 ↗	0.76 ↗	0.71 ↗	0.66 ↘	**Sud-Ouest**	0.82 ~	0.77 ↗	0.68 ~	0.62 ~
Ouest	0.80 ↗	0.75 ↗	0.74 ↗	0.72 ~	**Ouest**	0.84 ~	0.79 ~	0.72 ~	0.65 ↘

Different areas
of France

METEO FRANCE

Application :
Chronology of the scores over 3 months

Global score

Score for temperature

Score for weather type

Z500 RMSE
3 models

Verification of sensible weather 8/9

Summary :

> The score gives satisfaction according to the objectives.

> The visualisation gives a real time information about final forecast quality.

> The temporal evolution of the score shows bad forecasts situations.

> Final validation is in progress.

METEO FRANCE

METEO FRANCE

Context :

- TAC sub-group on extreme events verification

- Search for proper scores

- Have the possibility to give more attention to certain ranges of values

- References :
  - « Forecaster's Dilemna : Extreme Events and Forecast Evaluation », Lerch et al., 2015
  - « Comparing Density Forecasts Using Threshold and Quantile weighted Scoring Rules », Gneiting and Ranjan, 2008

METEO FRANCE

Usual CRPS calculation :

(a) Forecast PDF and Observed

(a) Forecast and Observed CDF
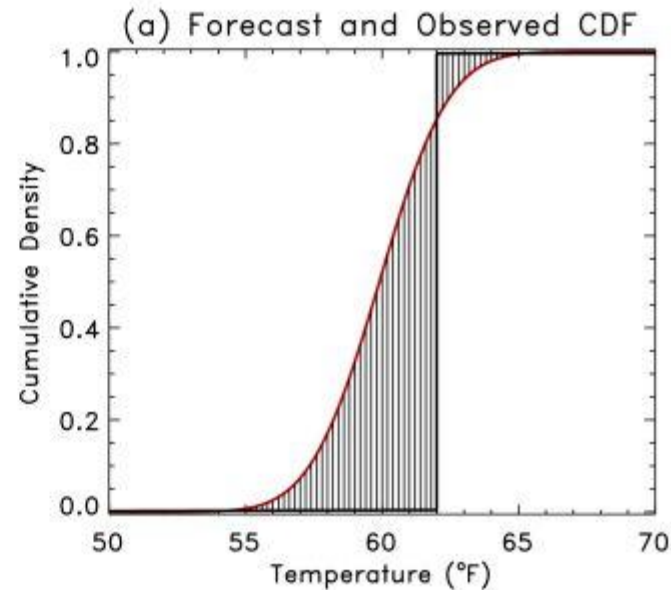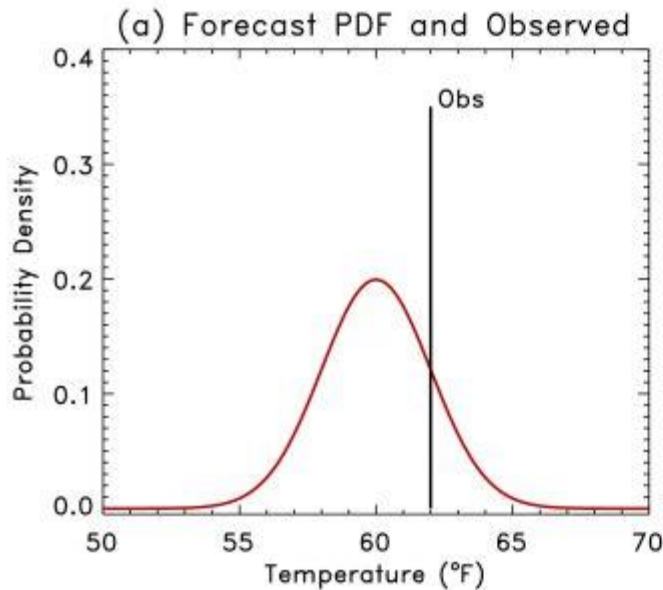
$$CRPS(forecast) = \frac{1}{ncases} \sum_{i=1}^{ncases} \int_{x=-\infty}^{x=-\infty} \left(F_i^f(x) - F_i^o(x) \right)^2 dx$$

METEO FRANCE

Modified CRPS calculation :

consider a given range of values → apply a weight function w(x)

$$CRPS(forecast) = \frac{1}{ncases} \sum_{i=1}^{ncases} \int_{x=-\infty}^{x=-\infty} w(x) \left(F_i^f(x) - F_i^o(x)\right)^2 dx$$

Different weight functions are tested :

Indicator left

Indicator right

Normal CDF left

Normal CDF right

threshold   *threshold*   *threshold*   *threshold*

METEO FRANCE
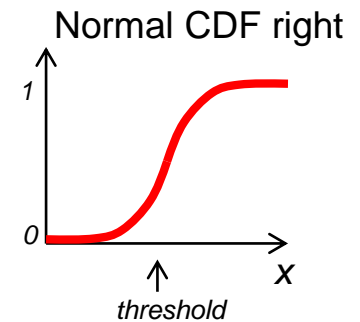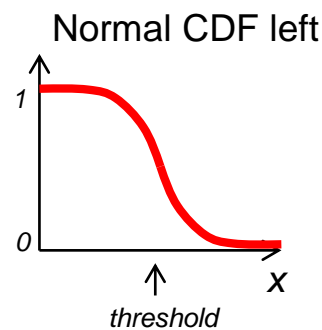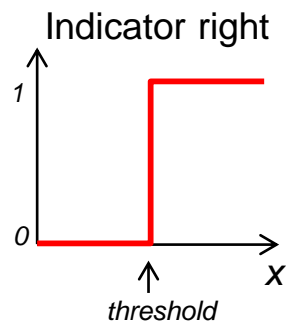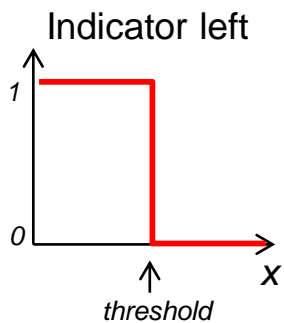
Data used :

- Parameters : 10m-windspeed, 6h-precipitations, 24h-precipitations

- PEARP ensemble system (based on ARPEGE), starting at 18h UTC.

- Time ranges from 0/6/24 to 108h

- Period : mai 2015 to april 2016

- Area : France

- Verification against synoptic observations

METEO FRANCE

First results :

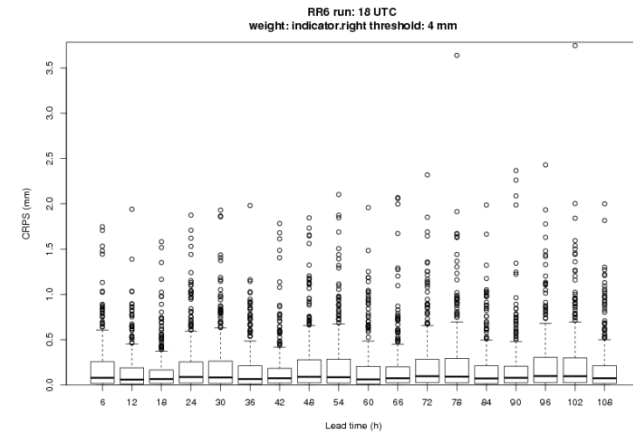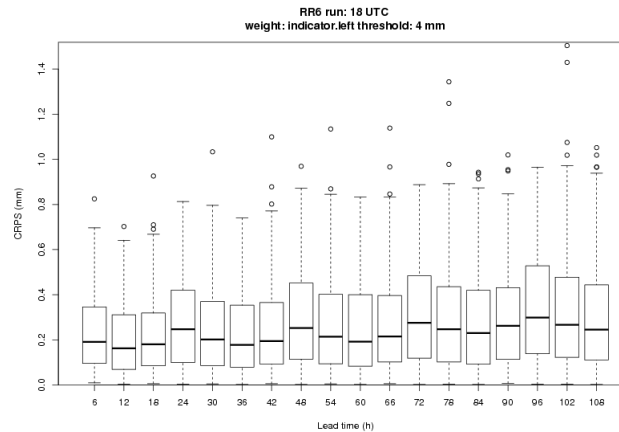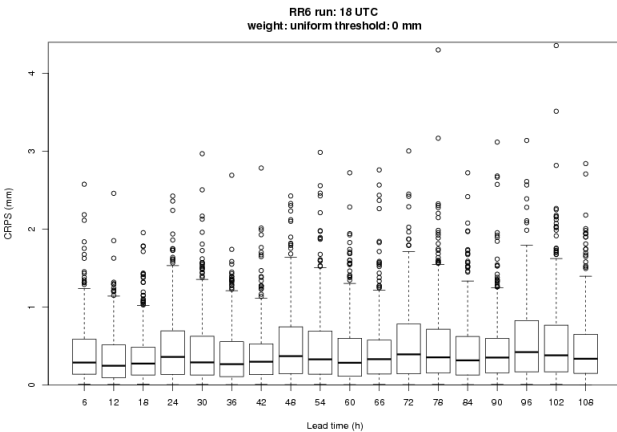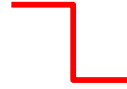Comparison uniform/left/right weight for windspeed – threshold 5 m/s
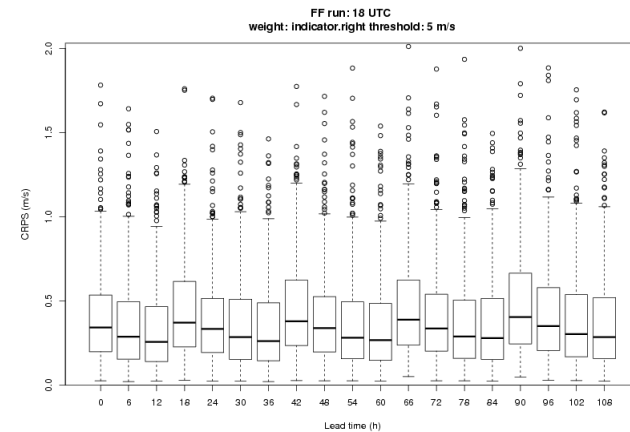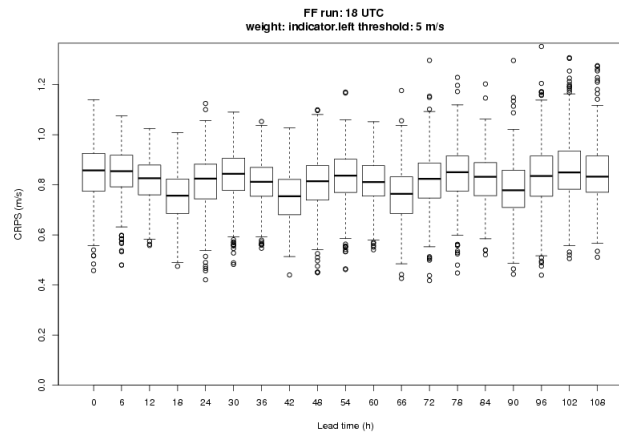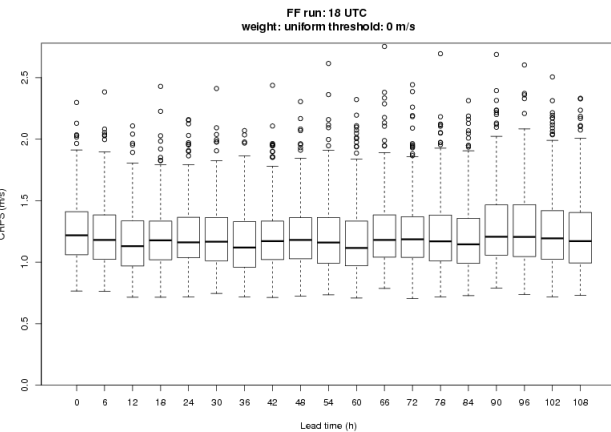
→ Daily cycle is more pronounced with weighted CRPS (clear difference for valid time 12h)

→ no tendancy with time range
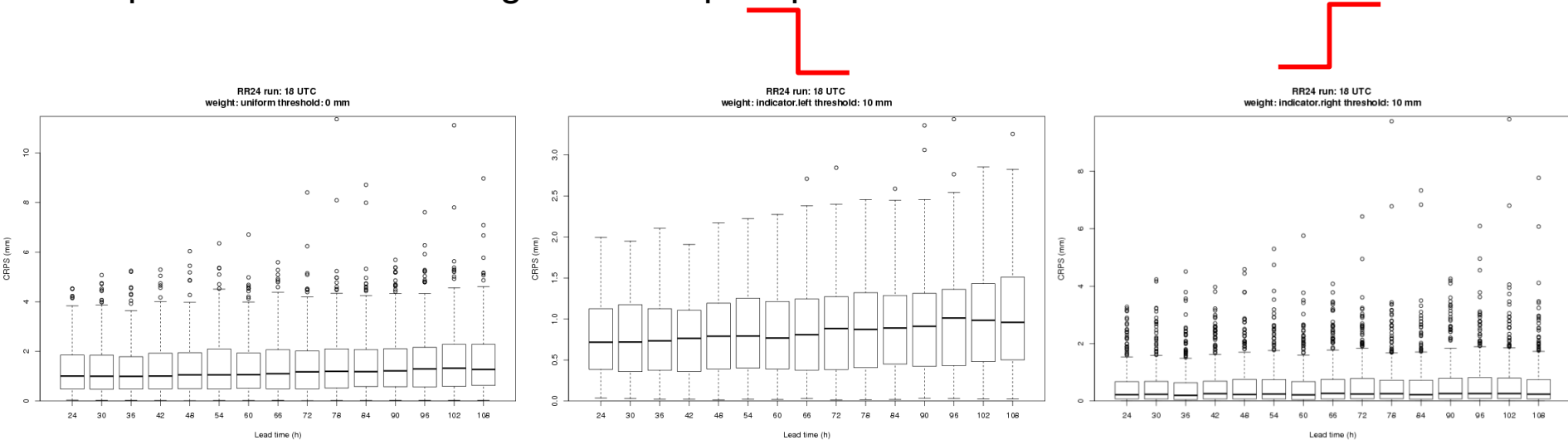
First results :

Comparison uniform/left/right for 6h-precipitations – threshold 4 mm

→ Light deterioration with time range

→ Impact of threshold / frequency of the event

METEO FRANCE
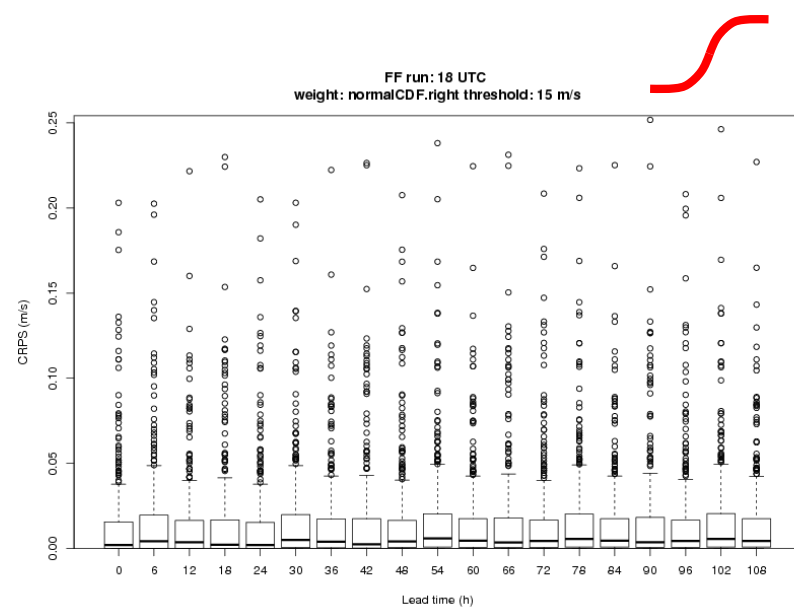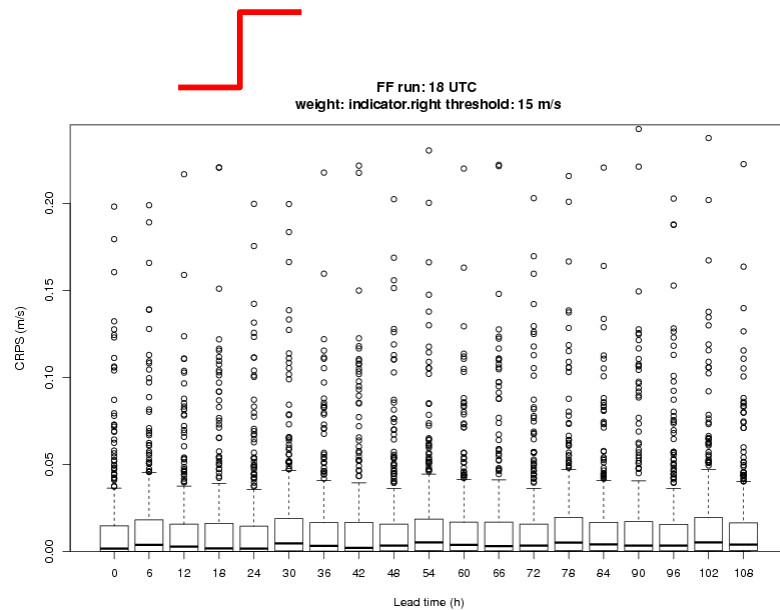
First results :

Comparison uniform/left/right for 24h-precipitations – threshold 10 mm

→ Light deterioration with time range

→ Impact of threshold / frequency of the event
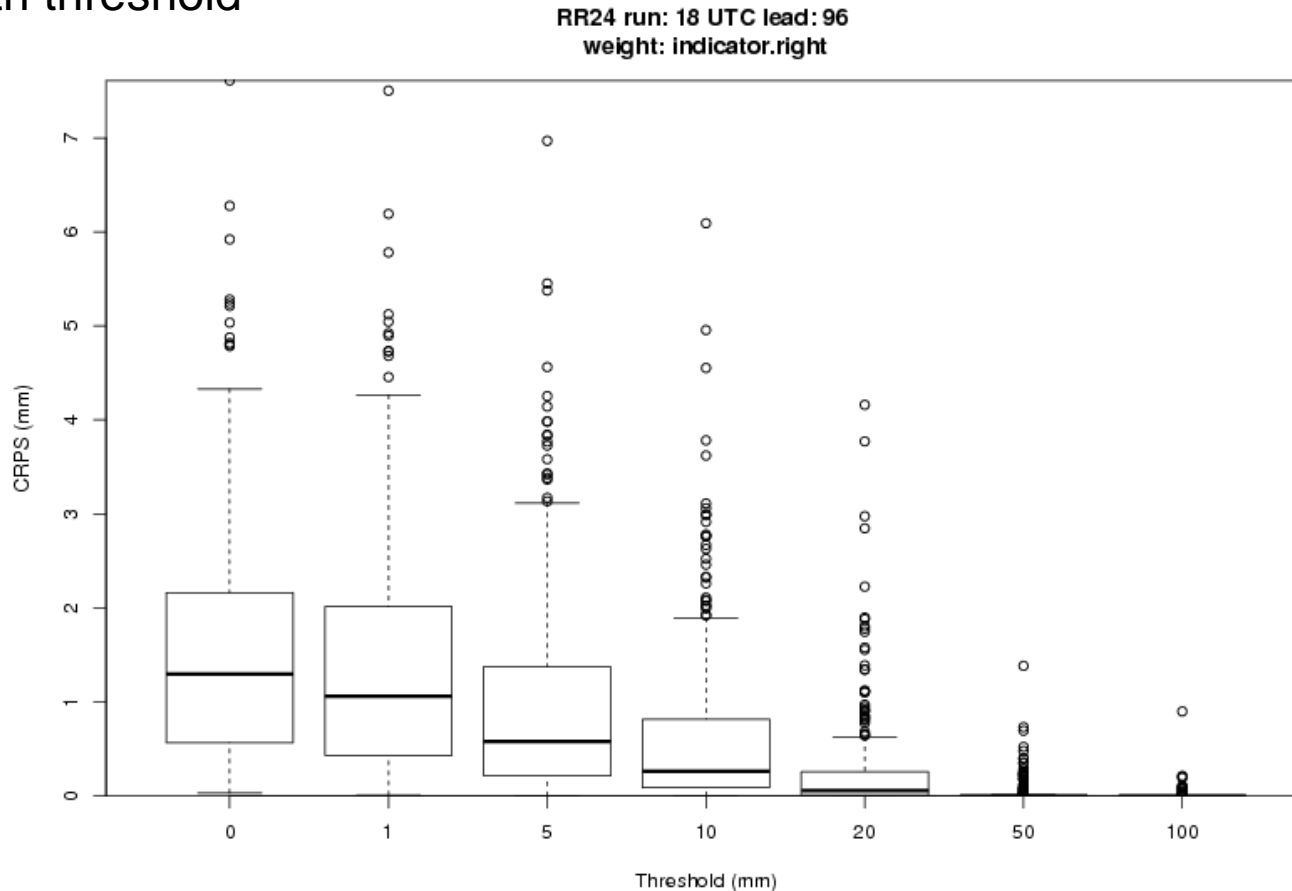
First results :

Comparison indicator and normal CDF weights for windspeed – threshold 15 m/s

→ indicator and normal CDF weights are very similar

First results :

Evolution with threshold

RR24 run: 18 UTC lead: 96
weight: indicator.right

→ This behavior is well known for the Brier score, and is named « degeneracy » of the score : this is linked with the frequency of the assessed event.

So what ?

➢ The weight method works, first results are encouraging and coherent with the Brier score behavior

➢ It can be used to compare different ensembles for a given range of event

➢ Problem of degeneracy will be further investigated :

- Test the skill version of weighted CRPS

- Use weight functions based on the quantiles, not the absolute forecast values

METEO FRANCE

→ *two current ways of investigation have been presented*

→ *first one tries to consider final user perception of the forecast error. It is nearly in operations.*

→ *second one aims to focus on extreme events. First results are encouraging.*

METEO FRANCE

Thank you for your attention !

METEO FRANCE