Verification of the nowcasting version of AROME-France

Isabelle Sanchez
Verification team, Météo-France
SRNWP- EWGLAM meeting, october 3-6 2016

AROME-NWC verification

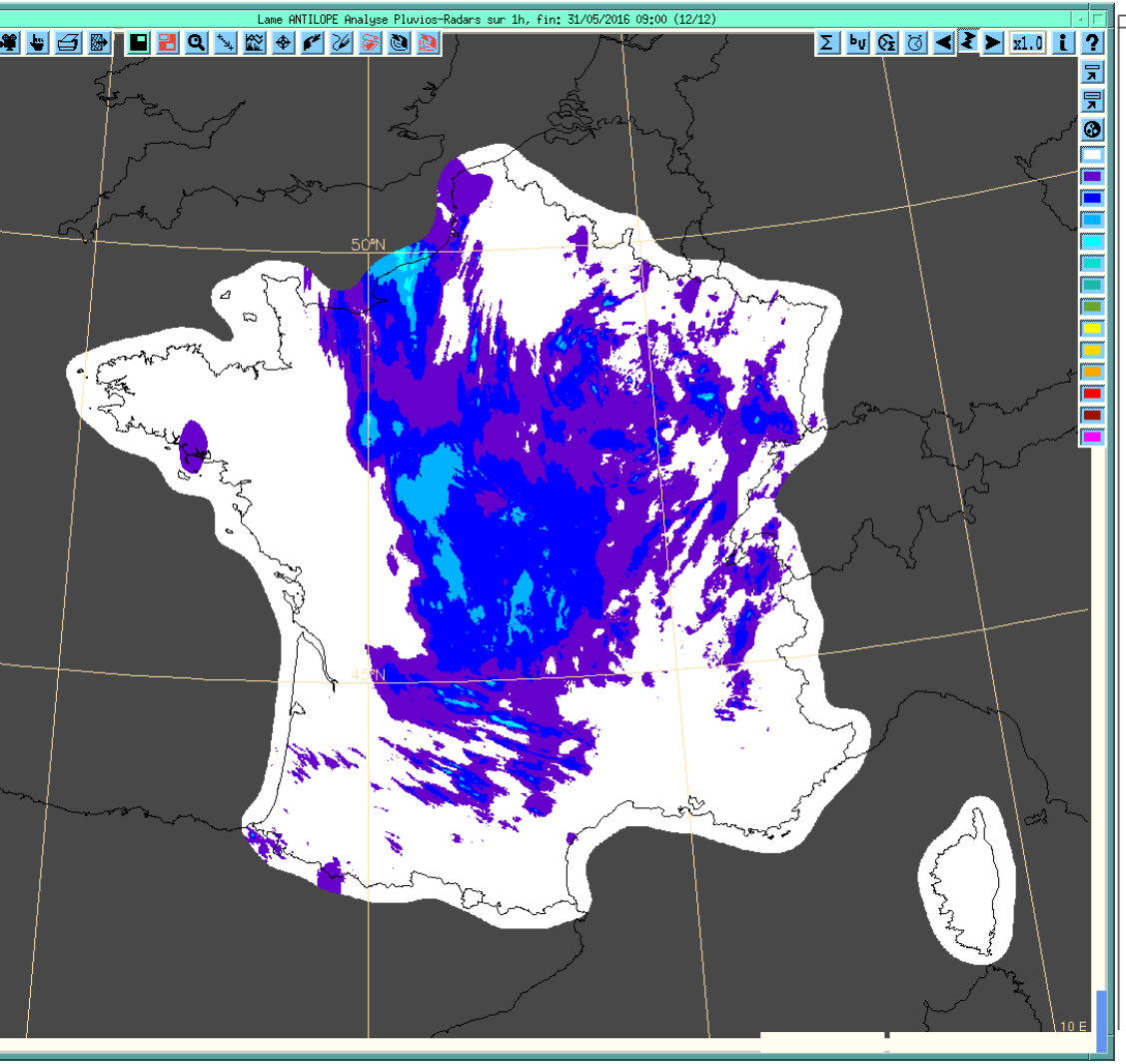
- Operational forecast verification
  - Outlines
  - 1 hour rainfall skill
- Rare event scores
  - Information brought by different indexes
- Perspectives

Operational verification
Outlines

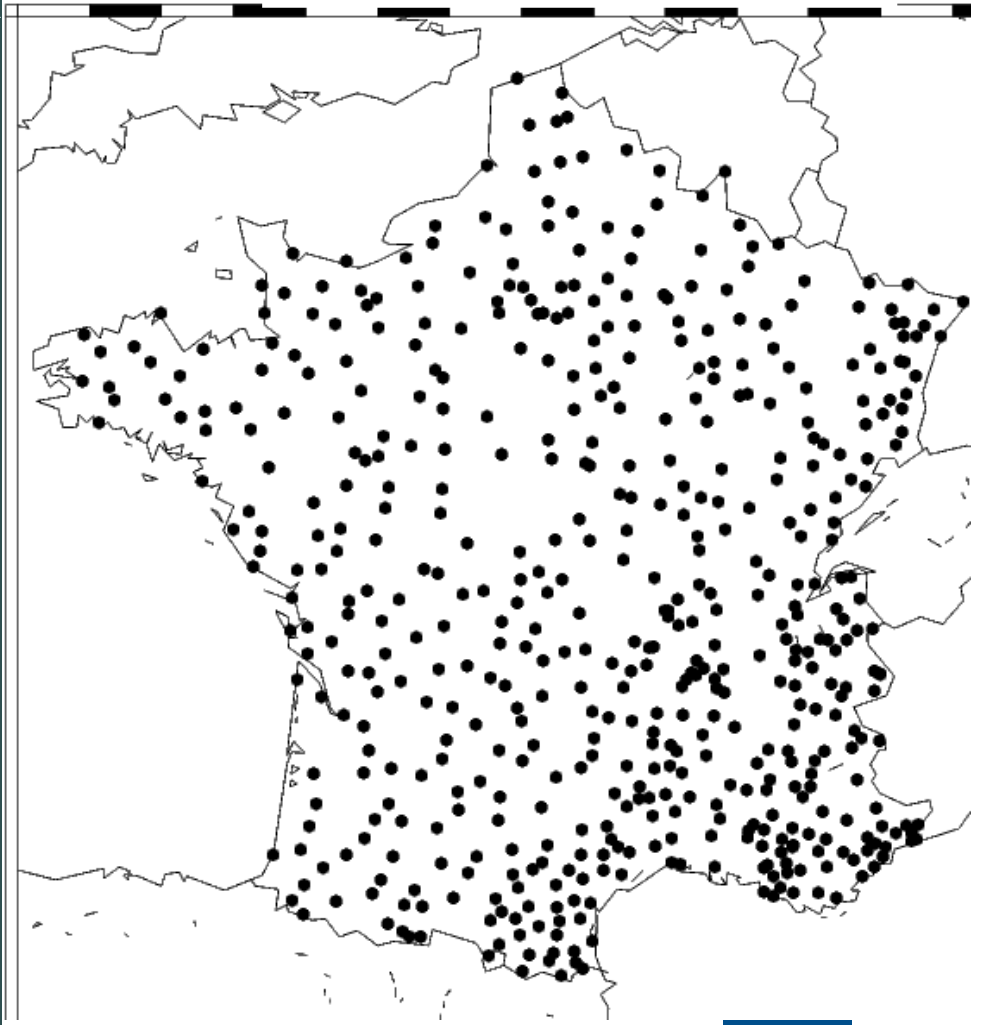
- Since march 2016: official starting date for AROME-NWC
- What is verified
  - 24 runs every day
  - 6 hours range
  - Wind force and Wind gust against surface observations
  - 2m Temperature against surface observations
  - 1h Precipitation against analysis ANTILOPE
- Comparison with AROME-France scores
- 3 month computation

References of verification for precipitation and wind

1 hour Rainfall analysis :
125,000 grid points verified



Wind observation network
720 points verified



1 hour rainfall verification basics

- Reference from analysis ANTILOPE
- 3 months computation + comparison with AROME-France scores
- Contingency tables thresholds 0.5 1 2 3 4 **5** 10 mm/h

Observed

Forecast	Yes	No	Total forecasted
Yes	*a* *(Hits)*	*b* *(False alarms)*	a + b
No	*c* *(misses)*	*d* *(correct negatives)*	c + d
Total observed	a + c	b + d	a + b + c + d

1 hour rainfall verification scores

- Usual scores from Contingency tables :

Obs / Forecast	Yes	No	Total forecasted
Yes	a	b	a + b
No	c	d	c + d
Total observed	a + c	b + d	a + b + c + d

$$\text{Bias} = \frac{\text{Number of YES forecasted}}{\text{Number of YES observed}} = \frac{a+b}{a+c}$$

Ability to forecast such events

$$\text{POD} = \frac{\text{Hits}}{\text{Number of YES observed}} = \frac{a}{a+c}$$

Proportion of observed event correctly forecasted

$$\text{FAR} = \frac{\text{False Alarms}}{\text{Number of YES forecasted}} = \frac{b}{a+b}$$

Proportion of forecasts actually did not occur

1 hour rainfall verification scores

- In addition to “classic scores” : Brier Skill Score

$$\text{Brier Score} = \frac{\sum \left(\text{Forcasted Probability} - \text{Observed frequency}\right)^2}{N}$$

$$\text{Brier Skill Score} = 1 - \frac{\text{Brier Score}\left(\text{for AROME} - \text{NWC forecasts}\right)}{\text{Brier Score}\left(\text{persistence}\right)}$$

- Uses an independent reference : persistence
- Allows comparisons between 2 models
- Summarises POD and FAR
- Used over neighbourhoods from 1 to 20 km

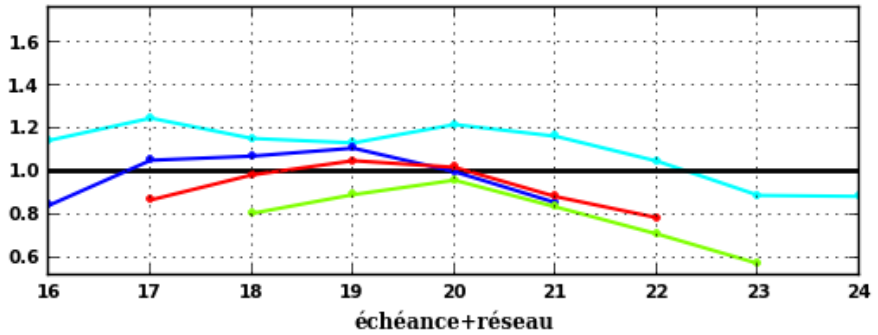
Also known as Fraction Skill Score with persistence as reference

Bias Pod and FAR for RR ≥ 5mm/h

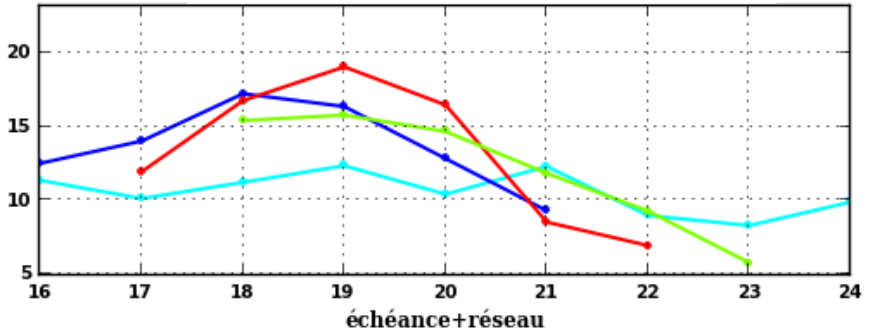
2nd quarter 2016 : april – mai - june

AROME run 12UTC
AROME-NWC r 15UTC
AROME-NWC r 16UTC
AROME-NWC r 17UTC

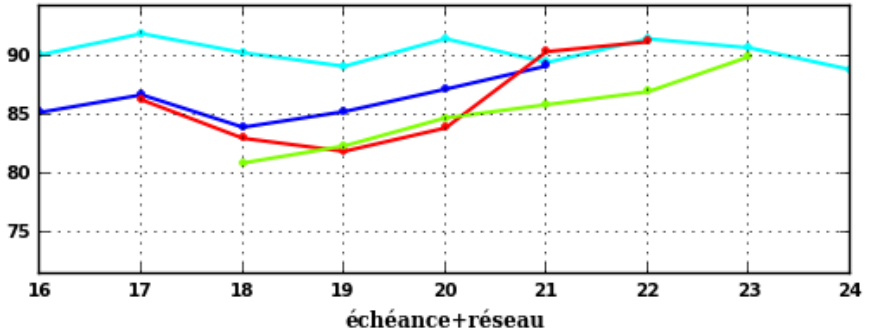
Bias



Probability Of Detection



False Alarm Rate



Improvements compared with AROME

Better Biases but

under-estimation
- at 1rst range
- in the early evenning

Better POD but

few detections by both
10% → 20%

Lower FAR but

High level
80 → 95%

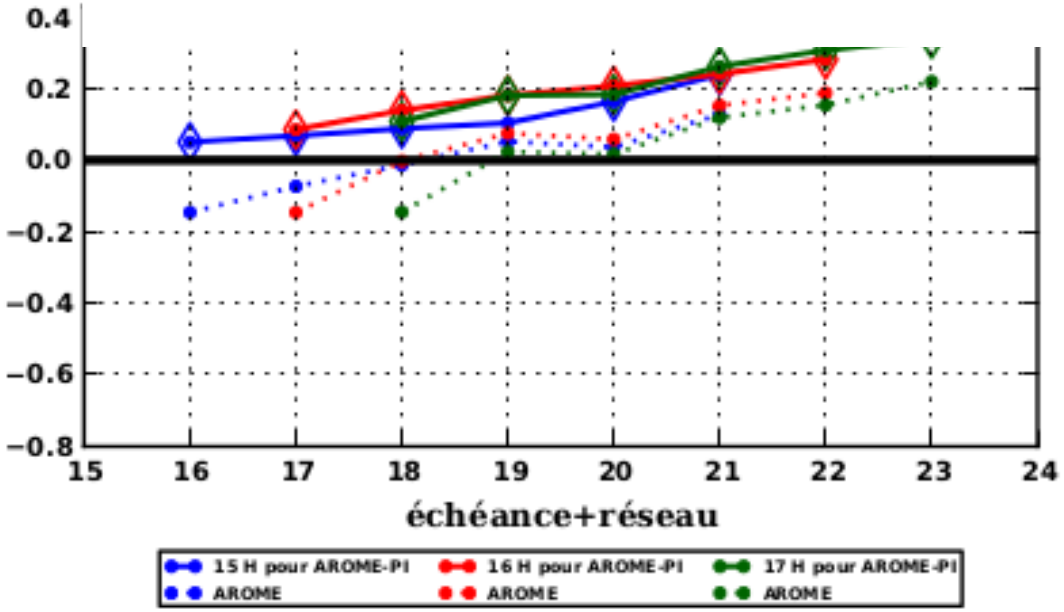
- SRNWP-EWGLAM 2016

METEO FRANCE

Brier Skill Score for RR ≥ 5mm/h

2nd quarter 2016 : april – mai - june

BSS against persistence 1km tolerance



Summarizes POD and FAR informations

BSS > 0 for AROME-NWC

BSS < 0 for 1rst hours of AROME

AROME-NWC BSS >AROME BSS

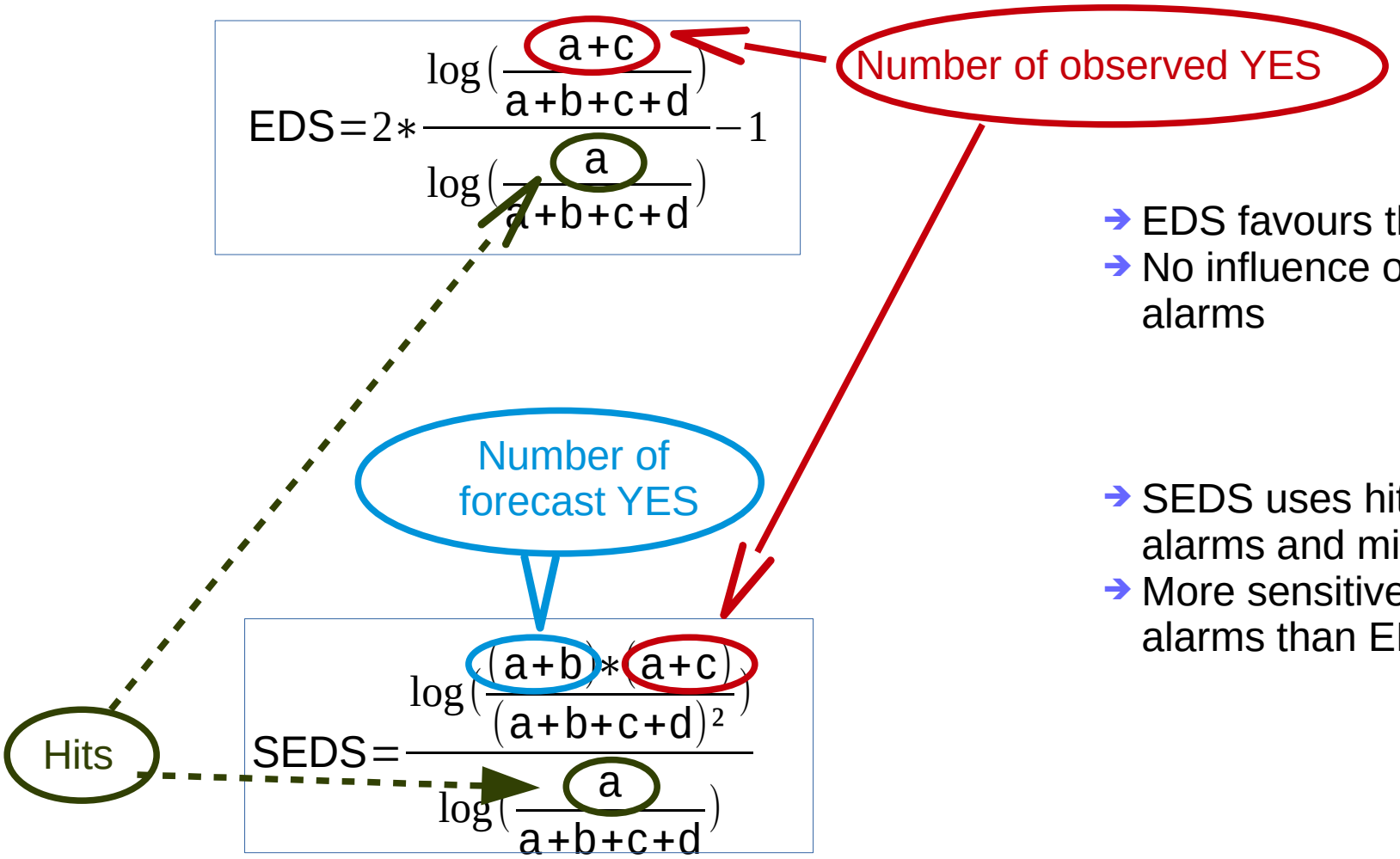
BUT Low values of BSS

Rare event scores

- Why
  - Rare events : frequency < 0.5% for RR ≥ 5mm/h
  - Low values of POD or BSS / high values of FAR
  - Sensitivity of classic scores to the climatology

- Needs
  - Found an index to measure improvements
    - AROME-NWC versus AROME
    - Reactive to software evolution
  - Necessity to be fair
  - Tested on AROME-NWC operable on other forecasts

Extreme Dependency Score / Symmetric EDS



- EDS favours the hits
- No influence of bias or false alarms

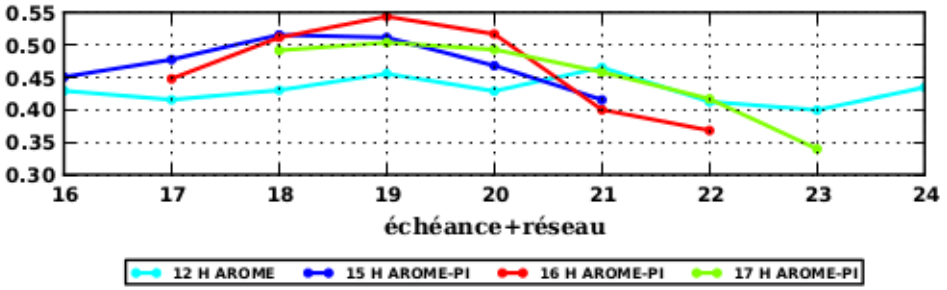
- SEDS uses hits, false alarms and misses
- More sensitive to false alarms than EDS

EDS versus SEDS

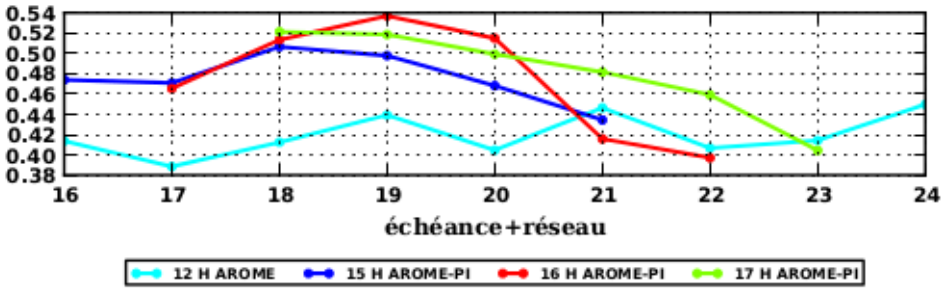
2nd quarter 2016 : april – mai - june

AROME run 12UTC
AROME-NWC r 15UTC
AROME-NWC r 16UTC
AROME-NWC r 17UTC

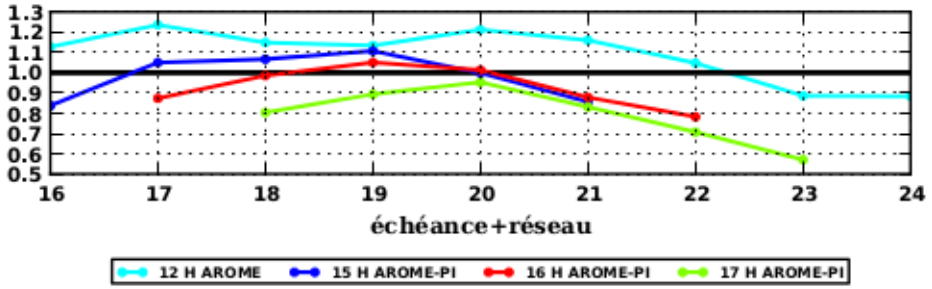
EDS



SEDS



Bias



For both scores
- Higher values / BSS or POD
- Quite the same ranking

EDS
- Same information as POD
- No bias / FAR information

SEDS
- Sensitive to bias
- Penalizes false alarms

Extremal dependency index and Symetric

$$EDI = \frac{\log(POFD) - \log(POD)}{\log(POFD) + \log(POD)}$$

$$SEDI = \frac{\log(POFD) - \log(POD) + \log(1-POFD) - \log(1-POD)}{\log(POFD) + \log(POD) + \log(1-POFD) + \log(1-POD)}$$

Where :

$$POD = \frac{\text{Hits}}{\text{Number of YES observations}} = \frac{a}{a+c}$$

$$POFD = \frac{\textit{False alarms}}{\textit{Number of NO observations}} = \frac{b}{b+d}$$

- Range -1 to 1
- 0 no skill
- 1 perfect score

- Independent of the event frequency (number of Yes observations) more equitable than EDS/SEDS
- EDI can be optimized for biased forecasts

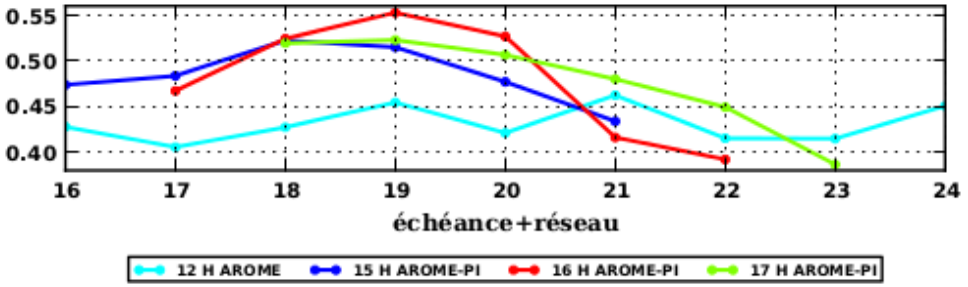
Reference : Ferro & Stephenson 2011; Extremal Dependence Indices: improved verification measures for deterministic forecasts of rare binary events.

METEO FRANCE

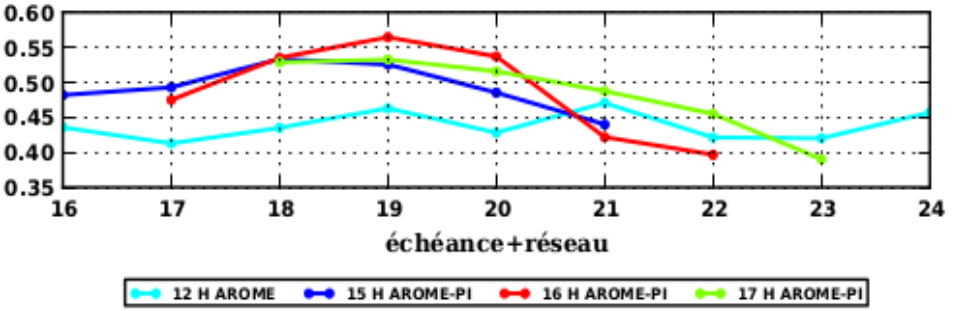
EDI versus SEDI

2nd quarter 2016 : april – mai - june

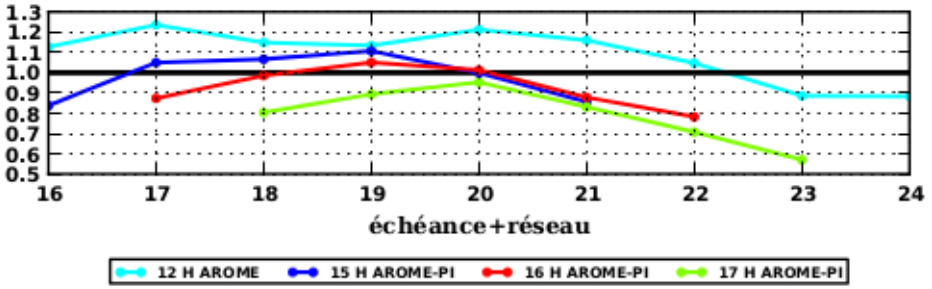
EDI



SEDI



Bias



- Higher values than BSS or POD
- Comparable to EDS and SEDS

- Exactly the same ranking for EDI and SEDI
- In this case very small differences between the two indexes
- Is it a particular case ?

Perspectives

- AROME-NWC verification :
  - Define a synthetic index ?
  - Longer term : low visibility and ceiling diagnostics

- Next steps for rare events : more questions than answers
  - Are the differences significant ?
  - Persistence behaviour for rare thresholds ?
  - Does SEDI show improvements in modelling ?
  - …