Ensemble prediction at convective scale with AROME

François Bouttier - réunion recherche MF/DGA - 20 mai 2016

The PEARP ensemble

The AROME-France-EPS system

Interpretation of convective scale EPS

The Nadine case

Downscaling ensembles

Higher resolution models are expensive, but provide better forecasts :

- improves ***parameters with important small-scale structures*** : convection-generated precip & gusts, low-levels parameters in mountainous regions

- ***more phenomena*** are explicitly simulated : cold pools, bow echoes, hail, valley fog, lakes, urban heat island, sea breezes, tornado-prone conditions...

Many meteorological institutes run their own limited-area ensembles.

Case of the IFS-ENS -> PEARP -> Arome-EPS complementarity :

- IFS-ENS : global dx=18km(cubic), late availability (~9h), 51 members

- PEARP : global dx=10km (linear), available earlier (~6h), 35 members

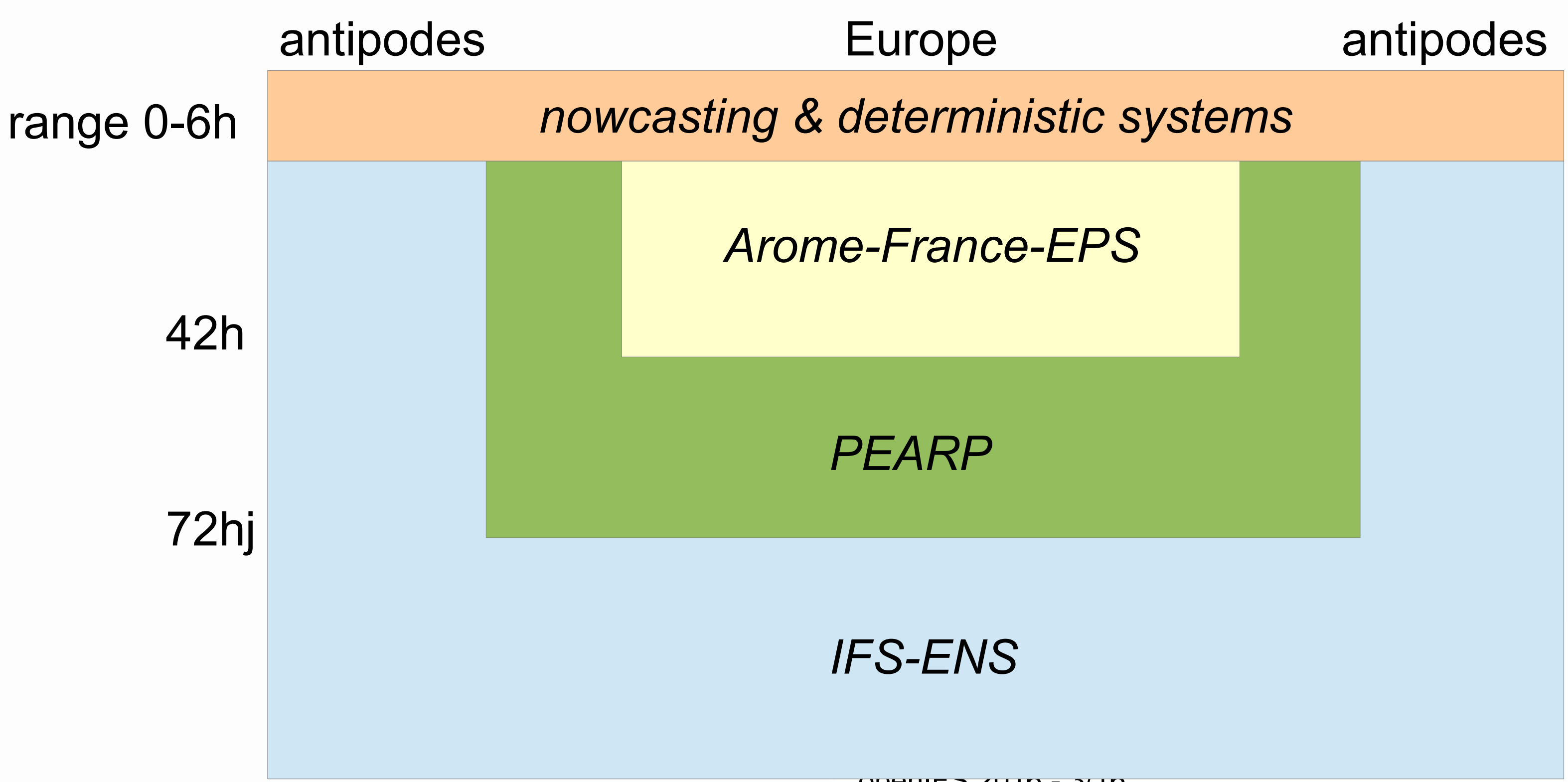
- Arome-France-EPS : regional dx=2.5km (linear), available after ~4h, 12 members

PEARP is a standalone global EPS (independent from ECMWF)

Arome-EPS is coupled to PEARP: similar lateral boundaries & initial conditions

Complementarity of ensemble prediction systems

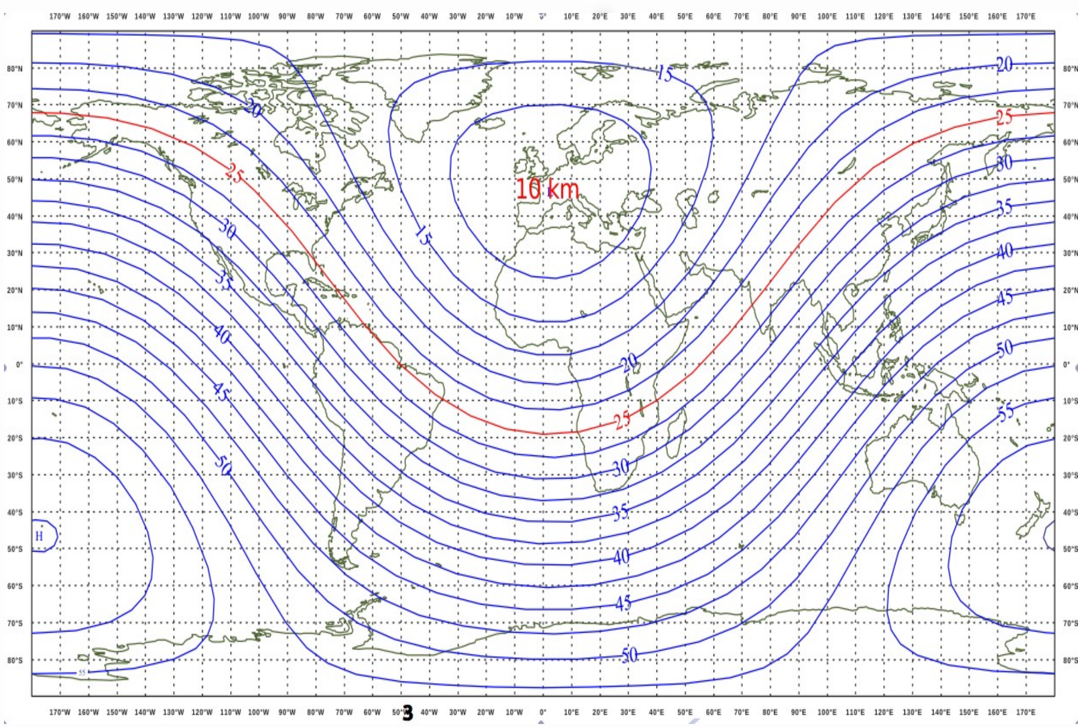
- IFS-ENS of ECMWF : 15-day range at dx=18km,résolution 18km, global, lower information at range<24h
- PEARP : variable resolution (10km over W.Europe, worse elsewhere), max range = 3 days
- Arome-EPS : 2.5km over W. Europe 36h, 12 membres, optimal information at range 12-42h



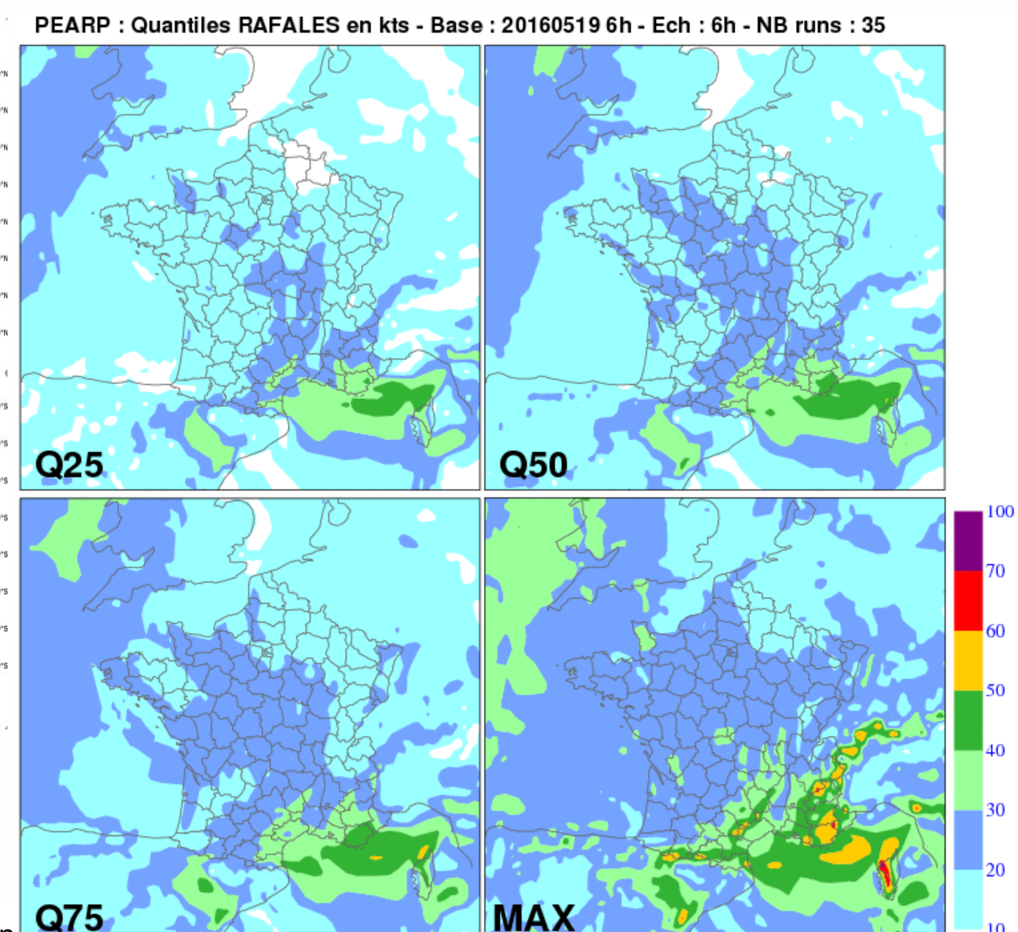
the PEARP ensemble prediction system

- uses the ARPEGE model (lower resolution than deterministic ARPEGE)
- perturbations : targeted singular vectors on extratropics & cyclonic areas, ensemble of data assimilations, multiphysics

PEARP resolution



a PEARP forecast : gust probabilities



the Arome-France-EPS system

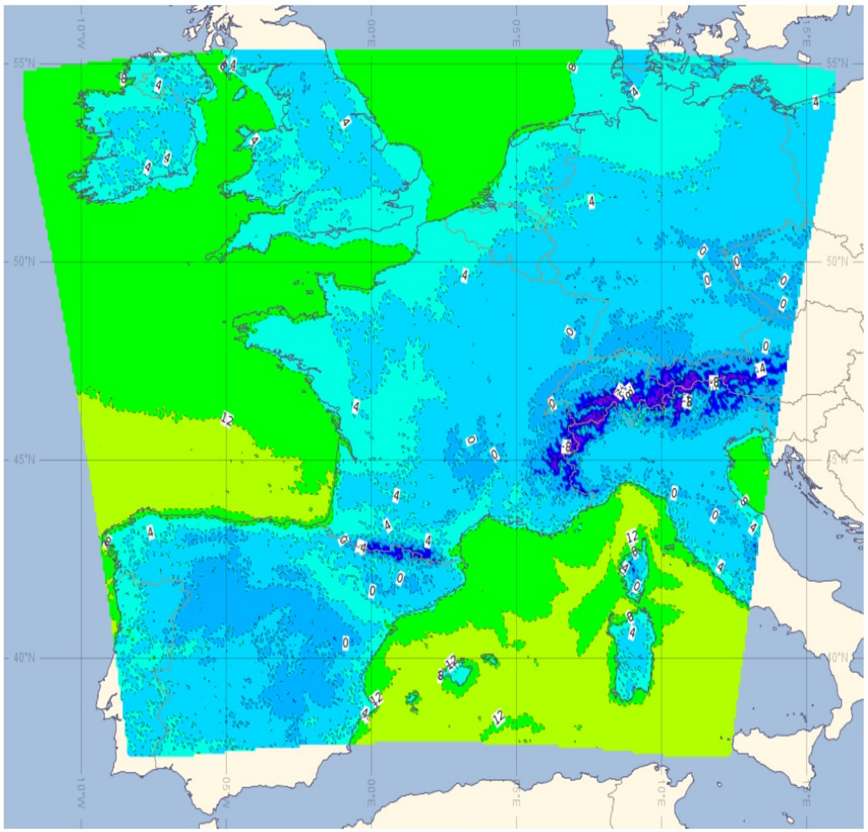
Components :

- model : like Arome-France-1.3kmL90, except dx=2.5km
- **LBC perturbations** : clustered from 34-member PEARP (the global MF ensemble)
- **IC perturbations** : PEARP perturbations centered on Arome-1.4km analysis
- **surface perturbations** : random IC perturbs on selected prognostic & physiographic fields
- **model perturbations** : simple LAM version of ECMWF SPPT stochastic physics
- (mostly identical to HarmonEPS in the Hirlam/Aladin world)

Status :

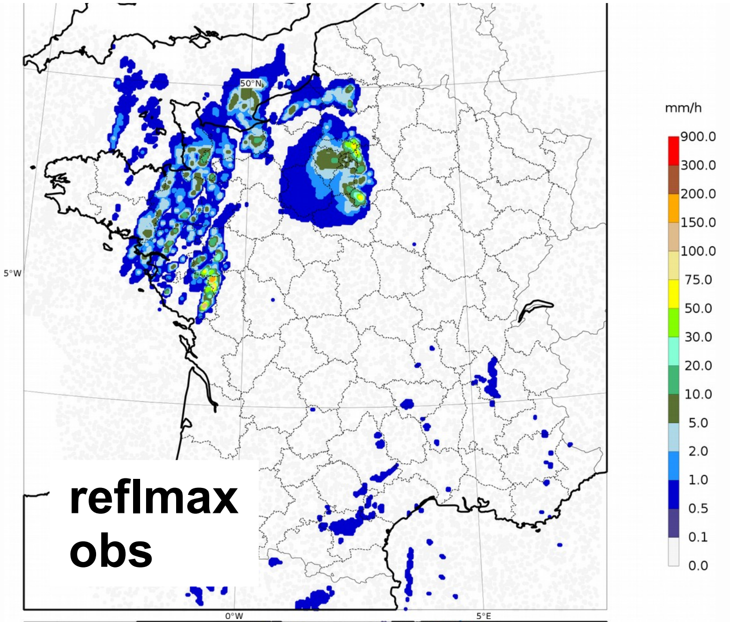
- August 2015 : preoperational daily runs
- Sept 2016 : real-time runs
- Dec 2016 : operational, 12members, 2 productions/day, 42-h range

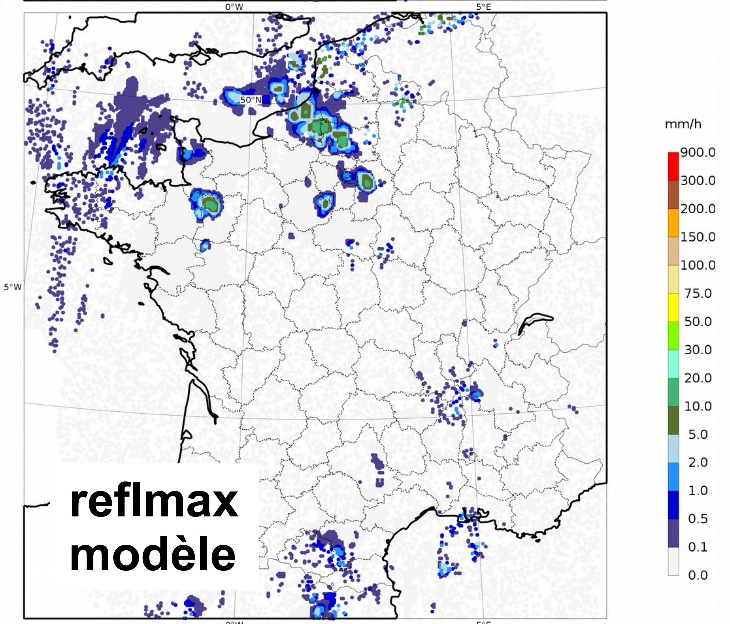
Arome-France 2016 domain



Example of Arome-EPS product : thunderstorms using radar reflectivities

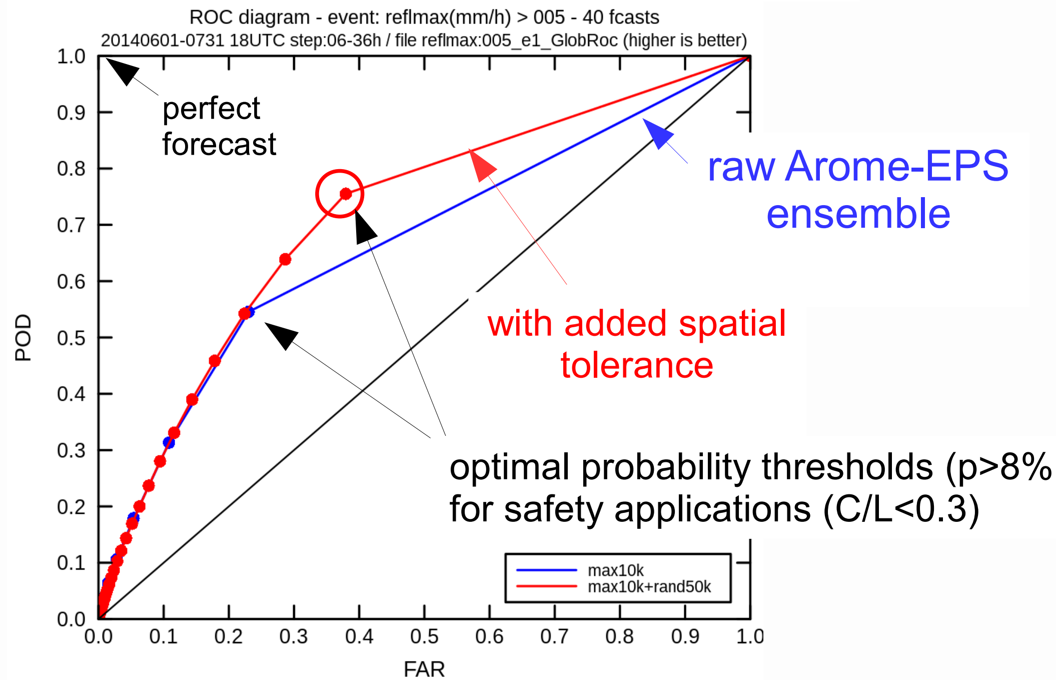
case study





Arome has detailed microphysics and simulates the largest convective cells, so it can be compared with radar images (lower-resolution models only predict convection-prone areas)

score over 2 months : ROC diagram



Comparison of ensemble strategies : IFS-ENS, PEARP, Arome-EPS

*** they all sample the ***same PDF*** (no trick to "optimize for extreme events") ***

different usable **domains & ranges** :

- IFS-ENS for exotic regions or longer ranges
- PEARP over Europe, 24-72h ranges. Can be compared to IFS or Arome.
- Arome-EPS on smaller domain, shorter ranges.

different **parameters and phenomena** :

- some are only available in AROME (eg : reflectivity to compare with radars)

different **ensemble size**:

- more IFS and PEARP members = less sampling error, but more model error
- tiny Arome-EPS = better forecasts, but large sampling errors

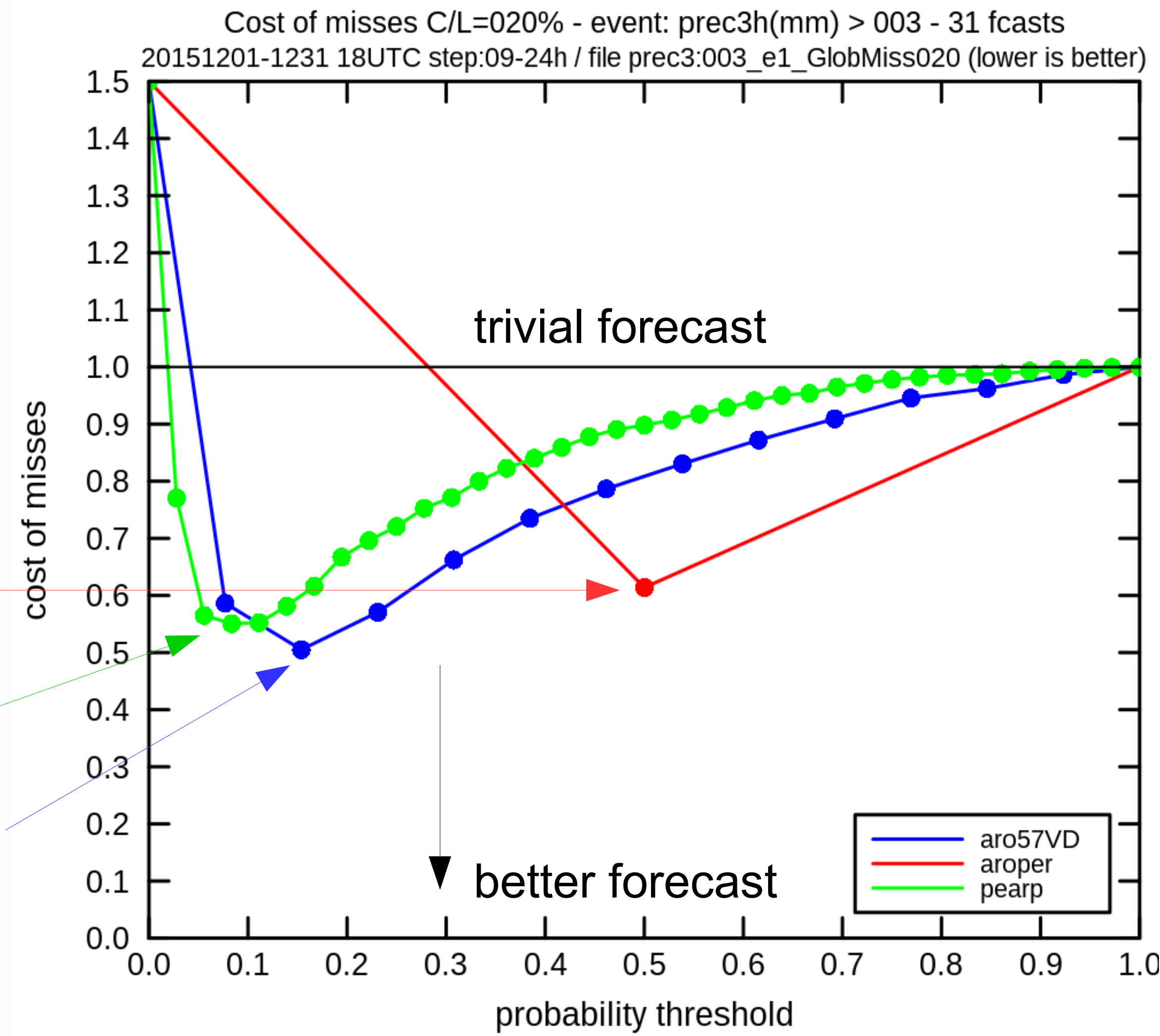
AROME wins for convection, heavy precip, fog, orography-related events. But not on other parameters. No system is always better than the others.

not the same **NWP infrastructure:** only IFS-ENS has **reforecasts to produce EFI** (extreme forecast indices), useful for untrained forecasters

FRANCE

Comparison of ensemble scores : PEARP vs Arome-EPS

Plot : weighed number of wrong forecasts measured by 5 ND + FA (we accept 5 false alarms for each non-detection), normalized by the best possible trivial forecast (always or never forecast).



Conclusions :

- Arome-deterministic is best for predicting the median
- PEARP is best for detecting very rare events, because it has more members
- Arome-EPS maximise user satisfaction, by using the Q85 quantile.

METEO FRANCE

Measuring EPS performance

(from simplest to most meaningful)

- **Reliability :** when we forecast an event with probability n%, it occurs n% of the time
- **Statistical resolution (or 'sharpness') :** an EPS system is more useful if it often predicts probabilities close to 0 or 1 (i.e. it "takes risks")
- **Usefulness :** does the ensemble improve decision-making ? Check false alarms & non-detection rates.
- **User ability** to benefit from ensemble output : requires education, training, understanding of what is being forecasted (e.g. to what extent false alarms are tolerated)

Common ensemble validation issues :

- too small ensemble spread on poorly predicted events (=unreliable spread)
- misinterpretation of PDFs (=lack of training)
- judging ensembles on case studies (PDF improvement is a statistical concept)
- trying to issue a deterministic forecast for highly uncertain events (=blaming the ensemble for unpredictability)
- misunderstood parameters (=point probabilities are not regional forecasts)
- misunderstood user needs (=issuing probabilities with the wrong non-detection/false alarm ratio)

METEO FRANCE

Interpreting the PDF of a small ensemble

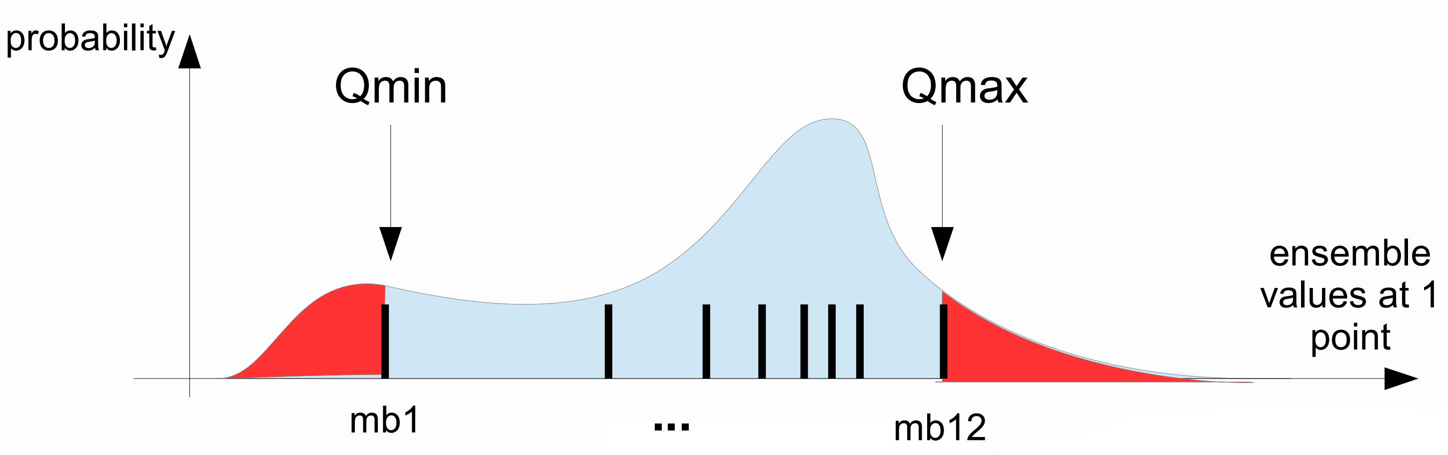
An ensemble is not a bracket :

In a perfect 12-member ensemble, any forecast parameter has probability 2/13=15 % of being outside the range of ensemble values.

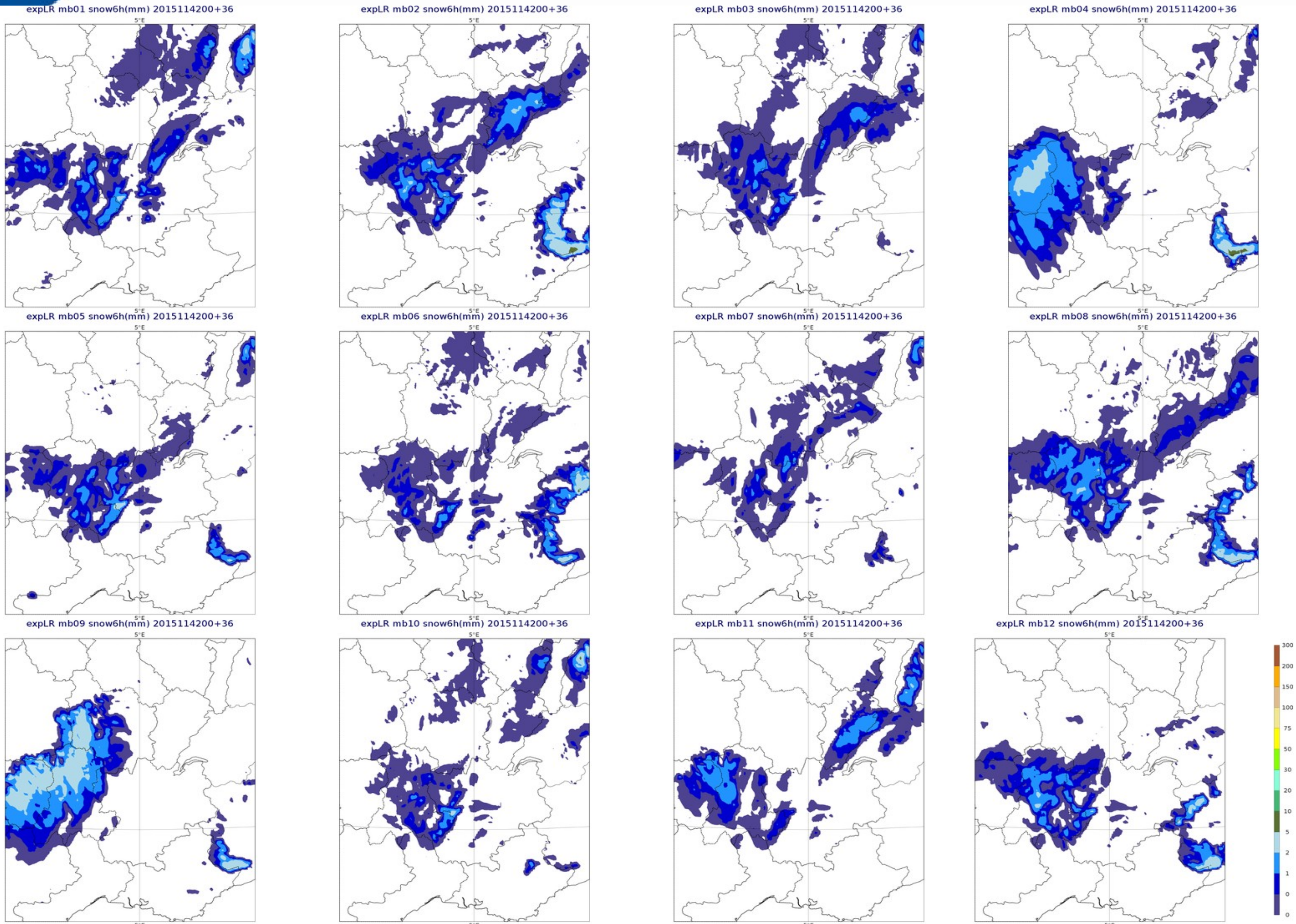
In practice it is even more, because of systematic model errors and lack of spread in ensembles.

Solutions :

- take *safety margins* around the range of ensemble-generated values & scenarios (e.g. timings & locations)
- avoid summarizing the ensemble by its "dominant synoptic scenario", because the information is in the ensemble *tails*.



Example of Arome-EPS forecast (snowfall, 12 members)

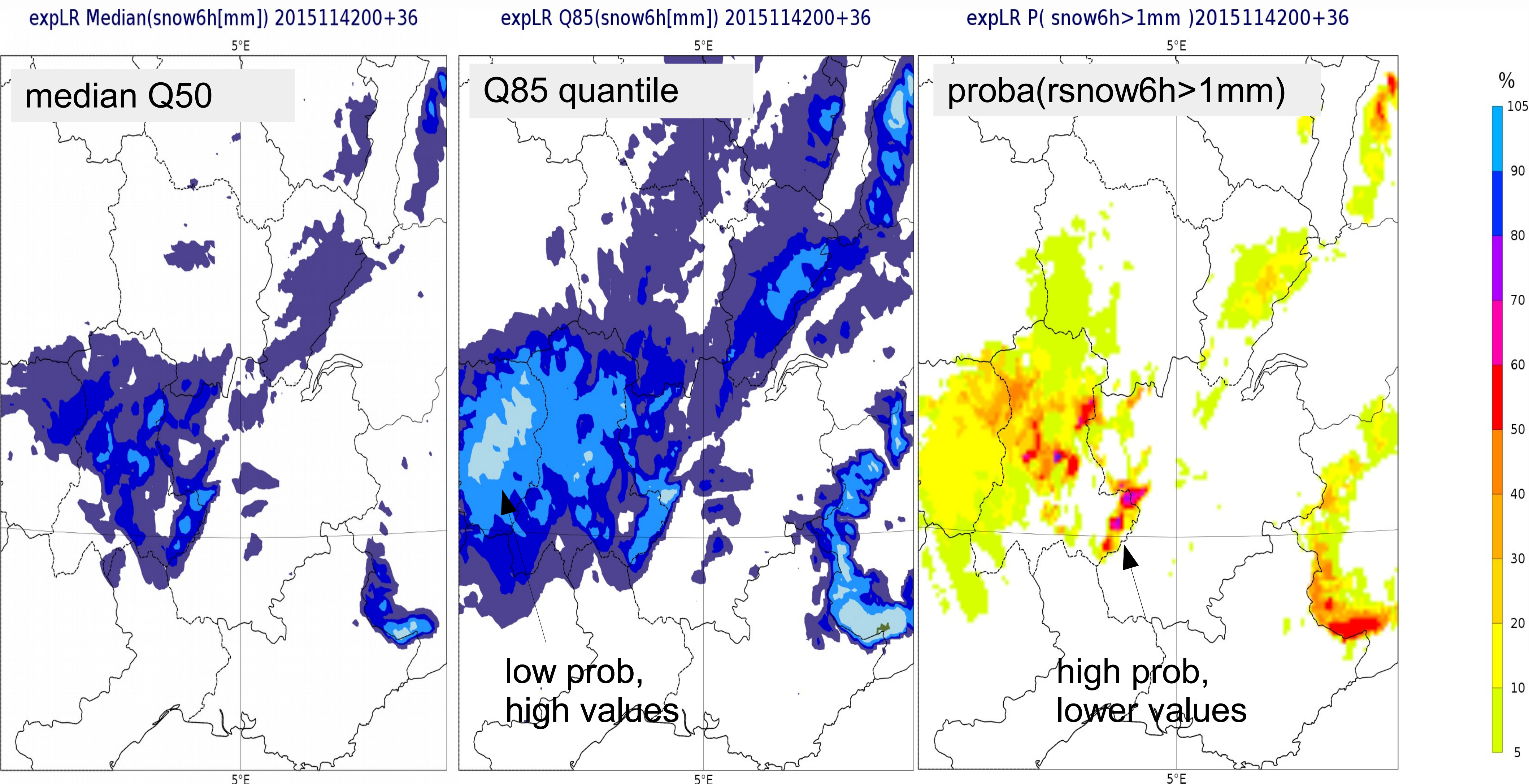


Maps of point probabilities (snowfall on same event)

Quantiles & probabilities are used to distinguish between

- likely small-impact events
- low-probability, high-impact events

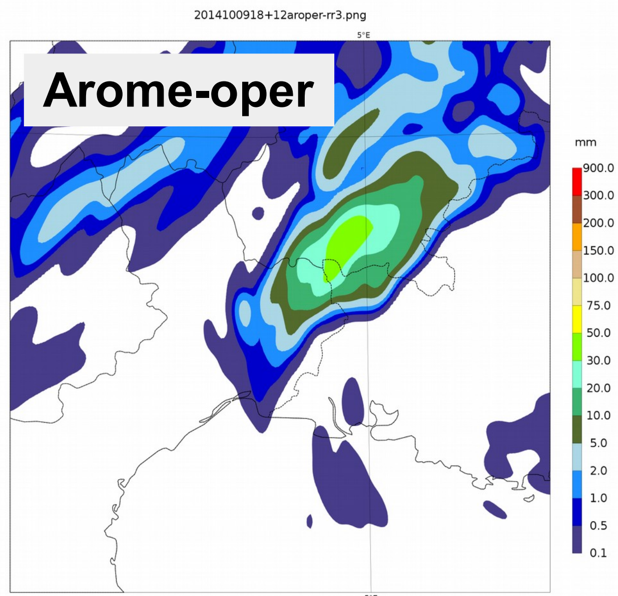
(not possible using the ensemble mean or medium quantiles)

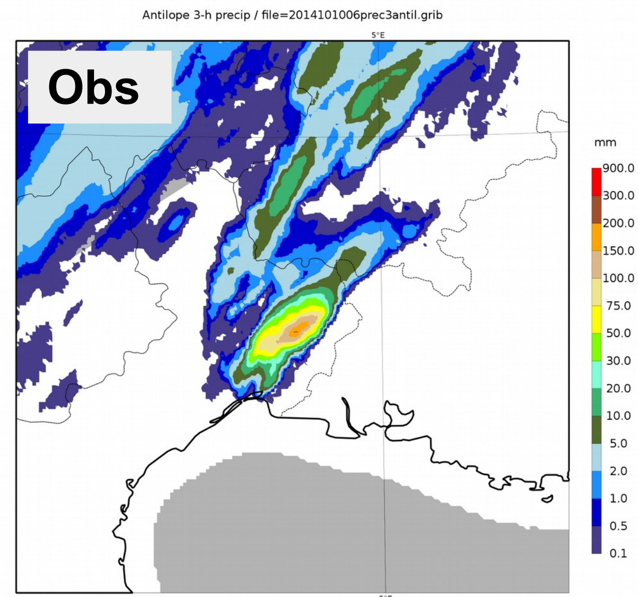


Forecasting extreme events

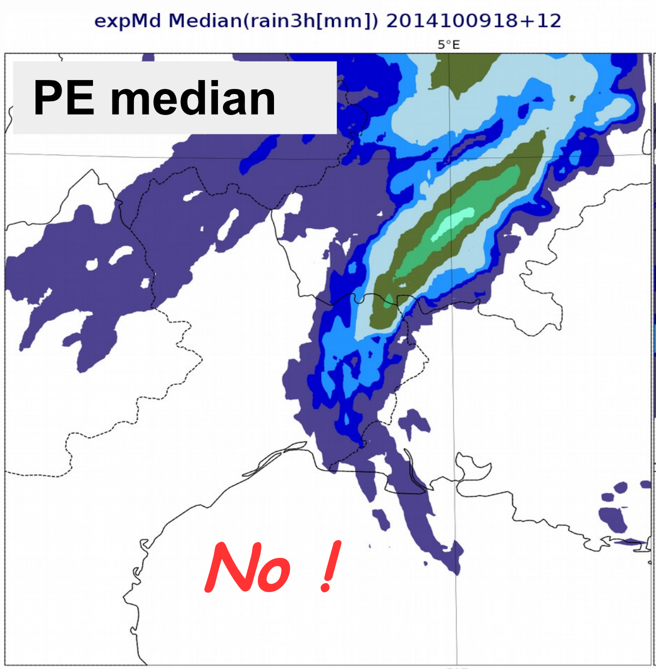
Need to use high quantiles because (1) users have a low cost/loss ratio, (2) the event is "spatially rare" (is is much smaller than the area at risk)

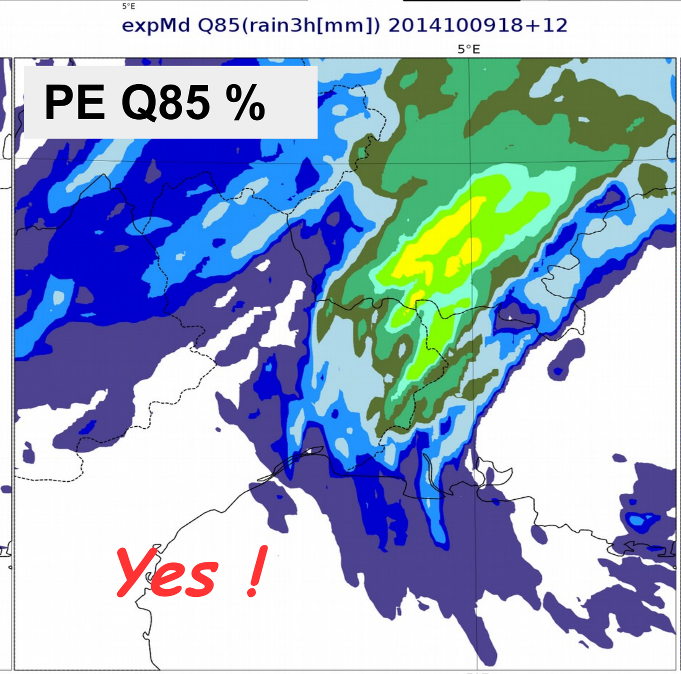
rr3h
base 9 oct 2014
step 12h

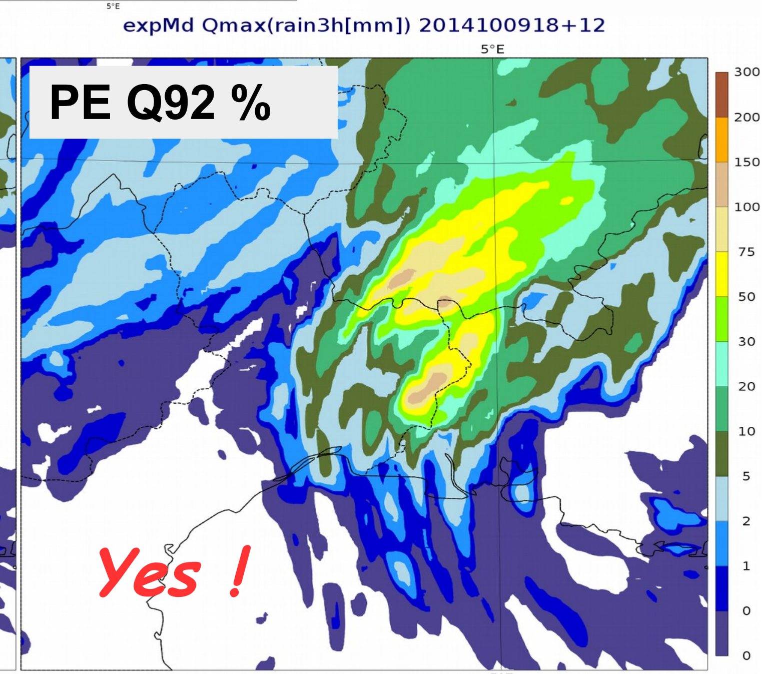




exp Md - plot e842

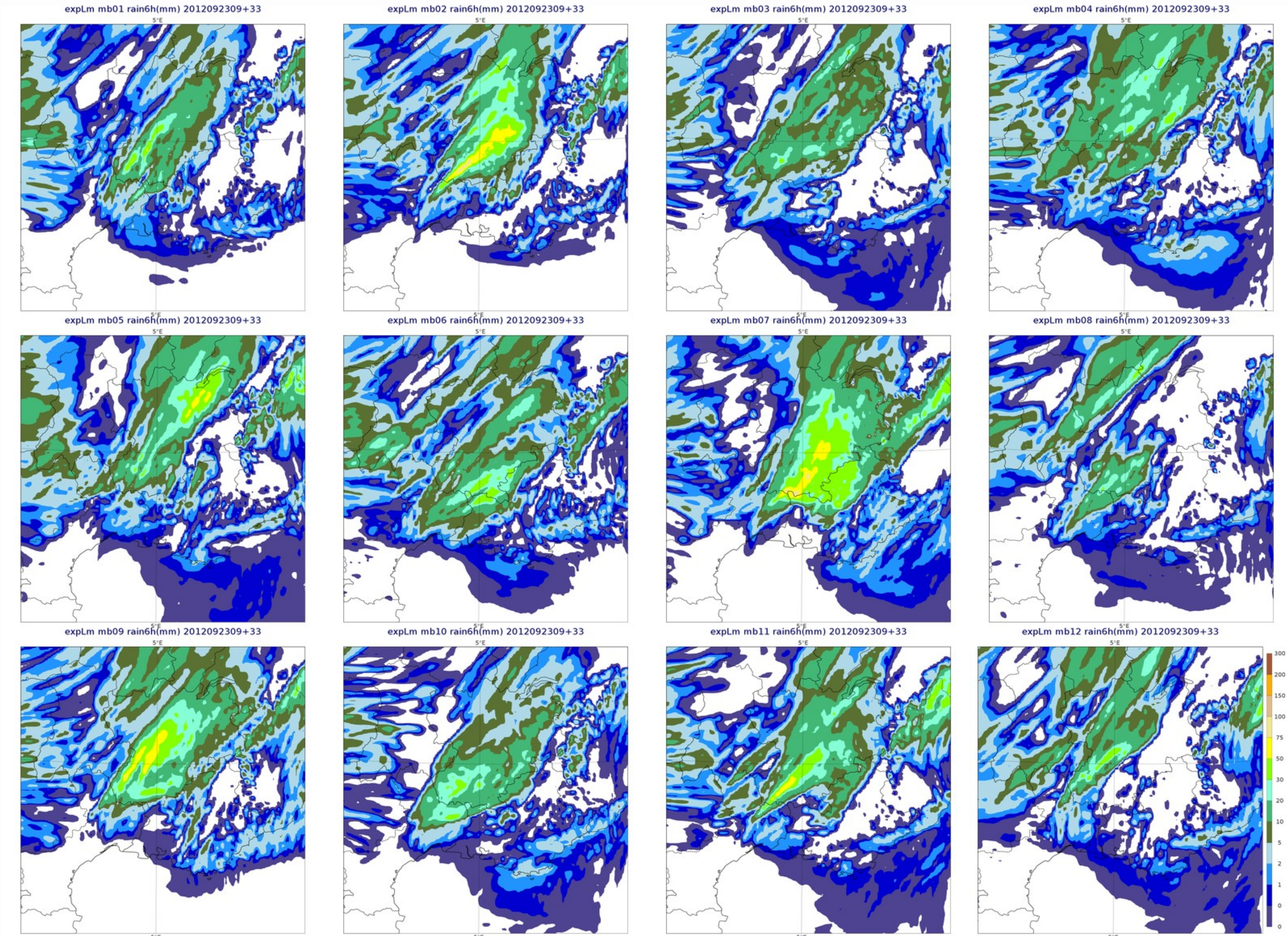




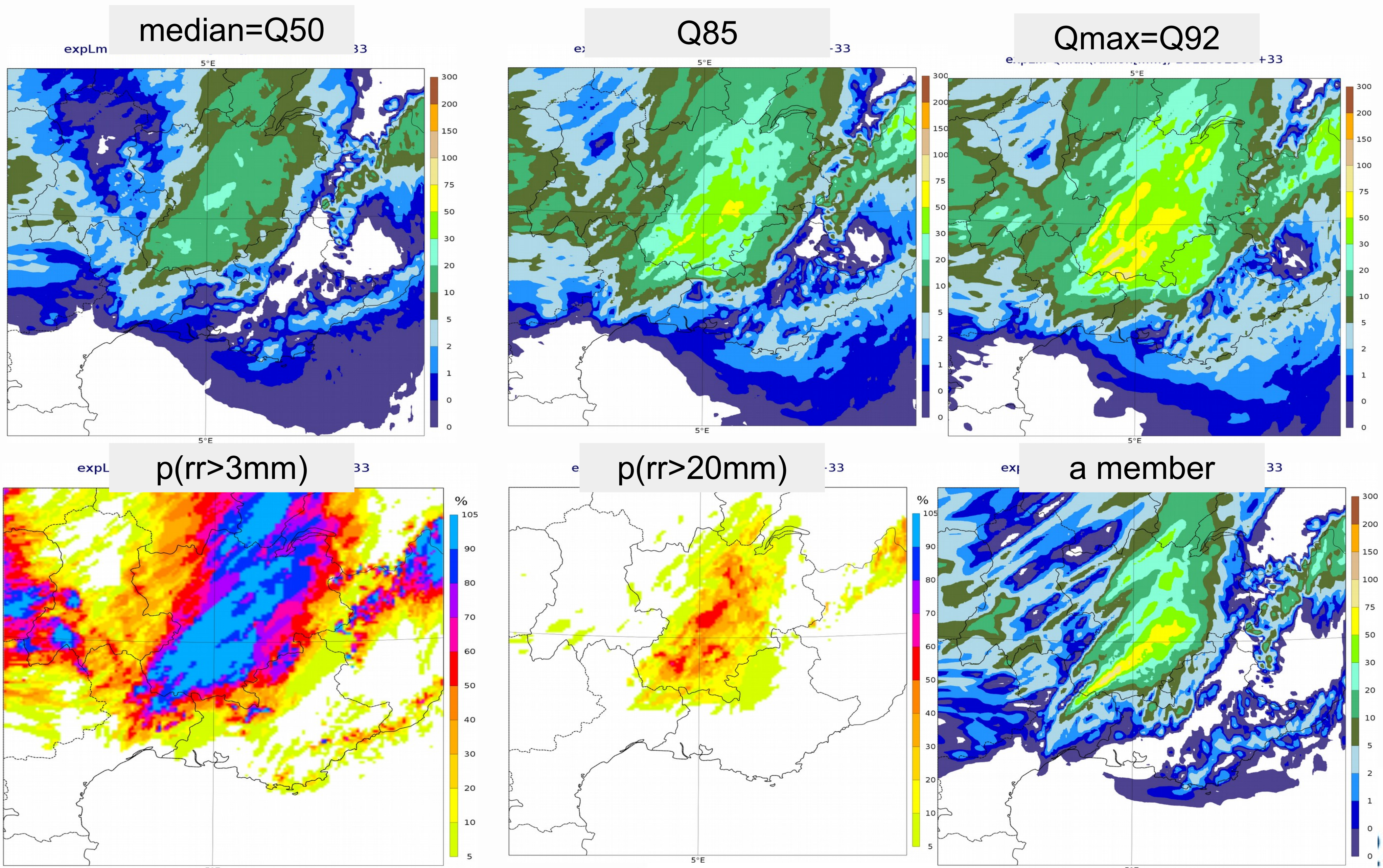


Nadine forecast, Arome-EPS (12members)
base=2012092300 range=33 param=6-hourly rain

all members



Nadine forecast, Arome-EPS (12members)
base=2012092300 range=33 param=6-hourly rain



Nadine forecast, PEARP (35members)
base=2012092300 range=33 param=6-hourly rain

Much smoother & weaker than Arome-EPS, similar location (because same synoptics)