Verification of precipitation forecast in HARMONIE-AROME

Javier Calvo, Gema Morales and Daniel Martín

40th EWGLAM & 25th SRNWP Meetings, 1st/4th Oct. 2018 Salzburg (Austria)

1. Questions

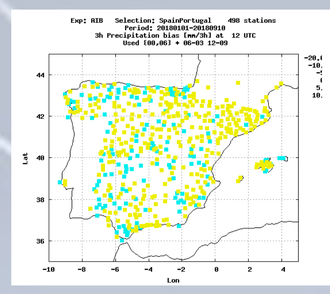
- Diurnal cycle of precipitation
- Spin up
- Double penalty issue
- Effective model resolution
- Objective point verification. Limitations
- Spatial verification
- How to decide that a model version is better than other

2. Experiments

The following model/experiments are included in the study

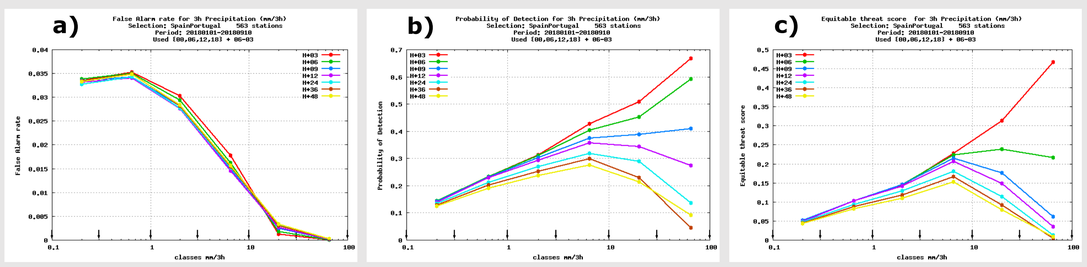
- cycle 38 and cycle 40 of operational HARMONIE-AROME. 2.5 km convection-permitting model. Both cycles differ in the assimilation and in the physics leading to significant differences in the precipitation forecasts but it is not straight forward to decide which version is better
- IFS/ECMWF. Deterministic High Resolution version.

3. Verification network

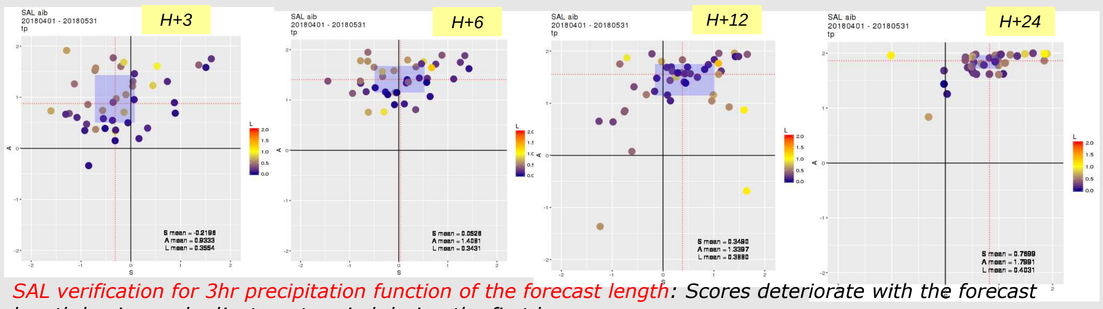


Bias of 12 hr accumlated precipitation. No clear regional behaviour is found. The resolution of the rain gauge network can also be seen in the plot.

4. Verification function of the forecast length

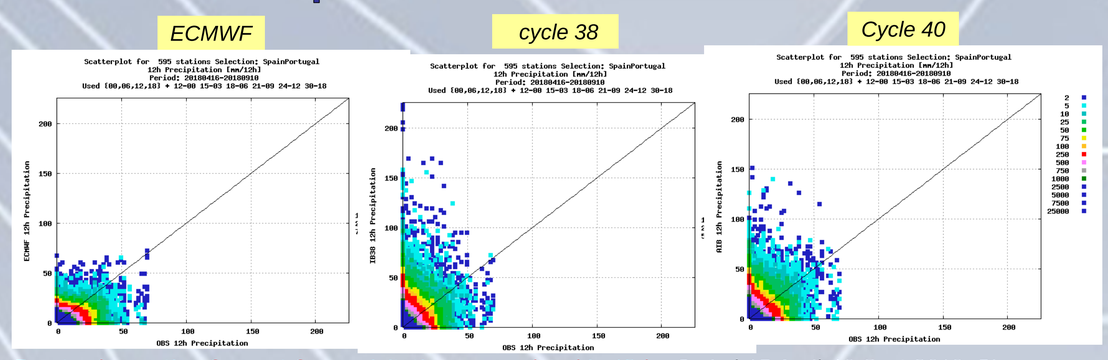


Categorical verification for different forecast lengths: (a) False Alarm Rate, (b) Probability of Detection and (c) Equitable Thread Score. Short range forecast are generally better but have bigger FAR

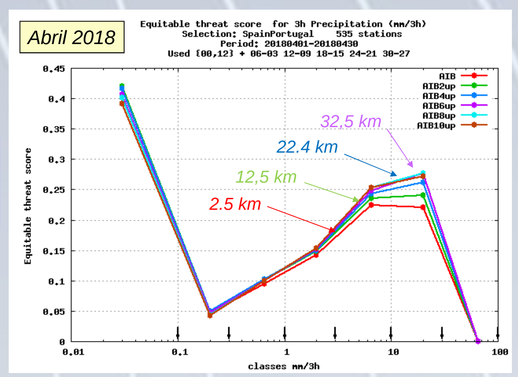


SAL verification for 3hr precipitation function of the forecast length: Scores deteriorate with the forecast length having and adjustment period during the first hours

5. Issues of the point verification

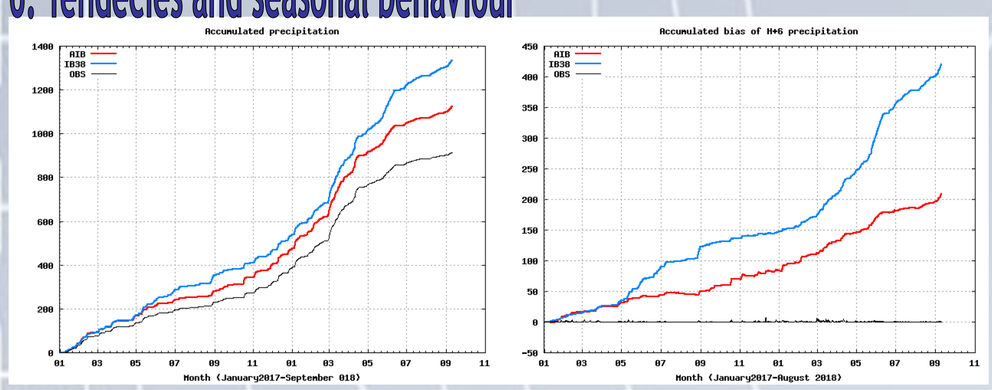


Events observation-forecast for precipitation accumulated in 12 hr. Period 15 April-10 Sept 2018. Big dispersion showing that the local effects are not well represented in the models. ECMWF tends to produce large areas of small precipitation and can not produce amounts above 60 mm/12h. HARM-AROME is able to produce big amounts but errors are important, specially the location errors. Even more, there is the problem of the representativiness of the observations with generally underestimate the maxima due to the poor resolution of the rain gauges network



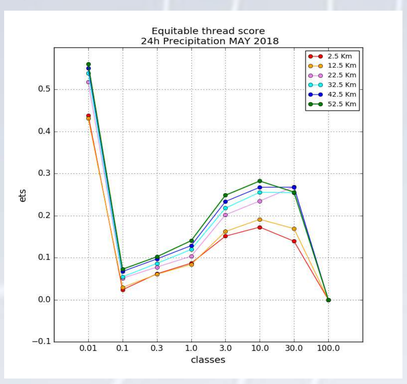
Double penalty issue. ETS for different categories using various upscaling lengths. ETS improves increasing the size of the averaging grid up to 32 km where it saturates. The original resolution of the model is 2.5 km and the observations are not upscaled.

6. Tendecies and seasonal behaviour



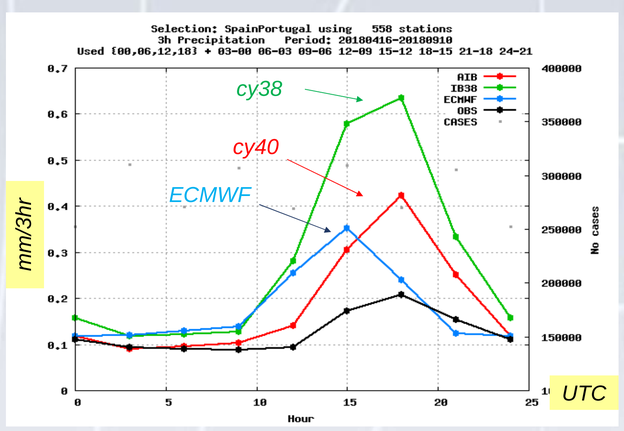
Accumulated ppt and acc. bias for 2 exps. and 2 hydrological seasons (2017 & 2018). cy38 and cy40 are compared with obs. Both exps seem to overestimate (cy38 more) the ppt but probably the obs und). cy40 underestimate the ppt due to their resolution. The differences are small in the cold months and increase in the convective period

7. Effective model resolution



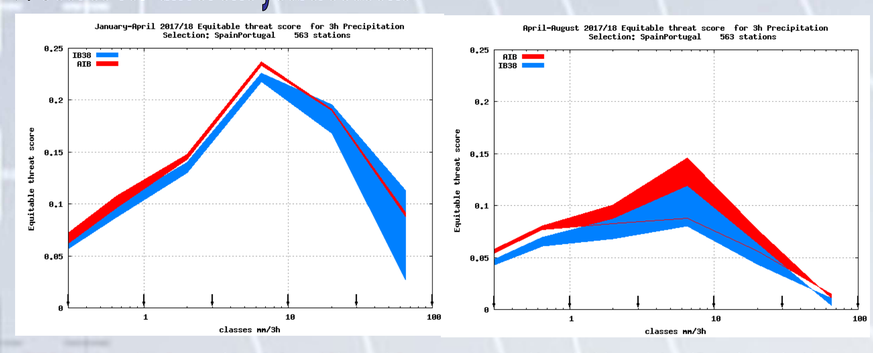
ETS of 24 hr ppt for different upscaling lengths (both forecasts and obs). The scores improve with the upscaling at least up to 50 km pointing out that the effective resolution may be at least of this order. The drawback in this sensitivity test is the poor resolution of the observation network.

8. Diurnal cycle of precipitation



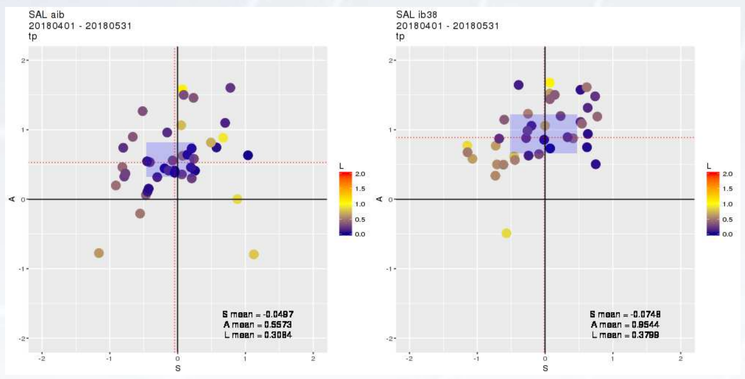
Diurnal cycle of precipitation. As expected, convection-permitting models reproduce better the diurnal cycle than models with parameterized convection (ECMWF). The maximun takes place between 15-18 UTC whereas in ECMWF occurs 3 hr earlier. All models tend to kill convection too quickly after the máximum

9. Scores for different years/seasons

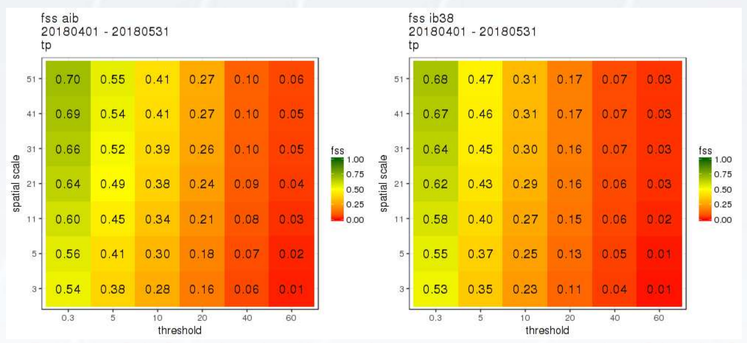


ETS 3hr ppt for 2 exp (Cy38 & cy40) and 2 seasons: Jan/Apr (left) and Apr/Sep (right). Harm-Arome 2.5 km forecast are compared with rain gauge ppt. The scores are significantly lower in the convective season. The shadding is plot between the curves corresponding to 2017 and 2018. The latter has been a very humid year with frequent ppt events and a lot of soil moisture in the ground. Cy40 performs better than cy38 specially for convection.

10. Spatial verification

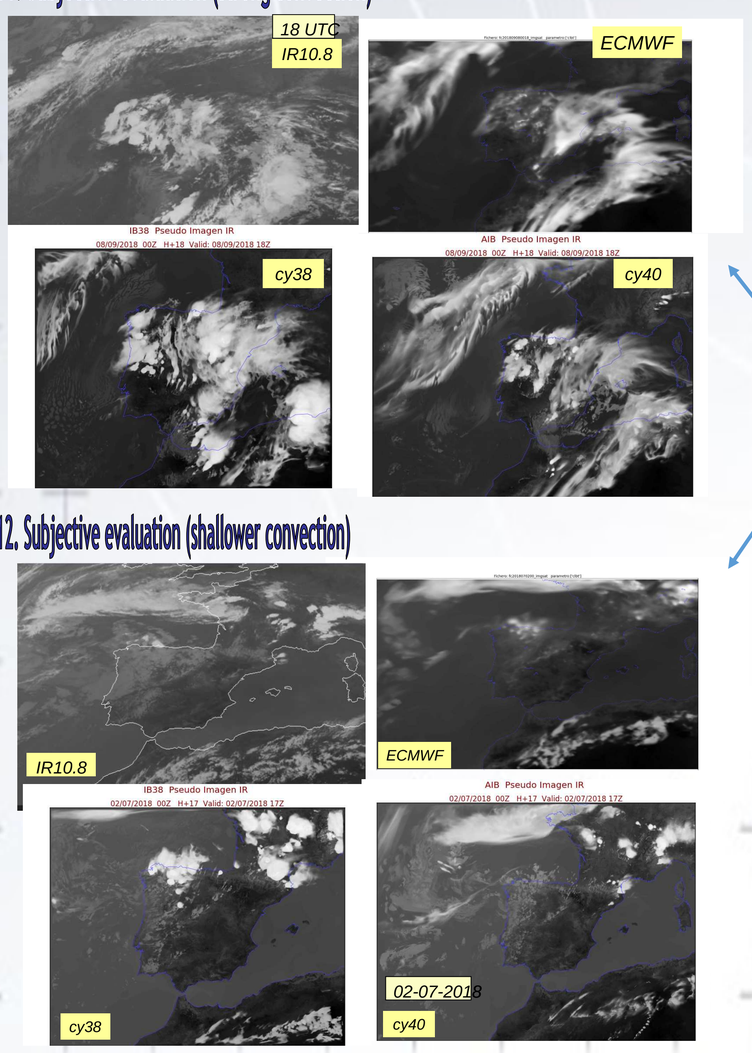


Spatial verification (SAL) for 24hr ppt for 2 exps cy40 (left) and cy38 (right). Both experiments show very small Structure error and a overestimation of the Amplitude, specially cy38 in agreement with point verification.



Fractional Skill Score function of the grid scale and the threshold (ppt/24h). for 2 exps cy40 (left) and cy38 (right). Scores improve with the length scale and seems to saturates around 40-50 km. Cy40 verifies better for all the thresholds.

11. Subjective evaluation (strong convection)



12. Subjective evaluation (shallower convection)

Simulated satelite images compared with MSG ones for 2 weather types: one with generalized intense convection (4 upper pannels) and another with shallower convection (4 lower pannels). Cy38 tends to produce more intense convection and generally, ECMWF underestimates the convective activity. The differences are larger in cases of lower active convection when cy38 tends to produce more false alarms.

13. Conclusions

- Verification of precipitation is a complex issue. From objective verification, many statistics can be computed but they do not give the same signal always. Spatial verification may complete point verification but its application is tricky and at the end is just another ingredient to help in the assessment of the forecasts quality. Resolution of the obs. is a key aspect for the spatial verification. We intent to use the radar analysis calibrated with gauges although for Iberian Peninsula is not a specially good product due to the complex orography and the variety of weather regimes.
- Good news is that convection-permitting models reproduce the diurnal cycle of precipitation much better than models with parameterized convection.
- The performance is significantly poorer for convective precipitation and just by looking at different model versions, we can evaluate the uncertainty in the convection representation.
- For precipitation, the effective model resolution is much lower than the model grid. In this study, we have seem that this effective resolution may be 50 km even for a 24 accumulation.
- To complete the evaluation we think is important to perform a subjective evaluation, specially to account for extremes.