European Commission

Will post-processing always improve my forecasts?

EGU General Assembly 2020
Sharing Geoscience Online

And some explanations to why it sometimes doesn't happen

Jon Olav Skøien

Peter Salamon

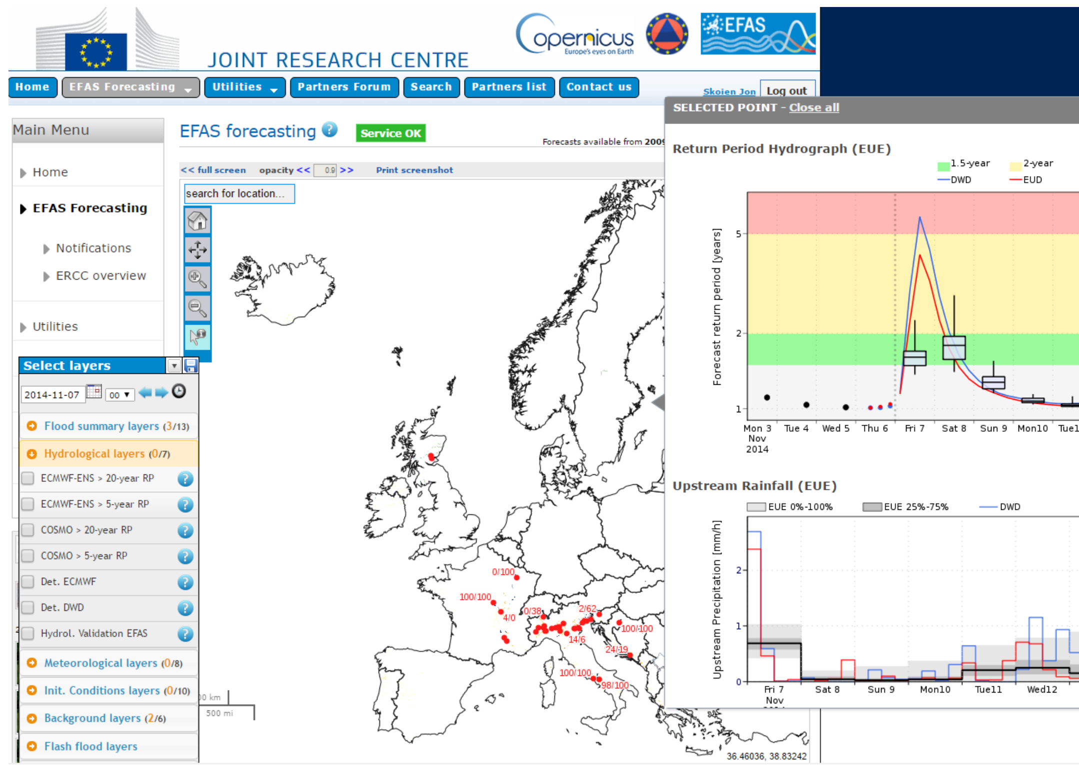
Fredrik Wetterhall

ECMWF

Joint Research Centre

Context

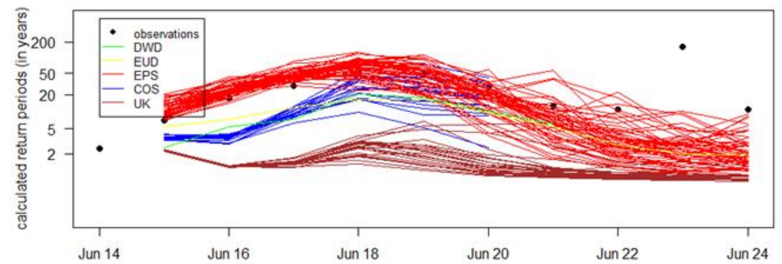
- European Flood Awareness system (EFAS) based on runoff ensemble forecasts
- Operational forecasts of flood levels for Europe
- 1-10 days probabilistic forecasts
- National HydroMet services as partners
- Part of Copernicus Emergency Management Service (EU)

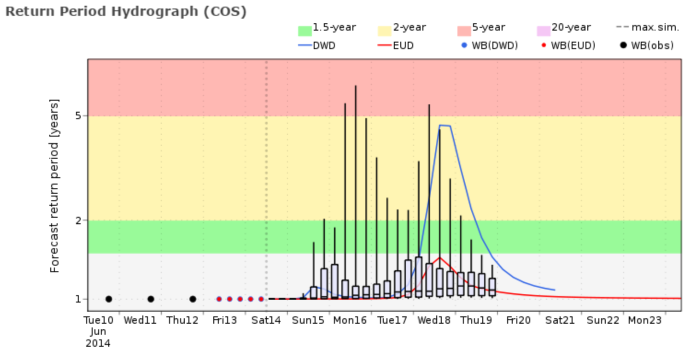


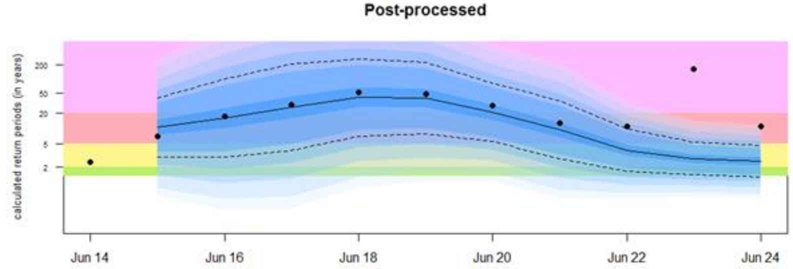
European Commission

Post-processing of ensemble forecasts

- Usually recommended
- Should reduce biases and dispersion errors (ensemble variance too large/too small)
- Here: Using Ensemble Model Output Statistics (EMOS) on continental scale hydrological ensembles (LISFLOOD), based on different meteorological forecasts (ensemble from ECMWF and COSMO, deterministic from ECWMF and DWD)
- NOTE: We're comparing with simulated runoff (based on observed precipitation), not runoff observations





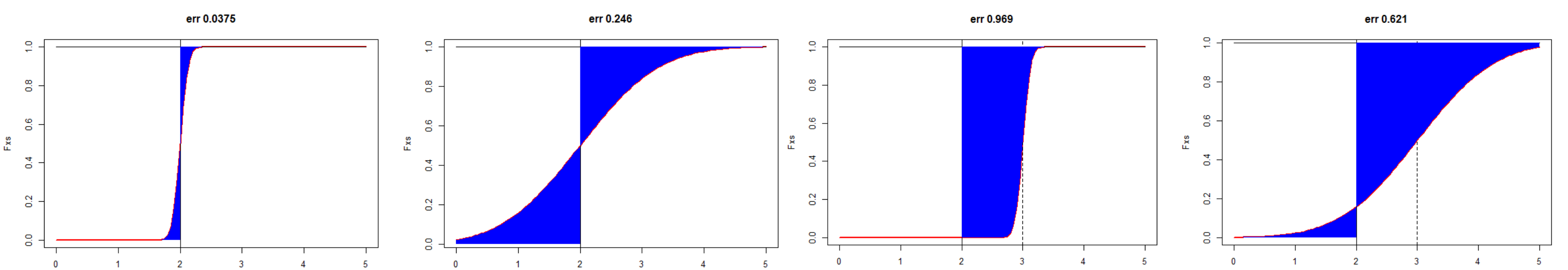


Calibrating with Ensemble Model Output Statistics (EMOS-method)

- Not all forecasters are equally good, best prediction as weighted mean
- Variance as function of ensemble variance - correct dispersion errors

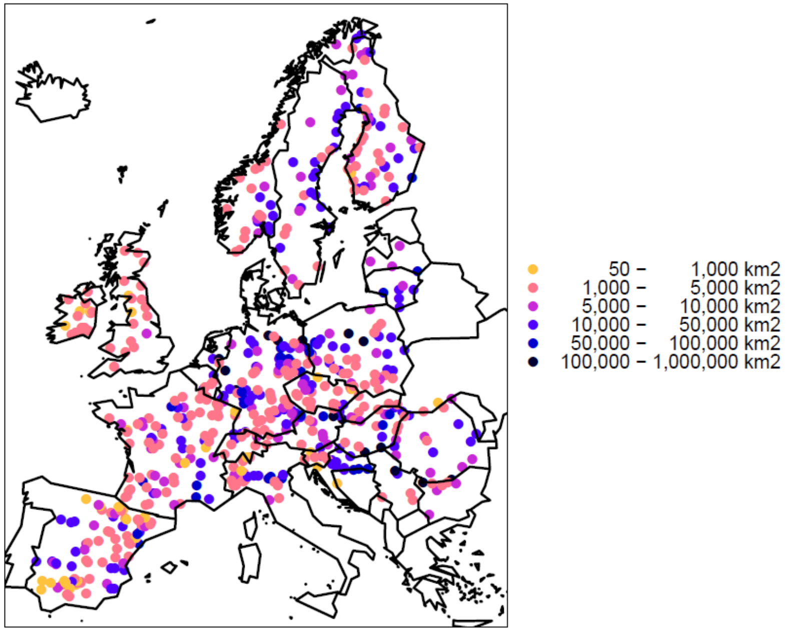
$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n + e$$
$$Var(e) = c + dS^2$$

- Minimizing Continuous Ranked Probability Score (CRPS) – which is a combination of prediction error and dispersion error (variance too high or too low – there should be a reasonable relationship between prediction variance and prediction error)
- CRPS-errors in blue below. The two panels to the left have correct forecasts, the two panels to the right with prediction errors. Panels 1 and 3 have low prediction variance, panels 2 and 4 have large prediction variance.



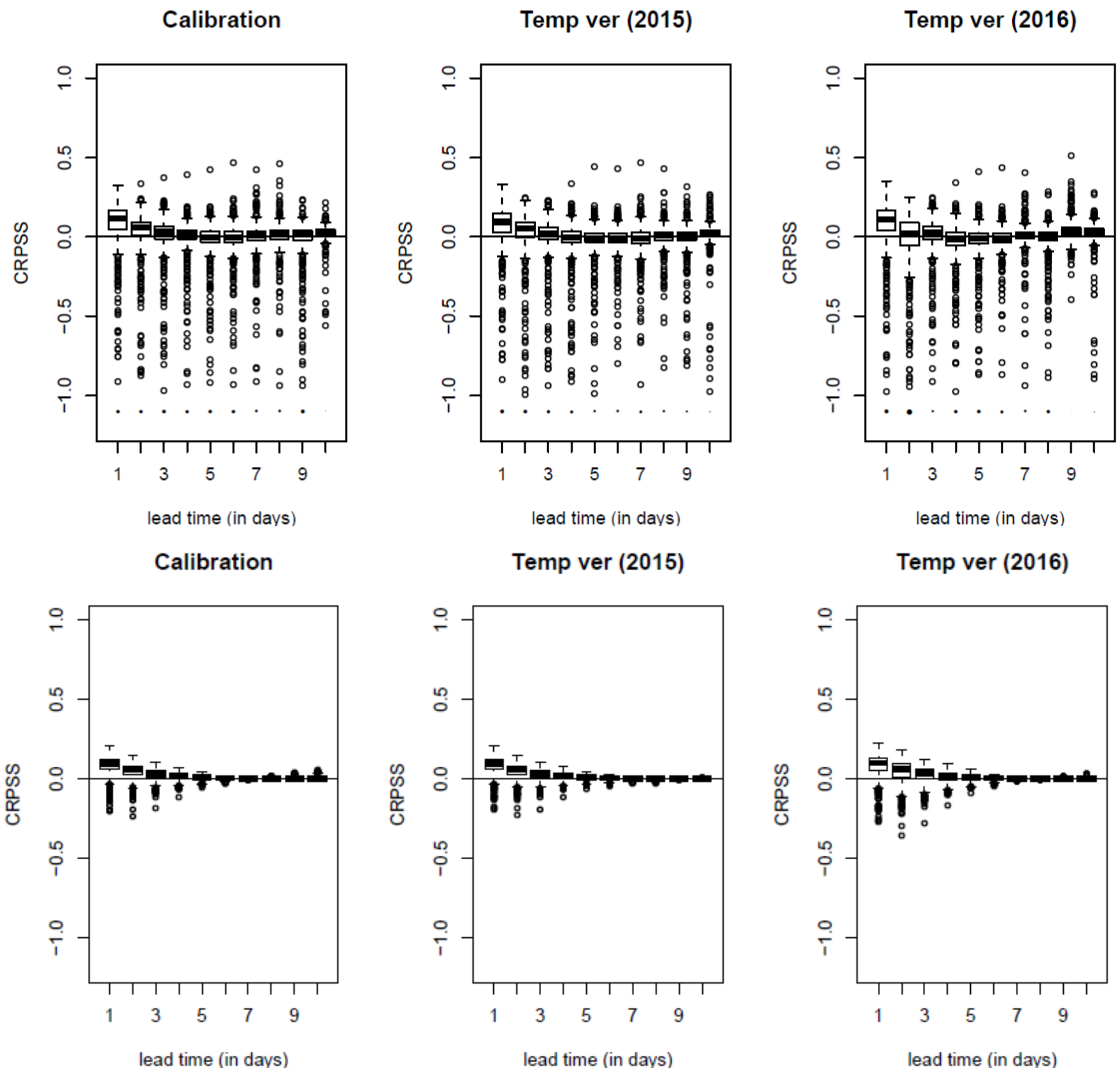
Calibration and validation

- Simulated runoff data from observed precipitation and from weather forecasts from more than 600 stations in Europe for 2015-2017
- Using Continuous Ranked Probability Skill Score (CRPSS) as a measure of the improvement from post-processing.
- CRPSS compares CRPS-errors for raw ensemble and post-processed ensemble.
- CRPSS is above 0 when the post-processed distribution is better than the raw ensemble (up to 1).
- Calibrating for each year separately. Post-processing ensembles from 2017 with calibrated parameters from 2016 as validation.



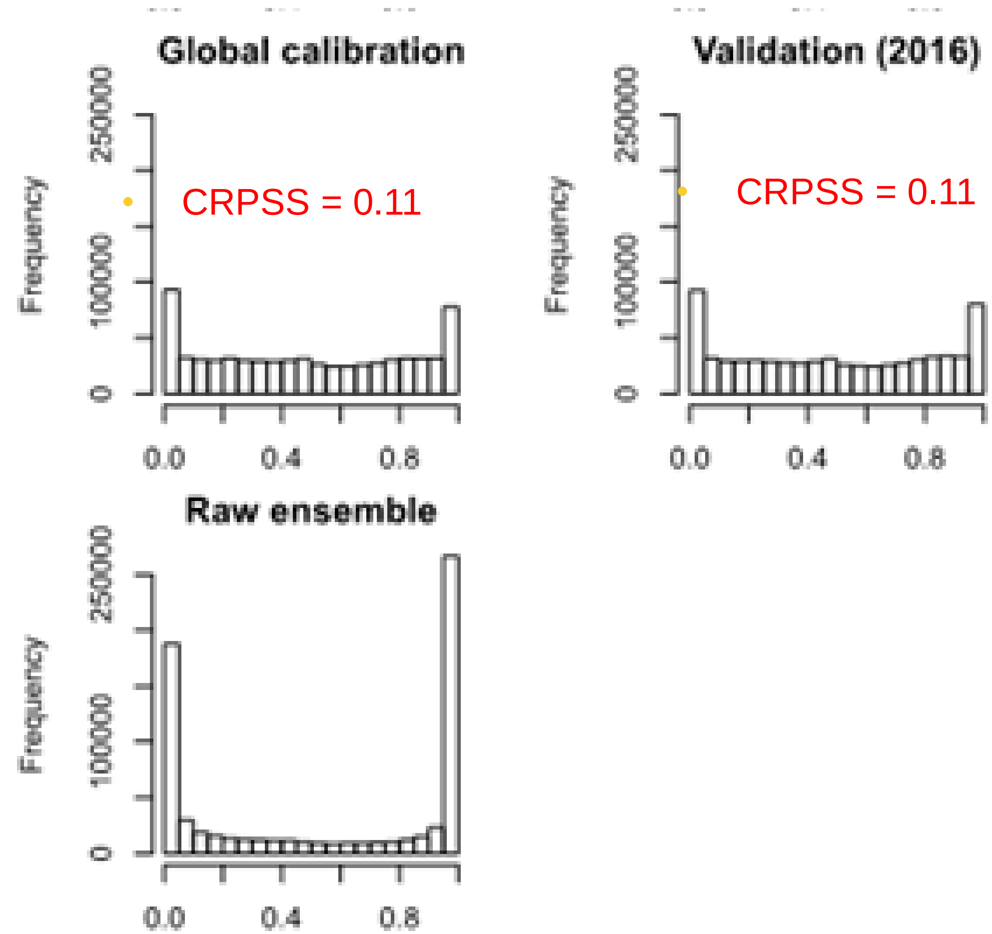
CRPSS-results

- Calibration and validation results from EMOS
- Only some improvement for short lead times
- Only variance inflation parameter fitted in the lowest panels – similar results



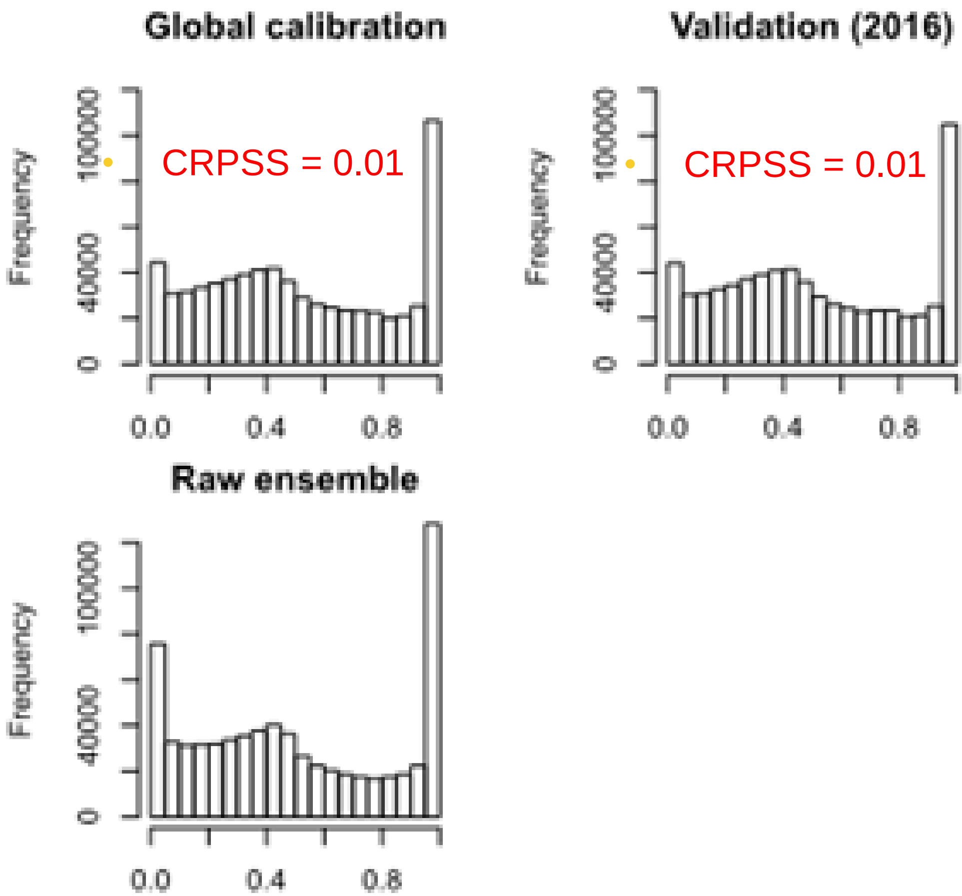
PIT-diagram of raw ensemble and post-processed distribution – 1 day lead time

- Probability integral transform diagrams is a check on how the observations plot in the predictive distribution of the ensemble. It should ideally have a uniform distribution.
- The raw ensemble is strongly underdispersed (bottom)
- PIT-diagrams of post-processed distribution are more uniform, validation almost as good as calibration
- CRPSS around 0.1 (~ 10% improvement)



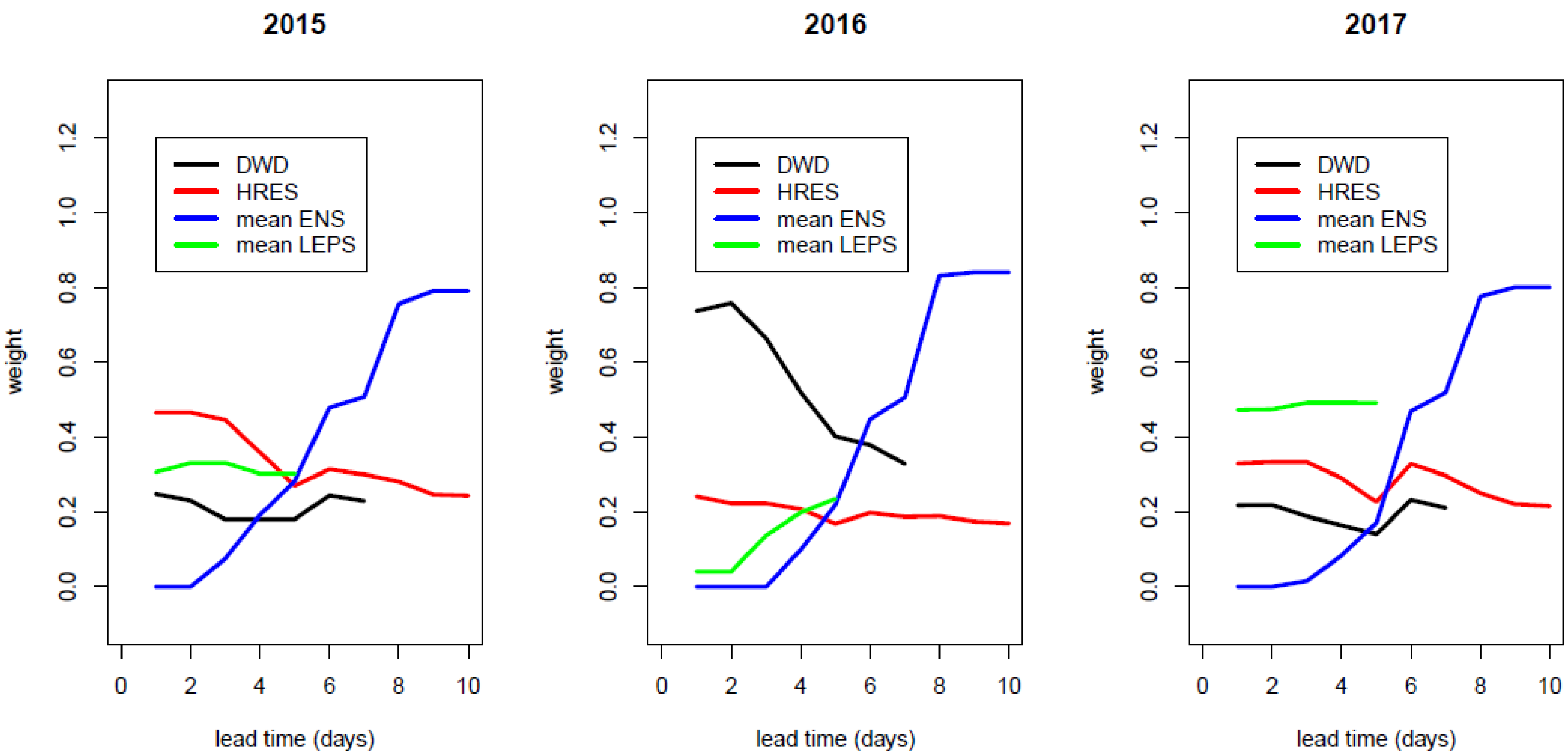
PIT-diagram of raw ensemble and post-processed distribution – 7 days lead time

- The PIT diagram of the raw ensemble is almost trimodal (bottom)
- The variance post-processing is not able to reduce the underdispersion without simultanously increasing the peak in the middle – best fit is to leave the variance as it is.
- CRPSS is close to zero
- The ensemble for 7 days lead time is mainly from ECMWF, so mixing methods as Bayesian Model Averaging (BMA) would probably not improve.



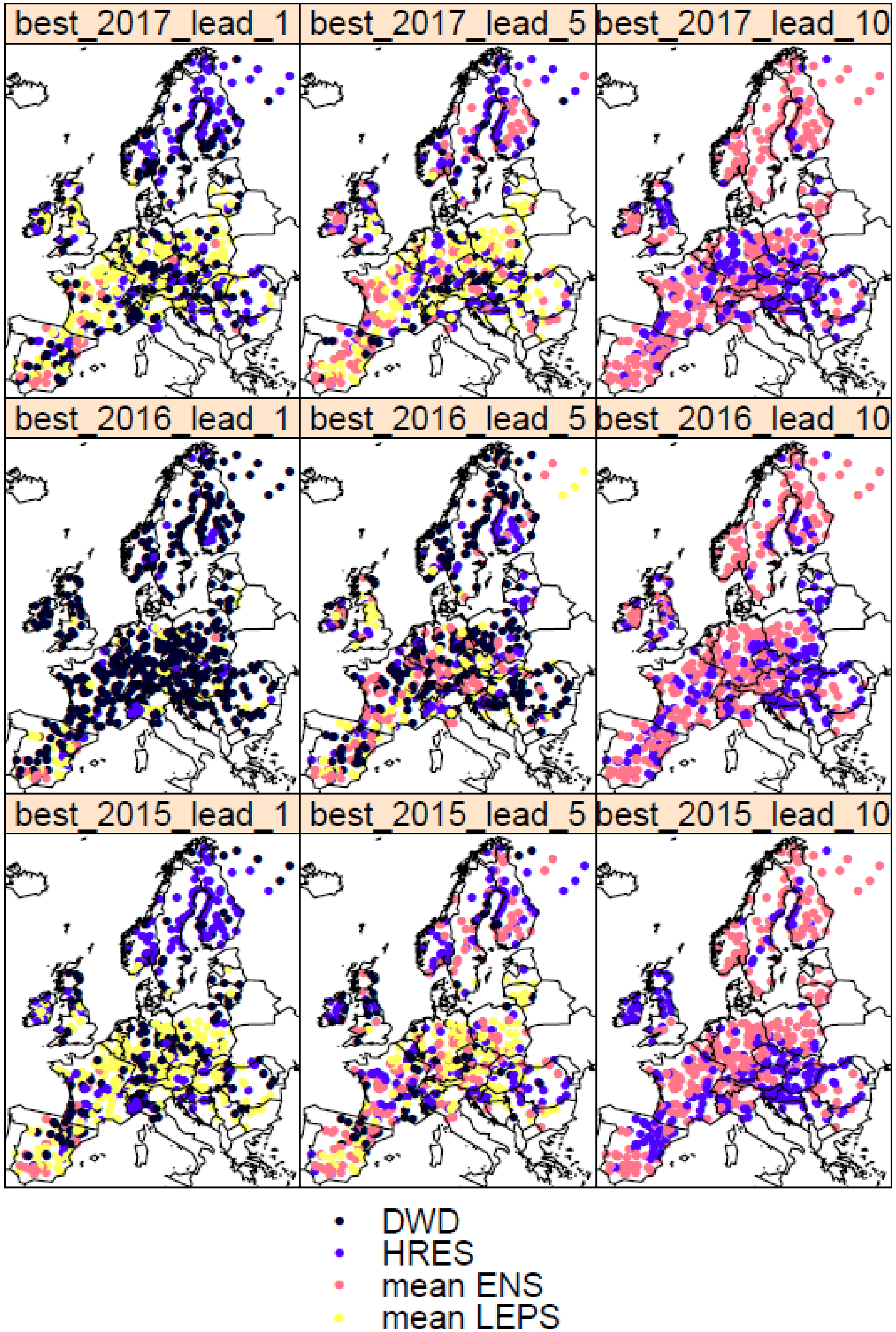
Weights change from year to year

- There is little consistency in the calibrated weights for each forecast between the different years
- The best set of weights for the calibration year is not necessarily the best set for the validation year



Analysing NSE

- We computed Nash-Sutcliffe efficiency for each forecast, each station and for each year.
- The figure shows which forecast gave highest NSE for each station and each year for lead times 1, 5 and 10 days.
- Some spatial patterns visible for individual years, but not between years



Conclusions

- Post-processing helps for lead time 1-3 days in our case, but not for the remaining days
- Bias correction seems unnecessary for this application. Maybe also as a result of using simulated runoff for comparison (this is done as EFAS is focusing on return periods, not on the runoff itself). However, there could still have been a bias between meteorological observations and forecasts.
- There are no forecasts that consistently give better results for different years. We have tried shorter calibration/verification periods, with same result.
- The variance is underestimated by the model for short lead times. Post-processing can help in this case.
- The ensembles for long lead times have a multi-modal distribution, which cannot be easily fixed by the solution we have tested here.

Keep in touch

EU Science Hub: ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

Eu Science Hub

European Commission

Thank you

European Commission