MOSAC 23 and SRG 15 Meetings

January 2019

MOSAC PAPER 23.9

IMPROVER - the new post processing and verification system

Nigel Roberts, Ben Fitzpatrick, Jonathan Flowerdew, Simon Jackson, Ken Mylne & Bruce Wright

1. Introduction

The post-processing component of the Numerical Weather Prediction (NWP) forecast chain is an essential part of our weather forecasting capability. It is the stage that has the greatest potential to provide the biggest quick gains in both information provision and improved forecast skill. Post processing has previously largely focussed on correcting coarse-resolution deterministic forecasts to provide better local information, but requirements are now radically changing. We will soon operationally run expensive "convection-permitting" ensemble UK forecasts (2.2km MOGREPS-UK) that are updated every hour and extend to five days, in addition to hourly deterministic convection-permitting UK forecasts (1.5km UKV) and the global ensemble MOGREPS-G. This introduces a host of new opportunities for better, more innovative, forecast products as well as bringing new difficulties. One difficulty is the serious danger of information overload for our operational meteorologists. To make use of this radical advance in NWP capability we should aim to fully exploit this new wealth of highly-detailed forecast information in a scientifically sensible way. That means extracting the important signals on appropriate time and space scales from multiple forecasts, whilst retaining consistency, improving accuracy and providing an indication of uncertainty. A further challenge is how to adapt to the way forecasts are presented and utilised these days. Forecasts used to be disseminated via television, radio and newspapers for general areas every few hours, nowadays the public expect hourly forecasts for their location of interest updated on their mobile phones or other devices as well as timely warnings of severe weather. The onus is on us to produce automated forecasts for any location in the UK (and potentially the globe) when we know there are inherent predictability issues at those local scales.

An investigation in how to proceed with post processing into the next decade and beyond was performed in 2015-16. This resulted in a strategy document for a new probabilistic post processing system in March 2016. The basis of the strategy was presented to MOSAC and endorsed prior to final release (Roberts & Mittermaier 2015). Later in 2016, work started on the development of the new probabilistic post processing system called IMPROVER, which is described here.

Key principles of the post processing strategy that are being built into IMPROVER are:

1. A single processing chain for each variable which operates on a standard grid with location-specific information extracted at the end.
2. A fully probabilistic system that allows for time-lagging and blending between models/ensembles and provides probabilistic forecasts out to two weeks.
3. Verification at every step through the chain so that the performance of each can be measured and optimised.
4. The ability to provide outputs to enable automated forecasts of "ordinary" weather for any location as well as more granular information about the possibility of high-impact weather for constructing warnings and risk estimation.
5. A modular software framework following modern professional software development practices.

A lot of progress has been made in the development of IMPROVER in the last two years. This paper will describe the current capability and outline the development timescales and some future aspirations.

2. Probabilistic processing chains

The construction of IMPROVER has focussed on building modular processing chains for all the key variables. The processing chains operate on NWP model output from multiple models that have already been re-projected on to the standard UK or global grids. The gridded forecasts can be processed through several steps as outlined in Figure 1. Some of these steps already have well developed functionality and others are still at a very early stage. A key aspect to highlight is that the infrastructure is already in place to pull through either ensemble or deterministic forecasts and convert those forecasts into probabilities (or percentiles), which can then be time-lagged and blended with other models to produce a probabilistic (or extracted deterministic) gridded or spot-location forecast for each variable. The "V's" represent verification at each stage. Level 1 defines individual forecasts on the standard grids, level 2 defines data that is internal to IMPROVER, and level 3 defines blended probabilistic output from IMPROVER. Only levels 1 and 3 data can be accessed by external users.

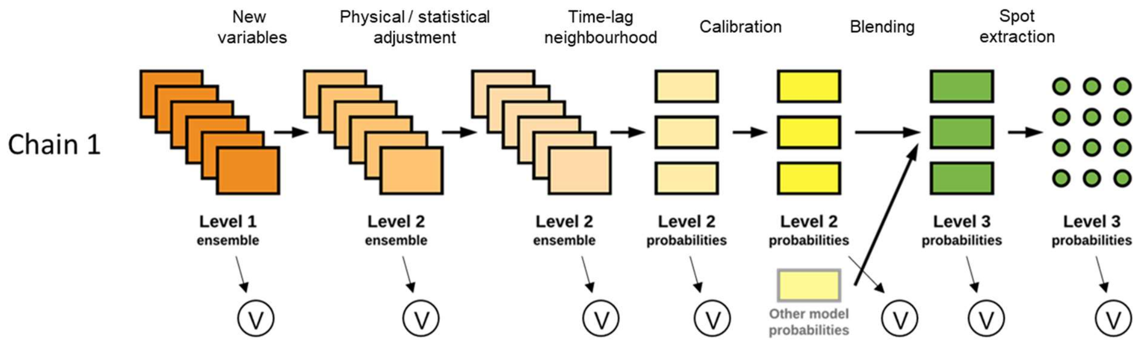

Figure 1. Schematic of the generic processing steps available in each chain. See text for more details.

The steps are as follows:

(1) New variables. This is where new variables can be computed from each individual forecast. These are either needed for further post-processing later in the chain or as derived quantities to be passed to users. Examples are: wet-bulb temperature and snow melting level, feels-like temperature, temperature lapse rate (Sheridan 2015) and diagnosis of convective or stratiform precipitation.

(2) Physical / statistical adjustments. This is where adjustments and corrections can be made to each individual forecast using either physical or statistical methods. It currently includes physically-based downscaling the wind speed and temperature to be more representative of more accurate topography.

(3) Time-lag-neighbourhood processing. This is the stage at which individual forecasts are converted to probabilities (or percentiles). It first involves applying “fuzzy” exceedance thresholds to obtain probabilities. Ensemble members are combined and time-lagging is applied to produce combined probabilities for each threshold. For example, for the hourly-cycling 3-member MOGREPS-UK, the three forecasts are combined by computing a mean probability and then the mean is taken over the six (flexible) most recent forecast cycles to construct probabilities from a time-lagged 18-member ensemble. Neighbourhood processing (Schwartz and Sobash 2017) is used to effectively increase the ensemble size and produce smoothly-varying probabilities. The neighbourhood processing is currently applied after the three members are combined. A variety of neighbourhood techniques have been introduced.

(4) Calibration. This is another location in the chain where statistical correction (see section 4.3) can be applied – this time to the ensemble probabilities.

(5) Blending. This is where probabilities from different models are combined (see section 3).

(6) Spot extraction. Here, probabilities or percentiles are extracted for spot locations. The spot extraction can adjust for altitude and use the most appropriate local grid square. Weather codes are generated on the grid (see section 4.4) and then used for spot locations. Deterministic outcomes can be extracted too.

3. Model blending

IMPROVER is designed to ingest and blend forecasts from multiple models/ensembles to produce seamless probabilistic forecasts (blending step in Figure 1). This use of multiple models is depicted in Figure 2. The coloured bars show the current models incorporated for the UK grid and the red boxes the blending periods, with the red arrows indicating flexibility in those blending periods. If we consider precipitation rates over the UK, an extrapolation nowcast is converted into probabilities and blended with probabilities from the hourly-cycling UKV up to a forecast length of 3-6h (see section 4.1). From around 6h to 12h the UKV is blended with increasingly weighted probabilities from the time-lagged MOGREPS-UK. Beyond 12h up to around 4 days the forecast is entirely comprised of MOGREPS-UK until blending into the time-lagged MOGREPS-G probabilities (projected on to the UK grid) for the remainder of the forecast out to seven days.

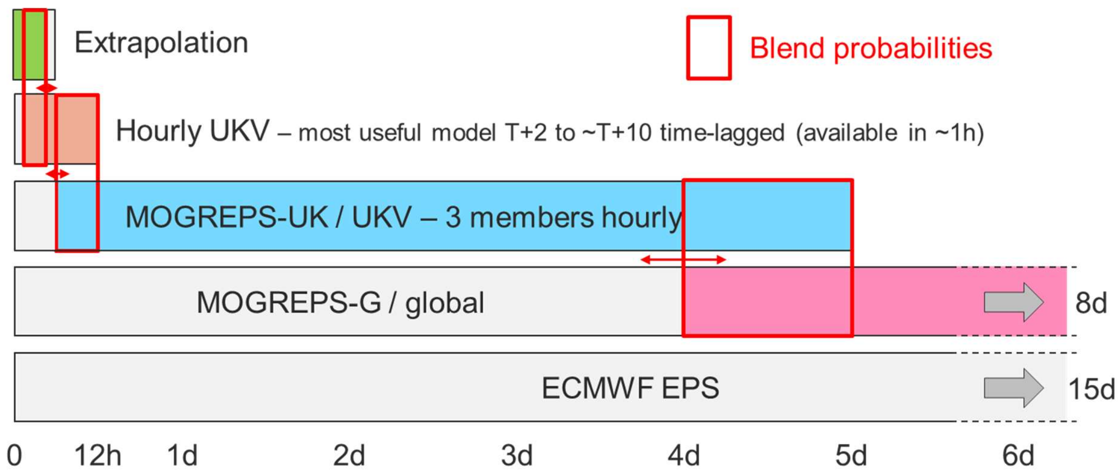

Figure 2. A schematic of the models blended in IMPROVER. See text for details.

For variables other than precipitation, very short-range forecasts are currently entirely based on the hourly-cycling UKV. Although Figure 2 shows no more than two models being blended in any forecast period, IMPROVER is designed to allow blending of more than two models at once and this will be examined in future. The current configuration does not yet incorporate ECMWF ensemble forecasts, but will next year and then forecasts will extend to two weeks based entirely on ECMWF for week two.

An example of the probability processing steps including model blending is shown in Figure 3. MOGREPS-UK members are converted into probabilities, then neighbourhood processed, then time-lagged and finally blended with a time-lagged UKV forecast.

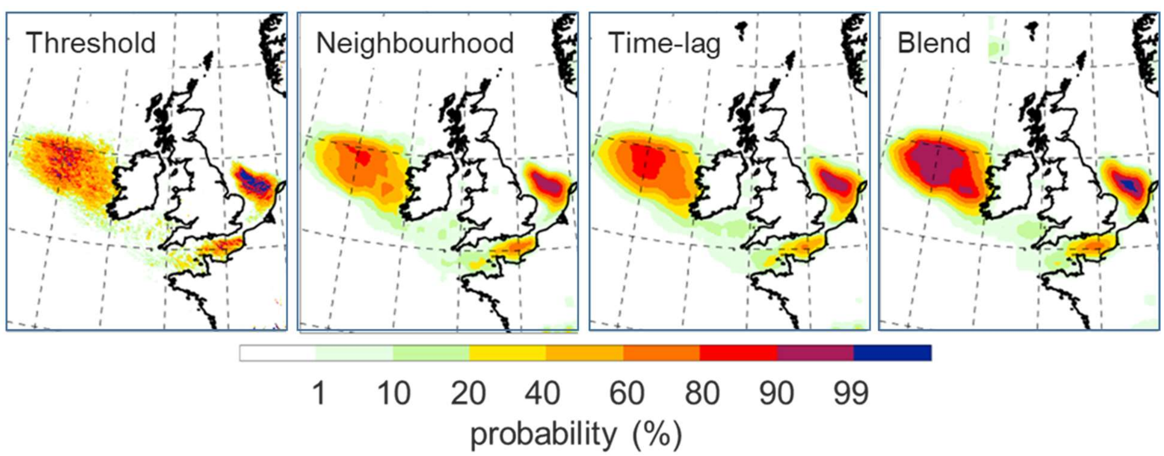

Figure 3. Probabilities of wind speed exceeding 20 m/s from MOGREPS-UK and UKV forecasts. Picture by Gavin Evans.

4. Some examples of capability and opportunities

4.1 Precipitation nowcasting

A fundamental aim of IMPROVER is to provide probabilistic precipitation forecasts from the very start of the forecast. For very short forecast ranges (0-2h) an extrapolation of radar precipitation typically outperforms the most recently available UKV forecasts (Simonin et al 2017). It is crucial the “what’s happening now” is as accurate as possible so that confidence is not undermined in the rest of the forecast; therefore radar extrapolation is an essential part. In order to incorporate radar extrapolation a separate nowcast suite called MONOW has been built. It ingests the latest national composite radar rates every 15 minutes, removes the inbuilt orographic enhancement, performs a 6-hour extrapolation using optical flow or model steering-level winds and re-inserts any orographic enhancement. This provides 6-hour extrapolation forecasts which are treated in the same way as another NWP model forecast in IMPROVER. The extrapolation forecasts are turned into probabilities in IMPROVER for key thresholds (e.g. rain/no rain) using neighbourhood processing and then blended with the time-lagged hourly-cycling UKV. The neighbourhood size used for the radar extrapolation is initially very small and increases with forecast length to account for increasing spatial uncertainty. Weighting is increased in favour of the UKV with forecast length and becomes entirely UKV after 3-6h. Many opportunities exist for optimising this process further.

4.2 Topographic neighbourhoods

The use of neighbourhood processing is a key part of the probabilistic approach. The principle is that there are scales below which a forecast has little predictive skill and, for any given grid square, the outcomes at nearby grid squares are equally likely and can be included as surrogate ensemble members. It effectively increases the ensemble size and filters out the noise in the probabilities, making blending possible. This works well for fields like precipitation, but is not so beneficial for fields that are strongly constrained by altitude, such as fog or temperature. For temperature we might apply neighbourhood processing to account for the uncertainty in the positioning of a front, or sea breeze, or outflows from showers, but it is not sensible to include grid squares with different altitudes or with different surface characteristics (e.g. land/sea) that are not representative. To address this problem, neighbourhood processing has been developed that incorporates topography. It only includes grid squares with similar elevation and land/sea type in the neighbourhood. Figure 4 shows an example of the topographic neighbourhood in action. The standard neighbourhood unphysically smears the probabilities in hilly areas, whereas the topographic neighbourhood retains a sharper distinction in those areas whilst keeping the same filtering in other areas (e.g. over the sea). This technique should be particularly useful for low visibility when fog is constrained in valleys or on hills or coasts.

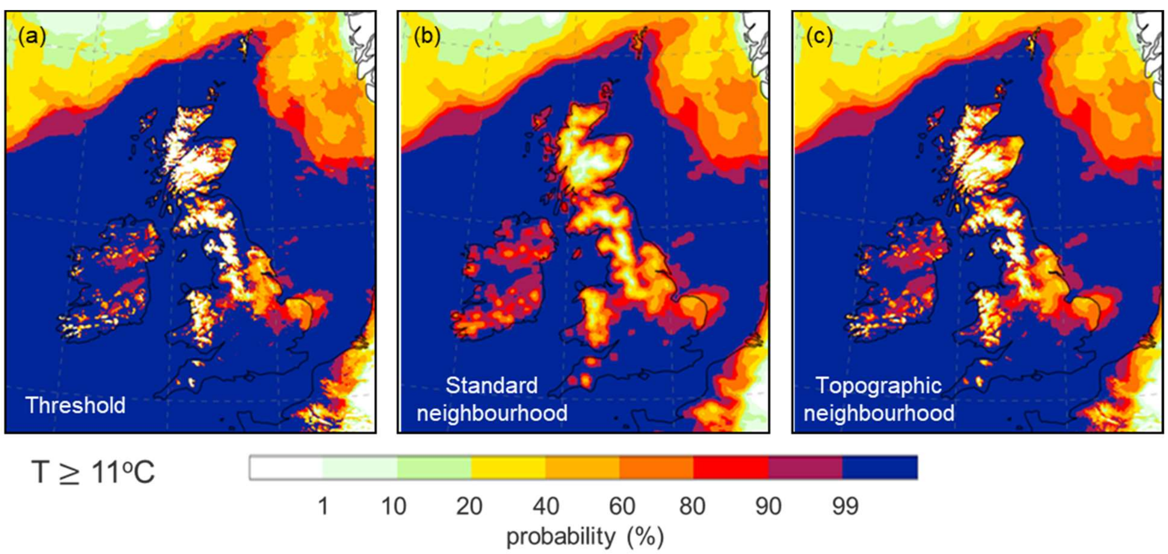

Figure 4. An example of the use of the topographic neighbourhood. Picture by Fiona Rust.

4.3 Statistical calibration

There has long been a need for statistical methods to correct biases in NWP model forecasts. In the past this was most needed to correct the unrepresentativeness of coarse-resolution deterministic forecasts for point locations. Although we are now using convection-permitting models and ensembles to better represent local detail, there is still a requirement for corrections to biases and ensemble spread using appropriate statistical methods. Statistical calibration can be applied at three stages in the processing chains; before conversion to probabilities, after blending or at the spot extraction. The method of Ensemble model Output Statistics (EMOS) (Gneiting et al 2005) was tested in a demonstration mode early in the project for temperature and wind speed and gave some positive results. It will now be incorporated after the blending stage using a rolling training period of around 15 days. Because IMPROVER operates on gridded data, the training has to be done against model analyses rather than observation sites and it will be interesting to see how this performs. There are further opportunities for use of EMOS at the spot extraction stage or introduction of other methods such as quantile mapping, Kalman Filter or machine learning (ML). ML is currently an area of research that needs collaboration with the academic community and ultimately incorporation into IMPROVER where its value can be demonstrated.

4.4 Gridded weather symbols

A significant benefit of generating smoothly-varying probabilities in space and time is the ability to automatically extract meaningful weather types on the grid – and therefore for any required location. Examples of maps of most likely weather type for the UK and the globe are shown in Figure 5. The approach provides new exciting opportunities for producing multiple possible weather types/symbols for situations when one is insufficient, such as a largely dry weather but with a small chance of a thunderstorm. Symbols are currently generated using a decision-tree approach, but this sort of categorisation is particularly suitable for machine learning and this has also been tested on short periods with promising initial results.

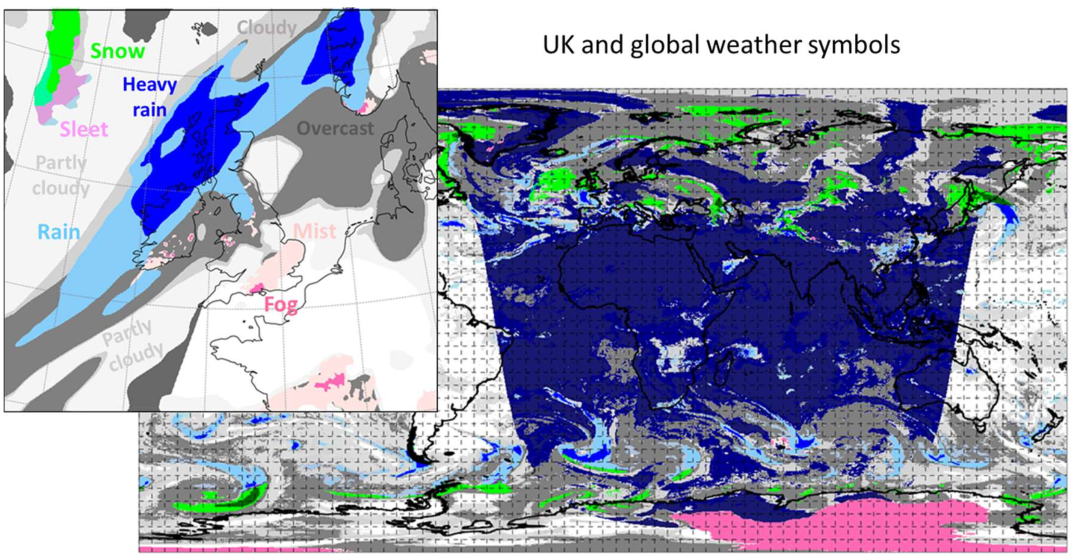

Figure 5. An example of weather types extracted from probabilities for the UK domain and the globe. Pictures by Caroline Jones

5. Verification, trialling and governance

One of the big advantages of IMPROVER compared to the current operational post-processing systems is integration with verification and a suite design which supports running of trials as well as operational running. Verification is key not just to the final assessment of the new post-processing system, but also to the development process. Proper assessment of the probabilistic forecasts at the heart of IMPROVER requires statistical evaluation against a corresponding distribution of outcomes assembled over many cases. Verification allows components to be tuned, alternatives to be compared, assumptions to be checked, and priorities for future work identified. To support these aims, IMPROVER allows the same configuration that will run operationally to be run in historic trial mode, as is common practice in NWP development. This allows new developments to be thoroughly tested and tuned in advance of operational implementation. By applying verification to each step of the processing chain, their impact on a variety of metrics can be independently assessed and optimised, retaining only those steps which improve forecast performance. In order to exploit the current operational verification capability, probability and percentile forecasts are converted back to ensemble members using part of the technique which has become known as Ensemble Copula Coupling (ECC; Flowerdew, 2014; Schefzik et al., 2013), allowing the full range of ensemble verification scores to be applied.

The first year of development delivered the basic trialling capability. Ensemble verification through the chain provides an assessment of the impact of each post-processing step, whilst verification of the final deterministic spot forecasts (extracted as the ensemble median) provides a comparison with the current operational system. Current developments include providing spot deterministic verification for each step of the chain, constructing an equivalent real-time verification for the upcoming real-time release, exploring the verification of precipitation rates (as opposed to accumulations) and adding verification that allows for the impact of observation errors. Future work will provide further verification tools to support the post-processing development, use these tools to optimise and assess the post-processing configuration, support the development of statistical post-processing, and potentially assess the performance of precipitation 'scenarios' for hydrology. Examples of verification from a single-month trial are shown in Figure 6. These show how skill varies with neighbourhood size for hourly precipitation accumulations and the benefit of the topographic neighbourhood for temperature forecasts.

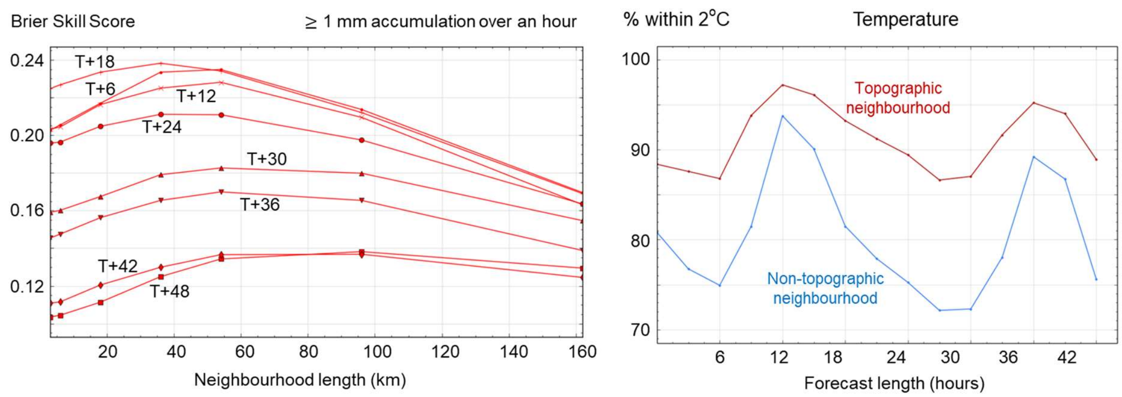

Figure 6. Examples of verification against spot locations used in IMPROVER for a month trial for November 2017. Left panel shows the Brier Skill Score (up is better) for different neighbourhood sizes and lead times for 1-hour precipitation accumulations from MOGREPS-UK. Right panel shows the end-of-chain percentage of forecasts that have a median temperature within two degrees of the observed value when using either the standard or topographic neighbourhood processing (for a non-optimised neighbourhood size).

The trialling capability will provide evidence of the benefits of upgrades to inform good governance in the same way as has been done for operational forecast model upgrades for many years. As IMPROVER becomes operational it will be brought under the same formal governance procedure as the NWP models, with approval of changes required for both the scientific performance and the suitability for operational use in serving customers. IMPROVER upgrades will be brought in through the Parallel Suite (PS) process and the aim is to include the IMPROVER suite in the final package trials for each PS, so that implementation decisions can be taken based on performance of both the NWP and Post-Processing components and of the complete end-to-end system.

6. Technical infrastructure

IMPROVER software consists of an open-source Python codebase that resides on GitHub (https://github.com/metoppv/improver) which implements meteorologically oriented algorithms for post-processing, plus a logistics or workflow layer ('Rose/cylc suites') to hook into NWP data and automate the production process. In line with the science and technology strategy, the code is pluggable and modular, with self-contained command line utilities and tasks that each produce debuggable and, crucially, verifiable output in NetCDF. IMPROVER leverages existing Scientific Python libraries such as Iris, dask, and numpy. There is a high focus on following modern professional software development practices including unit and acceptance test development, automated test running on proposed changes (Travis CI), two-stage review process, and automated style and defect reviewing (Codacy). The code is inherently amenable to parallel processing and use on highly parallel computers.

7. Timescales and Further Work

In March 2019 a first real-time running version will provide outputs that can be seen by downstream users. In 2020 an operational version will be released to allow a hand over from the use of current post processing systems to IMPROVER.

Beyond March 2019 the focus will be on the introduction of ECMWF ensemble forecasts into the processing chain. There will be further development of pre-operational scientific capability, including verification, and testing using trials and case studies, especially around precipitation and nowcasting and downscaling of MOGREPS-G & ECMWF forecasts to the UK grid. Statistical calibration will be incorporated using the existing EMOS code. One important area to develop further is the capability to produce regional probabilities of high-impact weather over time windows so that operational meteorologists can quickly identify the main areas of concern rather than having to sift through every forecast. All the while, work will progress on optimisation.

IMPROVER will offer a full probability distribution allowing the presentation of not just the most likely outcome but also the uncertainty and risks of other outcomes and providing the opportunity for forecasts to be adapted to the needs of individual users. It is important this capability is exploited and the hand over to users is properly managed. We also need to ensure that the outputs meet Public Weather Service performance targets and existing requirements.

In the longer term, the IMPROVER framework provides an unprecedented opportunity to develop new or enhanced post processing methods and create new innovative outputs.

8. Acknowledgements

This is a team effort from a committed group of people. Thanks to Paul Abernethy, Ben Ayliffe, Mark Baker, Laurence Beard, Anna Booton, Dan Brierley, Clare Bysouth, Rob Coulson, Sean Coultas, Ric Crocker, Gavin Evans, Roger Harbord, Aaron Hopkinson, Kathryn Howard, Teresa Hughes, Caroline Jones, Stephen Moseley, Tim Pillinger, Fiona Rust, Caroline Sandford, Tomasz Trzeciak and Mark Worsfold.

9. References

Jonathan Flowerdew (2014) Calibrating ensemble reliability whilst preserving spatial structure, Tellus A: Dynamic Meteorology and Oceanography, 66:1, DOI: 10.3402/tellusa.v66.22662

Gneiting, T., A.E. Raftery, A.H. Westveld, and T. Goldman, 2005: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Wea. Rev.,* **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1

Nigel Roberts and Marion Mittermaier (2015). MOSAC paper 20.19 on Post Processing.

Schefzik, R., Thorarinsdottir, T.L. & Gneiting, T., 2013. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. Statistical Science, 28(4), pp.616– 640. Available at: http://projecteuclid.org/euclid.ss/1386078881%5Cnhttp://arxiv.org/abs/1302.7149%5Cnhttp://projecteuclid.org/euclid.ss/1386078881.

Schwartz, C.S. and R.A. Sobash, 2017: Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Mon. Wea. Rev.,* **145**, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1

Sheridan, P. , Smith, S. , Brown, A. and Vosper, S. (2010), A simple height-based correction for temperature downscaling in complex terrain. Met. Apps, 17: 329-339. doi:10.1002/met.177

Simonin, D. , Pierce, C. , Roberts, N. , Ballard, S. P. and Li, Z. (2017), Performance of Met Office hourly cycling NWP-based nowcasting for precipitation forecasts. Q.J.R. Meteorol. Soc, 143: 2862-2873. doi:10.1002/qj.3136