A Vision for providing Global Weather Forecasts at Point-scale

Tim Hewson, ECMWF

Original Abstract

A Vision for providing Global Weather Forecasts at Point-scale

This presentation will provide a vision, based around current initiatives, of how post-processing and machine learning could work in tandem to downscale the ensemble output of current-generation global models, to deliver probabilistic analyses and forecasts, of multiple surface weather parameters, at point-scale, worldwide. Skill gains would be achieved by adjusting for gridscale and sub-grid biases. One particularly attractive feature of the vision is that observational data is not required for a site that we forecast for, although the more 'big data' that we use, worldwide, the better the forecasts will be overall.

The vision is based on four building blocks - or steps - for each parameter. The first step is a simple proof-of-concept, the second is supervised training, the third is hindcast activation and verification, and the fourth is real-time operational implementation. Here we will provide 3 examples, for 3 fundamental surface weather parameters - rainfall, 2m temperature and 100m wind - although the concepts apply also to other parameters too. We stress that different approaches are needed for different parameters, primarily because what determines model bias depends on the parameter. For some, biases depend primarily on local weather type, for others they depend mainly on local topography.

For rainfall downscaling, work at ECMWF has already passed stage 4, with real-time worldwide probabilistic point rainfall forecasts up to day 10 introduced operationally in April 2019, using a decision-tree-based software suite called "ecPoint", that uses non-local gridbox weather-type analogues. Further work to improve algorithms is underway within the EU-funded MISTRAL project. For 2m temperature we have reached stage 2, and ecPoint-based downscaling will be used to progress this within the EU-funded HIGHLANDER project. The task of 100m wind downscaling requires a different approach, because local topographic forcing is very strong, and this is being addressed under the umbrella of the German Waves-to-Weather programme, using U-net-type convolutional neural networks for which short-period high-resolution simulations provide the training data. This work has also reached stage 2.

For each parameter discussed we see the potential for substantial gains, for point locations, in forecast accuracy and reliability, relative to the raw output of an operational global model. As such we envisage a bright future where probabilistic forecasts for individual sites (and re-analyses) are much better than hitherto, and where the degree of improvement also greatly exceeds what we can reasonably expect in the next two decades or so from advances in global NWP.

This presentation will give a brief overview of downscaling for the 3 parameters, highlight why we believe heavily supervised approaches offer the greatest potential, illustrate also how they provide invaluable feedback for model developers, illustrate areas where more work is needed (such as cross-parameter consistency), and show what form output could take (e.g. point-relevant EPSgrams, as an adaptation of ECMWF's most popular product).

Contributors to the above initiatives include: Fatima Pillosu (ECMWF, ecPoint); Estibaliz Gascon and Andrea Montani (ECMWF, MISTRAL); Michael Kern and Kevin Höhlein (Technische Universität München, Waves-to-Weather).

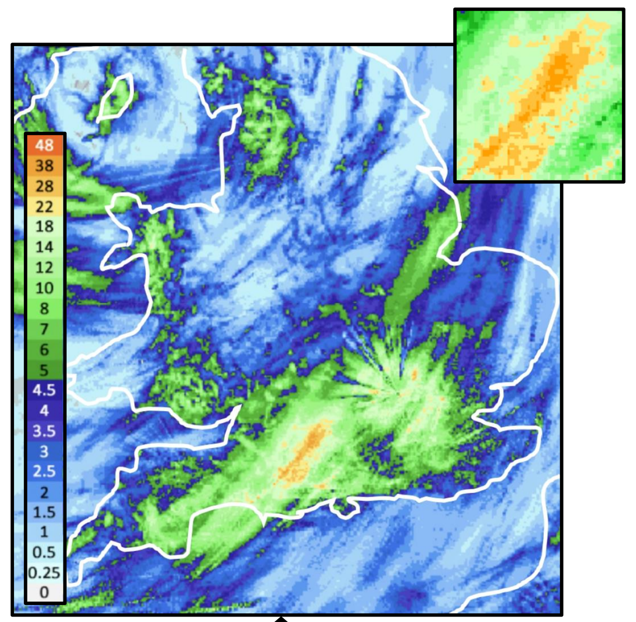
Material here is a rather cut-down version of what was intended for inclusion at the point of abstract submission.

Note: The "Vision" is not formally part of ECMWF's long term strategy, although ECMWF has used and will continue to use external project funding and collaboration to explore some of the ideas described.

Example 1 – Post-processing (PP) on the gridbox scale to provide probabilities for *points within* each gridbox

Uses: for e.g. Rainfall or 2m Temperature

Status: Global point forecasts for rainfall operational now for 1 year (experimental layers in ecCharts). 2m temp PP about to begin (HIGHLANDER project).



12 hour radar-derived rainfall totals 06-18UTC 30th April 2020 (c/o netweather.tv)

Main orange band in southern England (values up to 38mm) led to minor impacts, and some flood watches being raised

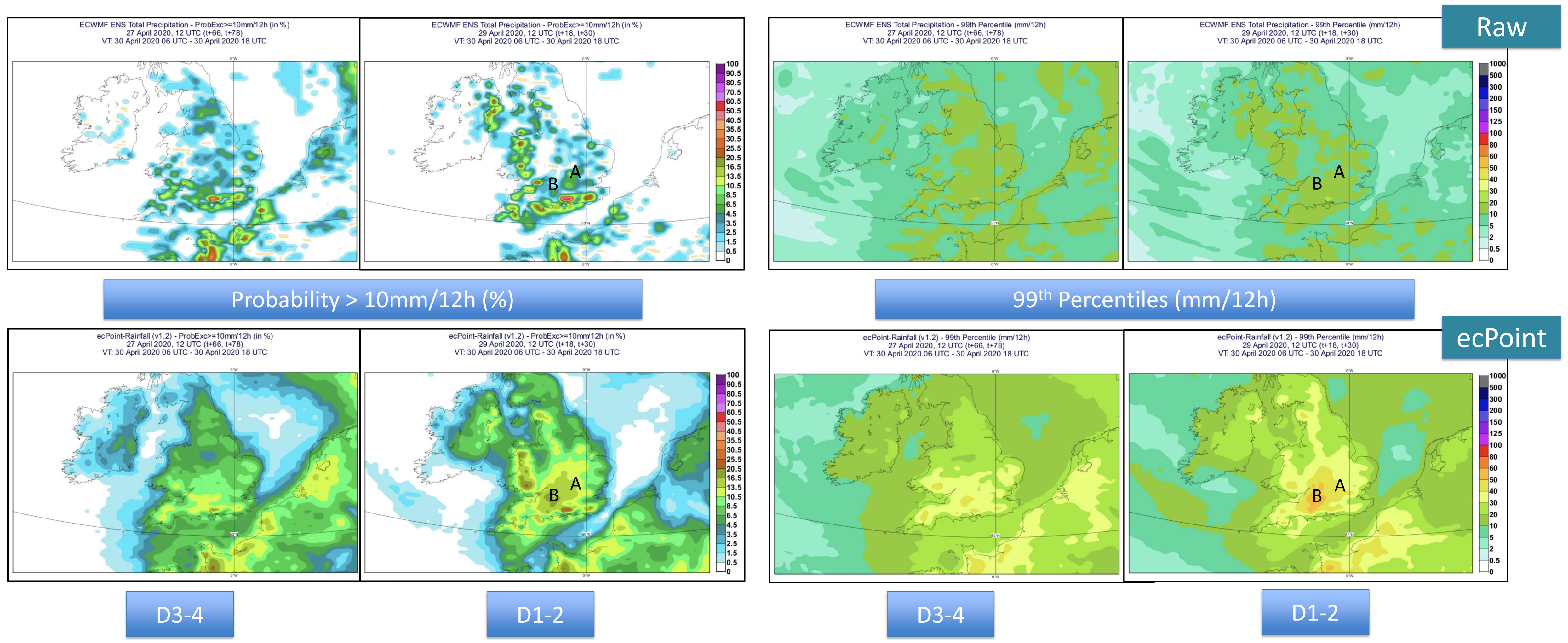
Could this event, in this area, have been foreseen using:

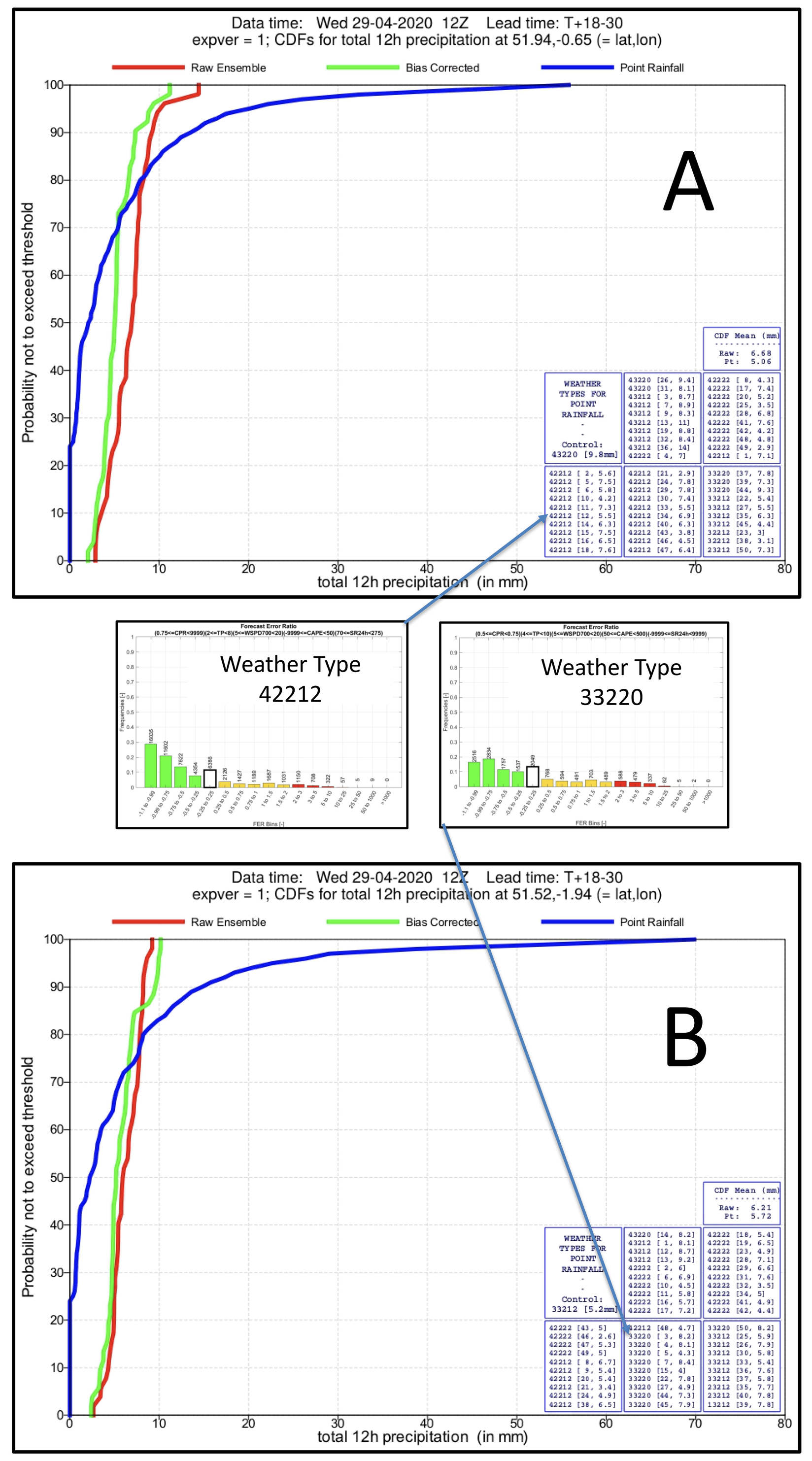
(i) The raw ECMWF ensemble (ENS) TOP ROW

(ii) Point rainfall (post-processed ENS) BOTTOM ROW

Download ecPoint paper preprint here:

https://arxiv.org/abs/2003.14397





In the Raw ensemble products (top row) the probabilities for >10mm/h are noisy and inconsistent over England and Wales, which does not inspire confidence, and nor are the highest probabilities focussed in the main affected area. Meanwhile the 99th percentile – in effect the highest value within the ensemble, denoted at each gridpoint - is bland and unhelpful. Values are too low to represent local observations (as represented by the radar) – widely 10-20mm (dark green) - and the pattern also suggests that almost anywhere in England and Wales is at risk. Comments apply to both forecasts (i.e. lead times) shown, and other forecasts produced inbetween (not shown).

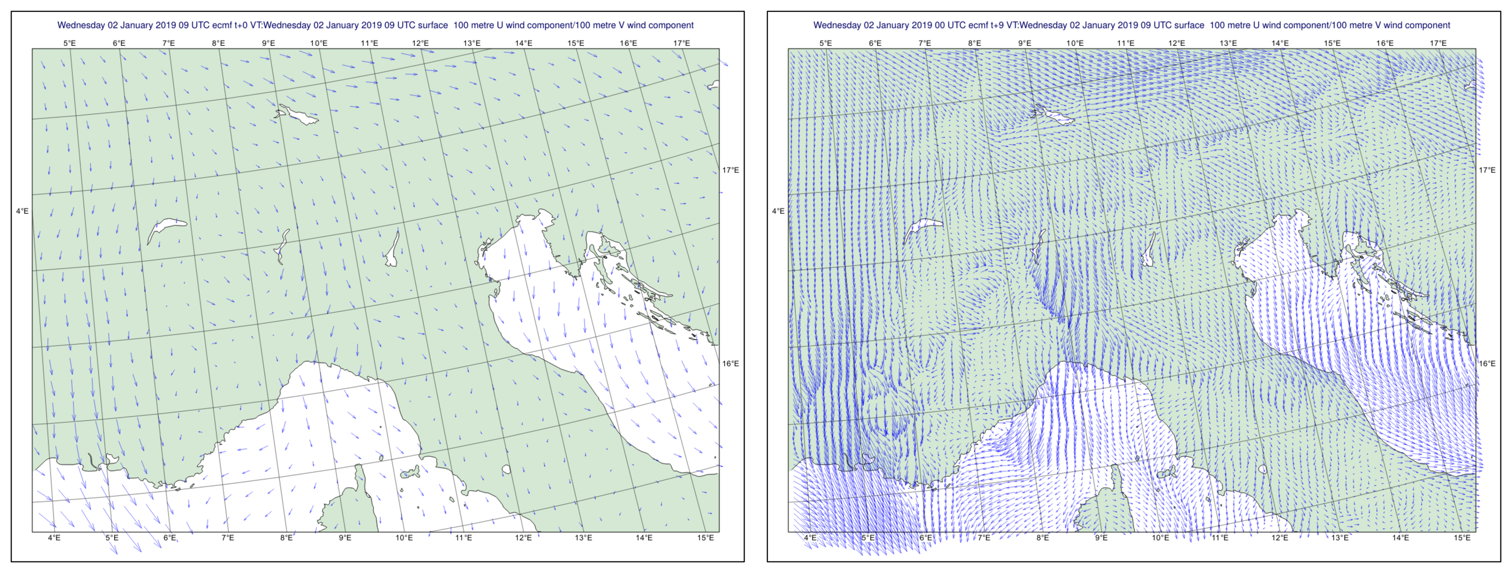
In ecPoint (point rainfall) (bottom row) probabilities are much less noisy and more consistent, and mostly they correctly highlight the high risk areas. Even more helpful is the 99th percentile field, which is spot on with its depiction of the main risk area at day 1-2 leads (see darker orange). Even at day 3-4 the post-processing was able to allocate a higher risk to this area. Indeed this region was shown to be at greatest risk in all the other forecasts inbetween (not shown).

So how did ecPoint post-processing create a signal for higher totals, localised in the right area? The reason is that ecPoint predicts different degrees of sub-grid variability, and anticipates grid-scale biases, with both such factors being a function of the "gridbox weather type" (trained using 1 year of global data). CDFs to the right illustrate how, for sites A and B (see maps), the post-processing first adjusts for gridscale bias (red to green), showing a notable reduction in totals at A, whilst at B there is a much smaller reduction, and even an increase for the wetter members. This is because at A the weather types are mostly strongly convective (>75% of rain is diagnosed as convective), whilst at B a sizeable number have a convective-dynamic mix. In training data the former types commonly associate with over-prediction, whereas the latter can be associated with under-prediction. Physically the reason may be that in the latter case the model is trying to develop organised rainfall out of convection (e.g. as in an MCS) but is struggling so to do. Two related weather type mapping functions are shown – note: ecPoint forecast total = (!+FER) * raw total, for each member. On CDFs green to blue denotes the addition of sub-grid variability: also very relevant for the map plots. Finally note that the raw ensemble mean at point A was *greater* than it was at point B (marked in box on CDFs).

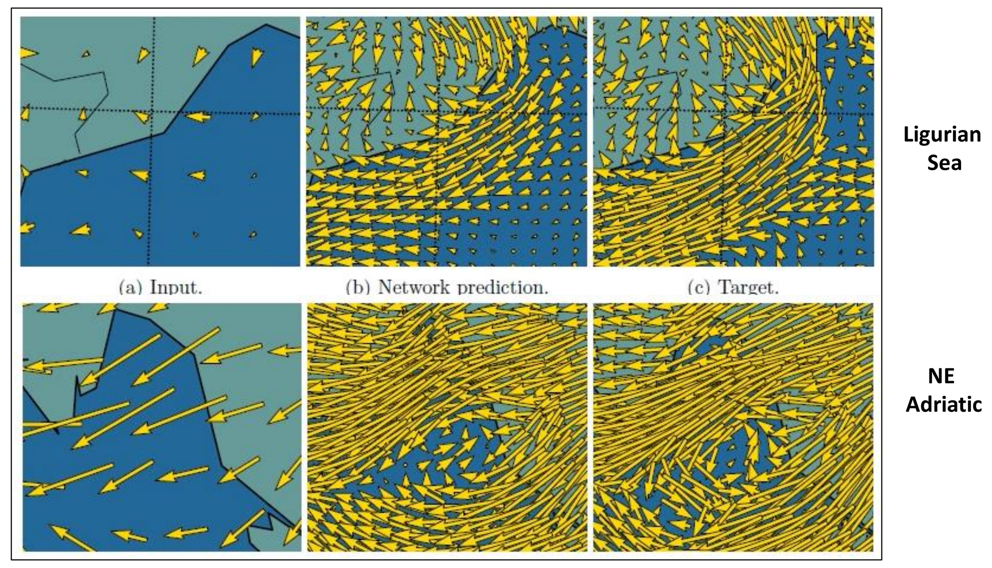
Example 2 – Post-processing down to 1-2km (?) grid scale, to provide forecasts *for specific points*

Uses: e.g. Low level winds

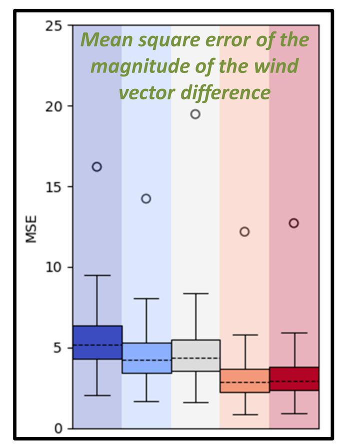
Status: Proof of concept work, using neural networks, to predict 100m winds at 9km resolution (ECMWF HRES), using 31km resolution data as input (ERA5)

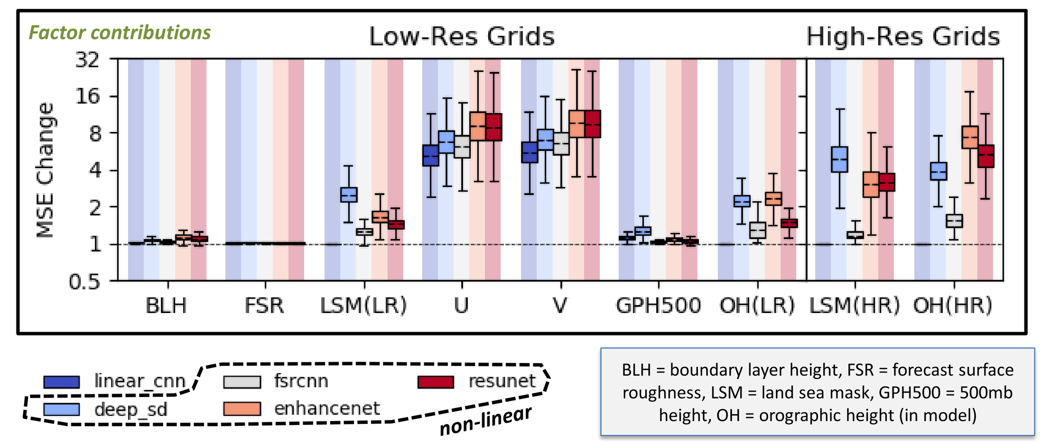


There exist relationships between wind fields in low resolution models and wind fields in high resolution models (presumed more accurate). To make real point forecasts, by downscaling, the relationships need to be established via training, and need to be robust. Topography and coasts will play a role, and other variables too.



With one year of training (hourly data), the high resolution wind field can be reproduced well in topographically complex regions, using U-net convolutional neural networks (non-linear)





BLH = boundary layer height, FSR = forecast surface roughness, LSM = land sea mask, GPH500 = 500mb height, OH = orographic height (in model)

In tests one linear algorithm delivers larger errors than four non-linear. Enhancenet and U-net (resunet) are best. Moreover, non-linear methods can make more use of other relevant training variables, such as boundary layer height, although unsurprisingly the key (variable) parameters, in the predictor dataset, are U and V. The key static parameters here are on the target high-res grid: orographic height and land-sea mask.
In some future operational system one could use a limited period 1-2km resolution offline global simulation for training and target, with ECMWF ensemble forecasts (now 18km resolution) providing input to downscale from. A key question would be the cost of real time running. Then various options would exist to operationalise.