Highlander

High performance computing
to support smart land services

D4.3 Downscaled datasets of Subseasonal forecast

Deliverable Lead	CINECA/ECMWF
Deliverable due date	2021/03/31 updated 2021/07/16
Status	DRAFT
Version	V1

Co-financed by the Connecting Europe
Facility of the European Union

Highlander
High performance computing
to support smart land services

DOCUMENT CONTROL PAGE

Title	D4.3
Creator	ECMWF
Publisher	Highlander Consortium
Contributors	Estibaliz Gascon (ECMWF), Tim Hewson (ECMWF), Augustin Vintzileos (ECMWF)
Type	
Language	EN
Rights	Copyright “Highlander Consortium”
Audience	Public
	X Restricted
Requested deadline	M12 (delayed to M18)

Highlander
High performance computing to support smart land services

Index

1 Introduction or Executive Summary

This document accompanies the first release of the post-processing output of statistical downscaled sub-seasonal forecast using our first calibration algorithm on the ecPoint technique (Sub-SEA), developed by ECMWF.

2 Sub-seasonal ecPoint products

The sub-seasonal ecPoint (Sub-SEA) are post-processed products delivering probabilistic forecasts at point-scale (instead of at grid-box scale) of global 24-hour precipitation and minimum, maximum and mean temperatures in 24h for weeks three and four of the forecast (days 16-30). They are based on the ECMWF ensemble (ENS) forecast system with 51 ensemble members (1 control and 50 perturbed members, see https://www.ecmwf.int/en/forecasts/documentation-and-support), run at ~0.4° horizontal resolution (~36km), with the final outputs for ecPoint precipitation and temperature provided in percentiles for each grid box.

These products are computed on the CINECA HPC facility of Galileo to create, in real time, probabilistic point rainfall and point temperature forecasts. In HIGHLANDER, initial focus has been be on lead times of 16-30 days, but shorter ranges may also be addressed in the future, notably days 11-15. Statistical downscaling benefits tend to be greater when working at lower resolution, so downscaling from 36km x 36 km in HIGHLANDER has the capacity to deliver even more improvement to raw model output than for other pre-existing ecPoint products from the medium-ranges (up to day 15) where ENS resolution is 18 x 18 km. This "statistical downscaling" is referred to convert from gridbox scale to point scale inside of that grid box (what can happen in any point inside of each gridbox?) and not to the usual resampling of the gridbox to a smaller gridbox size.

A key characteristic of the ecPoint output will be having much more reliable climatological representations for points in space than the raw model output - e.g. numbers of dry days, numbers of very wet days and amounts of rainfall on the very wet days (this definition can vary depending on the location and its climatology), all three of which are substantially under-forecast in raw model output at 36 km x 36 km resolution (at the expense of over-forecasting intermediate events). This facilitates greater uptake e.g. for agricultural planning, with associated benefits in both economic and resource availability terms. However, ecPoint is also developed for the first time in this project to improve forecasts of 2m air temperatures (maxima, minima or daily means). Although sub-grid variability (which ecPoint corrects for) tends to be much lower than it is for rainfall, temperature biases can still be relatively large, and ecPoint also corrects for these, including the extreme values. So again, there can be important benefits for users. In hot summer weather for example maxima can be under-predicted in raw model output by 3°C or more, which is relevant for agricultural applications such a crop management, as evapotranspiration is influenced by temperature as well as rainfall. Similarly, animal and human welfare, which relate firstly to actual temperatures, but also to metrics of "perceived temperature" which incorporate humidity as well, can also be better managed.

These products will be utilized by several specific Downstream Applications and (pre-)Operational Services (DAPOS) inside of the HIGHLANDER project.

2.1 EcPoint concept

The point rainfall and point temperature products aim to deliver probabilities for point measurements of rainfall or temperature respectively (within a forecast model gridbox). "ecPoint" (Hewson and Pillosu, 2020) is the name given to the post-processing philosophy and software package.

To understand the concept, we show an example for the ecPoint-Rainfall product. Let us first consider the radar-derived rainfall totals shown on the example in Figure 1. We assume that these are accurate, meaning that they indicate what would have been measured locally by rain gauges. Then consider the ECMWF ENS gridbox highlighted. Within this box, whilst the gridbox average rainfall total is about 17mm, the minimum and maximum rainfall amounts are about 2mm and 60mm respectively. This implies a lot of sub-grid variability. A completely accurate ensemble member forecast would predict 17mm. But clearly this of itself would give the user no idea that locally there was much more (and indeed much less) than this amount. And to cause flash floods, as were observed, probably a 17mm total, locally, would not have been enough. The point rainfall approach aims to estimate the range of totals likely within the gridbox, and indeed it delivers probabilities for different point values within that gridbox (albeit without saying where the largest and smallest amounts are likely to be).

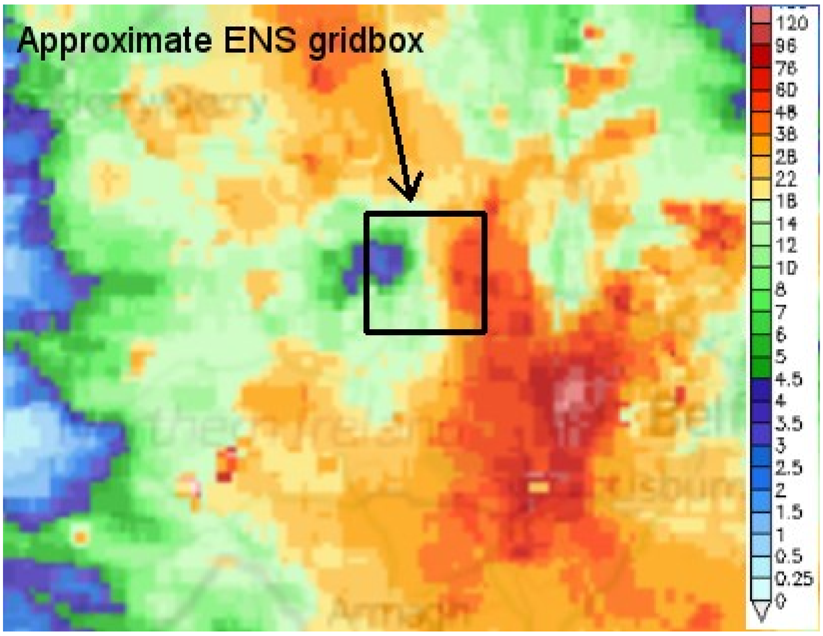

Figure 1. *Radar-derived rainfall totals over part of Northern Ireland for the 12h period ending 00UTC 29 July 2018. Scale is in mm. Flash floods occurred in some locations. Figure based on data from netweather.tv.*

4 main aspects of ecPoint can be highlighted:

- The new products are post-processed products from the ECMWF ENS system, which calibrates output fields from the ensemble, including the control run
- It delivers probabilistic forecasts of rainfall totals or temperature *for* **points** - i.e. that would be measured by a standard rain gauge or standard screen thermometer randomly located within a model gridbox (hence the word 'point')
- By way of comparison, the raw ENS delivers probabilities for *gridbox-average* rainfall or temperature (in the current ENS for sub-seasonal time ranges that means ~36 x 36 km boxes, from day 16 up to day 30)
- The term **rainfall** in the ecPoint-Rainfall name, means all precipitation types: rain, rain plus snow, snow etc., always converted into mm of rain equivalent.
- EcPoint is specifically designed to improve the reliability and discrimination ability of the forecast. In verification it has shown some particularly good results for large (and indeed zero) totals of precipitation (the ecPoint-Temperature verification will be developed later in the project).

There are two reasons why point rainfall or point temperature and raw ensemble distributions of these parameters commonly differ:

- Sub-grid variability - Figure 1 shows an example of large sub-grid variability - which itself varies according to the weather-type in a gridbox.
- Model biases, on the gridscale, also intrinsic to the weather type in a gridbox.

"Weather type" here refers to specific conditions inside a gridbox - e.g. (for rainfall) "mainly convective precipitation with strong mid-level winds, …". In practice, ecPoint post-processing creates, on the basis of the calibration results (see below), an "ensemble of ensembles", that is 100 new point rainfall realisations for each ensemble member, and so comprises, at one intermediate stage in the computations, 5100 equi-probable values of point rainfall totals within each ensemble gridbox. However, prior to saving, the values for each gridbox are sorted, and distilled down into 99 percentile fields (1, 2, ..., 99). As a set these percentile fields constitute the ecPoint-Rainfall and ecPoint-Temperature products.

Figure 2 shows examples of cumulative distribution functions (CDFs) for different sites on two different occasions, for the full ECMWF ensemble, illustrating the two types of adjustment that ecPoint makes, bias correction and the addition of sub-grid variability (here using a rainfall example). These adjustments are most in evidence on the first plot, whilst the second plot shows also that the point rainfall and raw model gridscale distributions may sometimes be quite similar.

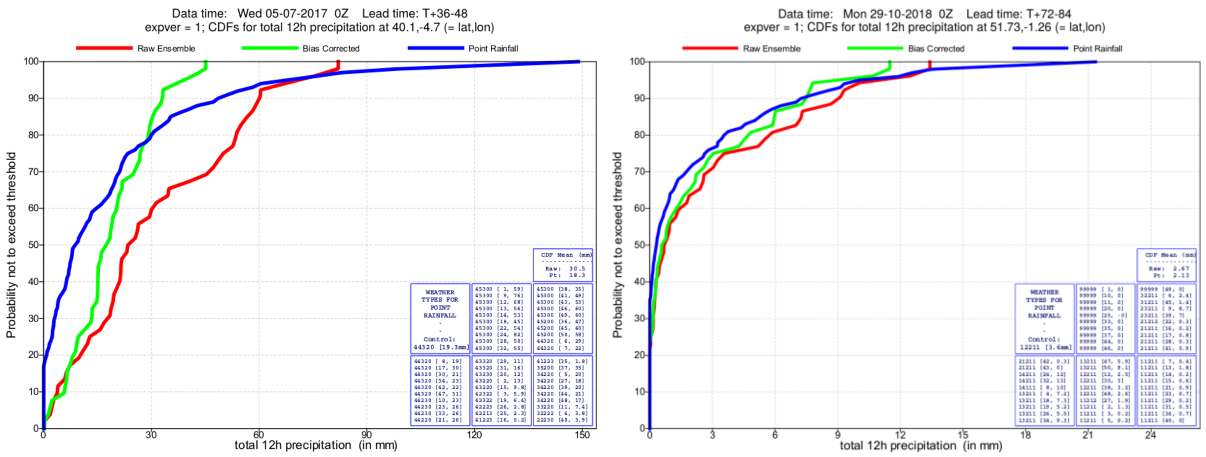

Figure 2. *Two CDF examples comparing raw gridscale (red), post-processed bias-corrected gridscale (green), and post-*

processed point rainfall (blue). The first plot is for a site in Spain in summer at day 2. The second plot is for a southern England site in autumn, at day 4. Note that for each case the areas to the left of the green and blue curves, that represent the mean gridbox rainfall over the ENS, should be the same. Note also that these plots show forecasts for 12h rainfall, whilst in HIGHLANDER we create forecasts for rainfall and temperature in 24h periods for days 15-30 of the forecast.

2.2 Calibration

As with any post-processing system ecPoint has to be calibrated. For this it uses short-range control run forecasts of 24h total precipitation and 24h minimum, maximum and mean temperatures covering one year (the "training period"), which are individually compared with rainfall or temperature observations, for the same times, within the respective grid boxes, all over the world. Here, we assume that the forecast errors due to incomplete physical process representation pertaining to sites inside a grid box (sub-grid variability) are similar for all the lead times, so we can develop these post-processing products using short-time forecast but with the sub-seasonal forecast horizontal resolution (36 x 36 km). The entire procedure is not described here but involves segregation according to grid box-weather-types, which each have different sub-grid variability structures and/or different bias corrections associated. The 24h point rainfall and temperature systems being introduced by ECMWF for HIGHLANDER incorporate a range of different predictors for each product, including dynamic weather-related model parameters (e.g. 100m wind speed, cloud cover), static model-related parameters (e.g. the land-sea mask, sub-grid orography), dynamic astronomical variables (e.g. related to solar radiation), and potentially also external variables (e.g. population density). The predictors for each variable are not included in this document yet, since they are "preliminary"; however, they will be re-considered during the project, and more research will be done to optimize them. It does not mean that the first selected predictors are wrong, and they definitely will improve the forecast; however, the ecPoint concept is reviewed and updated regularly to incorporate new and better predictors with the time, and we had very limited time to implement the first version, which means that there is scope for improvement during the project.

The **version 1.0.0** of most of the ecPoint sub-seasonal products are ready and described here:

- **ecPoint-TemperatureMin**: The 24h point minimum 2m temperature system for the extended-range forecast incorporated 118 weather types. The type definitions are currently based on the 10 following parameters: land-sea mask, standard deviation of the filtered subgrid orography, minimum temperature at 2m in 24h, difference between minimum temperature at 2m in 24h and skin temperature, wind speed at 100m, 24h clear-sky direct solar radiation, volumetric soil water at layer 1, snow depth, boundary layer height, total cloud cover.

- **ecPoint-TemperatureMax**: The 24h point maximum 2m temperature system for the extended-range forecast incorporated 114 weather types. The type definitions are currently based on the 10 following parameters: land-sea mask, standard deviation of the filtered subgrid orography, minimum temperature at 2m in 24h, difference between minimum temperature at 2m in 24h and skin temperature, wind speed at 100m, 24h clear-sky direct solar radiation, volumetric soil water at layer 1, snow depth, boundary layer height, total cloud cover.
- **ecPoint-TemperatureMean**: The 24h point mean 2m temperature system for the extended-range forecast incorporated 54 weather types. The type definitions are currently based on the 9 following parameters: land-sea mask, standard deviation of the filtered subgrid orography, mean temperature at 2m in 24h, difference between mean temperature at 2m in 24h and skin temperature, wind speed at 100m, 24h clear-sky direct solar radiation, volumetric soil water at layer 1, boundary layer height and total cloud cover.
- **ecPoint-Rainfall**: The 24h point rainfall system for the extended-range forecast incorporated 181 weather types. The type definitions are currently based on the 5 following parameters: convective precipitation ratio, total precipitation in 24h, wind speed at 700 hPa, maximum cape in 24h and 24h clear-sky direct solar radiation

For more information on the post -processing concept see: *https://confluence.ecmwf.int/display/FUG/Point+Rainfall*

2.3 Output formats

We will have available 24h periods from day 16 up to day 30, namely T+360h-T+384h, T+384h-T+408h, ..., T+696h-T+720h. Each file contains the 99 percentiles for temperature or precipitation for a single 24h period in grib format. Both ecPoint products will be available twice per week (Tuesday and Friday). ecPoint temperature products (minimum, maximum and mean) will be operational products from the beginning of April 2021, while ecPoint 24h rainfall will be ready at the end of spring 2021. We anticipate providing an upgraded version of the post-processing calibration later in the project, which should improve product accuracy.

The outputs have also been adapted to the different DApOS input requirements, so the areas are cropped, and the file converted to NetCDF accordingly for each case.

3. References

T. Hewson, F. Pillosu. (2020). A new low-cost technique improves weather forecasts across the world. arXiv preprint arXiv:2003.14397