Assessing the benefit of growing a decision tree to post-process rainfall

Fatima M.Pillosu1,2, Christel Prudhomme2,3,4, Elisabeth Stephens5, Hannah L. Cloke1,5,6,7

1 Department of Geography and Environmental Science, University of Reading, Reading, United Kingdom

2 Forecast Department, European Centre for Medium-range Weather Forecasts, Reading, UK

3 UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

4 Department of Geography and Environment, University of Loughborough, Loughborough, United Kingdom

5 Department of Meteorology, University of Reading, Reading, United Kingdom

6 Department of Earth Sciences, Air, Water and Landscape Science, Uppsala University, Sweden

7 Centre of Natural Hazards and Disaster Science, CNDS, Sweden

*Correspondence to*: Fatima M. Pillosu (fatima.pillosu@ecmwf.int)

**Abstract.**

**Keywords.** Keyword\_1, Keyword\_2, Keyword\_3, Keyword\_4, Keyword\_5

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

P!:

* One of the challenges in meteorology is predicting extreme rainfall events reliably.
* Due to the representitativeness problem of model rainfall forecasts (Zied’s ROC paper), the challenge of forecasting localized rainfall extreme events is even higher.
* Post-processing techniques can help at improving rainfall forecasts. From simple linear regressions to decision-tree-based machine learning techniques to neural networks, to deep learning, different techinques are used in meteorology to post-processed raw rainfall forecasts.
* A specific method will be used depending on the quantity of training data available, the complexity of the problem in hand, and computational power.

P2

* ecPoint-Rainfall is a post-processing technique that anticipates sub-grid variability and reduces biases in the rainfall forecasts. ecPoint-Rainfall is a decision-tree-based techinque that post-processes rainfall forecasts through mapping functions (MF) that vary according to a weather scenario at grid-box scale, calle weather type (WT). Each MF is significantly different from each other.
* The building of the decision tree is done in a semi-automatic fashion, using the software ecPoint-Calibrate. The significance of the difference between MFs is assessed using the 2-sided Kolmogorov-Smirnov test.
* The current operational version at ECMWF has 214 MFs.
* The skill of the post-processed forecasts was assessed by Hewson and Pillosu in a one-year period, global verification. Such verification assesses the two main features of weather forecasts: reliability and discrimination ability. The verification analysis showed that both verification features were improved by ecPoint, especially for extreme localized events (tp>=50 mm/12h).

P3

* The question unanswered in Hewson and Pillosu article is: how effective is building a decision tree with n significantly different MFs compared to a less expensive single-WT ecPoint-Rainfall? The differenciation of the post-processing applied to the rainfall forecasts is unique to ecPOint, while more commonly, other post-processing technqiues will apply the same post-processing to all grid-boxes. The single-WT ecPoint simulates the latter condition.
* How much forecasts are improved by adopting a multiple-WT ecPoint-Rainfall compared to a single-WT version?
* To estimate this difference, the reliability component of the Brier score will be used; ROC curves will be used to estimate the discrimination ability of the forecasts. Both ecPoint-Rainfall forecasts (single- and multiple-WT) will be compared to the raw forecasts from ECMWF which will be used as a benchmark.
* In particular, the discrimination ability of the operational configurations of the three systems will be evaluated, i.e. 51 ensemble members for raw ECMWF ENS, and 99 ensemble members for single- and multiple-WT ecPoint-Rainfall, will be evaluated using the trapezoidal ROC curve. The potential discrimination ability of the three forecasts will also be estimated using the binormal ROC curve.

# Data

Table 1 – This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1. This is the caption for table 1.

|  |  |  |
| --- | --- | --- |
| **Header 1** | **Header 2** | **Header 3** |
| Text example 1 | Text example 1 | Text example 1 |
| Text example 2 | Text example 2 | Text example 2 |
| Text example 3 | Text example 3 | Text example 3 |
| Text example 4 | Text example 4 | Text example 4 |
| Text example 5 | Text example 5 | Text example 5 |

# Methods

# Results

# Discussions

# Conclusions

# References