EMS Annual Meeting Abstracts
Vol. 20, EMS2023-199, 2023, updated on 21 May 2024
https://doi.org/10.5194/ems2023-199
EMS Annual Meeting 2023

Creating skillful and reliable probabilistic forecasts using machine learning

Mariana Clare[1] and Thomas Haiden[2]
[1]ECWMF, Forecast Department, Bonn, Germany (mariana.clare@ecmwf.int)
[2]ECWMF, Forecast Department, Reading, UK

Machine learning techniques are increasingly used in weather forecasting, to either improve results from numerical weather prediction (NWP) models or emulate some aspects of the forward integration part of NWP models using a purely data-driven approach. So far most of the techniques presented have focused on producing deterministic forecasts, as many standard machine learning techniques lack the ability to express uncertainty.

Whilst methods that attempt to fully emulate NWP models are both very complex and computationally expensive, in this talk I will show how much simpler and computationally cheaper techniques can be used to create a probabilistic forecast from a single global high-resolution forecast by post-processing. Specifically, I will focus on the relatively novel technique of a Bayesian Neural Network and show how it can predict the distribution of the forecast error relative to its own analysis. By adding these error distributions to the original forecast, we can create a probabilistic forecast. Time permitting, I will discuss how more advanced techniques can be applied to take these distributions and generate spatially consistent ensemble members.

This methodology is particularly useful for NWP models at very high resolutions where running an ensemble is too computationally expensive and for machine learning approaches where no uncertainty information is available. In this talk, I will show how this methodology can be successfully applied to both ECMWF's high-resolution forecast and a purely data-driven weather forecast model being run at ECMWF, for both the surface variable of 2m temperature and the atmospheric variable of Z500. These probabilistic forecasts have been verified using standard metrics and, in the case of the high-resolution numerical forecast, have as good as or better CRPS scores than the ECMWF ensemble forecast for the lead times tested. Moreover, these probabilistic forecasts are reliable with spread-skill ratios close to one. Hence, this novel machine learning post-processing technique has the potential to produce probabilistic forecasts that are valuable and useful to forecasters.