EGU24-8918, updated on 25 Jul 2024
https://doi.org/10.5194/egusphere-egu24-8918
EGU General Assembly 2024

Can Machine Learning Models be a Suitable Tool for Predicting Central European Cold Winter Weather on Subseasonal Timescales?

Selina M. Kiefer[1], Sebastian Lerch[2], Patrick Ludwig[1], and Joaquim G. Pinto[1]
[1]Karlsruher Institut für Technologie, Institut für Meteorologie und Klimaforschung Troposphärenforschung, Germany
[2]Karlsruher Institut für Technologie, Institut für Statistik, Germany

For many practical applications, e.g. agricultural planning, skillful weather predictions on the subseasonal timescale (2-4 weeks in advance) are key for making sensible decisions. Since traditional numerical weather prediction (NWP) models are often not capable of delivering such forecasts, we use an alternative forecasting approach combining both, physical knowledge and statistical models. Selected meteorological variables from ERA-5 reanalysis data are used as predictors for wintertime Central European mean 2-meter temperature and the occurrence of cold wave days at lead times of 14, 21 and 28 days. The forecasts are created by Quantile Regression Forests in case of continuous temperature values and Random Forest Classifiers in case of binary occurrence of cold wave days. Both model types are evaluated for the winters 2000/2001 to 2019/2020 using the Continuous Ranked Probability Skill Score for the continuous forecasts and the Brier Skill Score for the binary forecasts. As a benchmark model, a climatological ensemble obtained from E-OBS observational data is considered. We find that the used machine learning models are able to produce skillful weather forecasts on all tested lead times. As expected, the skill depends on the exact winter to be forecasted and generally decreases for longer lead times but is still achieved for individual winters and in the 20-winter mean at 28 days lead time. Since machine learning models are often subject to a lack of interpretability and thus considered to be less trustworthy, we apply Shapley Additive Explanations to gain insight into the most relevant predictors of the models' predictions. The results suggest that both Random-Forest based models are capable of learning physically known relationships in the data. This is, besides the capability of producing skillful forecasts on the subseasonal timescale, a selling point of the combination of physical knowledge and statistical models. Finally, we compare the skill of our statistical models to subseasonal state-of-the-art NWP forecasts.