README file

Title: Forecast dataset associated with "From Random Forests to Flood Forecasts: A Research to Operations Success Story"

Abstract: Because excessive rainfall is poorly defined and difficult to forecast, there is a need for tools for Weather Prediction Center (WPC) forecasters to use when generating Excessive Rainfall Outlooks (EROs), which are issued for the contiguous United States at lead times of 1--3 days. To address this need, a probabilistic forecast system for excessive rainfall, known as the Colorado State University-Machine Learning Probabilities (CSU-MLP) system, was developed based on ensemble reforecasts, precipitation observations, and machine learning algorithms, specifically random forests. The CSU-MLP forecasts were designed to emulate the EROs, with the goal being a tool that forecasters can use as a "first guess" in the ERO forecast process. Resulting from close collaboration between CSU and WPC and evaluation at the Flash Flood and Intense Rainfall experiment, iterative improvements were made to the forecast system and it was transitioned into operational use at WPC. Quantitative evaluation shows that the CSU-MLP forecasts are skillful and reliable, and they are now being used as a part of the WPC forecast process. This project represents an example of a successful research-to-operations transition, and highlights the potential for machine learning and other post-processing techniques to improve operational predictions.

Contact: Russ Schumacher, Department of Atmospheric Science, Colorado State University, russ.schumacher@colostate.edu

Recommended data citation: Schumacher, R.S., A. J. Hill, M. Klein, J. A. Nelson, M. J. Erickson, and G. R. Herman, 2021: Forecast dataset associated with "From random forests to flood forecasts: A research to operations success story". Colorado State University. Libraries.
http://dx.doi.org/10.25675/10217/222367

Format of data files: netCDF

Location where data were collected: Continental United States

Time period during which data were collected: 2018-03-15 to 2020-10-15 for day-2 and day-3 forecasts; 2019-03-15 to 2020-10-15 for day-1 forecasts.

File information:

List of files:
day1/csu_mlp_2019_gefso_day1_all.nc
day1/csu_mlp_2020_gefso_day1_all.nc
day2/csu_mlp_2017_gefso_day2_all.nc
day2/csu_mlp_2019_gefso_day2_all.nc
day2/csu_mlp_2020_gefso_day2_all.nc
day3/csu_mlp_2017_gefso_day3_all.nc
day3/csu_mlp_2019_gefso_day3_all.nc
day3/csu_mlp_2020_gefso_day3_all.nc

dataset_readme.pdf

All files (except the readme) are netCDF-formatted files with dimensions of time, latitude, and longitude, comprising one gridded forecast per day from a version of the CSU-MLP forecast system. Forecasts are on a 721x276 lat/lon grid over the continental US at 0.09-degree grid spacing. These files were generated by regridding CSU-MLP forecasts to the same grid used by the Weather Prediction Center for their Unified Flood Verification System.
Each file includes all forecasts for a particular lead time (day 1, 2, or 3) and version of the forecast system (2017, 2019, or 2020 version).

Variable information:
The primary variable in each file is the gridded forecast of the probability of excessive rainfall, and has a name similar to "csu_mlp_2020_gefso". The year (in this example 2020) reflects the version of the CSU-MLP system. Latitude, longitude, and time dimensions and coordinates are self-describing in the netCDF files.

Methods:
Regridding of the CSU-MLP forecasts was performed using wgrib2; combining individual forecast files into a single netCDF file and incorporating metadata was performed using the xarray package in python.

Date dataset was last modified:
Are there multiple versions of the dataset? No
Last modification: 2021-01-16