Preliminary methods to post-process accumulated week 3–4 precipitation forecasts for the Climate Prediction Center

Rochelle Worsnop[1,2], Michael Scheuerer[1,3], Thomas M. Hamill[2]
[1]CIRES, University of Colorado, [2]NOAA/ESRL/Physical Sciences Laboratory, [3]Norwegian Computing Center
Email: Rochelle.Worsnop@noaa.gov

Project Goals

Compare traditional statistical and neural network methods to post-process raw ensemble accumulated precipitation forecasts for the Climate Prediction Center

- First phase extends existing methodologies (Scheuerer et al. 2020; S20) described below for use with NOAA forecasts, across the full contiguous United States, and for all months.
- Second phase (not shown) will involve novel development to try to gain even more skillful forecasts 3-4 weeks ahead.

A. Background & Motivation

Current capabilities in operations

The current Week 3-4 Outlook at CPC is minimally post-processed. Our methods aim to improve guidance provided in figures like this →

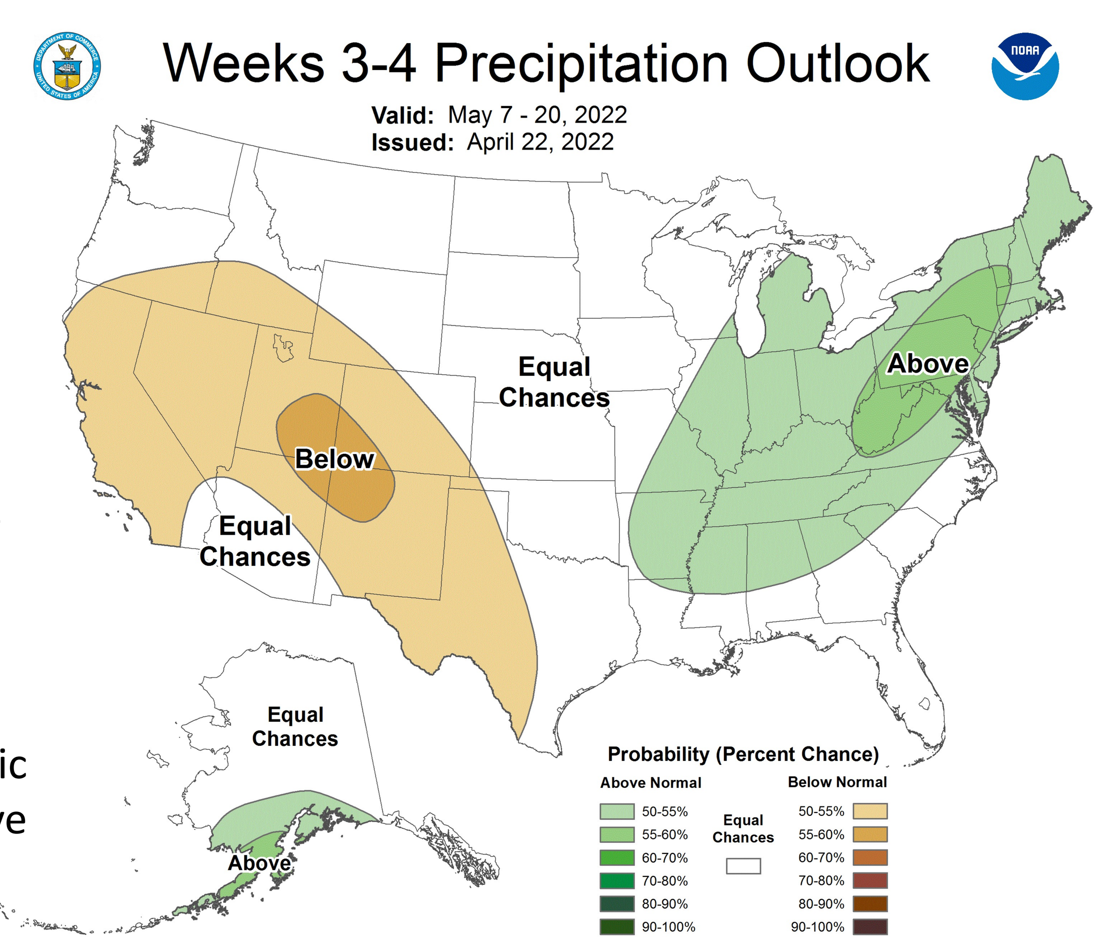


Figure from www.cpc.ncep.noaa.gov

Use of week 3-4 precipitation accumulation forecasts

- Managing water resources
- Monitoring ongoing & building droughts
- Identifying locations & times of enhanced or suppressed wildfire risk

Why ensembles need post-processing

Numerical model output inherently contains systematic model biases. Ensemble prediction systems additionally have errors in the magnitude of their ensemble spread.

Statistical post-processing can correct these errors by learning corrective relationships between past forecast and observations pairs. The corrections can then be applied to future forecasts so that they no longer suffer from the same deficiencies as the raw ensemble forecast.

B. Data Used in This Study

Forecast Dataset

(used as predictors in the statistical & neural network models)

- Global Ensemble Forecast System v12 (GEFSv12) reforecasts from 2000–2019
  Reforecasts: decades of past forecasts reran with the same model version
- 1x/week, 6-hourly forecasts initialized at 00 UTC and ran out to 35 days ahead
- Each reforecast is accumulated for a span of 7 and 14-days (Week 3, Week 4, Week 3-4)
- 0.5 deg horizontal grid spacing
- 11 ensemble members

Observed Dataset

(used to calibrate and evaluate the post-processing models)

- PRISM daily accumulated precipitation (PRISM 2022) from 1981–2019
- Observations were accumulated for a span of 7 and 14-days (Week 3, Week 4, Week 3-4)
- Upscaled with conservative regridding from 1/24th deg to 0.5 deg horizontal grid spacing

C. A More Traditional Statistical Post-processing Method

Use an established regression-based method which relates the smoothed and normalized raw ensemble mean to parameters of the predictive censored shifted gamma distribution, CSGD (μ, σ^2, δ) through a set of linear equations (ref. S15, S20) .

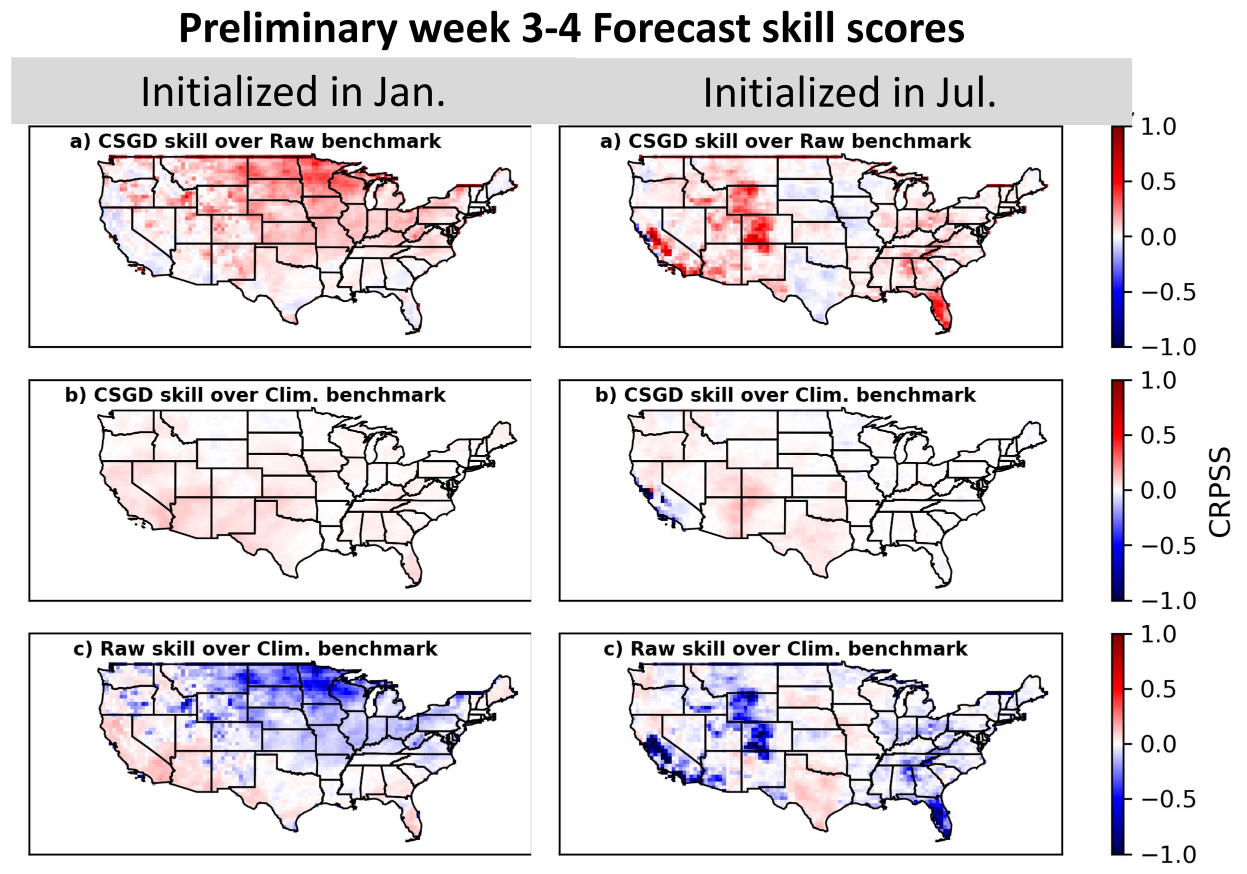
A CSGD is also fit to observed climatology and its parameters ($\mu_{cl}, \sigma_{cl}^2, \delta_{cl}$) are used as predictors too.

$$\mu = \frac{\mu_{\mathrm{cl}}}{a_1}\log\left\{1+\left[\left(\exp(a_1)-1\right)\left(a_2+a_3\bar{f}_{\mathrm{ano}}\right)\right]\right\}, \sigma = a_4\sigma_{\mathrm{cl}}\sqrt{\frac{\mu}{\mu_{\mathrm{cl}}}},$$

Select optimal regression coefficients based on minimizing the continuous ranked probability score (CRPS) of the CSGD over a training dataset.

Separate model is fitted for each month, lead time, and grid point. We use leave-one-year-out cross validation.

Preliminary week 3-4 Forecast skill scores



D. Artificial Neural Network Method

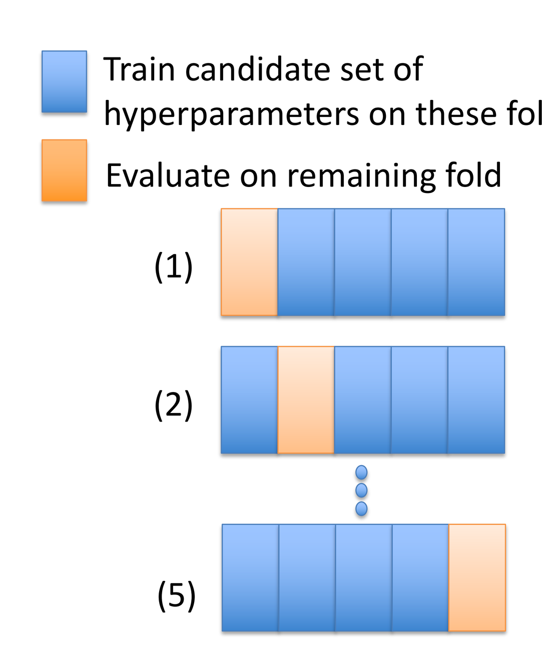
ANN is advantageous in two ways

1. We can include predictors that may interact in nonlinear ways
2. We can pool all grid points together to train the model allowing us to share regression parameters across grid points

Three main steps to generate full predictive CDF with an ANN

1. Partition the range of possible precipitation outcomes into *m*+1 categories based on location-specific quantiles of the observed climatology.
2. Predict probabilities for observed precipitation amounts to fall into each of those categories.
3. Interpolate the categorical probabilities to a full predictive CDF using the cumulative hazard function.

5-fold cross validation used for hyperparameter tuning



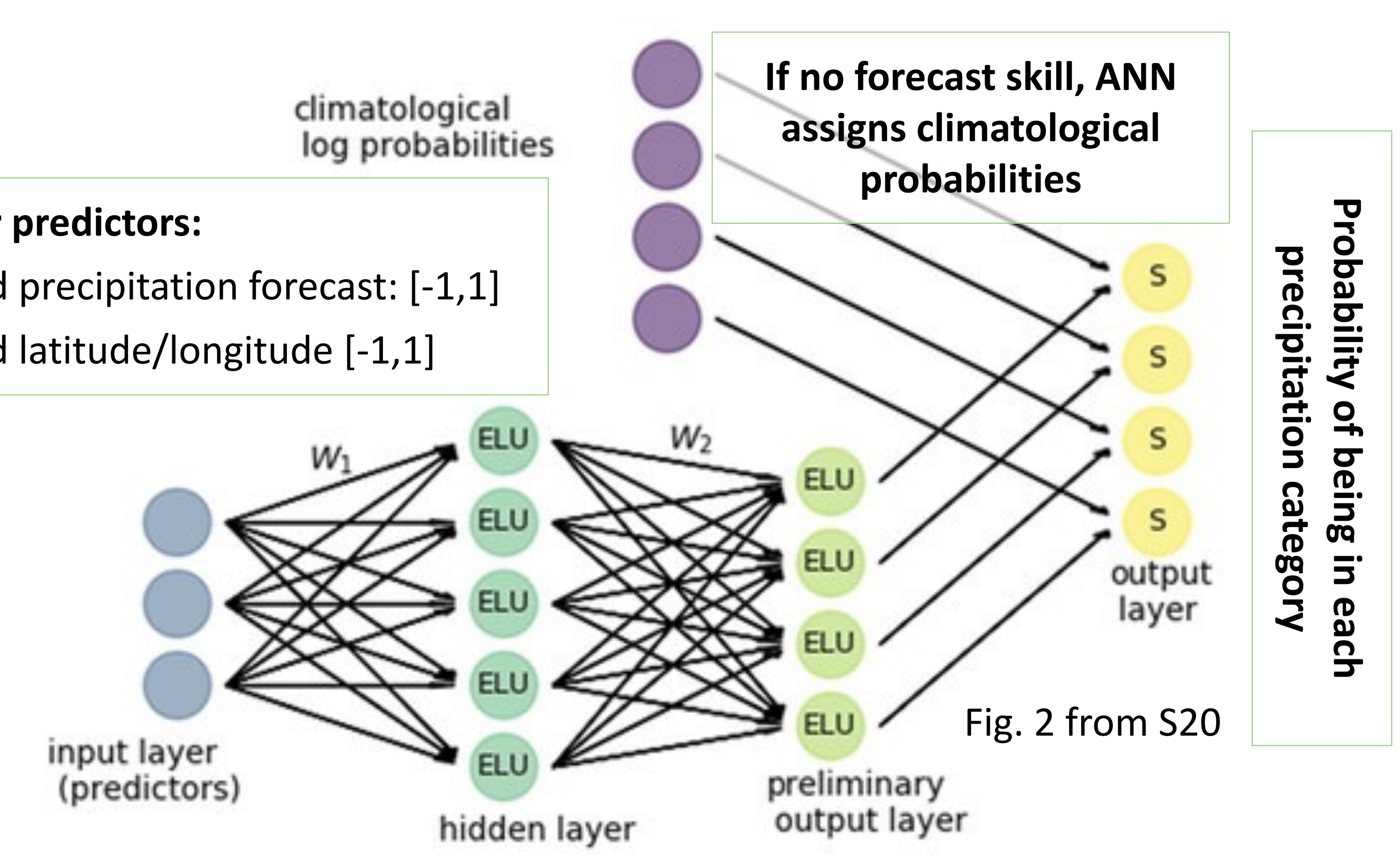


Fig. 2 from S20

E. References

GEFSv12: https://registry.opendata.aws/noaa-gefs-reforecast/

PRISM 2022: https://prism.nacse.org/

S15: Scheuerer, M., T. Hamill, 2015. Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.* DOI: 10.1175/MWR-D-15-0061.1

S20: Scheuerer, M., M. Switanek, R. Worsnop, T. Hamill, 2020. Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. *Mon. Wea. Rev. DOI:* 10.1175/MWR-D-20-0096.1