CIPS Experimental Analog-based Probability of Severe Guidance

1 March 2018

Overview

The CIPS Experimental Analog-based Probability of Severe Guidance provides a probabilistic threat of severe weather using historical environments (i.e., analogs) similar to the Global Ensemble Forecast System (GEFS) model forecast for the next eight days. The algorithm produces the probability of severe weather within 110 km of a grid point using a logistic regression model built from analogs and their related severe reports. The experimental guidance uses the 0000 UTC initialization of the GEFS and begins running at approximately 0600 UTC. The real-time output can be viewed here.

Finding Analogs

Analogs are historical environments, or in this case reforecasts, that are similar to the current GEFS mean forecast. The similarity between reforecast and GEFS model fields is computed by comparing twelve standard atmospheric fields (Table 1). The quality of the similarity is determined by the spatial correlation and the mean absolute difference between the GEFS forecasts and the historical analog over a domain (18 x 14 grid points on a 1 x 1 degree grid – see Illustration 1). Conceptually, the spatial correlation is assessing the similarity in a atmospheric field's pattern and the mean absolute difference is assessing the similarity in the amplitude of that pattern. The most similar analogs are found using the 0000 UTC reforecasts from years 1986-2016 and the assessment examines the same forecast hours. For example, 48-h GEFS forecasts are only compared to 48-h reforecasts from the GEFS. The analogs are found for a set of overlapping domains covering the CONUS (see Illustration 1) with only the top 100 analogs from each domain being passed on to the logistic regression procedure.

Logistic Regression

After determining the top 100 most similar analogs, the analogs and their associated severe weather reports are used to build a logistic regression model. Currently, the twelve standard atmospheric fields used in finding the analogs, in addition to surface-based CAPE and CIN, are predictors. In building the logistic regression model, a severe weather "hit" occurs if there are any severe reports within 110-km of a grid point. The model then analyzes the predictors' relationship to the severe reports and applies that relationship to the GEFS forecast fields, which results in a probability at each grid point. The regression analysis model is applied to all grid points within a sub-domain within each analog domain and the overlapping grids result in each grid point having 4 separate logistic regression probabilities (except along the edges of the CONUS were only 2 estimates are available – see Illustration 1). These probabilities are averaged together to create the final probabilistic value.

Some quality control is applied in calculating the resulting probabilities. First, if a grid point has fewer than 10 analogs (out of the top 100) with severe, the logistic regression is not performed. In these instances, we believe there are too few hits to robustly determine the coefficients of the logistic regression model. Second, if the average probability at a grid point is less than 10%, it is not displayed in the probabilistic graphics. These both help to reduce the "noise" that was observed in many of our test cases.

Issues

The experimental analog-based probability of severe guidance is built on the GEFS mean fields, therefore it is rare that probabilistic signals are found beyond day six. This is likely due to the spread in the GEFS member forecasts at those lead times. This spread can often result in a mean field that may not be representative of the individual members even though severe weather may be possible from many of the individual GEFS members. Consequently, as the GEFS member solutions start to converge on a large-scale pattern capable of severe weather, the GEFS mean is more representative of the members. More research is needed to determine when and why the probabilistic severe weather signal "locks in". Additionally, it would be interesting to examine the potential of severe weather for each individual member and then combine those probabilities for the resulting product. It is hypothesized that some members even 6-8 days out may show severe potential with would provide a probabilistic signal at extended ranges.

Mass Fields	Temperature Fields	Moisture Fields
Mean Sea Level Pressure	2-m Temperature	2-m Specific Humidity
850-mb Height	850-mb Temperature	850-mb Specific Humidity
500-mb Height	700-mb Temperature	700-mb Specific Humidity
300-mb Height	500-mb Temperature	Precipitable Water

Table 1. The fields used to assess the similarity between the current GEFS mean forecast and the GEFS mean reforecast (i.e., fields used to find analogs).

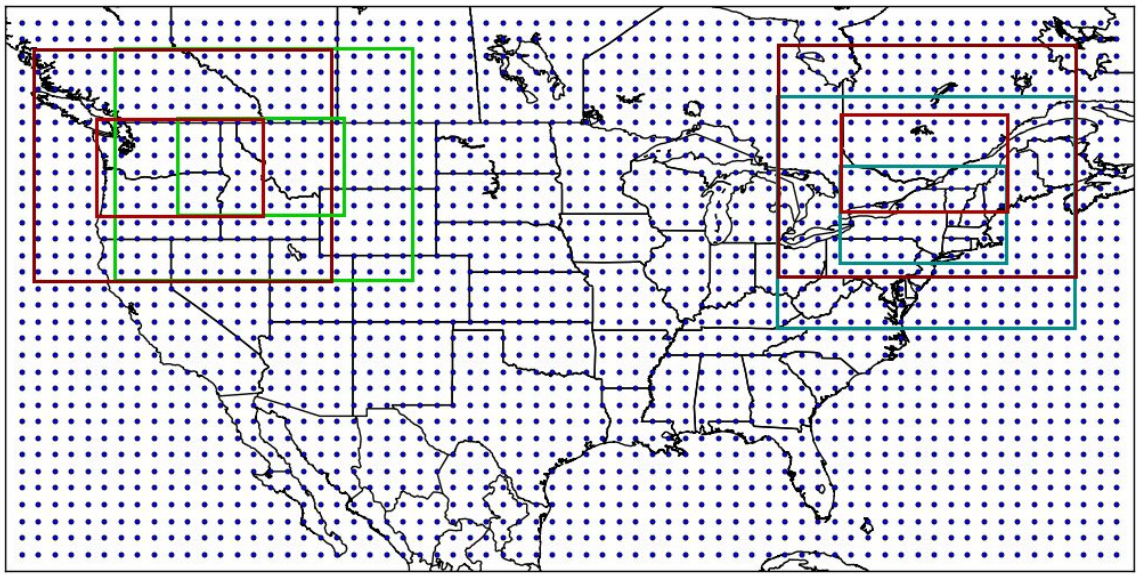

Illustration 1. Sample analog domains (larger boxes) and logistic regression domains (smaller boxes). The green and blue domains are examples of the overlap between adjacent domains.