64. PREDICTIONS OF SEVERE WEATHER ENVIRONMENTS BY THE CLIMATE FORECAST SYSTEM VERSION 2 MODEL SUITE

Adam J. Stepanek*
Purdue University, West Lafayette, IN

Robert J. Trapp
University of Illinois, Urbana, IL

Michael E. Baldwin
Purdue University, West Lafayette, IN

1. INTRODUCTION

The prospect for skillful long-term predictions of severe convective storms remains unclear. This serves as the motivation for an ongoing assessment of the NCEP Climate Forecast System Version 2 (CFSv2) suite of model products during climatological peaks in severe convective weather activity. Specifically, predictions from Climate Forecast System Reforecast (CFSRR) output have been analyzed for the spring (AMJ) months, and verified against output from the Climate Forecast System Reanalysis (CFSR).

Our particular focus remains on the predictability of convective environments at sub-seasonal lead times through the analysis of parameters with well-established correlations to severe weather – specifically convective available potential energy (CAPE) and deep-layer vertical wind shear (VWS, hereafter DLS) as highlighted in the work of Brooks et al. (2003a).

The purpose of this research is to evaluate the capacity of the CFSv2 to predict these environments through the establishment of a baseline methodology for analyzing the CFSRR, which is utilized from a predictive standpoint, and is subsequently compared to a climatological standard developed from the CFSR. Furthermore, this study aims to elucidate periods of notable skill through a variety of methods including proportion correct (PC), root-mean-square-error (RMSE) and Heidke skill score approaches. The methodology, which is predicated on cumulative distribution functions (CDFs) and area-under-the- curve (AUC) (calculated using basic trapezoidal methods using set limits of integration) techniques is described in section 2, while some of the initial key results are outlined in section 3. Finally, concluding remarks are made in section 4.

* *Corresponding author address:* Adam J. Stepanek, Purdue University – Department of Earth, Atmospheric, and Planetary Sciences, West Lafayette, IN 47907;
e-mail: astepane@purdue.edu.

2. DATA AND ANALYSIS

Twenty-nine years (1982-2010) of CFSRR output valid for individual month-long periods of April, May and June serve as the predictive element to this research. Three-dimensional pressure files from the CFSRR utilized in this research are available every five days, starting on 1 January, with members available at 0000, 0600, 1200 and 1800 UTC, and output extending forward through a 9-month period; this output and all other data in this study were obtained from the NOAA NOMADS data repository (https://nomads.ncdc.noaa.gov/).

An extensive baseline climatology is established using individual monthly (April, May and June) aggregates from the entire 32 years (1979-2010) of available CFSR data. Annual CFSR output serves as a proxy for verification data. As compared to archived sounding data using a nearest neighbor technique, the CFSR exhibits a negligible bias in regions of interest with regards to DLS, and only a mild, low bias across the western Great Plains with regards to CAPE. Here, CAPE is computed using a surface-based parcel method, and DLS is generated using the vector wind difference from the 1000-500 hPa layer, and represents a proxy for 0-6 km deep-layer wind shear. Note that the use of near-surface and mid-level wind components are designed to represent the shear that is effectively available for storm organization and any consequent rotation, yet differs from storm-relative helicity (SRH) in that there is no need to know, or estimate, storm motion a priori to performing the calculation.

Cumulative distribution functions of mean predictions and verifications (climatology) are generated using all available data from the respective model source for the individual months of April, May and June. As such, the mean predictive CDFs are comprised of either 3480 or 3596 individual values, while monthly climatologies consist of 960 or 992 members. Subsequently, individual predictions comprised of all applicable model members are aggregated using a lagged average forecasting technique developed by Hoffman and Kalnay (1983) where all forecasts are assigned equal weights regardless of the length of the lead. These predictions are based upon a specified lead-time prior to the start of a month of interest, with the exact lead predicated on availability within the 5-day period of the CFSRR. This can be further explained in that, for example, the 10-day prediction for the month of May (for any given year) would be the lagged average ensemble of all prediction members initialized on April 21st of the same year. Given the focus on sub-seasonal time scales, predictions for leads ranging from approximately day-10 to week-8 are analyzed in this research.

Although the predictive (CFSRR) and climatological / verification (CFSR) components of this study originate from the same model physics, it is found that that CFSRR predictions exhibit non-negligible levels of bias (relative to CFSR) in both CAPE and DLS, with CAPE displaying significant spatial variance to the bias. Specifically, a high bias exists in CAPE for many areas across the central United States, particularly across the central Great Plains/High Plains. A widespread low-amplitude bias in DLS exists across much of the central CONUS (Fig. 1).

To address this inherent bias, the *difference* in AUC between the mean prediction (climatology) and the prediction for an individual month-long period (individual monthly verification) is calculated for all years of available CFSRR output, resulting in scalar values that communicate information about the CAPE / DLS environment. These values, termed 'anomalous' AUC (A-AUC) are developed at each grid point over select areas prone to severe storms to further investigate measures of skill. The differencing is performed in a manner that allows positive (negative) A-AUC values to represent a prediction of an overall above- (below-) average period of the coincident parameter. Subsequent comparison to A-AUC values for the climatology/individual monthly verification output is the basis for determining if skill is present in the predictions.

To be able to generate meaningful comparisons of A-AUC values both at grid point level as well as over larger regions, it was imperative to standardize the limits of integration, ensuring that a single outlier in the data would not have undue influence when calculating area-under-the-curve. After creating CDFs for the mean prediction and climatology for a given month, the value for each parameter at the 95th percentile was determined for every grid point within the domain of interest. All values exceeding this measure for both the mean predictive and climatological CDFs as well as individual predictions and verifications are subsequently 'truncated' down to this value. Furthermore, the lowest CAPE or DLS value is always set to be zero, effectively normalizing the upper and lower bounds of integration for AUC calculation. This is necessary to eliminate any undesirable influence from extreme values, and allowing the A-AUC calculation to appropriately convey anomalous behavior for both CFSRR predictions and CFSR-based verifications.

3. **RESULTS**

The first skill-based analysis of CFSRR output was predicated upon the ideology that model output has the potential to be more influential when it has a strong signal relative to the mean. As such, grid points were analyzed for 'notably anomalous' behavior, i.e., grid points consisting of an A-AUC value greater than one standard deviation. An aggregate of all years exhibiting this behavior at each grid point were then compared to verification data to see if the correct 'sign' was predicted. A 'hit' was awarded if a prediction of above- (below-) average A-AUC verified with the proper sign, independent of magnitude. Figure 2 shows the 'notably anomalous' proportion correct for the central and eastern CONUS for May, with the predictive members originating 10-days prior to the start of the month (Fig. 2). Strong (weak) performance for DLS (CAPE) is noted over the Plains (Midwest) with isolated areas where positive skill (based upon values greater than 0.5, or a completely random expectation) overlap.

Composite predictions are considered for specified regions including the southern Great Plains (SGP), central Great Plains (CGP), Midwest (MW) and Southeast (SE), which exhibit similar climatological peaks in severe weather activity (Brooks et al. 2003a) (Fig. 2). Combining data points within each region, mean trends in skill with regard to lead time were analyzed by month and region (Fig. 3). Note that not all CFSRR output was available and, as such, there are some points missing in the trend lines. Positive skill (relative to a random prediction) in the model prediction of CAPE for a combination of all four regions is noted for all leads for April and most of June, the latter of which exhibits increasing skill with decreasing lead time (Fig. 3; upper left). A more notable positive trend in skill is seen in the regional average of DLS, with the greatest overall skill found in April and a mild decrease in skill moving into May and June (Fig. 3; upper right).

Viewing individual regions averaged over April, May and June (or the available subset), minimal trends can be noted in output for CAPE, although the SGP region exhibits consistently high skill (minimal skill in the MW region) (Fig. 3; lower left). Some of the most intriguing results are found in the DLS average for each region – whereas the data are somewhat chaotic beyond an approximate week-3 lead, a strong and consistent upward trend is seen from a lead time of ~21 days to a lead time of ~11 days, corresponding to increases of nearly 10% in terms of 'notably anomalous' proportion correct (Fig. 3; lower right). Furthermore, aside from the SE region, the CFSRR displays almost universally positive skill for deep-layer shear. Overall, when the specific focus is on notably above- or below-average predictions of CAPE/DLS, the CFSRR exhibits a generally positive measure of skill at lead times throughout the sub-seasonal temporal range.

Heidke skill score (HSS) values were then calculated for the entirety of the CFSRR period. Based upon a standard 3x3 contingency table, A-AUC calculations from both predictions and verifications were separated by thirds into categories of 'above-average', 'near-normal', and 'below-average' relative to a quasi-normal distribution of output. Binning was performed first at grid point level, and secondarily over the previously outlined regions of specified interest. For the purposes of this study, skill in terms of HSS is defined as any value greater than zero, given the inherent difficulty in long-range predictions. Focusing on the month of May, positive skill DLS predictions are noted for all regions, with the strongest CFSRR performance noted from the Central Great Plains into the Midwest (Table 1). Skill in CAPE prediction is much more muted, with positive values noted in the Plains, and negative skill in the Midwest and Southeast regions.

	CAPE	**DLS**
SGP	0.112	0.100
CGP	0.056	0.243
MW	-0.038	0.195
SE	-0.039	0.038

Table 1. Heidke Skill Scores for the month of May, based on 29 years of CFSRR model output

To better understand the meteorological contributors to periods of CFSRR skill, it is desirable to relate such skill to large-scale meteorological patterns and furthermore to phenomena such as the Madden Julian Oscillation (MJO) and El Nino-Southern Oscillation (ENSO), which have established teleconnections between the tropical Pacific Ocean and the CONUS midlatitudes. These efforts are founded upon studies showing the CFSv2 to possess improved predictability of the MJO beyond a week-2 lead by Zhang and van den Dool (2012), while high ENSO predictive skill was demonstrated by Saha et al. (2014).

Employing RMSE techniques using climatology as the predictive tool upon which model skill is based, analyses were performed within all regions of interest for individual years at varying lead times. At the regional level, individual years exhibiting the highest RMSE skill (based upon improving model skill with decreasing leads as well as average RMSE difference, relative to climatology, over all leads) were combined for each parameter. The aggregate of these years was subsequently used to generate plots of anomalous behavior of other variables, such as 500-hPa geopotential height and dew point at the .995 sigma level. Additional sorting was performed to separate periods of above- and below-average verification.

An example shown (Fig. 4) is for the SGP region for the month of May, comprised of three years (1997, 2001 and 2007) that exhibited predictive RMSE skill and coincided with an above-average verification of CAPE. The 500-hPa pattern shows anomalous ridging in the western CONUS (Fig. 4; upper left, upper right). Furthermore, near-surface dew point values show an influx of higher values across the eastern Plains (Fig. 4; lower left), with some mild positive anomalies across the SGP (Fig. 4; lower right). It must be stressed that severe thunderstorms are not insinuated by this setup; this is simply an illustration of the concurrent pattern associated with positive skill in terms of CAPE in the SGP region. A preliminary investigation into common threads between the three aforementioned years shows a strong (or strengthening) MJO in the western and/or central tropical Pacific Ocean – future efforts will focus on expanding this aspect of the study.

4. CONCLUSIONS / FUTURE WORK

The objective of this research was to establish a baseline for analyzing the skill of sub-seasonal predictions of convective environments. The work introduced here focuses on the reforecast element of the CFSv2, the CFSRR, and comparisons to the CFSv2 reanalysis product, which allows for the generation of an extended climatology and a basis for verification. A variety of methods employed to assess model skill have been applied to 'anomalous' area-under-the-curve calculations at grid points across the central and eastern CONUS, focusing on four regions of particular interest. The CFSRR has been shown to exhibit skill (compared to a baseline of random prediction or a prediction of climatology) at multiple leads for both CAPE and DLS, with deep-layer shear displaying greater overall skill than CAPE. Specifically, DLS was shown to have a strong signal of positive skill when averaged over the course of all available spring months (AMJ) out to an approximate week-3 lead.

Whereas this work provides a strong proof of concept into assessing skill of the CFSv2, additional tasks will be performed in order to better understand the association between model skill and the large-scale pattern, which are often driven in part by predictable, global-scale features such as the MJO and ENSO. An algorithm for aggregating years and regions of significant model skill is in development with the goal of identifying patterns and/or features associated with phenomena such as the MJO that can be exploited at the sub-seasonal predictive range. Furthermore, all analysis techniques will be applied to the operational CFSv2 (2011-present) to further analyze established connections between model skill and the large-scale environment. Finally, CAPE and DLS will be analyzed in tandem to better test the skill of the CFSv2 to predict an environment conducive to severe thunderstorm development. It must be noted that this work is designed to analyze environments supportive of severe storms, but does not guarantee convection initiation. Ultimately, the goal remains to elucidate when the CFSv2 exhibits skill and utilize this information to create skillful sub-seasonal forecasts of severe thunderstorm activity.

Acknowledgments

The lead author would like to acknowledge support from the NSF AGS 106226 award and NOAA/OAR award no. NA15OAR4590231 to allow for travel and presentation of this research.

REFERENCES

Brooks, H. B., C. A. Doswell III, and M. P. Kay, 2003a: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Hoffman, R. N., and E. Kalnay, 1983: Lagged-average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35,** 100–118.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate,* **27,** 2185-2208.

Zhang, Q., and H. van den Dool, 2012: Relative merit of model improvement versus availability of retrospective forecasts: The case of Climate Forecast System MJO prediction. *Wea. Forecasting,* **27,** 1045-1051.

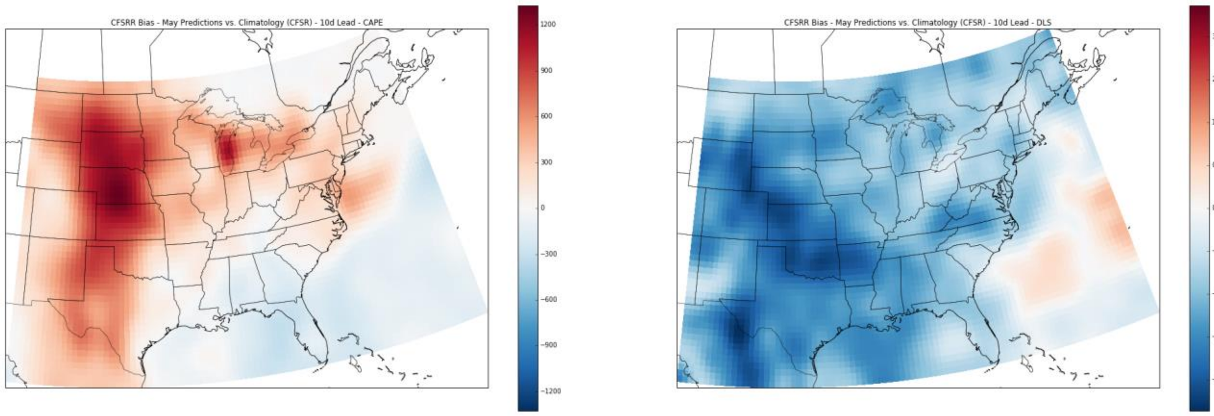

Figure 1. Analysis of bias in CAPE (left) and DLS (right) between the mean prediction from the CFSRR (May; 10-day lead) and a baseline climatology developed from the CFSR. Warm (cool) colors represent a high (low) model bias. Complimentary months exhibit similar spatial biases.

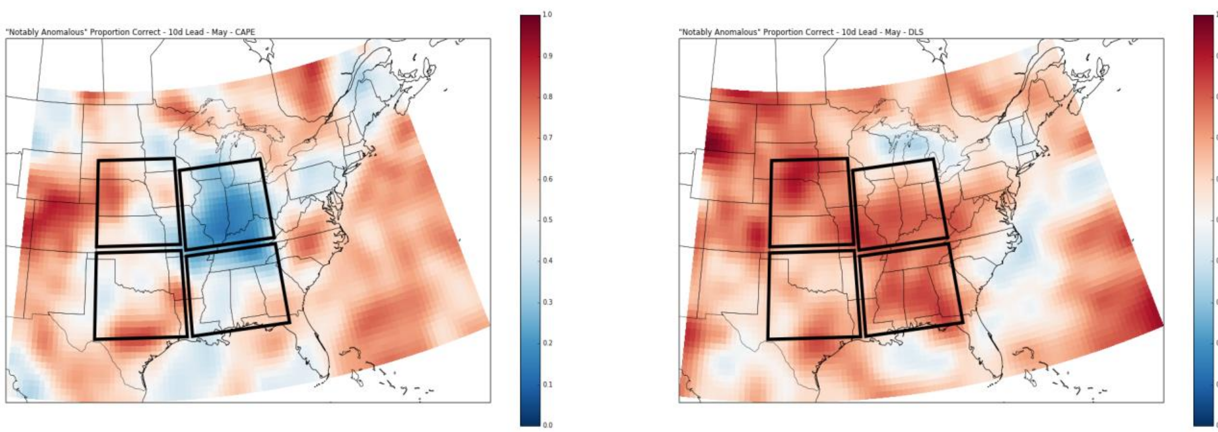

Figure 2. 'Notably Anomalous' Proportion Correct predictions for CAPE (left) and DLS (right) (May; 10-day lead). Warm (cool) colors represent positive (negative) skill compared to a random prediction. Black outlines define the following regions, clockwise from lower left: southern Great Plains (SGP), central Great Plains (CGP), Midwest (MW) and Southeast (SE).

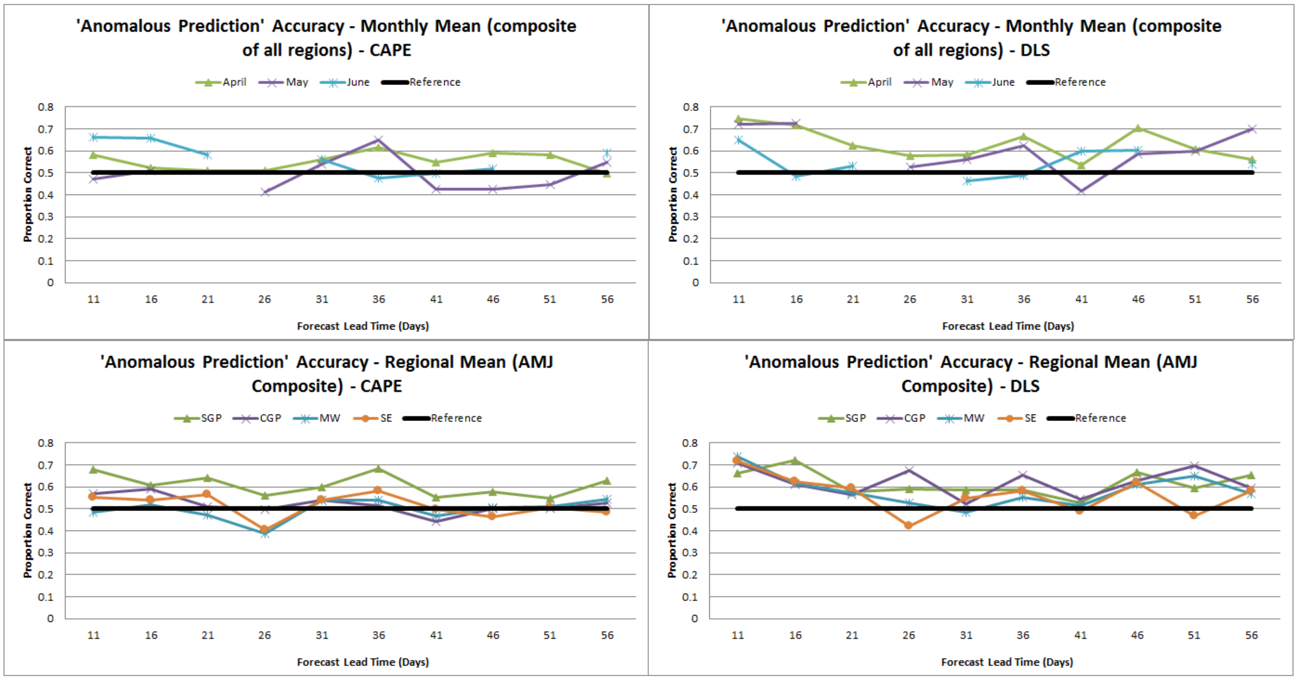

Figure 3. Time series plots (increasing lead time along the abscissa from left to right) for composites of 'notably anomalous' proportion correct skill analysis. CAPE (DLS) averaged over all regions for (available) individual months is shown in the upper left (upper right). CAPE (DLS) averaged for all (available) months for individual regions of interest is shown in the lower left (lower right).

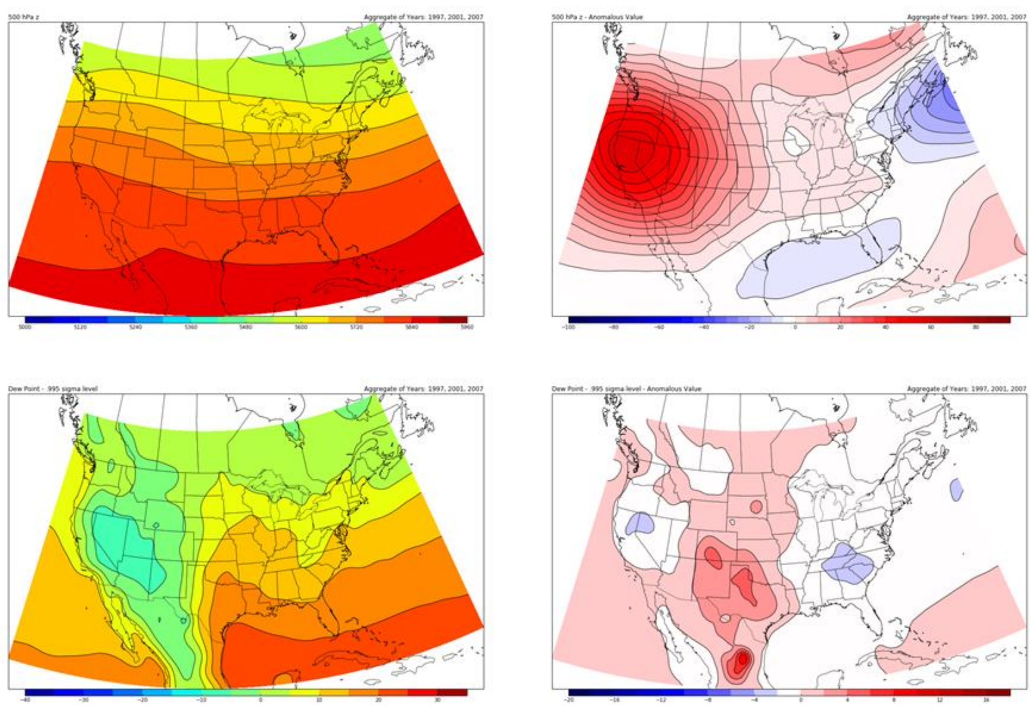

Figure 4. Composite (left column) and anomalous (right column) 500-hPa heights (top row) and .995 sigma dew point (bottom row) for years exhibiting a strong positive RMSE-based skill for CAPE in the SGP region during the month of May. All years shown exhibit above-average verification of CAPE (1997, 2001 & 2007).