AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL069150

Key Points:
- Objective climate regionalization identifies four major CONUS regions that are homogeneous in interannual warm season rainfall variability
- CFSv2 nearly matches the observed regions but systematically misplaces variability in some areas
- Regionalization can be used to map and potentially to address areas of poor forecast performance

Supporting Information:
- Supporting Information S1

Correspondence to:
S. K. Regonda,
satish.regonda@gmail.com

Citation:
Regonda, S. K., B. F. Zaitchik, H. S. Badr, and M. Rodell (2016), Using climate regionalization to understand Climate Forecast System Version 2 (CFSv2) precipitation performance for the Conterminous United States (CONUS), *Geophys. Res. Lett.*, *43*, doi:10.1002/2016GL069150.

Received 18 APR 2016
Accepted 28 MAY 2016
Accepted article online 1 JUN 2016

Using climate regionalization to understand Climate Forecast System Version 2 (CFSv2) precipitation performance for the Conterminous United States (CONUS)

Satish K. Regonda[1,2], Benjamin F. Zaitchik[1], Hamada S. Badr[1], and Matthew Rodell[3]

[1]Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, Maryland, USA, [2]World Bank, Washington, District of Columbia, USA, [3]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center (GSFC), Greenbelt, Maryland, USA

Abstract Dynamically based seasonal forecasts are prone to systematic spatial biases due to imperfections in the underlying global climate model (GCM). This can result in low-forecast skill when the GCM misplaces teleconnections or fails to resolve geographic barriers, even if the prediction of large-scale dynamics is accurate. To characterize and address this issue, this study applies objective climate regionalization to identify discrepancies between the Climate Forecast System Version 2 (CFSv2) and precipitation observations across the Contiguous United States (CONUS). Regionalization shows that CFSv2 1 month forecasts capture the general spatial character of warm season precipitation variability but that forecast regions systematically differ from observation in some transition zones. CFSv2 predictive skill for these misclassified areas is systematically reduced relative to correctly regionalized areas and CONUS as a whole. In these incorrectly regionalized areas, higher skill can be obtained by using a regional-scale forecast in place of the local grid cell prediction.

1. Introduction

The Climate Forecast System version 2 (CFSv2) [*Saha et al.*, 2014] is one of the most widely used dynamical seasonal forecast systems worldwide, in both research and operational mode, for a range of applications, for example, energy industry decisions [e.g., *Dutton et al.*, 2014], marine resources management [e.g., *Stock et al.*, 2015], new climate model [e.g., *Swapna et al.*, 2014] and products [e.g., *Kirtman et al.*, 2014] development, cyclone-related forecasts [e.g., *Choi et al.*, 2016; *Kim et al.*, 2015; *Zhan and Wang*, 2016], and hydroclimate variables [*Yoon et al.*, 2012; *Yuan et al.*, 2013, 2014, 2015; *Demargne et al.*, 2014; *Georgakakos et al.*, 2014; *Sheffield et al.*, 2014; *Tian et al.*, 2014; *Sharma et al.*, 2015; *Siegmund et al.*, 2015; *Zuo et al.*, 2015]. The performance of the model has been evaluated in hindcast mode for several regions [e.g., *Yuan et al.*, 2011; *Zhu et al.*, 2013; *Kumar et al.*, 2014; *Lang et al.*, 2014; *Silva et al.*, 2014; *Ma et al.*, 2015]. Performance varies by region and season, but results generally show that CFSv2 does show promise for forecasts up to 1 month lead [e.g., *Yuan et al.*, 2011, 2013; *Mo et al.*, 2012; *Roundy et al.*, 2015]. Model skill tends to degrade beyond the 1 month forecast horizon.

The degradation of CFSv2 forecast skill with increasing lead time is seen in Figure 1, which displays grid cell correlations between observed and CFSv2 hindcast precipitation for the warm season (May–September) for the Conterminous United States (CONUS). There is significant skill over much of CONUS at 1 month lead, but skill is limited at longer lead times. This is consistent with previous evaluations of CFSv2 for CONUS [e.g., *Zhu et al.*, 2013; *Roundy et al.*, 2015]. Focusing on the relatively high skill scores seen at 1 month lead, we see strong performance in many regions but pockets of low skill as well—notably in the upper Midwest, the Texas panhandle, and portions of the lower Mississippi basin.

The purpose of this paper is to investigate these regions of low skill in CFSv2 through objective climate regionalization [e.g., *Fovell and Fovell*, 1993; *Rao and Srinivas*, 2006; *Badr et al.*, 2015, and references therein]. This approach is motivated by the premise that a global climate model (GCM) forecast system like CFSv2 can only be expected to achieve strong predictions for locations that the model successfully places in the correct climate region, where the "climate region" is defined in terms of interannual variability in the variable of interest (in this case, precipitation). For example, if a particular location is known from observation to fall within a climate region for which precipitation is most sensitive to eastern Pacific Ocean variability, but circulation

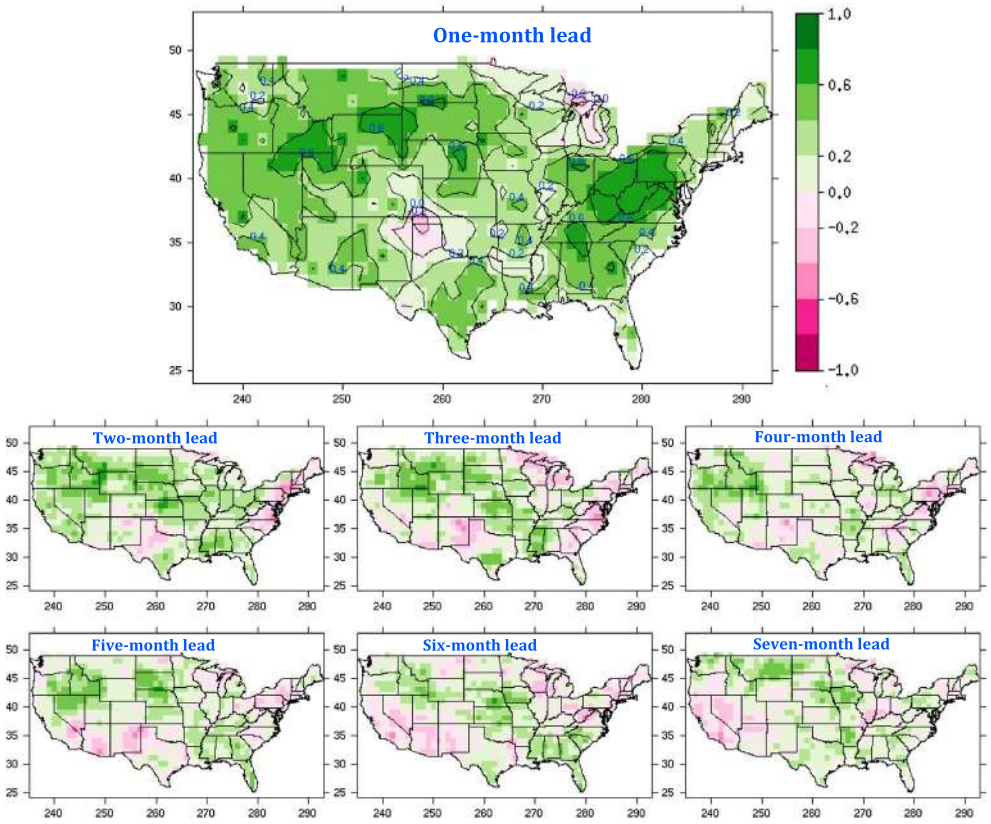

Figure 1. Grid cell to grid cell correlations between observed and CFSv2 hindcast precipitation for warm season (May through September) for different lead times; (top) 1 month lead time and (bottom right) 7 month lead time.

biases in the GCM erroneously place the location within a climate region that is more sensitive to Atlantic Ocean variability, then the prediction for that location will be unreliable even if the GCM captures the broad picture of teleconnections reasonably well. Objective regionalization can be used to identify these localized mismatches and, potentially, to explain spatial variations in model performance. A secondary aim of the paper is to identify ways in which an understanding of model performance based on regionalization might be used to develop more useful forecasts for areas in which grid-based skill metrics are low. Philosophically similar approaches were pursued by *Koster et al.* [2008] and *Roundy et al.* [2015].

2. Data

In this study we used precipitation data from the CFSv2 hindcast archive and the University of Delaware (UDel) gridded precipitation product.

The CFSv2 model is a second-generation coupled ocean-land-atmosphere GCM developed by the NOAA National Center for Environmental Prediction (NCEP). An archive of CFSv2 hindcasts for 1982–2010 is maintained to facilitate research and evaluation. The archive includes 24 ensemble members for each month (six different initialization days * 4 times of day), each run for a 9 month forecast and archived at 1° resolution, globally. We downloaded hindcasts from the International Research Institute (IRI) map room (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/.MONTHLY/.prec/dataselection.html) and subset to the CONUS.

Here we evaluate the CFSv2 24-member ensemble mean. A "1 month lead" forecast is defined as the ensemble average of all forecasts included in that month's CFSv2 mid-month hindcast release (See Appendix B in *Saha et al.*, 2014). We focus on the warm season (May–September). This period accounts for >50% of annual precipitation in many regions within CONUS, and it is known to be difficult to predict: warm season precipitation is often convective, and the coupling between local weather and large-scale atmospheric patterns tends

to be weaker than in the cold season, when precipitation variability over much of CONUS is dominated by synoptic dynamics.

In this study we use the University of Delaware observation-based gridded precipitation analysis (UDel) [*Matsuura and Willmott*, 2012] as the basis for evaluating CFSv2 hindcasts. UDel has a gridded resolution of 0.5° × 0.5°. It was selected for this study because it is based entirely on interpolated in situ observations, is spatially and temporally complete and quality controlled, and has global coverage that would make it possible to repeat the study outside of CONUS. The UDel data set has been used in many previous studies [e.g., *Wang et al.*, 2009; *Caldwell*, 2010; *Elguindi and Grundstein*, 2013]. A full description of the analysis technique, shortcomings, and advantages of UDel is found in *Matsuura and Willmott* [2012] as well as at http://climate.geog.udel.edu/~climate/. Here we resample UDel to 1° to match the resolution of CFSv2. A complementary analysis was performed using North American Land Data Assimilation System (NLDAS) precipitation [*Xia et al.*, 2012] as the basis for evaluation to confirm that the UDel global product behaves similarly to CONUS-only NLDAS.

3. Climate Regionalization

The goal of climate regionalization techniques is to identify regions that are homogenous in certain aspect of a climate variable(s) of interest. There is a rich and evolving literature that includes numerous statistical and conceptual techniques [e.g., *Fraley and Raftery*, 1998; *Jain et al.*, 1999; *Sheridan*, 2002]. These methods all define "regions" (which do not need to be spatially contiguous) that are homogenous with respect to the variable of interest (e.g., high intraregional correlation between members and regional mean) and are sufficiently distinct from other regions (e.g., low interregional correlation between regional means). These techniques are frequently described as "objective regionalization" algorithms, because an objective metric of similarity versus dissimilarity is applied to define regions. The methods are, however, subjective to some extent, in that the selection of the number of regions and level of acceptable intraregion homogeneity relative to interregion difference is subjective; there are some automated methods for making these decisions [e.g., *Badr et al.*, 2015], but ultimately, the choices depend on the purpose of the regionalization.

In this application, we regionalize the gridded observational and model data sets on the basis of interannual variability in warm season precipitation. Regionalization was performed for the period 1982–2010, as this is the full extent of the CFSv2 hindcast archive. Several forecast lead times were analyzed, but we focus on the 1 month lead forecasts, as they exhibit meaningful skill across much of CONUS. Lead time refers to the lead for each month: the May–September forecast comprises 1 month lead forecasts for each month in the season (i.e., May predicted from April model initializations and June from May initializations).

This choice of clustering variable (i.e., time series of gridded seasonal precipitation) means that regions are defined on the basis of coherent interannual climate variability rather than mean climate conditions. This is important, as coherent interannual variability suggests a common response to large-scale climate drivers. When regionalization is applied to observations, then the approach can help us to identify regions of common climate sensitivity and to identify mechanisms that drive variability for each region. When applied to the model, the technique allows us to examine whether the model connects large-scale drivers to local climate variability in a reliable way. For a high performing seasonal prediction system, we would expect that model regions strongly resemble observed regions and that dominant drivers of variability in each region also match observations. The further implication is that the model's grid cell to grid cell skill values are likely to be low in areas where the model's regionalization does not match observed.

Our rationale for this approach is that areas of common variability are probably sensitive to common large-scale climatic drivers. Regionalization, then, allows us to understand how different regions within CONUS respond differently to large-scale climate patterns. It also allows us to compare the regional coherence of observations to the regional coherence of the model. We applied different regionalization methods and chose Ward's hierarchical clustering [*Murtagh*, 1983; *Ward*, 1963], because its derived regions are distinct and approximately of similar size, which one can expect for low noise data, whereas other methods such as regional linkage [*Badr et al.*, 2014] can isolate the noise in very small regions for quality control. Ward's method calculates the sum of squared distances between different objects, pools objects with the smallest sum of squared distances under one cluster, and repeats it until all objects either are associated with an existing cluster or form a new cluster. The hierarchical clustering algorithm yields a dendrogram (a tree diagram) that illustrates the arrangement of objects into different clusters, e.g., Figure 2a. The *x* axis represents the

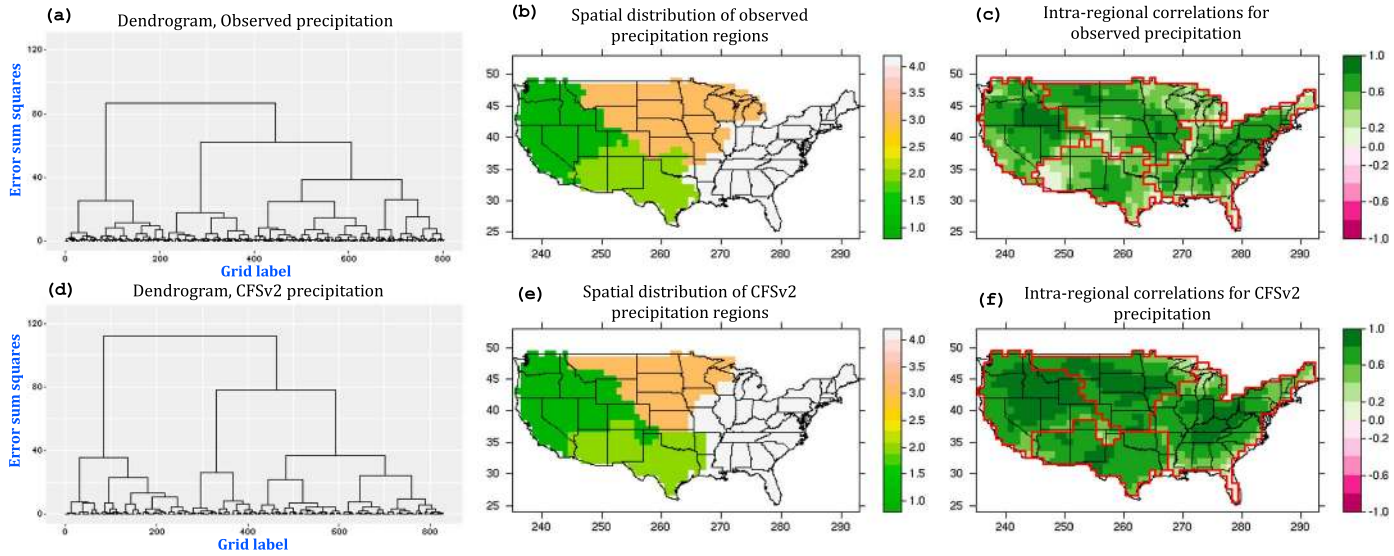

Figure 2. Regionalization results using Ward's method classification for warm season (May through September) precipitation, for (a–c) UDel observed and (d–f) CFSv2 1 month lead hindcasts precipitation data sets. Dendograms, spatial distribution of regions, and intraregion correlations are presented in the first, second, and third columns, respectively.

labels of objects used in clustering, where each branch from the bottom represents one object or a cluster if the dendrogram is cut at a higher level. The *y* axis represents the cost of merging two branches. In Ward's clustering, the merging cost is the error sum of squares in a cluster. While several factors play a role in choosing an optimum number of regions, the dendrogram suggests a number for which the merging cost does not decrease significantly. In this study, Ward's method was implemented using the HiClimR version 1.2.3 [*Badr et al.*, 2014] climate regionalization package [*Badr et al.*, 2015] for R [*R Core Team*, 2016].

4. Results and Discussion

Figures 2a–2c show results of Ward's method classification for interannual variability in May–September precipitation using UDel precipitation. As noted above, the selection of the number of regions is subjective. In this case we select four regions based on considerations of intraregional variance, i.e., how homogenous is the region after merging two smaller regions (see dendrogram in Figure 2a) and the objective of identifying regions that are associated with distinct patterns of large-scale climate variability. We refer to these regions as West, South-central, North-central, and East (Figure 2b; Figure S1, in the supporting information, shows sensitivity to data resolution). Maps of alternative cutoff values are provided in the supporting information (see Figure S2). Figure 2c shows intraregional correlations for each region (i.e., the correlation of each grid cell with the mean time series for the region it is assigned to). As expected, we see high correlations in most areas, indicating that the regions are relatively homogeneous (see also Table S1 in the supporting information). Correlation values do drop at some region edges and in particularly dry regions like Arizona have the lowest intraregional correlation overall. Correlations between regions are relatively low, indicating good separability of these four regions (Table S1). NLDAS results are generally similar to UDel (see Figures S2 and S3).

Figures 2d–2f repeat the regionalization process for CFSv2 1 month lead forecasts. The dendrogram for CFSv2 suggests a similar structure of grids and clusters as observations but with some differences (Figures 2a and 2d). The dendrogram for CFSv2 (Figure 2d) does not provide definite information on optimum number of clusters; i.e., when five regions are selected, the fifth region is very small relative to other regions, whereas when four regions are selected, the fourth region has large intraregional variance relative to other three regions but importantly the same number of regions as observations. Note that the cutoff value, which varies with several factors including data resolution, data structure, and region size, differs for observations (~25) and CFSv2 hindcast (~35). However, similar spatial patterns are observed for model and observations for a four region classification (Figures 2b and 2e). This is encouraging, as it suggests that the geographic

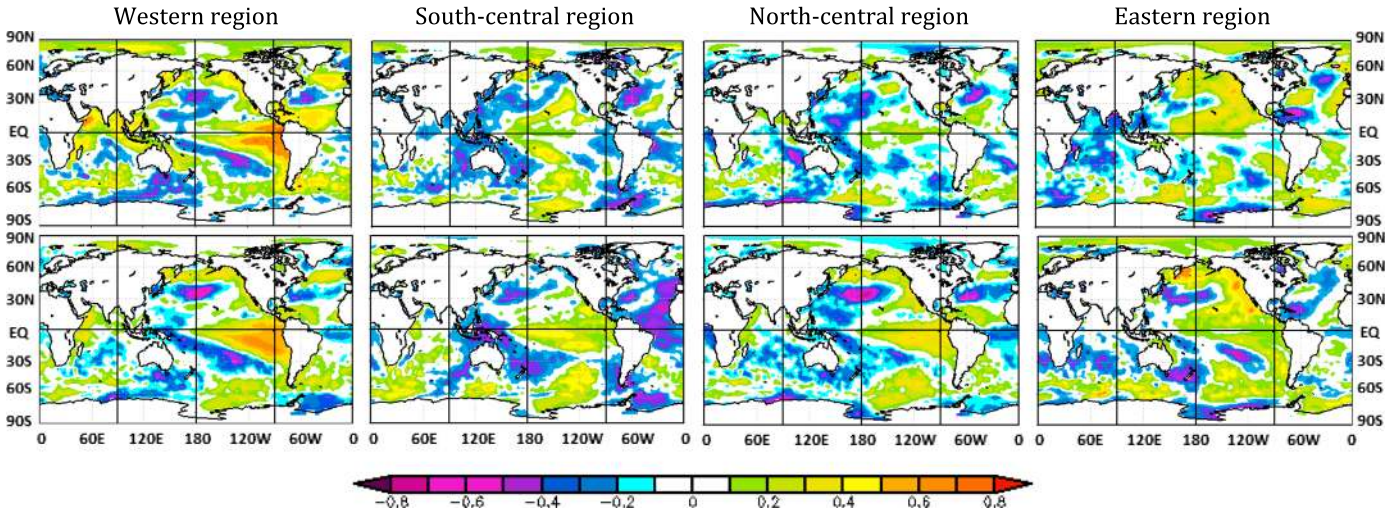

Figure 3. Concurrent correlations between global sea surface temperatures and regional mean warm season precipitation for all four regions that were derived from (top row) UDel observed and (bottom row) CFSv2 1 month lead hindcasts precipitation data sets. Correlations of each region, i.e., West, South-central, North-central, and East correspond to a column each from left to right, respectively.

distribution of climate sensitivities in CFSv2 resembles reality. Intraregional correlations are generally larger for CFSv2 than for observations (Figure 2f and Table S1), which is expected for a relatively smooth model result. Interregional correlations are low except for the correlation between West and North-central (Table S1). We keep these regions separate for this analysis in order to provide consistent comparisons with observations, but their high correlation suggests that in CFSv2 they are responsive to a similar set of large-scale climate modes.

To investigate the relationship between large-scale forcing and precipitation variability in each region, we examine correlations with concurrent global sea surface temperature (SST) patterns (Figure 3); SSTs are from NCEP Reanalysis [*Kalnay et al.*, 1996]. We find that the regions show distinct correlation patterns, with the West correlating strongly with eastern tropical Pacific SST variability, South-central correlating with western Atlantic and central Pacific variability, North-central correlating with the northern Pacific and Atlantic oceans, and the East correlating with Atlantic in a manner that contrasts with the North-central correlation pattern. We do not delve into details of physical mechanism, but focus on two points: (i) unique correlation pattern for each region—all four regions exhibited distinct correlation patterns with global SSTs and (ii) relatively similar correlation patterns between observation and CFSv2, suggesting that the model captures the general character of teleconnections for each region. We do note the somewhat overactive eastern tropical Pacific correlation for North-central and South-central in CFSv2 relative to observations and other small but systematic differences are also observed and warrant further study. Similar analysis with NOAA optimum interpolation SST v2 [*Reynolds et al.*, 2002] yielded similar results (Figure S4). In the supporting information we present correlations with NCEP Reanalysis 300 mb geopotential height (Figure S5) and time-lagged correlations with NCEP Reanalysis SSTs (Figure S6 and S7). The pattern of regional distinction and general similarity between CFSv2 and observations holds for these fields, though interesting differences between CFSv2 and observation do emerge in SST correlations at longer lead times. *Zhu et al.* [2013] present plots similar to Figure 3 for different variables for the northwestern U.S., and they appear to be broadly consistent with our findings.

Having described the general similarities between CFSv2 and observations, we turn to differences between the two regionalizations. Figure 4a maps areas of spatial disagreement between CFSv2 and UDel regions; mismatch between grid cells of the same region of CFSv2 and observations is defined as spatial disagreement and is shown in grey. Not surprisingly, these disagreements occur on borders between regions. The largest area of disagreement is in the placement of the border between East and North-central: in observations the Midwest is split between these regions, while in the model the East region extends across much of the Midwest. There is also notable nonoverlap in parts of the lower Mississippi basin, the Rocky Mountains, and a Southern Great Plains area that includes the Texas panhandle.

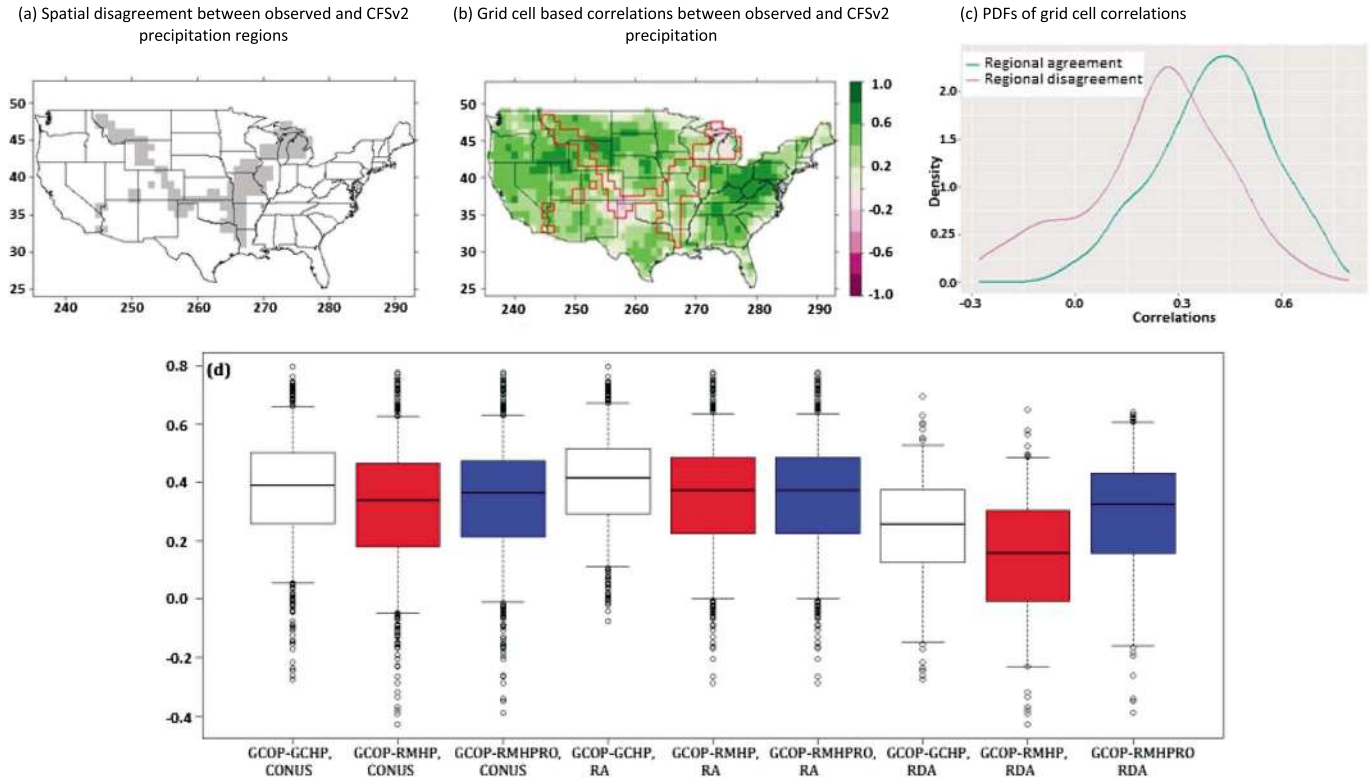

Figure 4. (a) Spatial disagreement between regions derived from UDel observed and CFSv2 1 month lead hindcast precipitation data sets; (b) map of grid cell to grid cell correlations between observed and hindcast precipitation data sets for the warm seasons overlaid with outline of region disagreement; (c) PDFs of correlations for grid cells of regional agreement (red) and grid cells of regional disagreement (green); (d) boxplot of correlations that are calculated using grid cell-observed precipitation and grid cell as well as regional mean precipitation hindcasts for all grid cells (first set of three boxplots), grid cells of regional agreement (second set of three boxplots), and grid cells of regional disagreement (third set of three boxplots). GCOP-GCHP: Grid Cell-Observed Precipitation and Grid Cell Hindcast Precipitation Correlations, GCOP-RMHP: Grid Cell-Observed Precipitation and Regional Mean Hindcast Precipitation (Region assigned based on Hindcast precipitation) Correlations, GCOP-RMHPRO: Grid Cell-Observed Precipitation and Regional Mean Hindcast Precipitation (Region assigned based on Observed precipitation) Correlations. RA, Grid cells of regional agreement; RDA, Grid cells of regional disagreement.

These differences allow us to test the hypothesis that areas in which CFSv2 regionalization does not match observations are more likely to be areas of poor model performance under standard grid-to-grid evaluation metrics. Figure 4b overlays an outline of region disagreement on the map of model correlation with observations—the same field shown in Figure 1. Visually, it does appear that zones of notably poor model performance—Texas panhandle, Upper Midwest, and parts of the lower Mississippi basin—lie in areas of regional disagreement. This pattern is tested visually by comparing the PDFs of correlations for grid cells of regional agreement and grid cells of regional disagreement (Figure 4c) and statistically by performing a *t* test on two sets of correlations that correspond to regional agreement and regional disagreement. We find that the two data sets are significantly different ($p<0.05$), and grid cell to grid cell correlations between CFSv2 1 month forecasts and observations are lower for grid cells of regional disagreement than they are for grid cells in areas of regional agreement and for the all-CONUS average (white boxes in Figure 4d). Resorting to regional mean correlations—i.e., averaging all grid cells within the region as defined by the model, as is often done when presenting seasonal forecasts over broadly defined regions of interest—does not improve performance. In fact, this averaging has no significant effect on correlations for CONUS on the whole or for grid cells that lie in areas of regional agreement, and it degrades performance for the areas of nonagreement (red boxes in Figure 4d). Correlation for areas of regional disagreement does improve somewhat, however, when grid cells are predicted using the regional mean as defined by the *observed* regionalization (rightmost three boxplots in Figure 4d)—for example, when grid cells in Missouri are predicted on the basis of CFSv2 forecast for North-central, even though CFSv2 places Missouri in the East region. This suggests that regionalization could be applied to interpret and, in some cases, to improve seasonal forecast performance in regions where the model fails to reproduce observed regional associations of climate variability.

5. Conclusions

Objective climate regionalization is a valuable but underutilized method for understanding variability in observations and models. Here we applied an established regionalization method to examine the spatial distribution of interannual warm season precipitation variability in the CFSv2 hindcast archive relative to a gridded observation-based precipitation analysis for the CONUS. This yielded insights relevant to the interpretation and application of CFSv2. First, there is general agreement in regions and associated large-scale SST patterns between CFSv2 1 month forecasts and observations. This suggests that at 1 month lead time the model does capture the spatial character of CONUS climate variability and relevant teleconnections, though model performance drops off at longer forecast horizons. Second, there are areas at the border between regions for which CFSv2 regionalization does not match observations, and model performance in these areas is systematically worse than its performance in areas of regional agreement. This supports the expectation that a GCM-based forecast system depends on the model's ability to connect local conditions to the correct set of large-scale forcings. It also argues for caution when applying CFSv2 in areas of regional disagreement, and it suggests that a similar analysis could be useful when evaluating and applying any gridded, GCM-based forecast system. Third, we find that there is some potential to make clever use of CFSv2 output to improve predictions for areas in which low performance is associated with disagreement in regionalization. For example, in some cases the regional mean time series—as defined by the region in observations rather than in the model—yields higher correlation with observations than grid-to-grid correlations or regional average correlations based on the model's regions. Thus, regionalization aids in interpreting climatic variability in observations and models, and it suggests areas for model improvement and parameters for effective utilization of a model given its deficiencies. Previous studies have demonstrated the scientific benefits of philosophically similar approaches [e.g., *Koster et al.*, 2008; *Roundy et al.*, 2015]. Formalizing and operationalizing such approaches would require additional research.

The study also raises questions about CFSv2. For example, why is it that the model pushes the East climate region too far into the Midwest and Plains states, and can that tendency be corrected through improved parameterization or initialization? Does the remarkably high correlation between West and North-central regions in CFSv2 relative to observation indicate that the model is missing a process that distinguishes these regions? Does the intrusion of the West region into the Texas Panhandle in CFSv2 indicate that the Rocky Mountains are not a sufficient barrier in the model? Regionalization on its own cannot address these questions, but it can help to identify them and to diagnose model realism as the development of the system continues. This includes evaluation of new dynamics and incorporation of corresponding parameterizations in the model [e.g., *Chowdary et al.*, 2015; *Jiang et al.*, 2013]. Correction of model in the first month most likely decreases propagation of errors over the time and consequently increases model performance for longer forecast horizons.

Acknowledgments

This work was supported by NASA Applied Sciences Program award NNX15AD50G. CFSv2 monthly precipitation hindcasts downloadable from the International Research Institute (IRI) map room (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/.MONTHLY/.prec/dataselection.html), whereas UDel and NLDAS precipitation data sets are available at http://climate.geog.udel.edu/~climate/ and http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php, respectively. NCEP Reanalysis Derived data and NOAA_OI_SST_V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/ and spatial correlation plots developed from the Web site at http://www.esrl.noaa.gov/psd/data/correlation/.

References

Badr, H. S., B. F. Zaitchik, and A. K. Dezfuli (2014), HiClimR: Hierarchical climate regionalization, Comprehensive R Archive Network (CRAN). [Available at http://cran.r-project.org/package=HiClimR.]

Badr, H. S., B. F. Zaitchik, and A. K. Dezfuli (2015), A tool for hierarchical climate regionalization, *Earth Sci. Inf.*, 1–10.

Caldwell, P. (2010), California wintertime precipitation bias in regional and global climate models, *J. Appl. Meteorol. Climatol.*, *49*(10), 2147–2158.

Choi, W., C. H. Ho, J. Kim, H. S. Kim, S. Feng, and K. Kang (2016), A track pattern-based seasonal prediction of tropical cyclone activity over the North Atlantic, *J. Clim.*, *29*(2), 481–494.

Chowdary, J. S., A. Parekh, S. Ojha, C. Gnanaseelan, and R. Kakatkar (2015), Impact of upper ocean processes and air–sea fluxes on seasonal SST biases over the tropical Indian Ocean in the NCEP Climate Forecasting System, *Int. J. Climatol.*, doi:10.1002/joc.4336.

Demargne, J., et al. (2014), The science of NOAA's operational hydrologic ensemble forecast service, *Bull. Am. Meteorol. Soc.*, *95*(1), 79–98.

Dutton, J. A., R. P. James, and J. D. Ross (2014), A probabilistic view of weather, climate, and the energy industry, in *Weather Matters for Energy*, pp. 353–378, Springer, New York.

Elguindi, N., and A. Grundstein (2013), An integrated approach to assessing 21st century climate change over the contiguous US using the NARCCAP RCM output, *Clim. Change*, *117*(4), 809–827.

Fovell, R. G., and M. Y. C. Fovell (1993), Climate zones of the conterminous United States defined using cluster analysis, *J. Clim.*, *6*(11), 2103–2135.

Fraley, C., and A. E. Raftery (1998), How many clusters? Which clustering method? Answers via model-based cluster analysis, *Comput. J.*, *41*, 578–588.

Georgakakos, K. P., N. E. Graham, T. M. Modrick, M. J. Murphy, E. Shamir, C. R. Spencer, and J. A. Sperfslage (2014), Evaluation of real-time hydrometeorological ensemble prediction on hydrologic scales in Northern California, *J. Hydrol.*, *519*, 2978–3000.

Jain, A. K., M. N. Murty, and P. J. Flynn (1999), Data clustering: A review, *ACM Comput. Surv.*, *31*(3), 264–323.

Jiang, X., S. Yang, Y. Li, A. Kumar, W. Wang, and Z. Gao (2013), Dynamical prediction of the East Asian winter monsoon by the NCEP Climate Forecast System, *J. Geophys. Res. Atmos.*, *118*, 1312–1328, doi:10.1002/jgrd.50193.
Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–470.
Kim, H. M., E. K. Chang, and M. Zhang (2015), Statistical–dynamical seasonal forecast for tropical cyclones affecting New York state, *Weather Forecasting*, *30*(2), 295–307.
Kirtman, B. P., et al. (2014), The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601.
Koster, R. D., T. L. Bell, R. H. Reichle, M. J. Suarez, and S. D. Schubert (2008), Using observed spatial correlation structures to increase the skill of subseasonal forecasts, *Mon. Weather Rev.*, *136*, 1923–1939, doi:10.1175/2007MWR2255.1.
Kumar, S., P. A. Dirmeyer, and J. L. Kinter III (2014), Usefulness of ensemble forecasts from NCEP Climate Forecast System in sub-seasonal to intra-annual forecasting, *Geophys. Res. Lett.*, *41*, 3586–3593, doi:10.1002/2014GL059586.
Lang, Y., A. Ye, W. Gong, C. Miao, Z. Di, J. Xu, Y. Liu, L. Luo, and Q. Duan (2014), Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China, *J. Hydrometeorol.*, *15*(4), 1546–1559.
Ma, F., A. Ye, X. Deng, Z. Zhou, X. Liu, Q. Duan, J. Xu, C. Miao, Z. Di, and W. Gong (2015), Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China, *Int. J. Climatol.*, doi:10.1002/joc.4333.
Matsuura, K., and C. Willmott (2012), Terrestrial precipitation: 1900–2010 gridded monthly time series (1900–2010) (v. 3.01 added 6/14/12), Univ. of Delaware.
Mo, K. C., S. Shukla, D. P. Lettenmaier, and L.-C. Chen (2012), Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction?, *Geophys. Res. Lett.*, *39*, L23703, doi:10.1029/2012GL053598.
Murtagh, F. (1983), A survey of recent advances in hierarchical clustering algorithms, *Comput. J.*, *26*, 354–359.
R Core Team (2016), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [Available at https://www.R-project.org/.]
Rao, A. R., and V. V. Srinivas (2006), Regionalization of watersheds by hybrid-cluster analysis, *J. Hydrol.*, *318*(1), 37–56.
Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim.*, *15*, 1609–1625.
Roundy, J. K., X. Yuan, J. Schaake, and E. F. Wood (2015), A framework for diagnosing seasonal prediction through canonical event analysis, *Mon. Weather Rev.*, *143*, 2404–2418.
Saha, S., et al. (2014), The NCEP climate forecast system version 2, *J. Clim.*, *27*(6), 2185–2208.
Sharma, S., P. Srivastava, X. Fang, and L. Kalin (2015), Long-range hydrologic forecasting in El Niño Southern Oscillation-affected coastal watersheds: Comparison of climate model and weather generator approach, *J. Hydrol. Eng.*, 06015006.
Sheffield, J., et al. (2014), A drought monitoring and forecasting system for sub-Sahara African water resources and food security, *Bull. Am. Meteorol. Soc.*, *95*(6), 861–882.
Sheridan, S. C. (2002), The redevelopment of a weather-type classification scheme for North America, *Int. J. Climatol.*, *22*, 51–68, doi:10.1002/joc.709.
Siegmund, J., J. Bliefernicht, P. Laux, and H. Kunstmann (2015), Toward a seasonal precipitation prediction system for West Africa: Performance of CFSv2 and high-resolution dynamical downscaling, *J. Geophys. Res. Atmos.*, *120*, 7316–7339, doi:10.1002/2014JD022692.
Silva, G. A., L. M. Dutra, R. P. da Rocha, T. Ambrizzi, and É. Leiva (2014), Preliminary analysis on the global features of the NCEP CFSv2 seasonal hindcasts, *Adv. Meteorol.*, *2014*, 695067.
Stock, C. A., et al. (2015), Seasonal sea surface temperature anomaly prediction for coastal ecosystems, *Prog. Oceanogr.*, *137*, 219–236.
Swapna, P., M. K. Roxy, K. Aparna, K. Kulkarni, A. G. Prajeesh, K. Ashok, R. Krishnan, S. Moorthi, A. Kumar, and B. N. Goswami (2014), The IITM Earth system model: Transformation of a seasonal prediction model to a long term climate model, *Bull. Am. Meteorol. Soc.*, *96*, 1351–1367.
Tian, D., C. J. Martinez, and W. D. Graham (2014), Seasonal prediction of regional reference evapotranspiration based on Climate Forecast System version 2, *J. Hydrometeorol.*, *15*(3), 1166–1188.
Wang, S. Y., R. R. Gillies, E. S. Takle, and W. J. Gutowski (2009), Evaluation of precipitation in the Intermountain Region as simulated by the NARCCAP regional climate models, *Geophys. Res. Lett.*, *36*, L11704, doi:10.1029/2009GL037930.
Ward, J. H., Jr. (1963), Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, *58*, 236–244.
Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.
Yoon, J. H., K. Mo, and E. F. Wood (2012), Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States, *J. Hydrometeorol.*, *13*(2), 463–482.
Yuan, X., E. F. Wood, L. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophys. Res. Lett.*, *38*, L13402, doi:10.1029/2011GL047792.
Yuan, X., E. F. Wood, J. K. Roundy, and M. Pan (2013), CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States, *J. Clim.*, *26*(13), 4828–4847.
Yuan, X., E. F. Wood, and M. Liang (2014), Integrating weather and climate prediction: Toward seamless hydrologic forecasting, *Geophys. Res. Lett.*, *41*, 5891–5896, doi:10.1002/2014GL061076.
Yuan, X., J. K. Roundy, E. F. Wood, and J. Sheffield (2015), Seasonal forecasting of global hydrologic extremes: System development and evaluation over GEWEX basins, *Bull. Am. Meteorol. Soc.*, *96*, 1895–1912.
Zhan, R., and Y. Wang (2016), CFSv2-based statistical prediction for seasonal accumulated cyclone energy (ACE) over the western North Pacific, *J. Clim.*, *29*(2), 525–541.
Zhu, J., B. Huang, Z. Z. Hu, J. L. Kinter III, and L. Marx (2013), Predicting US summer precipitation using NCEP Climate Forecast System version 2 initialized by multiple ocean analyses, *Clim. Dyn.*, *41*(7–8), 1941–1954.
Zuo, Z., S. Yang, R. Zhang, D. Xiao, D. Guo, and L. Ma (2015), Response of summer rainfall over China to spring snow anomalies over Siberia in the NCEP CFSv2 reforecast, *Q. J. R. Meteorol. Soc.*, *141*, 939–944, doi:10.1002/qj.2413.