# High resolution decadal precipitation predictions over the continental United States for impacts assessment

KAUSTUBH SALVI[1], GABRIELE VILLARINI[1], GABRIEL A. VECCHI[2*]

[1]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa, USA

[2]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA

[*] Current affiliation: Geosciences Department, Princeton University, Princeton, NJ, USA

Manuscript submitted to

*Journal of Hydrology*

Revised 6th Jun 2017

*Corresponding author:*

Kaustubh Salvi, IIHR-Hydroscience & Engineering, The University of Iowa, 100 C. Maxwell Stanley Hydraulics Laboratory, Iowa City, 52242, Iowa, USA. E-mail: kaustubhanil-salvi@uiowa.edu

## Abstract

Unprecedented alterations in precipitation characteristics over the last century and especially in the last two decades have posed serious socio-economic problems to society in terms of hydro-meteorological extremes, in particular flooding and droughts. The origin of these alterations has its roots in changing climatic conditions; however, its threatening implications can only be dealt with through meticulous planning that is based on realistic and skillful decadal precipitation predictions (DPPs). Skillful DPPs represent a very challenging prospect because of the complexities associated with precipitation predictions. Because of the limited skill and coarse spatial resolution, the DPPs provided by General Circulation Models (GCMs) fail to be directly applicable for impact assessment. Here, we focus on nine GCMs and quantify the seasonally and regionally averaged skill in DPPs over the continental United States. We address the problems pertaining to the limited skill and resolution by applying linear and kernel regression-based statistical downscaling approaches. For both the approaches, statistical relationships established over the calibration period (1961-1990) are applied to the retrospective and near future decadal predictions by GCMs to obtain DPPs at ~4km resolution. The skill is quantified across different metrics that evaluate potential skill, biases, long-term statistical properties, and uncertainty. Both the statistical approaches show improvements with respect to the raw GCM data, particularly in terms of the long-term statistical properties and uncertainty, irrespective of lead time. The outcome of the study is monthly DPPs from nine GCMs with 4-km spatial resolution, which can be used as a key input for impacts assessments.

**Keywords:** Decadal predictions, continental United States, statistical downscaling, precipitation

## 1. Introduction

Water is the most abundant natural resource that manifests itself in diverse states and forms such as precipitation, snowfall, surface and ground water, rivers, lakes, oceans. Despite such a wide availability, most of our activities pertaining to water are constrained because of a lower percentage of available fresh water, upcoming problems concerning water usage, and even the solutions adopted to negotiate the problems. For instance, to address elevated food requirements of the rising population (population more than quadrupled in 100 years), measures such as doubled crop land area, six-fold increase in irrigated area (*Freydank and Siebert*, 2008) are implemented, leading to rise in global water use (i.e., withdrawal) by nearly 8 times with a steep increase at a rate of 15% per decade between 1960 and 2010. Such measures have put additional constraints on the available fresh water resources and the exacerbating situations demand careful planning and mastering of the available water resources. However, design and implementation of water resources planning measures is a difficult task because precipitation represents a key input. Precipitation is the most important and equally complex climate variable to understand and foresee mainly because of extreme variability, revealed by precipitation patterns at different spatio-temporal resolutions. The variability in precipitation might lead to high-intensity disasters such as flooding and mudslides (*Trenberth et al*., 2007), or lower-intensity, longer-duration events such as droughts (e.g., *Gray*, 2009; *Henry et al*., 2004; *Hunter et al*., 2011). Such catastrophes not only lead to heavy infrastructure damages and pose a serious threat to life but also lead to sensitive issues such as migration, which has been a common strategy to avert the consequences of weather events and/or a changing climate (e.g., *Nawrotzki et al*., 2012, *Kniveton et al*., 2008; *McLeman and Smit*, 2006).

The background discussed up to this point highlights the importance of precipitation and the associated problems, and leads us to concluding that such problems can be mitigated only through robust planning policies. The success of such measures would be highly dependent on the knowledge of future precipitation, which itself represents an extremely challenging task. Our current best knowledge of future precipitation and the associated uncertainties can be obtained from projections from General Circulation models (GCMs) produced in support of the fifth generation of Coupled Model Intercomparison Project (CMIP5) framework. This includes results from a standard set of experiments by state-of-the-art climate models developed by different teams of experts worldwide. This "ensemble of opportunity" (commonly used terminology to represent the suite of GCM data; e.g., *Sanderson et al.*, 2015) is the best available source of information on the range of physically plausible climate evolutions over the next century. Representative Concentration Pathways (RCPs) from the CMIP5 suite provide us with future projections of climate variables as a possible response to different radiative forcings. These projections solve our problem to a certain extent and provide us with the climate projections as a possible response to these forcings. However, there is a three-fold problem associated with the use of GCM projections in the planning process: (1) the RCP scenarios are driven by end-conditions, where the trajectories of different climate variables are plausible responses to the corresponding radiative forcing at the end of the 21st century. Such projections form a better way of understanding the long-term future of climate variables, but they cannot be used for short-term planning policies. This is because realistic weather conditions that we encounter in our day-to-day life are driven by their initial rather than their boundary conditions. Also, it is established that long-term behaviors of climate variables may differ from their short term properties (e.g., *Cane*, 2010; *van Oldenborgh et al*., 2012). (2) A complex variable such as precipitation is difficult to predict and the skill exhibited by the GCM in projecting them is limited. (3) The coarse spatial resolution at which the GCM projections are available cannot be used for impacts assessment over limited areas. CMIP5 simulations themselves tend to resolve the first problem by providing decadal scale forecasts that are initialized every year or intermittently to provide predictions for the next 10-30 years. These predictions are known as decadal predictions, which range from 10 to 30 year lead times (*Meehl et al*., 2009) and are forced with observed initial and boundary conditions. Initialized decadal predictions, while still in their infancy, have been subject to increasing attention by the scientific community. These are called upon by different names, including "a new kid on the block," "the fascinating baby that all wish to talk about" (*Goddard et al*., 2012), "high profile predictions" (*Goddard et al*., 2013). Several scientists are currently working on evaluating: (i) different aspects of this new product such as initialization strategies (*Meehl et al*., 2014), assessment of skill in decadal predictions (*Meehl et al*., 2009), gap between decadal climate predictability and predictions, impacts of initialization (*Branstator and Teng*, 2012), opportunities and challenges, importance of ocean observations for successful decadal predictions (*Hurrell et al*., 2009), positive phase of the inter-decadal Pacific oscillations (*Meehl et al*., 2016), role of sea ice, land surface, stratosphere, and aerosols in decadal-scale predictability (*Bellucci et al*., 2015a), and comparison between initialized and non-initialized predictions (*Fyfe et al*., 2011); (ii) performances of individual GCM with respect to the decadal predictions e.g. Flexible Global Ocean-Atmosphere-Land System model, Grid-point Version 2 (FGOALS-g2) (*Bin et al*., 2012), coupled Earth System model of the Max Planck Institute for Meteorology (MPI-ESM) (*Muller et al*., 2012), MIROC4h and MIROC5 (*Mochizuki et al*., 2011); and (iii) the skill of decadal predictions for different climate variables and quantities, including surface temperature (*Smith et al*., 2007; *Choi et al*., 2016; *Salvi et al*. 2017), Sahelian precipitation (*Gaetani and Mohino*, 2013), hurricane activity (*Smith et al*., 2010; *Vecchi et al*., 2013), and regional surface climate (*Kim et al*., 2012; *van Oldenborgh*, 2012; *Doblas-Reyes et* *al*., 2013; *Goddard et al*., 2013; *Caron et al*., 2014; *Meehl et al*., 2014; *Bellucci et al*., 2015b). However, despite recent progress, we still have to comprehensively quantify the quality of GCM precipitation predictions, including over the United States; moreover, even though CMIP5 simulations have outperformed CMIP3 in terms of spatial resolution, the resolution is still not high enough for impacts assessment. With this background, the goal of this work is to obtain high-resolution precipitation decadal predictions over the continental United States (CONUS).

The specific objectives of this study involve: 1) the evaluation of raw GCM precipitation skill over CONUS; 2) the enhancement of the skill using two data-driven approaches; and 3) the development of a dataset of decadal predictions of precipitation over CONUS at a spatial resolution of ~4 km. The two data-driven approaches that are implemented in this study are transfer function based statistical downscaling (SD) methodologies. Transfer function based downscaling methodologies rely upon establishing statistical relationships between coarse resolution climate variables (predictors) that are relatively well understood in terms of underlying physics and hence, well simulated, and the fine resolution climate variable of interest that needs to be downscaled, i.e. predictand (precipitation in this study). The established relationships are applied to GCM predictions to obtain downscaled data. We use two SD approaches: (1) linear regression (LR) based, which assumes a parametric form of the relationship, with the variation between predictors and predictand that is assumed to be linear; and (2) kernel regression (KR) based, which is a non-parametric form of regression. The premise behind the selection of two methodologies that are categorized under the same cluster (transfer function approach) is explained in section 3. These methodologies are applied to the predictors by nine GCMs to obtain improved decadal precipitation predictions at fine spatial resolution (~4 km). The manuscript is organized as follows. Details about the study region and the data used are in Section 2. Section 3 provides details about the data-driven approaches, which are applied here to enhance the prediction skills and different evaluation metrics. The results of the evaluation of raw and processed GCMs' skill are discussed in Section 4. We conclude this study with summary and discussion in Section 5, followed by concluding remarks.

## 2. Study region and data

Figure 1 shows the spatial variations of precipitation over the study region for the period 1961-2014, obtained using the Parameter–Elevation Regressions on Independent Slopes Model (PRISM) data product (additional information about the PRISM data product can be found in Section 2.1). The figure also provides the details about the seven regions we focus on: Northeast (NE), Midwest (MW), Southeast (SE), Great Plains North (GPN), Great Plains South (GPS), Northwest (NW), and Southwest (SW). These regional delineations are largely based on the National Climate Assessment Report (*Karl et al.*, 2009) and then implemented in different studies (e.g., *Pryor and Schoof*, 2008). Similar to the recent studies, which use slightly modified delineations (*Schoof et al.*, 2010; *Kunkel et al.*, 2013; *Mutiibwa et al.*, 2015), we divide the Great Plains into two regions GPN and GPS. The division of the Great Plains into two visually homogeneous regions (homogeneous in terms of spatial variation of average precipitation) is an additional advantage for the improvement in performance of SD. In this study, we focus on the seven regions in Figure 1 and consider each one as an individual unit for: 1) applying statistical downscaling methodologies and 2) quantifying prediction skill in regionally and seasonally averaged spatio-temporal framework. Based on Figure 1, it is clear that the spatial distribution of the average precipitation over CONUS shows extreme variability, in particular in areas, where precipitation is mainly influenced by orography. This large spatial variability distinguishes NW and SW regions from the others, where a relatively smooth transition is encountered. We envision capturing the spatial variability through downscaling methodologies to prove the worthiness of the downscaled precipitation for impacts assessment.

### 2.1 Observed precipitation data

We select PRISM precipitation dataset at monthly temporal and ~4km spatial resolution as the baseline dataset. The PRISM precipitation data is used in the establishment of the statistical relationships and also as the reference data for the evaluation of the decadal precipitation predictions skills. PRISM (*Daly et al*., 2002; *Hijmans et al*., 2005) represents high-resolution climate observations that are derived from a wide range of monitoring networks to illustrate short- and long-term climate patterns. Point observations (precipitation measurements), obtained from the monitoring networks, a digital elevation model (DEM), and other spatial datasets to generate gridded estimates of annual, monthly and event-based climatic parameters (*Daly et al*., 1994) are utilized to generate the high-resolution dataset. Here, we use the monthly PRISM precipitation product from 1961 to 2014 over CONUS. There are specific limitations of the dataset such as the problem with the choice of the domains for the regressions (*Widmann and Bretherton*, 1999) and temporal discontinuities because of heterogeneous station network (*UCAR*, 2016). However, the above mentioned positive features of this dataset (e.g. the high spatial resolution for impacts analysis, and positive performance in regions of complex terrain) make PRISM an appropriate product for our study.

### 2.2 Reanalysis data

We use two SD approaches to obtain high-resolution predictions. The success of the data-driven approaches such as SD is highly dependent on the choice and quality of the predictors. Usually, standard reanalysis data products are validated for their quality and hence, used as sources of predictors. Reanalysis data are gridded data for climate variables over the entire globe, derived by applying data assimilation techniques over past data, including global radiosonde data, a Comprehensive Ocean-Atmosphere Data Set (COADS) that comprises a collection of surface marine data, aircraft data, surface land synoptic data, satellite sounder data, Special Sensing Microwave/Imager surface wind speeds, and satellite cloud drift winds. Here, we use reanalysis data for establishing statistical relationships with the predictand and to remove systematic errors from the GCM predictions. Different reanalysis data products are available at different spatial and temporal scales, including the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA40; *Uppala et al*., 2005), ERA-Interim (*Dee et al*., 2011), National Center for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR; *Kalnay et al*., 1996), and the Japanese 25-year ReAnalysis (JRA-25; *Onogi et al*., 2007), and JRA-55 (*Ebita et al*., 2011; *Kobayashi et al*., 2015) from the Japan Meteorological Agency. Statistical relationships are better established if a sufficient sample size is available: NCEP/NCAR reanalysis provides longer records of reanalysis climate variable data compared to other available products, with positive assets such as fixed state-of-the-art assimilation scheme, inclusion of more observations, better quality control (*Bromwich and Fogt*, 2004), better accessibility and spatial coverage over the regions where observed data are not available, and better categorization of climate variables based on the influences of observations over the climate variable. 'Categorization of climate variable' is an important feature of this product, allowing the users to identify each reanalysis climate variable with a category such as

'A', 'B', and 'C' (*Kalnay et al*., 1996), where category 'A' variables show higher accuracy compared to the other ones. As mentioned before, the success of SD also relies upon the choice of predictors, which represent synoptic scale circulation patterns over a study region, should be well simulated by climate models and should be associated with precipitation processes (*Wilby et al*., 2004). The categorical distinctions of predictors in the case of NCEP/NCAR data serve as a guideline to choose appropriate climate variables as predictors. Here, we consider four surface level climate variables, such as mean sea level pressure (psl), temperature (tas), near surface zonal (uas) wind, near surface meridional (vas) wind, and five pressure level climate variables (at 500hPa), including specific humidity (hus), geopotential height (zg), temperature (ta), Uwind (ua), and Vwind (va). Most of the predictors that we select belong to the category 'A'.

## 2.3 GCMs used

As a part of the CMIP5 program, a large number of GCMs provide the monthly decadal prediction of precipitation and are initialized either each year or intermittently. We use a set of nine GCMs for the analysis and Figure 2 illustrates three important details about the corresponding GCM data, i.e. the name and spatial resolution, the years of initialization, and the availability of the predictions based on different lead times. The lead time, which represents the time lag between the year of the initialization and the year of the prediction, varies from 1 to 30 years. However, we restrict our analysis to the first 10 annual lead times mainly because of two reasons. First, the availability of the decadal predictions for longer lead times is too small to provide a conclusive estimate of skill. Secondly, the influence of initial conditions, with which the models are forced, reduces with lead times (*Meehl et al*., 2009). It is important to note that the initialization patterns differ across the GCMs. Hence, we cannot get the same number of

GCMs contributing to the ensemble for a particular initialization. The bottom-right panel in Figure 2 summarizes the number of GCMs available for each and lead time; for instance, certain initializations (e.g. 1961, 1965) have their contributions from all the GCMs, whereas there is only one GCM which provides the predictions that are initialized in the year 2013-2015. It is clear that the presence of a higher number of GCMs provides additional confidence in the predictions, which we lack for the most recent years based on the CMIP5 models.

## 3. Methodology and skill assessment

The SD methodologies that we use can be categorized under the broad umbrella of the transfer function approaches (e.g., *Ghosh and Mujumdar*, 2006; *Kannan and Ghosh*, 2013; *Srinivas et al*, 2014). We use LR and KR based SD models to establish statistical relationships between predictors and predictand. These methodologies are applied to each season separately. For both the approaches, the overall methodology remains the same. The two approaches differ from each other over the models that we use to link coarse level climate variables with the fine resolution predictand. The reason behind the selection of two approaches lies in the complexity of the climate variable that we intend to capture (precipitation). Precipitation is a highly complex climate variable and there is a possibility that different attributes of precipitations that are important in the planning process such as total rainfall, seasonal variations, extremes, may not be captured by a single SD approach. Hence, we use two transfer function approaches. Two datasets originating from the two approaches are validated thoroughly to specify the positive attributes of each product, which would help in planning. These methodologies are well established and have been applied successfully, for instance, for rainfall projections with kernel regression approach (e.g., *Kannan and Ghosh*, 2013; *Salvi et al*., 2013) and monthly rainfall projections with linear regression approach (e.g., *Kannan et al*., 2014; *Shashikanth et al*., 2014). Here, we briefly describe the specific steps that are carried out while applying the methodologies, and additional details are provided in the supplementary material and in the literature (*Kannan and Ghosh*, 2013; *Kannan et al*., 2014). We select 30 years (1961-1990) as calibration period and 24 years (1991-2014) as validation period. We justify the selection of 1961-1990 as calibration period by carrying out sensitivity analysis. We select three different time slices as calibration periods (each of 30 years) and obtain the predictions for the validation period. Based on this sensitivity analysis, the skills for different seasons and regions vary with calibration period and we cannot identify a particular calibration period which provides the best results in all the cases. Hence, we select 1961-1990 as the calibration period for this study. The details can be found in the supplementary material (Sections S3, S-Figure 20). The downscaling approach begins with the identification of 'zone of predictors' (e.g., *Salvi et al*., 2013, 2015), which represents an imaginary area (usually rectangular or square in shape) that surrounds the study region. It is assumed that the predictors in the zone of predictors influence the precipitation in the corresponding region. Because we work with seven regions, we have seven zones of predictors. S-Figures 1-9 show the spatial extent of each zone of predictors, obtained using a correlation analysis approach (e.g., *Kannan and Ghosh*, 2010; *Salvi et al*., 2013). After the selection of the zone of the predictors, we proceed to the establishment of the statistical relationship between reanalysis climate variables over the zone of predictor and individual precipitation at each pixel in the region. To deal with the multicollinearity (i.e., correlation among predictors) and multidimensionality, we select and apply principal component analysis (PCA) as a dimensionality reduction technique. PCA is a form of orthogonal transformation that converts the correlated predictors into non-correlated principal components (PC) and arranges them in descending order of variability explained by each one. There is no single criterion about the selection of the number of PCs which would need to be retained in establishing the statistical relationship. A number of methods have been proposed in the literature to select the number of PCs, including the selection of PCs with eigen value greater than unity (*Kaiser*, 1960), Akaike information criterion (*Akaike*, 1974), cumulative percent variance (*Malinowski*, 1991), and variance of the reconstruction error (*Qin*, 1998; *Valle et al*., 1999). In this study, we mainly use the "percentage of variance explained" criterion to select the number of PCs, which involves the selection of the first few PCs for model calibration/validation depending upon the desired level of variability retained. As the 30-year calibration period provides us with a sample size of 90 (30 years × 3 months per season), we select the first three PCs of each predictor, leading to a total of 27 PCs (9 predictors × 3 PCs). For LR, we put an additional filter over these 27 PCs and select only those PCs (out of 27) which show a statistically significant relationship with the observed precipitation at each pixel. S-Figure 10 shows that on average between 3 and 10 PCs in a 95% confidence interval are used to establish the statistical relationship after the application of the second filter. The established statistical relationship is assumed to be time-invariant and applied to obtain: 1) precipitation with NCEP/NCAR reanalysis data for the validation period, and 2) decadal scale climate predictions with bias corrected GCM predictions. The details about the methodology can be found in the supplementary material (Sections S1-2).

Three sets of results (i.e., ensemble raw GCM predictions and ensemble GCM predictions after statistical downscaling with both the approaches) are evaluated with respect to different metrics. The performance of raw GCM predictions and the downscaled precipitations (with LR and KR) are evaluated over the validation period (1991-2014). These assessments are carried out over 'lead-wise' seasonally and regionally averaged time series for both the raw and the downscaled precipitation and we refer to this as the 'spatio-temporally averaged framework.' The regional delineations in Figure 1 are used to obtain spatially averaged time series to represent the respective regions. To obtain seasonal averages, we choose four seasons: spring (MAM), summer (JJA), fall (SON), and winter (DJF) and the complete year (YEAR). This results in the representative time series with 24 data points (one representative value for each season over the validation period of 24 years). This is different from some of the studies (*Singh et al*., 2012; *Gaetani and Mohino*, 2013; *Tiwari et al*., 2014), which retained monthly values even in the seasonal time series, resulting in 72 data points for each season (24 years × 3 months per year): one of the major limitations of such approach is the enhanced correlation due to the capturing of the seasonality of precipitation (e.g., November is wetter than October which is also wetter than September) rather than the correct values, potentially leading to an erroneous impression about the skill possessed by the predictions. Hence, we retain the prior validation approach (one representative value per region per season per year). This approach provides a much more direct information about the skill of the models (whether the raw GCM or after the statistical downscaling) for each season without having to rely on the climatology to achieve good skill. The evaluation of regionally and seasonally averaged time series is carried out considering three metrics. Each metric represents a specific quality that we intend to quantify for decadal predictions. The first metric represents a linear association between observation and predictions known as skill score (SS). *Murphy and Winkler* (1992) proposed a decomposed version of the skill score in terms of potential skill and biases (conditional and unconditional) as shown in equation 1:

$$SS = \rho_{fx}^2 - \left[\rho_{fx} - \left(\frac{\sigma_f}{\sigma_x}\right)\right]^2 - \left[\frac{(\mu_f - \mu_x)}{\sigma_x}\right]^2 \quad (1)$$

where, $\rho_{fx}$ is the correlation between predictions and observations; $\sigma_f$ and $\sigma_x$ are the standard deviation, while $\mu_f$ and $\mu_x$ are the mean of the predictions and the observations, respectively. The first term of the decomposition (right-hand side) is the potential skill (PS) of the forecasts without any bias. The second term is known as the slope reliability (SREL) and is a measure of the conditional bias. The third term is a standardized mean error (SME), a measure of the unconditional bias (*Hashino et al*., 2007). We refer to this correlation-based metric as "Memory-based" metric (*Dirmeyer et al*., 2016). The second metric represents the skill pertaining to the long term statistical properties such as mean, standard deviation (STDEV), extremes ($95^{th}$ percentile), and root mean square error (RMSE), which we refer to as "Long-term statistics based" metrics. The third metric represents the quantification and visual assessment of inter-model uncertainty, obtained in terms of the envelope of nine GCM predictions, and we refer to this as "Uncertainty-based" metric. We also compare the statistically downscaled GCM predictions with both approaches on a decadal scale as well: instead of comparing the predictions lead-wise for 1991-2014, we compare them decade-wise e.g. 1961-1970, 1961-1971. This is the format in which GCMs provide the data and our motive in comparing the two products in this manner is to identify the skill metrics which retain the consistency across different initializations. We call this comparison as "decadal framework" and we make these comparisons for regionally averaged time series (one time series for each region with 12 months × 10 years = 120 data points).

## 4. Results

### 4.1 Skill of the raw GCMs' predictions

Regionally and seasonally averaged raw GCM precipitation predictions (ensemble) for different lead times (in years, from 1 to 10) are evaluated for Memory-based and Long-term statistics metrics and the results are presented in Figure 3. Overall, GCM predictions are characterized by low PS values, ranging from 0 to 0.3. For most of the lead times, the PS values are as low as 0.1. For different regions, patches of slightly elevated PS values are encountered (e.g., GPS, NE, SW at the yearly scale; GPS, MW, NE, SE, SW in winter; NW in spring; NE, SW in summer; MW, NW in fall). Surprisingly, the higher PS values do not always correspond to the lead-1 predictions. This finding is contradictory to our pre-conceived belief that smaller differences between initialization and year of prediction should provide the best results, assuming the highest influence of initial conditions for shorter lead times. The absence of conditional bias, revealed in terms of lower SREL values, is the only positive aspect of the raw GCM predictions. This is consistent for all regions and seasons and across all lead times. However, the presence of unconditional bias in terms of higher SME values for most of the regions brings down the overall skills (SS). This is mainly visible for GPN, SW (over all seasons), MW and NE over winter and SE over summer. As a result, the values of the actual skill SS turn negative in magnitude. The fall season lacks unconditional biases and hence, positive SS are observed for GPS, MW, NE, and NW for certain lead times. Raw GCM predictions do not fare well when compared with observed precipitation based on long term statistical properties. The comparison of statistical properties, expressed as absolute difference between ensemble predictions and PRISM data, shows high differences for all the statistical properties for almost all regions and seasons over different lead times. In fact, high difference in mean and standard deviations is the prime cause for higher unconditional biases (SME) leading to poor skill. Except for the fall season, large differences in mean indicate that the GCMs lack the capability to capture the total amount of precipitation, and large differences in STDEV indicate their failure to capture the variability. Equally small skill values are revealed for extremes and RMSE, except for the extremes at the yearly temporal scale, where raw GCM predictions can reproduce the observations reasonably well. This analysis confirms the statement that usually GCM predictions for a climate variable such as precipitation cannot be directly used because they possess low skills.

### 4.2 Uncertainty quantification for the raw GCM predictions

The nine GCMs enable us to assess and quantify the uncertainty in precipitation predictions (Figure 4). These envelopes can be considered as a surrogate for the uncertainty in GCMs. Here, we discuss the results for lead-1, while the details about other lead times are shown in the supplementary material (S-Figures 11-19). All the plots show a wide uncertainty envelope (gray envelope) indicating that different GCM predictions, averaged over different seasons and regions differ from each other significantly. Along with the envelope, we compare the ensemble time series (shown by the blue line) with the observed data. Most of the regions show deviations between the ensemble time series and the observed precipitation data, representing biases in the predictions. One prominent feature of the observed data is high variability, which is indicated by large fluctuations (e.g. NE region in spring and summer). Ensemble raw GCM predictions fail to capture the variability present in the observed precipitation time series. The results in Figures 3-4 show that the skills pertaining to the original GCM predictions are mediocre across the different evaluation metrics used here, with limited potential skill, large unconditional biases and uncertainty.

Bias correction techniques, such as ‘Bias Correction and Spatial Disaggregation’ (BCSD) (Wood et al., 2004) and ‘Quantile mapping method’ (Li et al., 2010), have been extensively used in the literature to obtain skillful rainfall projections. However, it is important to understand that the removal of the bias requires the correction of the statistical properties/quantiles of GCMs projections to match the cumulative distribution function (CDF) of the observed data over historic (calibration) period at each pixel. It is obvious that these methods would result in “bias-free” projections with respect to the historical period; however, they cannot add any skill on their own unless the raw GCM projections are skillful. Therefore, the inherent skill in the GCM predictions represents a prerequisite for the application of such bias correction methods: bias correction methodologies should be used only for those climate variables that are reasonably well simulated by the GCMs otherwise it might lead to unrealistic predictions. Based on the results in Figure 3, the raw GCMs do not possess the skills to predict precipitation. Also, the bias correction does not involve any regional level modelling. Therefore, we choose two downscaling approaches (LR and KR based SD) in the present study, which involve regional level modelling, and our expectation is that the obtained data are more realistic. The subsequent sections provide the details of the methodical validations and the skill-evaluation exercise of the downscaled precipitation.

### 4.3 Validation of the downscaling models

Figure 5 shows the results of the validation based on Memory-based and Long-term statistics for the LR and KR-statistically downscaled products that are obtained with NCEP/NCAR forcings over the validation period (1991-2014). As discussed in section 3, the validations are illustrated in the spatio-temporally averaged framework for each season and region. Downscaled precipitation with LR shows better Memory-based skills as compared to KR. This is clear from the PS plots over different regions, where PS for LR approach are higher than those for KR for most of the seasons and regions. High PS values (above 0.5 leading to values of the correlation coefficient larger than 0.7) are found for all the regions, except for NE over different seasons. The methodology shows a very good performance for such a highly complex variable to predict. The LR is capable of removing the conditional (SREL) and unconditional biases (SME) for all the regions and seasons, resulting in positive values of SS. In fact, there are hardly any losses in the skill because of the presence of biases. The KR methodology shows a comparable performance in removing the biases. However, this model is not able to capture the season-to-season variations in the observed data, resulting in lower values of PS as compared to the LR. This is reflected in SS as well, where the KR shows negative SS for some cases (e.g., yearly SS for GPN, MW). The second set of validation metrics (i.e. Long-term statistics based validations) shows a complete reversal of skill. The comparison over variability, expressed in terms of absolute differences in standard deviation and extremes clearly shows that the KR model outperforms LR. The two models reveal similar skill when compared with respect to the absolute difference in mean and RMSE. Overall, the validation results highlight the capability of the two SD models in capturing some important properties of observed precipitation (i.e., the LR captures the season-to-season variability and mean, while the KR captures long term variability and extremes). The specific skill shown by the two methodologies can be an important guideline in selecting the downscaled product for a particular study or application.

Having established the credibility of downscaling models with the reanalysis forcings, we move ahead with the evaluation of the decadal prediction skills for precipitation obtained with the GCM forcings. Before building the statistical relationships with respect to the nine predictors, we assess the skills of the GCM forcings. We have carried out a separate analysis for each one of them (refer to section S4 of the supplementary material) and found out that the decadal predictions of the predictors are reasonably good compared to the raw precipitation skills (Figure 3), and can be included as predictors in the SD models.

### 4.4 Evaluation of the LR-based GCM predictions

Figure 6 shows the Memory-based and Long-term statistics based skill-evaluation plots for the precipitation downscaled with LR at the regionally and seasonally averaged scale. Compared to the raw GCM predictions, we do not see significant improvements in PS. We do, though, observed instances of relatively high PS across all seasons and for different regions. Similar to the raw predictions, these instances do not follow the influences of initial conditions and hence, in some cases, higher skill is observed at long lead times (e.g., NE in summer for lead period 7-8). As far as conditional and unconditional biases are concerned, there are marked improvements with respect to the raw predictions, especially for SME. The LR model is able to reduce the impact of these biases, resulting in positive SS. Overall, positive skill is observed mainly for the winter and spring seasons. The predictions obtained with LR show better linear association with the PRISM data as compared to its raw counterpart. The best part of the downscaled precipitation with LR is the skills with which the long term mean is captured. For most of the regions and seasons, the absolute difference between observed data and predictions is small (~5-7 mm/month). The same methodology fails to perform at the same level in capturing other long term statistical properties, such as variations in terms of STDEV, extremes and RMSE, leading to no significant improvements as compared to the raw predictions. Pixel-wise skills for downscaled precipitation, obtained by applying LR based downscaling approach for different lead times from 1 to 10 (years) in the form of US maps are shown in the supplementary material (from S-Figure 21 to S-Figure 30).

**4.5 Uncertainty quantification for LR-based GCM predictions**

The LR methodology showed the capability in closing the gap between PRISM and GCM predictions with respect to the mean precipitation. Here, we assess the improvements with respect to the uncertainty-based metrics, focusing on the results corresponding to the lead-1 (Figure 7), while those for the other lead times are shown in the supplementary material (S-Figures 31-39). The results for all regions and seasons lead us to two critical improvements in the LR based downscaled product as compared to the raw GCM predictions. The first improvement is in terms of reduced uncertainty. This is evident from the GCM envelopes (gray bands), which are considerably smaller than the raw predictions (Figure 4). The second improvement is the reduction in the gap between observed time series (black line) and ensemble mean predictions (blue line). This gap represents biases in the predictions and here it is completely eliminated (e.g., compare the uncertainty envelope for GPN in Figure 4 and Figure 7). Along with these improvements, Figure 7 brings out some of the shortcomings of the LR approach as well. For all regions and seasons, the blue line, representing the ensemble of all GCM predictions, is able to trace the mean trajectory of the PRISM data, even though it completely fails to capture the large fluctuations. As far as uncertainty-based metrics are concerned, they are the exact replica of the validation shown in Figure 6, where the long term mean is captured with a high accuracy and the fluctuations are too small.

**4.6 Evaluation of the KR-based GCM predictions**

We evaluate the skill of the decadal precipitation predictions obtained using the KR approach and present the results in Figure 8. The KR approach is able to slightly increase the values of PS in the predictions, even though these increases are small compared to the raw GCM predictions. For most of the regions and seasons, the PS is around 0.15, and the higher skills are not necessarily for the 1-year lead. One such example is SW region, which shows higher PS at different lead times (lead-7 for yearly, lead-8 for winter) other than lead-1. The real enhancement in the downscaled product is observed in the complete removal of conditional and unconditional biases. The effect of this suppression is visualized in SS, where positive residual skill is present for different lead times. Positive values of SS represent an indirect indication that the KR approach shows the credibility to retain one-to-one association between the downscaled GCM product and PRISM precipitation at least for some of the lead times. This is mainly observed for the regions: MW, NW, SE, SW (at the yearly scale), GPN, MW, NE, SE, SW (over winter); GPN, MW, NE, SW (over spring); GPN, GPS, MW, NE, NW, SW (over summer); and GPS, MW, NE, NW, SE (over fall). The second positive attribute of KR approach is its ability to capture long term statistical properties, especially the mean. The approach shows good skill in capturing the long term mean (Figure 8), with the absolute difference between observed precipitation and GCM predictions that is about ~5 mm/month. Similar to the skill displayed by LR, the KR approach is not able to capture either the standard deviation, extreme and RMSE, except for the yearly scale. For other seasons, we observe cases in which the methodology captured STDEV and EXTREMES with excellent skills (e.g., SW in summer and fall; GPN in year and winter). Pixel-wise skills for downscaled precipitation, obtained by applying KR based downscaling approach for different lead times from 1 to 10 in the form of U.S. maps are shown in the supplementary material (from S-Figure 40 to S-Figure 49).

### 4.7 Uncertainty quantification for KR-based GCM predictions

We evaluate the downscaled precipitation with the KR approach in terms of GCM envelopes (Figure 9). Similar to the prior assessments (Figures 4 and 7), we restrict our discussion to lead-1 predictions only, with the evaluations for the other lead times in the supplementary material (S-Figures 50-58). Overall, here are two positive attributes associated with the KR, similar to what is discussed for LR. There is reduced uncertainty as shown by the reduction in the width of the GCM envelopes (gray band) as compared to the raw predictions in Figure 4. Also, ensemble GCM predictions (blue line) runs through the central tendency of the PRISM precipitation (black line), eliminating the biases between the two. The methodology is therefore able to capture the long term mean but not the fluctuations, similar to what was discussed for the LR methodology. If we compare the two downscaled products, we find that the LR approach performs better in reducing the uncertainty as compared to the KR methodology, evident from the width of the GCM envelope that is smaller. However, the ensemble GCM predictions for the downscaled precipitation with LR show little or no fluctuations at all, while those with the KR approach shows slightly higher degree of year-to-year variability.

### 4.8 Further validations of decadal predictions

In the previous sections (4.1-4.7), we provided a thorough evaluation of the raw GCM predictions and the statistically downscaled products obtained with the LR and KR. The focus of the validation was on the assessment of the prediction skill in a seasonally and regionally averaged spatio-temporal framework, for lead times from one to ten years. Such a lead-wise quantification of skill was possible because of the availability of several retrospective predictions, which enabled us to construct lead-wise datasets for validation. However, GCM predictions are not provided in a lead-wise format. Hence, the spatio-temporally averaged framework validation approach considered in this study cannot be used to assess GCM predictions in the form of single initialization of 'ten-year predictions' (decade). Here we modify the validation approach to be tailored to single initializations. We obtain regionally averaged time series of decadal predictions, consisting of 120 values (12 months × 10 years = 120 values) and compare them with the PRISM data to evaluate their performance. The motivation behind this type of evaluations lies in the fact that the decadal predictions are initialized either yearly or intermittently to provide the predictions for ten years. Therefore, we may not get the advantage of data availability to evaluate lead-wise skills. This procedure is carried out for each initialization (e.g., there are nine GCMs that are initialized with respect to the observations for the year 1961 to provide the predictions for 1961-1970). We obtain the ensemble of these nine GCMs and generate a regionally averaged times series (120 values) and estimate the skill by comparing the time series with regionally averaged PRISM observations for 1961-1970. The number of GCM predictions available at different initializations differs based on the data availability (e.g., for the predictions initialized in 1962 only five GCMs are available and we use just those five for obtaining the ensemble mean). A similar step is repeated for predictions that are initialized from 1961 onwards until 2005. We stop at the initialization year 2005 because the validation period for the present study is restricted to 2014 (which is the last year for the decadal predictions that are initialized in 2005). We evaluate the skill of decadal predictions using three metrics: 1) PS (representing the one-to-one association between predictions and PRISM data), 2) absolute percentage difference in mean (representing the total precipitation over a decade), and 3) absolute percentage difference in extremes ($95^{th}$ percentile, representing difference in extremes). We evaluate both data products (with LR and KR) and the results of this analysis are illustrated in Figure 10 in terms of radar plots. Each plot consists of 45 spokes, each representing a single decade (e.g., 1961-1970, 1962-1971) and labeled with the year in which the predictions was initialized. For regions GPN, MW, and NW there is high PS and the performance is consistent over all the decades, irrespective of the calibration/validation period. Both the downscaled precipitation products show similar PS, indicating that the methodologies are able to capture the one-to-one association over the mentioned region with a high confidence. GPS shows reasonable skills over the training period, even though the values of PS decline over the testing period. Both the downscaled products perform poorly over the regions NE, SE, and SW. The LR- and KR-SD precipitation predictions fare well in terms of capturing long term mean, represented in terms of absolute percentage difference and expressed as a fraction. GPN, MW, and SE are the regions where both the products show excellent skill in capturing observed mean (maximum difference smaller than 8%). Over the other regions, this difference is restricted to less than 15%, which makes these products a viable input data for impact studies. As far as PS and mean are concerned, the performance of both products is comparable. The real differences manifest themselves when we focus on the extremes. For all the regions, the extremes are better captured with the KR methodology. For most of the regions, the differences for KR based predictions are about 30%.

## 4.9 Decadal predictions up to 2024

Here we present future precipitation predictions over 2015-2024 (Figure 11). We use lead-10 data to ensure maximum future coverage. Each plot shows possible future variations in seasonally and regionally averaged precipitation obtained with the LR and KR approaches, with reference to the past baseline precipitation (average precipitation of 1981-2010) for a particular season and region. There are certain common features shared by almost all the plots. Ensemble future predictions (for both products) do not show a strong trend, indicating lack of variations in average precipitation. Also, these predictions are close to the baseline precipitation implying almost no change in future precipitation predictions as compared to the past. Future predictions with both the downscaled products are along a comparable trajectory. However, these statements are valid until 2022. We observe certain erratic behavior for the years post 2022. Also, the 'in-sync' behavior between the two data products until 2022 fades away for the last three years and we encounter differences between the two. However, this behavior can be attributed to the lack of predictions over these years. We just have a single GCM (CanCM4), which provides the data until 2015, significantly decreasing our confidence in the predictions after 2022. Also, as previously discussed, the predictions follow the trajectory, tracing the mean of the observed data but lack the variability in the observations. Assuming the same pattern to manifest in the future, it is expected to have much larger variability as compared to what is being depicted in the figure.

## 5. Discussion and conclusions

This study evaluates and enhances the skill of initialized decadal precipitation predictions over the continental United States (CONUS) and generate high resolution (~4km) precipitation prediction products by applying statistical downscaling (SD) methodologies. The downscaled precipitation products that are obtained with Linear Regression (LR) and Kernel Regression (KR) approaches are evaluated by imposing a stringent validation framework to understand the positive/negative attributes of these predictions. Raw General Circulation Models (GCM) predictions, when evaluated over different metrics, showed lack of skill (SS) because of the presence of high biases and high uncertainties, indicating that GCM predictions differ from one another, resulting in wide envelopes. Hence, it can be concluded that raw GCM predictions cannot be directly used for impacts assessments studies.

The validation of LR and KR models is performed to assess the credibility of two SD approaches over downscaled precipitation data, obtained with NCEP/NCAR forcings. Results show the strength of both the SD models in capturing some of the important properties of the observed precipitation. LR model shows superiority in capturing seasonal variations for most of the regions, showing positive residual skill (Skills Score) thank to the reduction in biases. On the other hand, the KR model captures the long term variability (standard deviations, extremes). These differences in performance support our need for the use of two SD approaches that can be categorized under the transfer function approach. The validation results serve as guidelines in selecting the data product based on the study that one intends to carry out: for instance, to study seasonal variations and formulate the policies for the same, LR-precipitation data are slightly more reliable and KR-precipitation data can be used to study extremes. Both methodologies, however, are capable of capturing a complex property of the observed data such as the spatial variability due to orography. Figure 12 shows the comparison of mean precipitation over different seasons (as discussed in section 3), obtained from the PRISM data and the statistically downscaled products (both with LR and KR) with the reanalysis forcings over the testing period (1991-2014). These methodologies are capable of capturing the complex spatial patterns of mean precipitation, especially near the U.S. West Coast. The credibility of the methodology to capture the spatial variability is extremely important from the point of view of impact assessment. This is because, even though we assess the skill of the predictions in a spatio-temporally averaged framework, impact assessment studies often demand reasonable skill over a much finer resolution within a region of interest. At such a small scale, the local effects like orography play an important role in controlling the spatial variability. Hence, a methodology that captures such local effects is likely to produce skillful downscaled products suitable for impact assessment studies.

The skill revealed by the downscaled GCM predictions is generally independent of lead times. This conclusion holds against the preconceived notion that smaller lags between initialization and predictions (smaller lead times) lead to higher skill. It is difficult to explain this inconsistent behavior. It can be only argued that the GCMs used for predictions might suffer from the shock of the initial conditions and hence, require some time before the model converges to provide better result. However, this statement is just a hypothesis and assessing its validity/ falsifiability is out of the scope of the present study. Both the approaches show reasonable skill in capturing important properties of the observed data. Similar to the skill revealed by SD predictions with NCEP/NCAR data, the GCM precipitation predictions with the LR approach show better skill in terms of PS and mean, whereas those with the KR approach show better skill with long term variability and extremes. The marked reduction in uncertainty as indicated by the reduction in the width of the GCM envelope is an indication that the downscaled precipitation data from different GCMs bear a close resemblance with each other. The comparison of the two products in a decadal validation framework shows that the downscaled precipitation with the KR-approach appears to capture the climatology better when compared to that with LR-approach. Future precipitation predictions, however, do not show a significant trend. The lack of visually significant trends in future precipitation can be taken as a word of caution. This is an indication that we need robust planning policies to cater to the increasing water demands, especially after having understood that the total precipitation is not likely to change significantly over the next few years.

Having discussed all the positive features of the methodology and downscaled data, we want to highlight some of the limitations of the present study as well. The methodology starts with the identification of zone of predictors for each of the seven regions. The seven regions are formed by merging few states and they do not necessarily represent meteorologically homogeneous regions. This may result in reducing some of the skill of the downscaled precipitation. The assumption of stationarity is a common limitation for all data driven approaches, where the statistical relationship is assumed to be time invariant. However, as we are applying the relationship which is established in the near past (1961-1990), there is a fair chance that the violation of the assumption will not negatively affect the results. The variability possessed by the observed precipitation is generally not captured by our methodologies. This is evident from the narrowness of the GCM envelopes. The accuracy with which the predictions are able to capture the mean does not hold for variability and extremes. Hence, we have high confidence in the total amount of precipitations that the products predict. However, the seasonal variations and extremes, revealed by the downscaled precipitation should be considered with care.

### *Acknowledgments*

This work is funded in part by the Broad Agency Announcement (BAA) Program and the Engineer Research and Development Center (ERDC)–Cold Regions Research and Engineering Laboratory (CRREL) under Contract No. W913E5-16-C-0002. We thank the Associate Editor, Dr. Pegram and an anonymous reviewer for their comments and suggestions.

## References


Akaike, H. (1974), Information theory and an extension of the maximum likelihood principle. In Proceedings 2nd International Symposium on Information Theory; Petrov and Caski, Eds., pp 267-281.

Bellucci, A., and Co-authors (2015a), Advancements in decadal climate predictability: The role of nonoceanic drivers, Rev. Geophys., 53, doi:10.1002/2014RG000473.

Bellucci, A., and Co-authors (2015b), An assessment of a multi-model ensemble of decadal climate predictions. Climate Dyn. 2787–2806, doi:10.1007/s00382-014-2164-y.

Bin, W., and Co-authors (2012), Preliminary Evaluations of FGOALS-g2 for Decadal Predictions. Advances in Atmospheric Sciences, VOL. 30, NO. 3, 2013, 674–683.

Branstator, G., and H. Teng (2012), Potential impact of initialization on decadal predictions as assessed for CMIP5 models, Geophys. Res. Lett., 39, L12703, doi:10.1029/2012GL051974.

Bromwich, D. H., and R. L. Fogt (2004), Strong trends in the skill of the ERA-40 and NCEP-NCAR reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958 – 2001, J. Clim., 17, 4603 – 4619

Cane, M. (2010), Climate science: Decadal predictions in demand, Nature Geoscience 3, 231 - 232 (2010), doi:10.1038/ngeo823

Caron, L. P., C. G. Jones, F., and Doblas-Reyes (2014), Multi-year prediction skill of Atlantic hurricane activity in CMIP5 decadal hindcasts. Climate Dyn., 42, 2675–2690, doi:10.1007/s00382-013-1773-1.

Choi, J., S. W. Son, Y. Ham, J. Lee, and H. Kim (2016), Seasonal-to-Interannual Prediction Skills of Near-Surface Air Temperature in the CMIP5 Decadal Hindcast Experiments, Seasonal-to-Interannual Prediction Skills of Near-Surface Air Temperature in the CMIP5 Decadal Hindcast Experiments. J. Climate, 29, 1511–1527.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteor., 33, 140-158.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris (2002), A knowledge-based approach to the statistical mapping of climate. Climate Research 22: 99–113.

Dee, D. P., and Co-authors (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137(656), 553-597

Dirmeyer, P. A., and Co-authors (2016), Confronting Weather and Climate Models with Observational Data from Soil Moisture Networks over the United States, J. Hydrometeor., DOI: 10.1175/JHM-D-15-0196.1

Doblas-Reyes F. J., and Co-authors (2013), Initialized near-term regional climate change prediction. Nat. Commun., 4, 1715, doi:10.1038/ncomms2704.

Ebita, A., and Co-authors (2011), The Japanese 55-year Reanalysis “JRA-55”: An Interim report, Sola, 7, 149–152, doi:10.2151/sola.2011-038

Freydank, K., and S. Siebert (2008), Towards mapping the extent of irrigation in the last century: Time series of irrigated area per country, Frankfurt Hydrol. Pap. 08, Inst. of Phys. Geogr., Univ. of Frankfurt, Frankfurt am Main, Germany.

Fyfe J. C., W. J. Merryfield, V. Kharin, G. J. Boer, W. S. Lee, and K. von Salzen (2011), Skillful predictions of decadal trends in global mean surface temperature. Geophys. Res. Lett., 38, L22801, doi:10.1029/2011GL049508

Gaetani, M., and E. Mohino (2013), Decadal Prediction of the Sahelian Precipitation in CMIP5 Simulations, Journal of Climate, volume 26, DOI: 10.1175/JCLI-D-12-00635.1

Ghosh, S., and P. P. Mujumdar (2006), Future rainfall scenario over Orissa with GCM projections by statistical downscaling. Current, 90(3).

Goddard, L., J. W. Hurrell, B. P. Kirtman, J. Murphy, T. Stockdale, and C. Vera (2012), Two time scales for the price of one (almost). Bull Amer Meteor Soc 93:621–629. doi:10.1175/BAMS-D-11-00220.1

Goddard L, and Co-authors (2013), A verification framework for interannual-to-decadal predictions experiments, Clim. Dyn., 40(1–2), 245–272, doi:10.1007/ s00382-012-1481-2

Gray, C. L. (2009), Environment, land, and rural out-migration in the southern Ecuadorian Andes. World Development, 37(2), 457–468.

Hashino T, A. A. Bradley, and S. S. Schwartz (2007), Evaluation of bias-correction methods for ensemble streamflow volume forecasts, Hydrol. Earth Syst. Sci., 11, 939–950, 2007

Henry, S., B. Schoumaker, and C. Beauchemin (2004), The impact of rainfall on the first out-migration: A multi-level event-history analysis in Burkina Faso. Population and Environment, 25(5), 423–460.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis (2005), Very High Resolution Interpolated Climate Surfaces for Global Land Areas, International Journal of Climatology, Int. J. Climatol. 25: 1965–1978.

Hunter, L. M., S. Murray, and F. Riosmena (2011), Climatic variability and U.S. migration from rural Mexico. Boulder: University of Colorado, Institute of Behavioral Sciences.

Hurrell, J. W., and Co-authors (2009), "Decadal Climate Prediction: Opportunities and Challenges" in Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.45

Kaiser, H. F. (1960), The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kalnay, E., and Co-authors (1996), The NCEP/NCAR 40-years reanalysis project, Bull. Am. Meteor. Soc., 77(3), 437-471.

Kannan, S., and S. Ghosh (2010),Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output,, Stochastic Environmental Research and Risk Assessment, Springer.

Kannan, S., and S. Ghosh (2013), A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin, Water Resour. Res., 49, 1360-1385, doi:10.1002/wrcr.20118.

Kannan, S., S. Ghosh, V. Mishra, and K. Salvi (2014), Uncertainty Resulting from Multiple Data Usage in Statistical Downscaling, Geophysical Research Letter, 41, 4013–4019, DOI:10.1002/2014GL060089

Karl, T. R., J. M. Melillo, and T.C. Peterson (2009), Global Climate Change Impacts in the United States. Cambridge University Press.

Kim, H-M., P. J. Webster, and J. A. Curry (2012), Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. Geophys. Res. Lett., 39, L10701, doi:10.1029/2012GL051644.

Kniveton, D., K. Schmidt-Verkerk, C. Smith, and R. Black (2008), Climate change and migration: Improving methodologies to estimate flows. Brighton: International Organization for Migration.

Kobayashi, S., and Co-authors (2015), The JRA-55 reanalysis: general specifications and basic characteristics. Journal of the Meteorological Society of Japan 93: 5–48, doi:10.2151/jmsj.2015-001.

Kunkel, K. E., and Co-authors (2013), Monitoring and understanding changes in extreme storms: State of knowledge, Bull.Am.Meteorol.Soc.,doi:10.1175/BAMS-D-12-00066.1, in press.

Li, H., J. Sheffield, and E. F. Wood (2010), Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching, J. Geophys. Res., 115, D10101, doi:10.1029/2009JD012882.

Malinowski, E. R. (1991), Factor Analysis in Chemistry; Wiley-Interscience: New York.

McLeman, R., and B. Smit (2006), Migration as an adaptation to climate change. Climatic Change, 76(1–2), 31–53.

Meehl, J. A., and Co-authors (2009), DECADAL PREDICTION Can It Be Skillful? Bull. Amer. Meteor. Soc., 90, 1467–1485.

Meehl, J. A., and Co-authors (2014), Decadal climate prediction: An update from the trenches. Bull. Amer. Meteor. Soc., 95, 243–267, doi:10.1175/BAMS-D-12-00241.1.

Meehl, G. A., A. Hu, J. M. Arblaster, J. Fasullo, and K. E. Trenberth (2016), Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation, Journal of Climate, DOI: 10.1175/JCLI-D-12-00548.1

Mochizuki, T., Y. Chikamoto, M. Kimoto, M. Ishii, H. Tatebe, Y. Komuro, and T. T. Sakamoto (2011), NOTES AND CORRESPONDENCE Decadal Prediction Using a Recent Series of MIROC Global Climate Models, Journal of the Meteorological Society of Japan, Vol. 90A, pp. 373--383, 2012. 373 DOI:10.2151/jmsj.2012-A22

Müller, W. A., J. Baehr, H. Haak, J. H. Jungclaus, J. Kröger, D. Matei, D. Notz, H. Pohlmann, J. S. von Storch, and J. Marotzke (2012), Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, Geophys. Res. Lett., 39, L22707, doi:10.1029/ 2012GL053326.

Murphy, A. H., and R. L. Winkler (1992), Diagnostic verification of probability forecasts, Int. J. Forecasting, 7, 435–455.

Mutiibwa, D., S. J. Vavrus, S. A. McAfee, and T. P. Albright (2015), Recent spatiotemporal patterns in temperature extremes across conterminous United States. J. Geophys. Res. Atmos., 120, doi:10.1002/2015JD023598.

Nawrotzki, R. J., R. Riosmena, and L. M. Hunter (2013), Do Rainfall Deficits Predict U.S.-Bound Migration from Rural Mexico? Evidence from the Mexican Census, Popul Res Policy Rev (2013) 32:129–158, DOI 10.1007/s11113-012-9251-8

Onogi, K., and Co-authors (2007), The JRA-25 Reanalysis. J. Meteor. Soc. Japan, 85, 369-432.

Pryor, S. C., and J. T. Schoof (2008), Changes in the seasonality of precipitation over the contiguous USA. J. Geophys. Res., 113, D21108, doi:10.1029/2008JD010251.

Qin, S. J. (1998), Dunia, R. Determining the number of principal components for best reconstruction. In IFAC DYCOPS'98, Greece.

Salvi, K., S. Kannan, and S. Ghosh (2013), High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment, J. Geophys. Res. Atmos., 118, 3557-3578, DOI:10.1002/jgrd.50280

Salvi, K., S. Ghosh, and A. R. Ganguly (2015), Credibility of Statistical Downscaling under Nonstationary Climate, Climate Dynamics, doi: 10.1007/s00382-015-2688-9

Salvi, K., G. Villarini, G.A. Vecchi, and S. Ghosh, Decadal temperature predictions over the continental United States: Analysis and enhancement, Climate Dynamics, doi:10.1007/s00382-017-3532-1, 2017.

Sanderson, M. G., and Co-authors (2008), A multi-model source-receptor study of the hemispheric transport and deposition of oxidised nitrogen, Geophys. Res. Lett., 35, L17815, doi:10.1029/2008GL035389, 2008.

Schoof, J. T., S. C. Pryor, and J. Surprenant (2010), Development of daily precipitation projections for the United States based on probabilistic downscaling, J. Geophys. Res., 115, D13106, doi:10.1029/2009JD013030.

Shashikanth, K., C. G. Madhusoodhanan, S. Ghosh, T. I. Eldho, K. Rajendran, and R. Murtugudde (2014), Comparing Statistically Downscaled Simulations of Indian Monsoon at different spatial Resolutions, Journal of Hydrology, Volume 519, Part D, Pages 3163:3177, 10.1016/j.jhydrol.2014.10.042

Singh, A., M. A. Kulkarni, U. C. Mohanty, S. C. Kar, A. W. Robertson, and G. Mishra (2012), Prediction of Indian summer monsoon rainfall (ISMR) using canonical correlation analysis of global circulation model products Meteorol. Appl. 19: 179–188, DOI: 10.1002/met.1333

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model. Science, 317, 796–799, doi:10.1126/science.1139540.

Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife (2010), Skillful multi-year predictions of Atlantic hurricane frequency. Nat. Geosci., 3, 846–849, doi:10.1038/ngeo1004.

Srinivas, V. V., B. Basu, D. N. Kumar, and S. K. Jain (2014), Multi-site downscaling of maximum and minimum daily temperature using support vector machine, Int. J. Climatol. 34: 1538–1560 (2014), DOI: 10.1002/joc.3782

Tiwari, P. R., S. C. Kar, U. C. Mohanty, S. Kumari, P. Sinha, A. Nair, and S. Dey (2014), Skill of precipitation prediction with GCMs over north India during winter season Int. J. Climatol. (2014) DOI: 10.1002/joc.3921

Trenberth, K. E., and Co-authors (2007). Observations: Surface and atmospheric climate change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), Climate change 2007: The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. New York: Cambridge University Press.

UCAR 2016, https://climatedataguide.ucar.edu/climate-data/prism-high-resolution-spatial-climate-data-united-states-maxmin-temp-dewpoint, 25th February 2016, 10 AM

Uppala, S. M., and Co-authors (2005), The ERA-40 re-analysis. Quart. J. R. Meteorol. Soc., 131, 29613012.doi:10.1256/qj.04.176

Valle, S., W. Li, and S. Joe Qin (1999), Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods, Ind. Eng. Chem. Res. 1999, 38, 4389-4401.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger (2012), Skill in the trend and internal variability in a multi-model decadal prediction ensemble, Clim. Dyn, 38.7 (2012): 1263-1280.

Vecchi, G. A., and Co-authors (2013), Multi-year predictions of North Atlantic hurricane frequency: Promise and limitations, Journal of Climate, 26(15), 5337-5357, 2013.

Widmann, M., and C. S. Bretherton (1999), Validation of Mesoscale Precipitation in the NCEP Reanalysis Using a New Gridcell Dataset for the Northwestern United States, J. Climate, 13, 1936–1950. doi: http://dx.doi.org/10.1175/1520-0442(2000)013<1936:VOMPIT>2.0.CO;2

Wilby, R. L., and Co-authors (2004), The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), Prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA). (http://ipccddc.cru.uea.ac.uk/guidelines/ StatDownGuide.pdf)

Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Clim. Change, 62(1–3), 189–216.

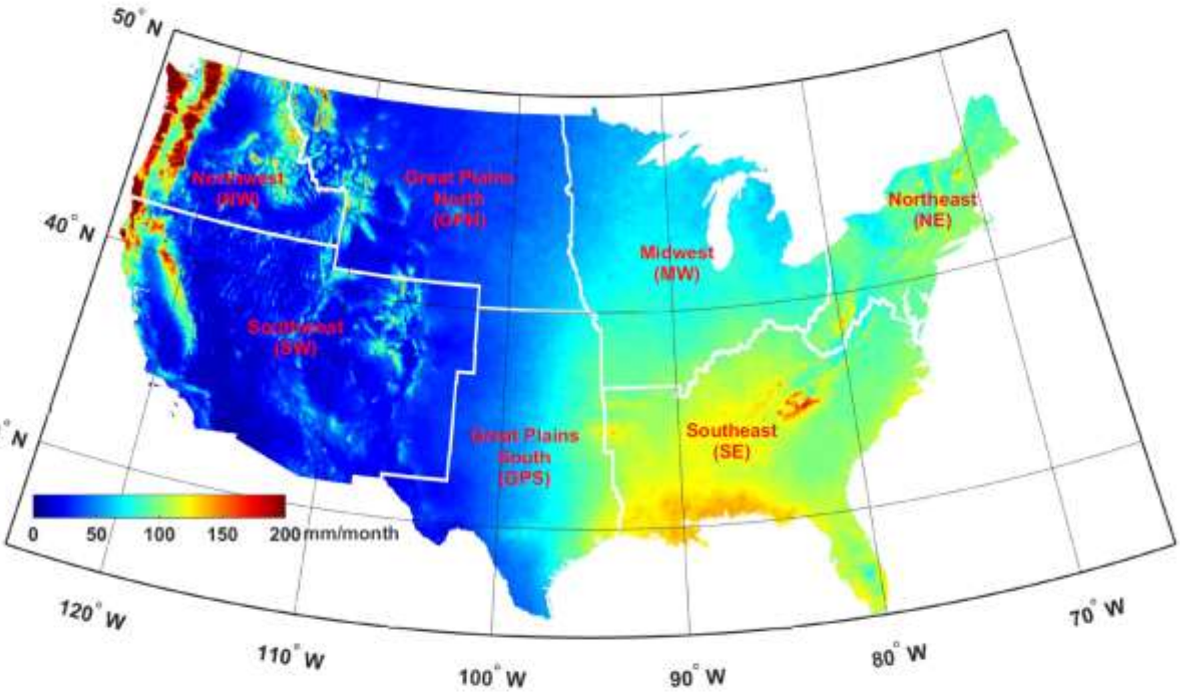


Figure 1: Map showing the monthly averaged precipitation for the 1961-2014 period based on the PRISM data. The map also shows the seven regions considered in the analyses.

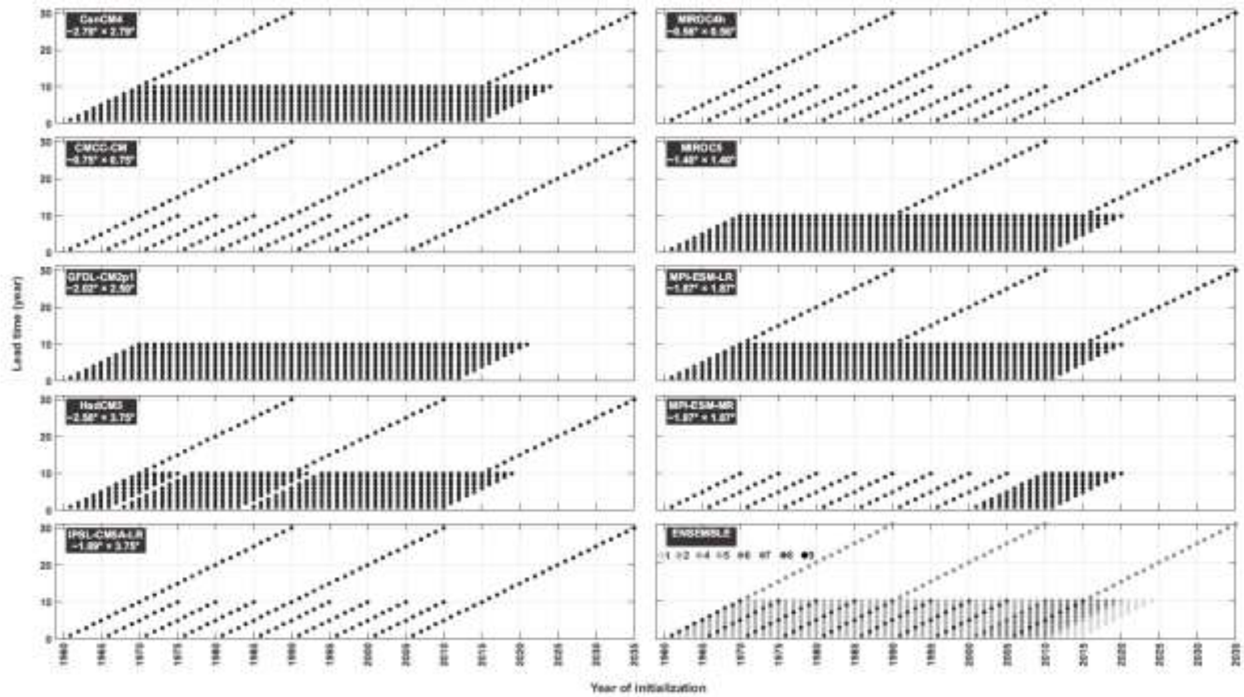


Figure 2: Graphical data inventory plots, illustrating the year of initialization and lead-wise prediction availability for nine GCMs along with their spatial resolution, mentioned in the top left corner of each plot. 'Ensemble' plot shows the number of GCMs available for different initializations and lead times for obtaining ensemble mean and uncertainty.

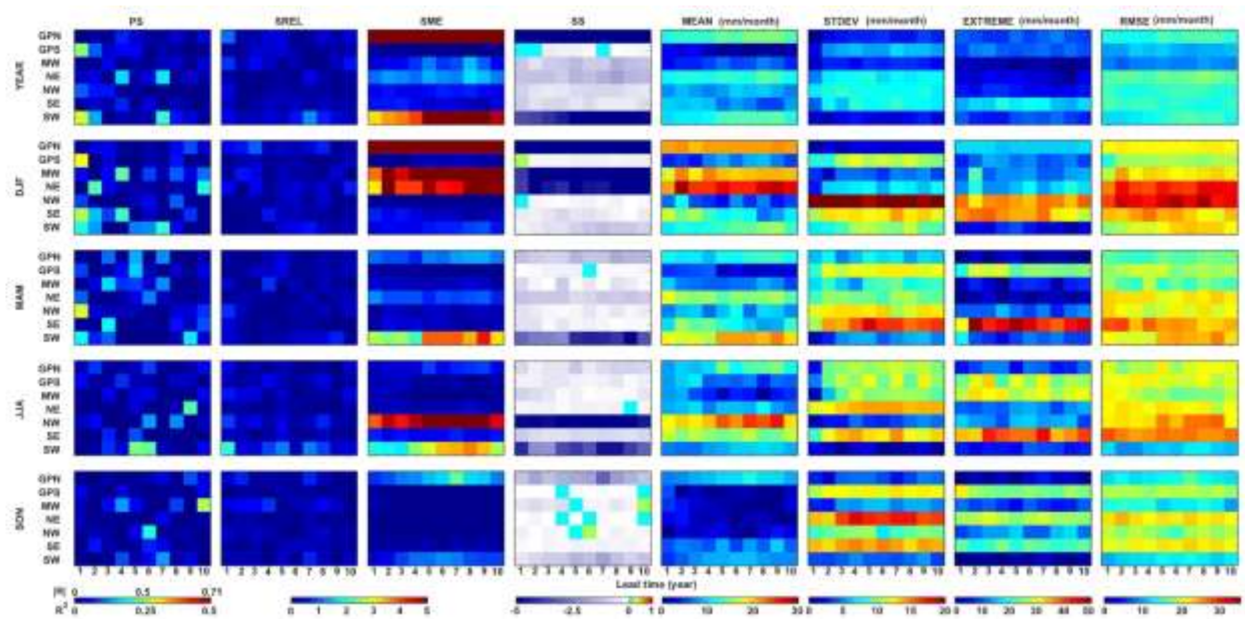


Figure 3: 'Memory-based' and 'Long-term statistics based' skill metrics for seasonally and regionally averaged ensembles of *raw* GCM predictions over lead times 1-10. The results are for the average of all available GCMs.

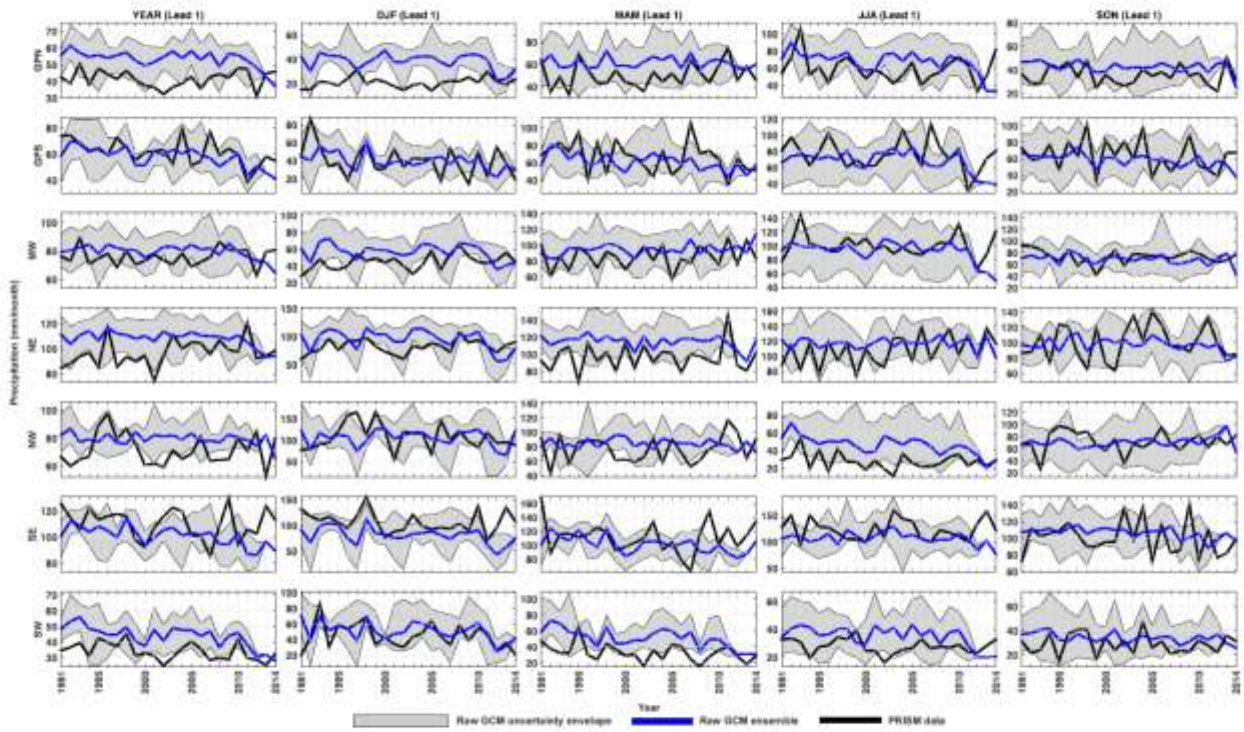


Figure 4: ‘Uncertainty-based’ skill metrics in terms of seasonally and regionally averaged envelope as a measure of uncertainty, obtained with 1-year lead time predictions. Wide envelopes (gray band) and high deviations of the ensemble GCM predictions (blue line) with respect to observed time series (black line) for some of the regions imply poor performance by the GCMs.

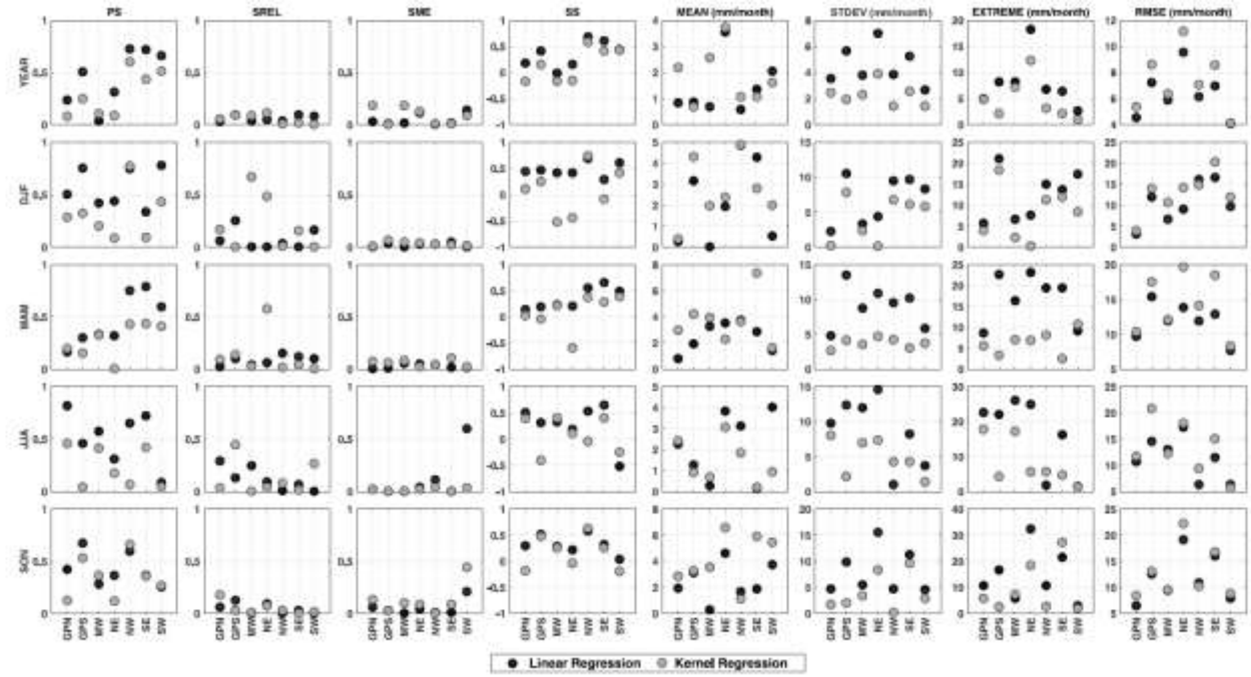


Figure 5: ‘Memory-based’ and ‘Long-term statistics based’ skill validations of LR and KR based SD methodologies applied to NCEP/NCAR reanalysis predictors over 1991-2014. The downscaling methodologies show increasing skill in extracting the relationship between NCEP/NCAR predictors and PRISM precipitation, evident from the positive values of SS, extremely low biases, and lower differences in long term statistical properties.

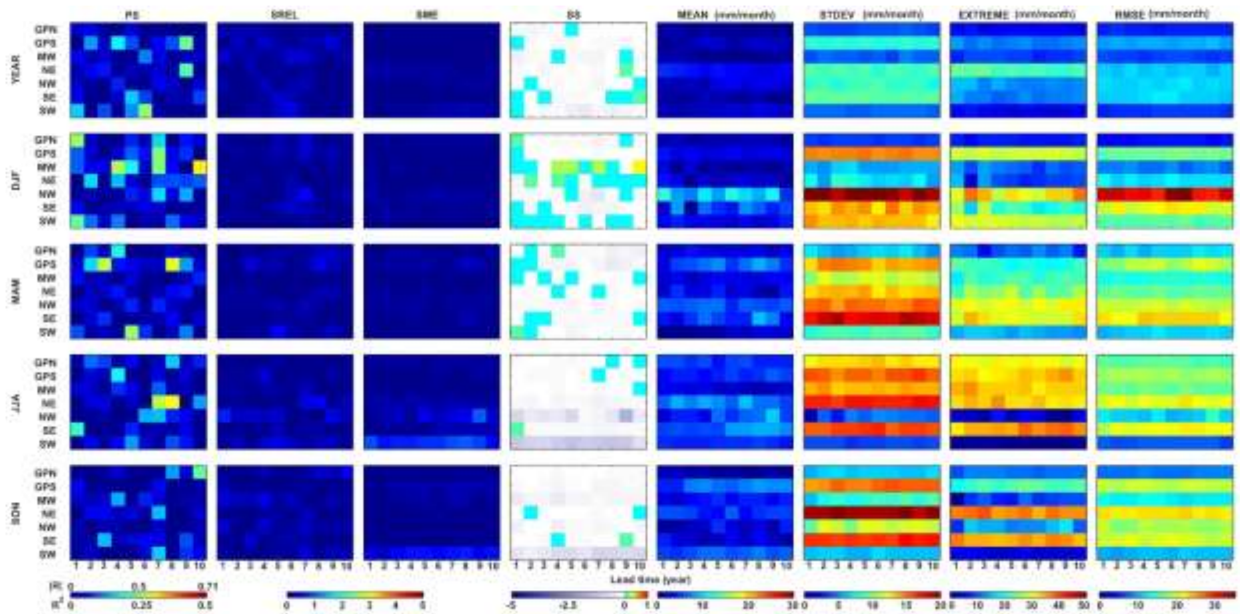


Figure 6: Same as Figure 3 but for seasonally and regionally averaged ensembles of LR based *statistically downscaled* GCM predictions.

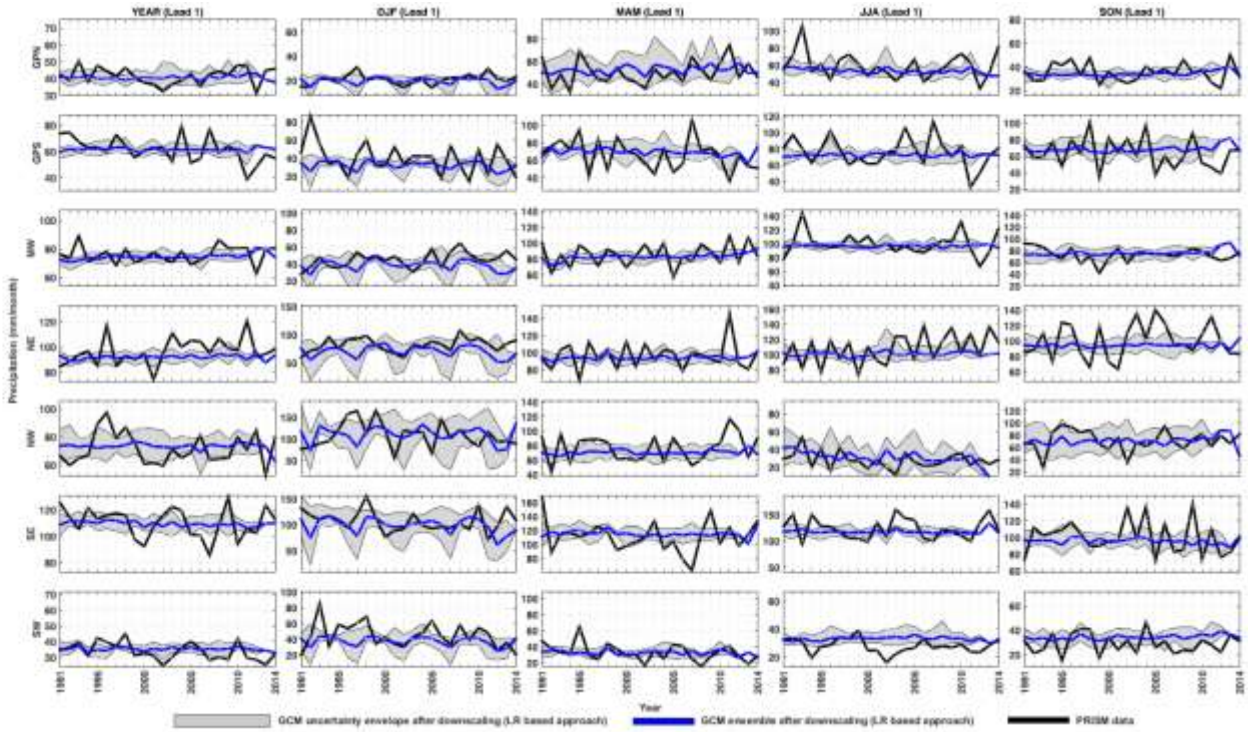


Figure 7: Same as Figure 4 but for the ensembles of LR based *statistically downscaled* GCM predictions.

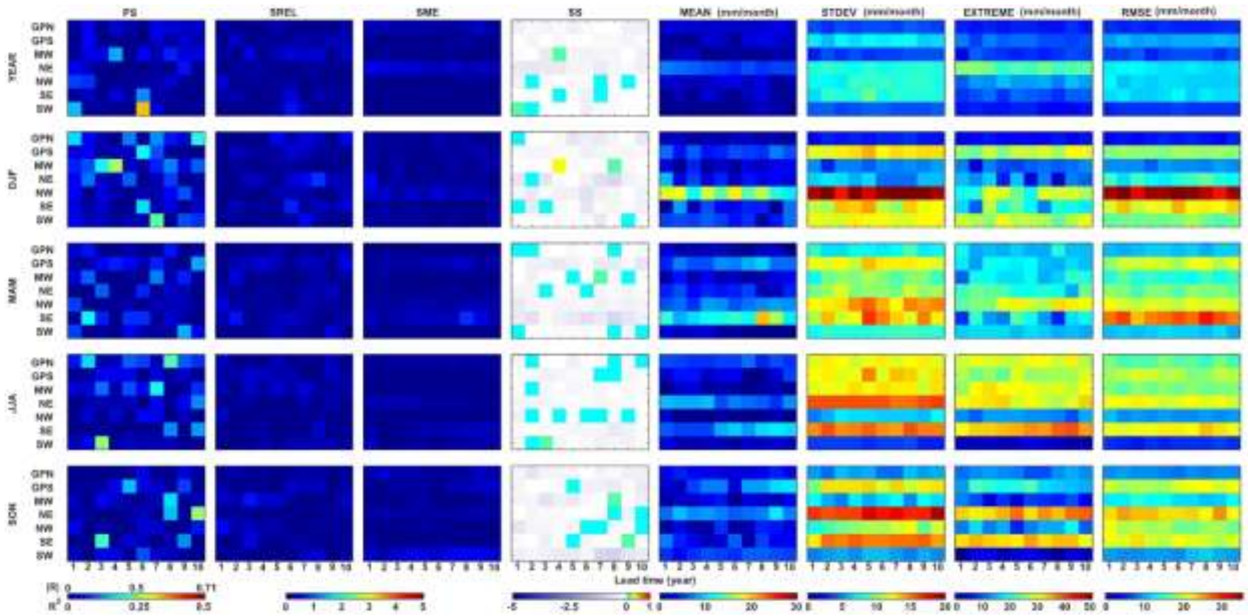


Figure 8: Same as Figure 3 but for seasonally and regionally averaged ensembles of KR based *statistically downscaled* GCM predictions.

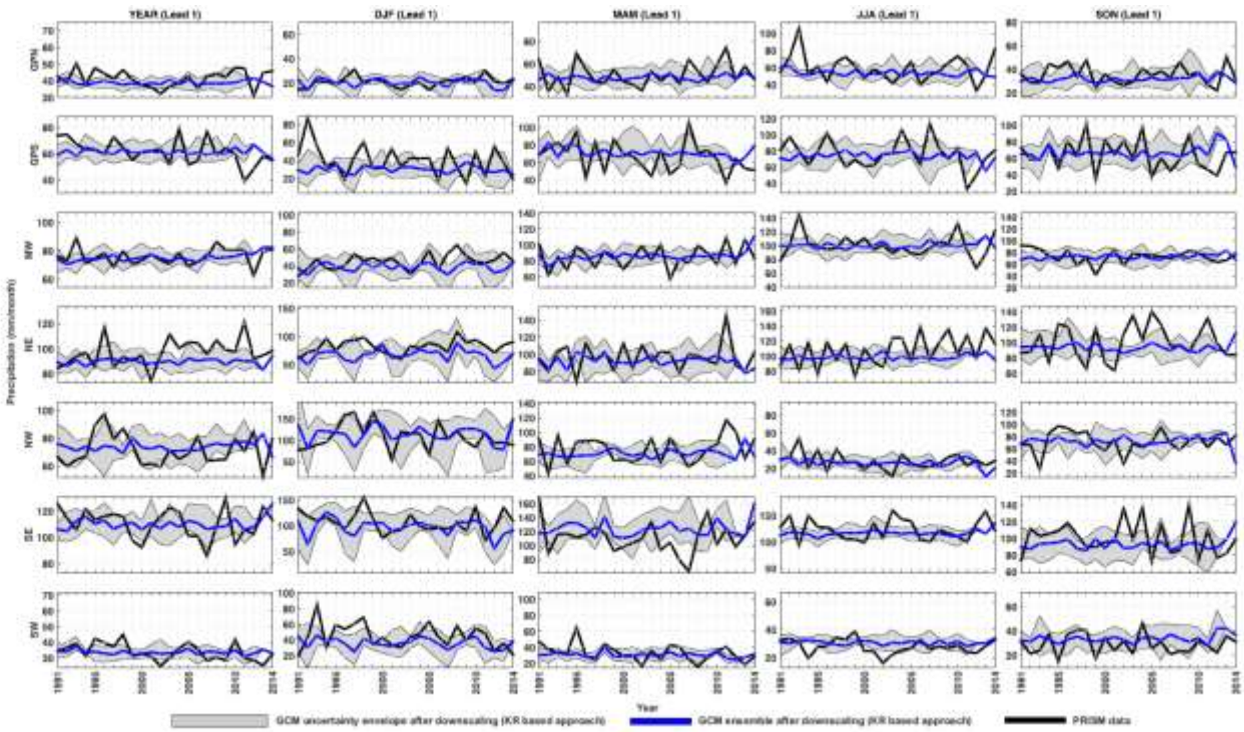


Figure 9: Same as Figure 4 but for the ensembles of KR based *statistically downscaled* GCM predictions.

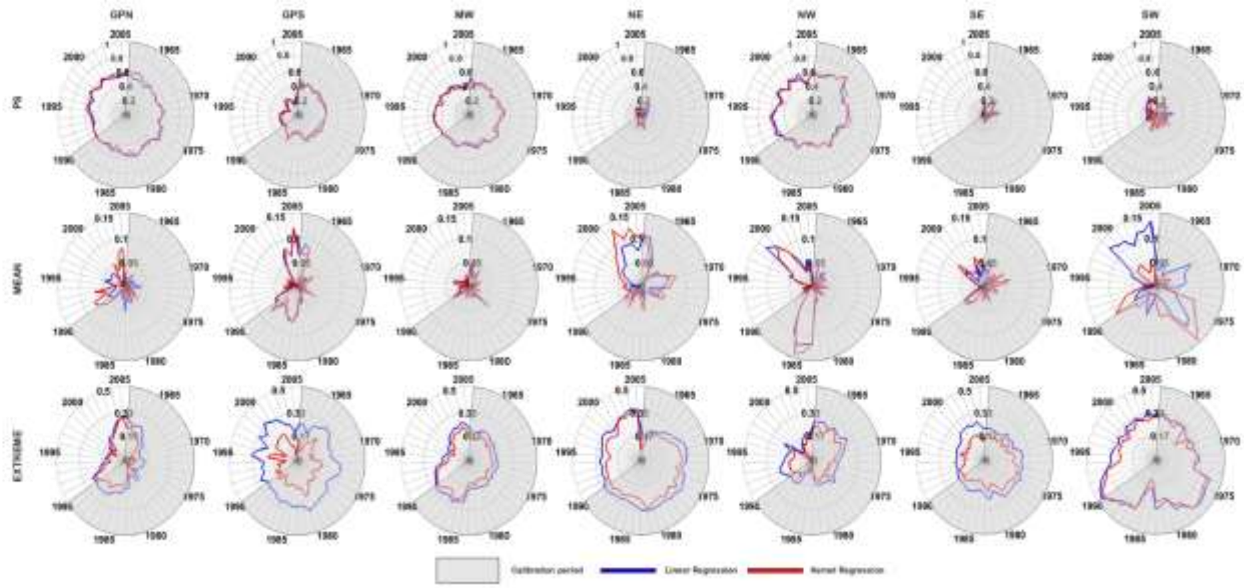


Figure 10: Radar plots comparing the skill between statistically downscaled precipitation skills (LR based, shown by blue color and KR based shown by red color) in the decadal framework. Each spoke represents the year of initialization. The gray region shows the skills for the years where year of initialization belongs to the calibration period (1961-1990).

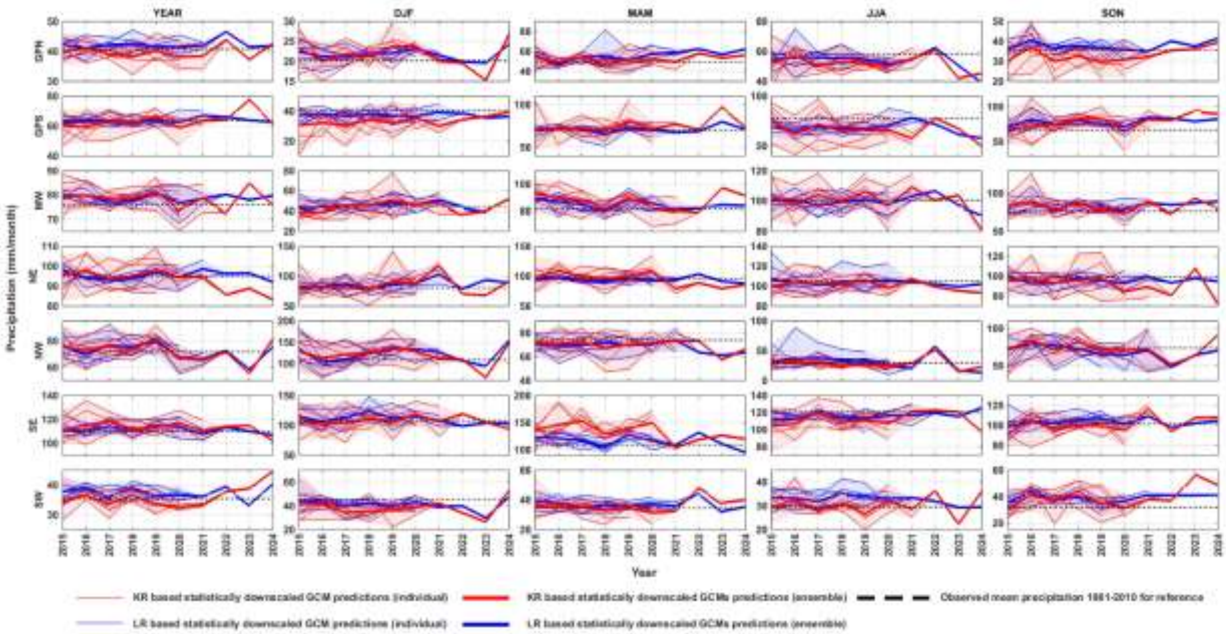


Figure 11: Precipitation predictions for the near future decade 2015-2024. Seasonally and regionally averaged ensemble of LR based statistically downscaled precipitation data (thick blue line), surrounded by individual bias corrected GCM prediction trajectories (thin blue lines) and envelope (light blue shaded region) showing the uncertainty associated with it. Red lines/shaded region are the exact counterparts of the LR based statistically downscaled precipitation but obtained with KR based SD methodology. The black dashed line represents the 30-year mean (1981-2010) to be used as reference for comparing future predictions.

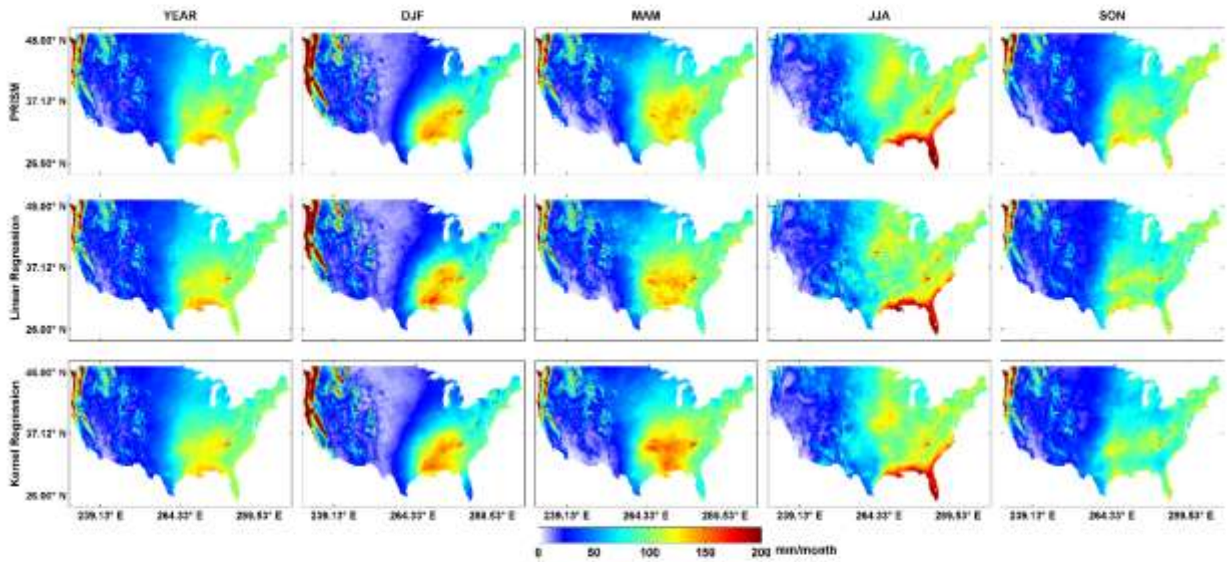


Figure 12: Comparison between PRISM mean precipitation (1991-2014) with statistically downscaled mean precipitation with linear and kernel regression based approach, applied to NCEP/NCAR reanalysis data predictors over different seasons. The methodologies are able to capture the orographic effects influencing spatial pattern of mean precipitation. The marked spatial variations in mean precipitation in PRISM (e.g., western United States) are captured by the two approaches.