Separating weather and climate using a spatially-scalable precipitation model with optimized subseasonal-to-seasonal statistics

Daniel J. Short Gianotti[1], Guido D. Salvucci[2], and Bruce T. Anderson[2]

[1]Massachusetts Institute of Technology
[2]Boston University

November 22, 2022

Posted on 22 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10503866.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Abstract

We present a kernel auto-regressive (KA) method which can be used to represent the daily to multi-day auto-correlation structure of precipitation time series, using information both in the occurrence and intensity of measured rainfall events. The method is able to capture a larger fraction of the memory in multiple time series than commonly-used occurrence-based Markov chain models, even when the intensity distribution is allowed to be conditioned on the Markov state. The KA method is less sensitive to the spatial scale at which the data is reported, as it is not strictly reliant on patterns of wet and dry days for providing correlation. Output from the KA model can be used as weather generator model simulations, as empirical representations of process structure, as representation of weather/climate variability partitioning, or as climatological null models against which observations can be compared for statistical significance. The KA method demonstrates improvements in each of these over classic occurrence Markov chain models and daily independent climatology, in both representations of interannual precipitation variability and in downstream water balance variables. We provide climate null confidence intervals for precipitation trends (driven largely by autumn increases), and decompose variability into trend, interannual, and weather components (in increasing order of magnitude) for the Contiguous United States.

Separating weather and climate using a spatially-scalable precipitation model with optimized subseasonal-to-seasonal statistics

Daniel J. Short Gianotti[1*], Guido D. Salvucci[2], and Bruce T. Anderson[2]

[1]Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[2]Department of Earth and Environment, Boston University, Boston, Massachusetts, USA.

Key Points:

- A spatially-scalable kernel method captures more daily-scale precipitation memory than Markov chain models.
- The method outperforms classic models more dramatically as occurrence frequency (scale) increases.
- Weather-scale precipitation variability dominates climate-scale variability and trends in mesic regions.

*Parsons Laboratory, Massachusetts Institute of Technology, 15 Vassar St., Cambridge, MA 02139, USA

Corresponding author: Daniel J. Short Gianotti, `gianotti@mit.edu`

Abstract

We present a kernel auto-regressive (KA) method which can be used to represent the daily to multi-day auto-correlation structure of precipitation time series, using information both in the occurrence and intensity of measured rainfall events. The method is able to capture a larger fraction of the memory in multiple time series than commonly-used occurrence-based Markov chain models, even when the intensity distribution is allowed to be conditioned on the Markov state. The KA method is less sensitive to the spatial scale at which the data is reported, as it is not strictly reliant on patterns of wet and dry days for providing correlation. Output from the KA model can be used as weather generator model simulations, as empirical representations of process structure, as representation of weather/climate variability partitioning, or as climatological null models against which observations can be compared for statistical significance. The KA method demonstrates improvements in each of these over classic occurrence Markov chain models and daily independent climatology, in both representations of interannual precipitation variability and in downstream water balance variables. We provide climate null confidence intervals for precipitation trends (driven largely by autumn increases), and decompose variability into trend, interannual, and weather components (in increasing order of magnitude) for the Contiguous United States.

Plain Language Summary

Weather generator models (WGMs) create realistic weather data which can be used for statistical climate analyses and determination of probabilities of weather events. Most WGMs represent precipitation occurrence (whether it rains) and intensity (how much it rains) separately, which can neglect some of the day-to-day interplay in these phenomena. Here we demonstrate a new WGM which combines occurrence and intensity processes, called a *kernel autoregressive* (KA) model. Because it combines occurrence and intensity, data from the KA method at different spatial scales (weather station, climate model) can be compared directly. The KA method also outperforms advanced versions of the most common WGMs. This makes the KA model superior for partitioning variability due to weather and longer term climate variability (El Niño, climate change, etc.). Even though weather fluctuations are large compared to longer climate signals and trends, roughly a quarter of the US shows changes in precipitation that are larger than would be expected for weather fluctuations.

1 Introduction

There is no shortage of existing weather generator models (WGMs) for any number of weather variables and for any number of specialty purposes (Wilks & Wilby, 1999; Ailliot et al., 2015). Weather generators are used by researchers, practitioners, businesses, and agencies for estimating natural resource availability, forecasting hazard risk, understanding fundamental meteorological processes, and driving other complex natural systems models. The basic motivations that these models share in common is the desire to represent some probabilistic structure of weather variables and a need for simulations of weather that meet basic statistical criteria.

Among the classes of WGMs are those that represent single versus multiple variables (e.g., precipitation, temperature, radiation, etc.), those that represent some level of physical process detail versus purely statistical methods, and those that assume some level of climate process stationarity versus those that represent process variability at climate time-scales. The difference in model form is dependent on the use the WGM will play: the classic “Richardson-type” WGM for precipitation represents daily rainfall, typically fit as twelve distinct parameterizations to represent the seasonal cycle, with a single-lagged Markov-chain representation of occurrence and a parameterized univariate distribution for intensity (classically exponential, but more typically gamma) (Richardson, 1981; Wilks & Wilby, 1999). This is useful for representing the scale of seasonal variability and for driving other physical models that may not require any sort of long-term change analysis. A model focused on sub-seasonal-scale extreme events will likely carefully fit more complex distributions to the tails of the distribution and require attention to daily-to-monthly-scale auto-correlation of these extremes (Koutsoyiannis, 2004; Min et al., 2011), while a WGM used in downscaling output from a global climate model (GCM) may focus mainly on a spatial covariance structure, conditioned on the state of multiple climate indices or a given mean value (Wood et al., 2004).

In this study, we propose a method for the stochastic simulation of precipitation to fit a specific set of criteria:

1. First, we are interested in a WGM for use as a *climatological null* model (von Storch & Zwiers, 2013) — that is, an entirely probabilistic data model that represents processes on weather time-scales as well as possible, using only lagged local precipitation as a predictor, while explicitly not representing variability due to processes on climate time-scales. Interannual variability will of course occur in these WGM time series, but we will attempt to optimally represent the interannual variability due to "weather-scale" processes, processes that would be deemed "stochastic" and due to "internal" system variability at climate time scales.
2. Secondly, we are interested in a model that can explicitly be used at *multiple spatial scales*. Due to the dependence of occurrence probability on the spatial scale of observations, weather generation methods used in earlier studies of weather and climate variability (Madden et al., 1999; Katz & Zheng, 1999) are best suited for scales at which occurrence probability is far from either zero or unity and are not applicable for inter-comparison between different spatial scales. The method proposed in this paper also provides a foundation through which climate and weather variability can be compared among global climate models and gridded observational datasets.
3. Beyond these, the method should be able to serve the purpose of any other weather generator model for applications which necessitate the proper representation of daily-to-weekly memory or auto-correlation structure.

The motivation for a "weather-only" representation is to create climatological nulls for separating variability on weather and climate time-scales. This is crucial for observationally-based (as opposed to model-based) potential predictability studies (Gianotti et al., 2013; Short Gianotti et al., 2014; Anderson et al., 2015b, 2015a, 2016). These observationally-based approaches are a necessary counterbalance to predictability modeling studies which must assume optimal internal representation of weather-scale statistics. Both approaches are necessary to bound our estimates of forecast skill for Earth System Processes (National Acadamies of Science, Engineering, and Medicine, 2020), to recognize forecast avenues of opportunity or diminishing return (Mariotti et al., 2020), and to properly bound the ways in which weather-scale vs climate-scale precipitation variability impact downstream Earth System Processes (Short Gianotti et al., 2020). The desire for spatial scalability is to allow for comparison of weather and climate variability between observed and modeled data sets.

In this study, we focus on the model itself, its representation of variability, its scaling behaviors, and its influence (relative to classic WGMs) on downstream process representation – specifically on surface soil moisture dynamics.

1.1 Climatological Null Models

Forecast skill is often measured relative to climatology (e.g., Heidke and Brier Skill Scores), and that climatology is typically enumerated as the probability distribution of a single variable for a given time period, marginalized over all states of the Earth system (including climate states, atmospheric states, secular trends, representations of spatial teleconnections, land surface conditions, etc.). For this model, we wish to explicitly acknowledge the daily-scale temporal correlation structure inherent in precipitation data by modeling it rather than marginalizing over it. This representation of a climatology with serial correlation serves two major purposes.

First, by representing precipitation as a data generating process which can include auto-correlation we create a more stringent baseline for quantifying weather forecast skills than daily-independent climatologies. When we test to see if a variable serves as a skillful predictor for precipitation, we compare the forecasts to climatology because we want to determine if that variable contains any useful information not already hidden in the precipitation data itself. If lagged precipitation values are more skillful than another predictor, those lagged precipitation values should be used in place of (or in conjunction with) that predictor. Thus, a weather generator model with appropriate memory structure is a stronger reference climatology (null model) for skill score calculations and assessment of predictor utility.

Second, by representing the auto-correlation of daily-scale precipitation, we explicitly start separating stochasticity from processes on weather time-scales and climate time-scales. Climate and weather are often difficult to extricate from one another, partially due to conflicting definitions. Climate is sometimes defined as “average,” “expected,” or “marginal” weather; sometimes as boundary conditions acting upon the atmosphere; and sometimes as low-frequency processes (as compared to high-frequency weather). Weather, similarly, can refer to the atmospheric state, that atmospheric state with some low-frequency climate signal removed (i.e., as anomalies from a slowly varying climate signal), or broadly anything with persistence shorter than the atmosphere’s chaotic time-scale on the order of weeks. By explicitly representing auto-correlation in precipitation data, we characterize atmospheric persistence as partially deterministic, in the same sense that modelers represent the climate state as partially deterministic by calculating the annual seasonal cycle explicitly in weather generator models. Thus, our model is not only a more strict climatology for weather forecasts, but also a null model for climate variability in that it represents some interannual-scale variability via weather time-scale processes.

Since the probability of precipitation is highly dependent on the spatial scale at which an observation is made, we would expect the performance of occurrence-driven data models to diminish as spatial scale increases. Specifically, the class of chain-based occurrence models, often used in stochastic climatological simulations, may represent a robust climatological null when using station data accumulated over the time-scale of the model, but display significant “underdispersion” at longer temporal accumulation periods; this underdispersion is expected to become more pronounced at larger spatial scales, due to the models’ inability to represent useful predictive auto-correlation in occurrence when it rains nearly every day. Similarly, models which represent auto-correlation in intensity are of limited utility at small spatial scales.

2 Methods

2.1 Overview and Data

To capture the correlation structure of daily precipitation without decoupling occurrence and intensity processes, we combine an inverse-CDF transformation of each day’s data with a generalized non-parametric auto-correlation model using Gaussian kernels. The transformation is known as a “rank-based inverse normal” transformation (Akritas, 1990; Cai et al., 2016), and it allows us to work in an unbounded domain, reducing some of the common complications inherent in both bounded and zero-inflated data. It also allows us to provide correlation structure between wet and dry days in the same manner that we represent the correlation structure between serial wet days. The kernel model is used to represent the joint probability density of m-day series of precipitation values without relying on the assumption that the covariance structure is multivariate normal (as in the typical AR time-series paradigm), or even that it follows any specific family of parametric distributions. By using Gaussian kernels, the kernel model is a specific instance of the broad class of Gaussian process models, common particularly in machine learning applications due to their flexibility and somewhat analytically tractable nature (Rasmussen & Williams, 2006).

We use precipitation data at three scales: station data from the Global Historical Climatology Network (Menne et al., 2012), 1/4 degree gridded data from the Climate Prediction Center's (CPC) Unified Gauge-based Analysis of Daily Precipitation over the Continental United States (Chen & Xie, 2008), and a 1 degree gridding of the same CPC data (U.S. Climate Prediction Center, 2015). In each case the data is from the years 1948–2004, inclusive.

2.2 Fitting

The model fitting procedure is shown in Figure 1. To preserve the seasonal patterns of precipitation occurrence, intensity, interannual variability, and short-term correlation structure, we fit a model to each day of the year for a given location. For any given day's model, we use data drawn from a width-p window around that day (the "pooling window"; $p = 31$ days in the subsequent analysis) to improve our estimates of the serial correlation (see Figure 1a). For each of the days in the pooling window, using all N years of observations, we use that day's observation and the $m-1$ previous observations to form a length m vector of serial precipitation. Each of these vectors represents a single point in an m-dimensional space to which we will fit a joint probability density function (PDF) of precipitation and its lags. The marginal distributions are simply empirical histograms of daily observed precipitation, and the conditional distribution of the m^{th} dimension given the other $m-1$ is the probability density for a single day given that you have just observed a specific $m-1$ days of precipitation.

The most common means of quantifying the correlation structure of serial data is the auto-regressive (AR) model, the simplest member of the ARCH/ARFIMA/ARMAX families. The AR(1) model fits a bivariate normal distribution to 2-dimensional vectors of observations, usually either maximizing the likelihood of the joint distribution or the likelihood of the conditional distribution. Since daily precipitation clearly does not fit the assumptions of normality, a typical AR-type model is inappropriate. The multivariate normal (MVN) distribution of the typical AR model can of course be replaced with other multivariate parametric models, or can be represented more empirically using a multivariate binned histogram (or probability mass function) to capture exotic distributions (examples given in Wilks & Wilby, 1999). However, for zero-inflated data (such as precipitation from weather stations), the size of the bin has a strong impact on the correlation structure of the model. Smaller bins will assign more likelihood weight to occurrence processes, and larger bins will assign relatively larger weight to intensity, and any finite bin width is effectively an arbitrary trade-off in the role of occurrence in the model.

Even more problematic than selecting a bin size, is that for any bin size the climatological occurrence frequency has a very large impact on the joint distribution (and thus model parameter likelihoods), making comparisons of parameters or simulations between different locations or the same location at different spatial scales (any gridded scale or point measurements) impossible. The same issue arises for other parametric distributions (such as a multivariate gamma): datasets with more dry days will lead to huge inflation of likelihood weight towards those identically-valued dry days, essentially forcing a continuous model to emulate a Bernoulli model as best as possible to maximize the zero-inflated likelihood.

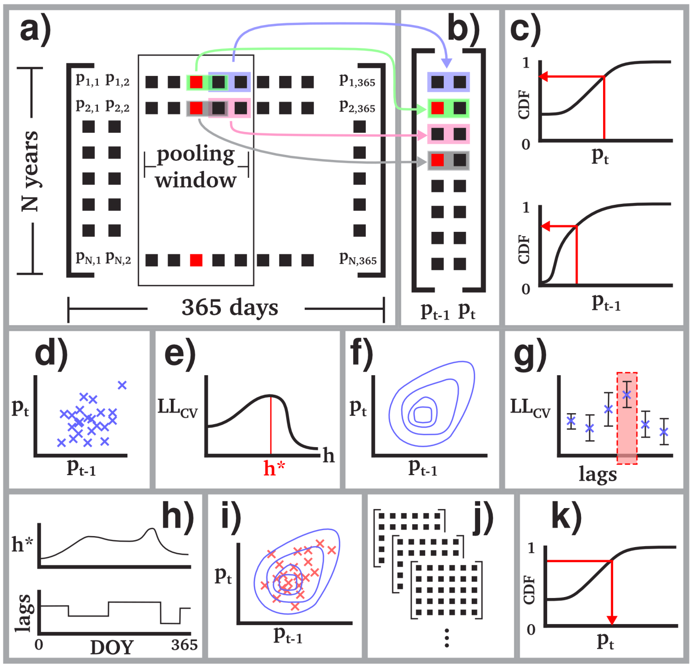

Figure 1. A schematic showing the steps for fitting the kernel-auto-regressive weather generator and simulating precipitation data. a) N years of daily data form the basis for 365 daily models. For each day (e.g, the column in red), a pooling window is used to optimize covariance estimation. b) All m-day serial vectors of observed precipitation from the pooling window are used to form an m-dimensional (2D shown) empirical distribution of precipitation p_t and the preceeding day's precipitation p_{t-1}. (c) Each column (day) is rank-transformed so that all marginal distributions of the joint distribution (d) are exactly Gaussian. (e) A bandwidth, h, is selected using cross-validation to create a kernel density (f) from the observations. (g) Selecting models with zero lags (1D) to five lags (6D) using cross-validation for each day yields 365 selected model dimensions and 365 optimal bandwidths (h), which together comprise the model for the location. (i) Simulating one day at a time using the corresponding kernel model and conditioning on the previous $m - 1$ days produces ensembles (j) of N-year stochastic precipitation data in the CDF-transformed domain, which are then back-transformed (k) for analysis.

To circumvent this inherent dependency of the simplest probability models on occurrence frequency, we transform our data into an unbounded domain and "un-inflate" our zero-inflated data. For our $(N{\cdot}p)$ by m matrix of observations for a given day (see Figure 1a–b, in which $m = 2$, or a 1-lag model), we transform each of the m columns through a rank-based univariate inverse normal CDF $\Phi^{-1}(\cdot)$, assigning the smallest observed value $\Phi^{-1}(1/(Np+1))$ and the largest observed value $\Phi^{-1}(1-1/(Np+1))$ so that each column of the transformed matrix is exactly normally-distributed (Figure 1c). Duplicate values (notably zeros) can be assigned random relative ranks (so as to be asymptotically uncorrelated with each other), and are handled as special cases when calculating likelihoods. Zeros, for example, will comprise the left tail of a univariate distribution, in randomly-assigned order.

In the CDF-transformed domain (Figure 1d), each dimension of the data is marginally normal, but the joint distribution is not necessarily MVN. To allow for as flexible a representation of the covariance structure as possible, we represent the joint distribution between the m days of serial observations using a kernel density. Since all dimensions of our data are scaled identically, we use a simple spherical Gaussian kernel, which has one scalar parameter — the bandwidth, h; using more complex multivariate kernel bandwidths would impose unwanted additional covariance structure beyond that directly represented by the empirical relationship between precipitation and its lags. We select the optimal kernel bandwidth for that day of year and for each number of lags ($1 \leq m \leq 6$ in this analysis) using cross-validation (Figure 1e). We perform a nested grid search of possible bandwidths and use a leave-out 20% repeated-random-subsampling cross-validation scheme. The likelihood to be optimized is that of the validation data using the full joint PDF of the training data kernel model. By selecting a bandwidth, we have selected a probability model for our data (Figure 1f).

Once we select an optimal bandwidth for each potential number of lags, we then pick the optimal number of lags using a second cross-validation step (Figure 1g). The entropy of the joint distribution scales with the dimension m, and so the comparison between models of differing dimension is scaled by the dimension of the model. Alternatively, one could compare the univariate conditional likelihood of the last day's precipitation given $m{-}1$ previous days for a more prediction-focused approach to model selection. The model with the highest mean likelihood across all repeated subsampling cross-validations is selected, and the dimension of that model becomes the dimensionality of the kernel model for that day of year. The dimension and bandwidth are the two critical parameters for each daily model, and the full model for a dataset at any location is specified by 365 dimension values and 365 corresponding bandwidth values (Figure 1h).

When determining bandwidth, likelihoods are calculated as typical for a Gaussian kernel model. Given N d-dimensional kernel means in the $N{\times}d$ matrix $\mathbf{T}$, a bandwidth h, and a d-dimensional vector x at which to calculate the density or likelihood, the likelihood function is

$$f(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{h(2\pi)^d}}\exp\left[-\frac{1}{2h}(\mathbf{T}_i - x)(\mathbf{T}_i - x)'\right] \tag{1}$$

where the summation is over all N kernel means, or equivalently all N rows of $\mathbf{T}$. The log-likelihood for a set of M observations is then

$$LL = \sum_{i=1}^{M}\ln f(x_i) \tag{2}$$

When calculating likelihoods used to determine the appropriate number of lags, since all dry day zero-values are equivalent, we force the distance $\mathbf{T}_i{-}x$ to be zero in any dimension where both $\mathbf{T}_i$ and x correspond to dry days.

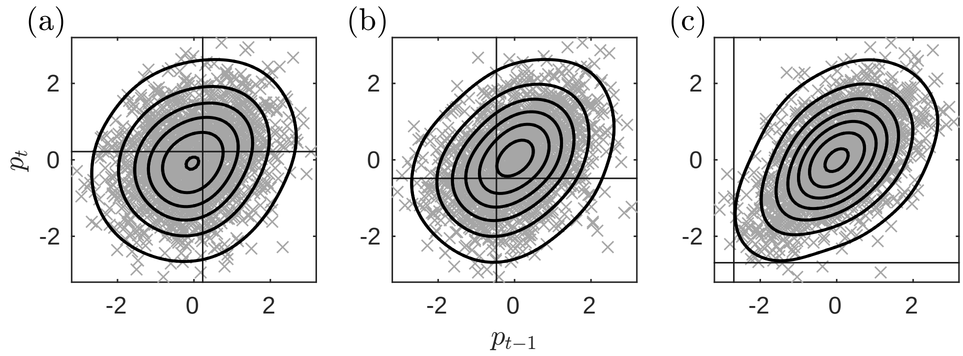

Figure 2. Joint probability densities for daily precipitation, p_t, and the previous day's precipitation, p_{t-1}, for day of year 180 (June 29) at Fairhope, AL using a 31-day pooling window. The spatial scale of the data increases from left to right: (a) shows the joint density for a single GHCN station, (b) for the co-located CPC Unified 1/4° gridded data, and (c) for the 1° CPC Unified data. Individual two-day observations are represented as "x-es" in Gaussian z-score units, contours show equal density levels, and vertical/horizontal lines show the cut-off threshold for precipitation occurrence: points above and right of the lines are wet days, below and left are dry days, and the dry-dry two-day pairs have no covariance structure. The marginal distributions are, by design, identically $\mathcal{N}(\mu = 0, \sigma = 1)$. As the occurrence probability increases (from left-to-right), the "wet quadrant" covariance structure becomes the more dominant feature of the joint density as a whole.

Figure 2 shows the 1-lag (2-dimensional) joint distribution for precipitation at Fairhope, AL at three different spatial scales for day of year 180 (June 29). In Figure 2a (station data), the majority of observations are dry days (left and/or below the threshold lines), which are uncorrelated with each other, but still provide the appropriate covariance between occurrence processes and intensity processes. At the 1/4°-scale (Figure 2b), occurrence frequency is higher than 50%, and the positive correlation structure of wet-day/wet-day pairs is more evident. Additionally, since vertical cross-sections give the conditional distribution of p_t given p_{t-1}, we can see that the heaviest wet events (the upper-most points in Figure 2b) tend to occur after other heavy wet days. At the 1°-scale (Figure 2c), dry days are rare and two day dry spells are non-existent in the observational data (and highly unlikely in the kernel model, though not impossible). The relatively symmetric PDF shows that the light-then-heavy pattern is essentially as probable as the heavy-then-light pattern, and that dry days are likely to be followed by light precipitation days.

2.3 Simulation

Simulation of precipitation is performed in the CDF-transformed domain, where the correlation structure is more simply represented, then transformed back through an inverse CDF transformation to the domain of the actual observations. The key step in the simulation process is conditioning the model for the given day of the year on the simulated values for the previous $m-1$ days so that the daily correlation structure is maintained.

The dimensionality of the model changes from day to day, but the maximum number of days used in conditioning is one less than the maximum dimensionality of the model over all days of the year. To keep track of this, a vector of length $\max(m) - 1$ is used as a buffer to store the relevant conditioning data. Since we initially have no data to condition with, the buffer is set to a random draw of climatological values for the appropriate days of the year, and then a one-year burn-in period is used (and later discarded) to represent proper correlation statistics.

For each day of the year, the marginal probability of producing the $m-1$ values in the buffer (marginalizing over the single dimension representing precipitation on the current day) is determined for all of the data-point/kernels in the joint PDF. This can be done either in the full data/kernel space, or can be thought of as a two-step process: (1) first randomly selecting a single data point/kernel and then (2) simulating a random point from that kernel's conditional distribution. With Gaussian kernels, the second of these approaches is computationally simpler. If the model is one-dimensional for that day (no lags/memory), the marginal probability used to select a single kernel is uniformly 1/N, where N is the number of data points/kernels. Otherwise, a single data point/kernel is selected stochastically with weightings based on the marginal probabilities. Following this, the conditional probability is determined for the previous $m-1$ days' rain. Since the kernels are Gaussian and spherically symmetric, the conditional PDF is a univariate normal distribution, the conditional mean is simply the m^{th} (last/current) value of the data point used as the multivariate mean, x_{ij}, and the (scalar) conditional variance is just the bandwidth, h_i, squared. Thus, the simulated precipitation (in the CDF-transformed domain) is just a random draw from $\mathcal{N}(\mu = x_{ij}, \sigma^2 = h^2)$, where i corresponds to day of the year, and j corresponds to the j^{th} (last) entry in an m-dimensional vector, x, representing the selected observed m-day precipitation data point/kernel.

Following simulation, the data are transformed back to the observational domain by interpolation using the original data and its CDF-transformed values. Before transformation, the simulated data are re-standardized to ensure proper variance representation (see Appendix A for further details). Any values below the no-rain cut-off in the original data are converted to zeros, and any values larger than the largest value in the observational data set need to be extrapolated. We use the tail of a gamma distribution to fit the extrapolated values. The gamma distribution is fit to the wet days for that day of year, we align the z-score of the largest observed value in the CDF-transformed observational data with the corresponding quantile of the gamma distribution; the extrapolated values are mapped to the appropriate part of the upper tail by normal-to-gamma quantile matching.

Simulations can be run for as many years as necessary to calculate asymptotic statistics, or can be run in independent ensemble modes (e.g., in multiples of the observational record length) for statistical assessment of climatological phenomena.

3 Results

The kernel-auto-regressive model (KA) was fit to data at three locations: Fairhope, Alabama; Blue Hill, Massachusetts; and Forks, Washington. In each location, separate annual models (each comprised of 365 daily models) were fit for each of the three data sources (GHCN, CPC-1/4°, and CPC-1°). In addition, an advanced chain-based model — referred to as the "Occurrence Markov Chain" or OMC model (Short Gianotti, 2016), also comprised of 365 daily submodels — and a no-memory, "zero-lag" (ZL) occurrence/intensity model were fit to the same datasets for model comparison. The OMC model uses a variable order Markov chain to represent the auto-correlation in the occurrence process and a flexible five-parameter gamma-gamma mixture model to represent intensity, also with a 31-day pooling window for parameter estimation. The chain order (number of lags) is determined for each day of the year using the corrected Akaike Information Criterion (Hurvich & Tsai, 1989), and the parameters for the intensity model are selected for each day of the year by maximum likelihood estimation (see Short Gianotti et al., 2014, for further details). The zero-lag model uses the same distribution family for intensity as the OMC model, but daily occurrence does not depend on the previous days' precipitation and simply follows the climatological probability of occurrence for that day of year (within the 31-day pooling window).

Each of the three models was used to simulate 1000 57-year ensembles of stochastic precipitation at each of the three spatial scales at each location. By design, all of the models asymptotically reproduce the probability of occurrence, mean intensity, and variance of intensity for every day of the year. Each of the models create interannual variability stochastically, but none of them represent climate variability processes, and thus are likely to be "under-dispersed" in their representation of accumulated totals relative to the observations (Katz & Parlange, 1998; Gianotti et al., 2013; Short Gianotti et al., 2014; Anderson et al., 2015a, 2015b). Additionally, the kernel-auto-regressive model and the OMC model each represent serial correlation (although the OMC only represents correlation in occurrence), and so precipitation totals accumulated over multi-day-to-multi-year periods will likely be more under-dispersed for the zero-lag model than for the KA or OMC models. If accumulating precipitation over multiple days (or weeks, months, years, etc.) the mean accumulated totals from the simulations match the observations asymptotically for each model. The KA model is able to represent any processes captured in the OMC model, but with more flexibility, and the ZL model is explicitly a restricted version of the OMC model with no memory, so we would expect the KA model to be most able to represent complex variability structure, followed by the OMC model, and then the ZL model.

Figure 3 shows empirical distributions of December–February seasonal precipitation for the observed data and simulations from the three models for each of the nine datasets (three locations times three spatial scales). At the station level (top row), while all models capture the observed seasonal means, all models similarly miss some of variability characteristics of the observations, presumably because none of them represent interannual variability other than through zero-to-six day correlation structure. Any variability caused by slower processes (such as climate modes) will not be well represented. Notably, for Forks, WA (the wettest location), the OMC model outperforms the zero-lag model, and the kernel-auto-regression model outperforms both of the simpler models. At the 1/4-degree scale (CPC-1/4°, second row), the same pattern holds, but with more notable performance improvements for the kernel-auto-regressive model. This is not surprising, as the increased frequency of occurrence makes the proper modeling of the intensity auto-correlation more important for characterizing the patterns of synoptic scale precipitation events. At the largest spatial scale (CPC-1°, third row), the KA model's performance is enhanced further, while the OMC model is effectively no better than the zero-lag model. The existence of auto-correlated memory structure encoded in precipitation intensity is evident, particularly for wet locations and at coarse spatial scales.

To investigate the role of the temporal accumulation scale as well as the spatial scale, we can compare the interannual variance of accumulated totals over a range of accumulation period lengths (effectively comparing the variance of the PDFs in Figure 3 for different sub-seasonal to annual windows). Figure 4 shows the interannual variance for each model/location/spatial-scale as a function of accumulation period, scaled (divided) by the period length, and averaged over the annual cycle. In each of the nine plots, the zero-lag model (a basic climatological null) shows essentially no response to accumulation period; this is because with no serial correlation, the observations are independent, and so the variance of the sum of the precipitation is equal to the sum of the (averaged) daily variances, which is constant. These lines lie at the same value as the annual average of the 365 daily variances from the observations. In the upper two rows (GHCN and CPC-1/4°), the OMC model represents more interannual variability than the ZL model for periods longer than a single day, but at the 1°-scale (third row), the daily occurrence probability is effectively 1, and so there is no useful memory structure in occurrence for improving the multi-day variability representation. The KA model consistently outperforms the OMC and ZL models, but seems to asymptote around 30 days, while the observa-

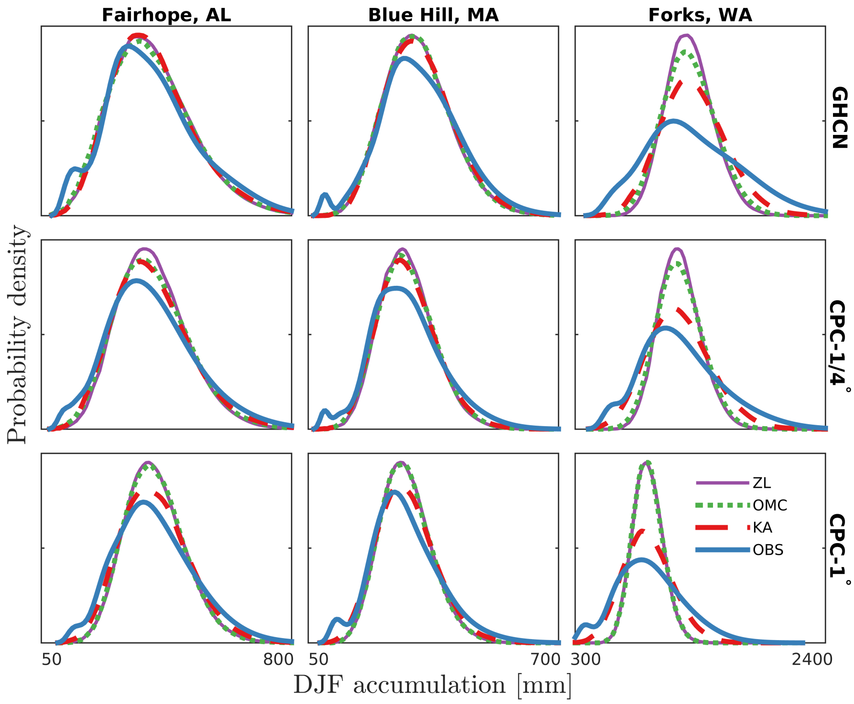

Figure 3. Probability density functions (PDFs) of December–February precipitation totals at Farihope, Alabama; Blue Hill, Massachusetts; and Forks, Washington: observations and three models: a zero-lag model with no daily-scale correlation byond climatology (ZL), a Markov chain based model which represents memory in occurrence processes (OMC), and the kernel auto-regressive model which represents memory in occurrence and intensity (KA). Blue lines correspond to observations (OBS) from each of the three data sets (GHCN, CPC gridded at 1/4°, and CPC gridded at 1°). All three models underestimate the variability of accumulated precipitation although they each are fit to optimally represent precipitation at the daily scale. At the wettest location (Forks, WA) and at larger spatial scales (lower two rows) the KA model's ability to represent the serial correlation in both intensity and occurrence enhances its ability to represent the 57-year distribution of winter precipitation totals.

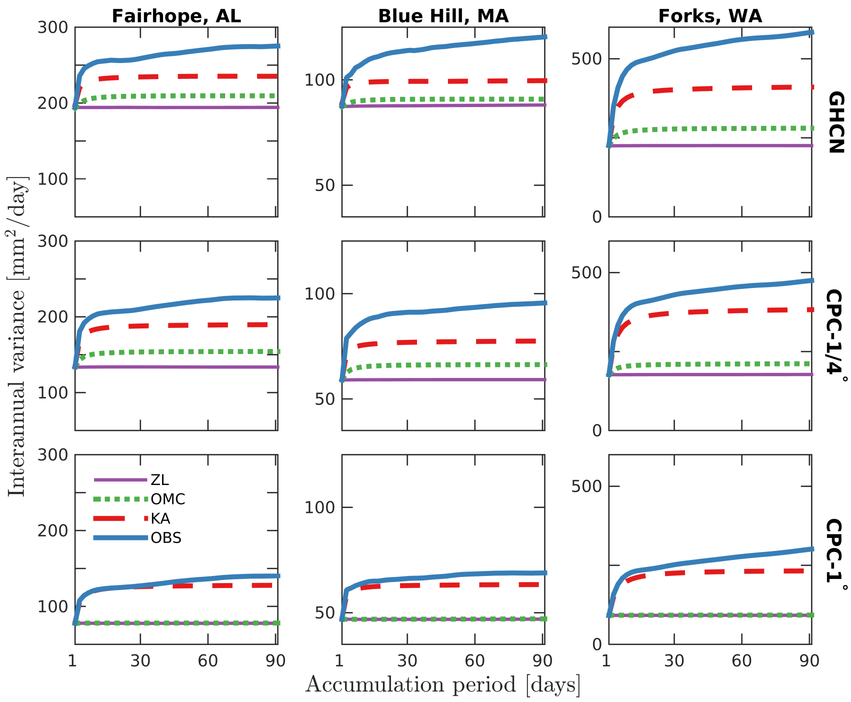

Figure 4. Comparison of different models' abilities to represent the variability of precipitation as a function of both spatial scale and temporal scale. As in Figure 3, ZL is a zero-lag model with no daily-scale correlation beyond climatology, OMC is a Markov chain based model which represents memory in occurrence processes, and KA is the kernel auto-regressive model which represents memory in occurrence and intensity. Blue lines correspond to observations (OBS) from each of the three data sets (GHCN, CPC gridded at 1/4°, and CPC gridded at 1°). At larger spatial scales, the OMC model's occurrence-based memory structure is no better than the climatological null (ZL model). The KA model, alternatively, seems to represent more of the observed varaibility at larger spatial scales, suggesting that either short-term "weather-scale" variability is more dominant at larger spatial scales relative to longer-term "climate-scale" variability, or that the model fit is more effective at larger spatial scales for a fixed-length data record. Variance values are scaled by the accumulation period, and averaged across the annual cycle.

tional data continues to increase in variability. This is not surprising, as the KA model does not represent any explicit drivers of variability at those time scales, and we know that there are earth system processes that would lead to variability at those scales (ranging in time scale from the Madden Julian Oscillation to the El Nino Southern Oscillation, multidecadal oscillations, and secular trends).

Comparing different rows, we see that there is less variability in the observations at larger spatial accumulation scales (roughly a factor of two difference in Var(OBS) between GHCN and CPC-1° at all three locations), and that the KA model represents more of the observed variability at larger spatial scales than at smaller spatial scales (by comparing the distance between the KA and OBS lines relative to the ZL null).

3.1 Water Balance Impacts

To see the impacts of better representation of memory processes in WGMs, we can use simulated precipitation time series to drive a water balance model. In this example application, we use the method of Akbar et al. (2019), which prescribes evapotranspiration and drainage losses as a function of surface soil moisture (a "bucket model" formulation), which in turn is driven by precipitation. Evapotranspiration follows a sigmoidal function of surface soil moisture; drainage is represented by a Clapp-Hornberger power law relation. Parameters, including the thickness of the surface layer are determined through a adjoint approach using surface brightness temperature data from the Soil Moisture Active/Passive (SMAP) satellite mission (Entekhabi et al., 2010; O'Neill et al., 2016). The parameter estimation seeks to minimize the combined errors in soil moisture retrievals from the brightness temperature data and a precipitation-driven water balance.

Using our observed and simulated time series of precipitation (CPC1) from each of the WGMs at Fairhope, AL, we obtain daily time series of surface soil moisture and fluxes from the surface. Figure 5a shows probability densities (PDFs) of mean annual volumetric surface soil moisture (unitless) as simulated from precipitation forcings from observations, and each of the ZL, OMC, and KA models. Parameters for the model (see Akbar et al., Equations 3 and 4) are $a = 0.423$ mm/day, $b = 1.43$, $c = 81.04$ mm/day, $d = 20.0$, $dz = 395.9$ mm, and porosity = 0.46.

The different precipitation forcings—despite having identical mean precipitation rates—lead to differences in the annual average surface soil moisture of about 3%, with the observations and KA model showing slightly lower soil moisture values on average than the simpler ZL and OMC precipitation models. This corresponds with higher modeled evapotranspiration rates for the observations (roughly 1.5% higher for the observations and KA model relative to ZL and OMC), but also noticeably different distributions of evapotranspiration (Figure 5b). The probability of annual ET being less than 1250 mm is 12.3% and 11.5% for the observations and KA model respectively, and 5.6% and 5.0% for the ZL and OMC models. Similarly, the probability of annual ET exceeding 1750 mm is 12.3%, 9.6%, 3.8%, and 3.9% for the observations, KA model, OMC model, and ZL model, respectively. This suggests that the improvement in representing interannual precipitation variability in the KA method substantially improves our ability to represent interannual variability in downstream processes, such as soil moisture and surface fluxes, particularly in representing extreme events such as drought or flood conditions.

3.2 Confidence intervals for trend estimation

As a demonstration of the utility of climatological nulls, we look at an example of trend detection in the daily precipitation data. Using the full 1/4° gridded CPC time series for Blue Hill, MA we first calculate a 365-day annual climatology for mean precipitation. This climatology is effectively identical for the observations and for simula-

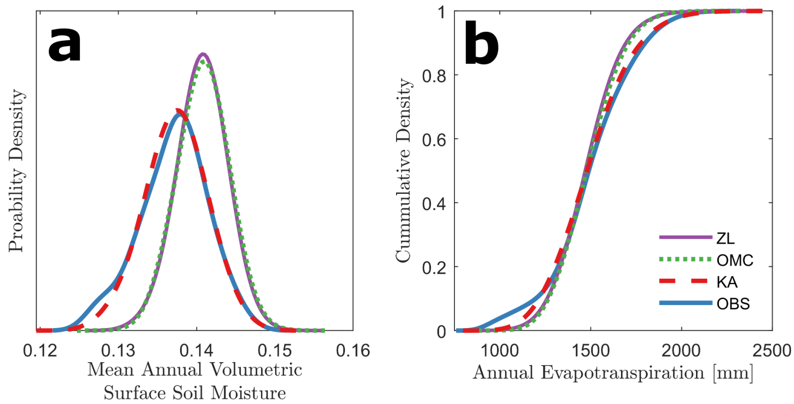

Figure 5. Distributions of surface soil moisture and evapotranspiration at Fairhope, AL driven by observations and weather generator models at coarse spatial scale (CPC1). a) Probability density functions for mean annual soil moisture as output from a daily-scale surface water balance model. Despite all models matching the interannual mean and variance for precipitation for each day of the year, the KA model better represents memory processes than the ZL or OMC models, and thus better incorporates stochastically-driven variability in mean annual soil mositure and better follows the soil moisture PDF as driven by precipitation observations. b) Cummulative Density Functions (CDFs) of annual evapotranspiration driven by the surface water balance model again show a better match between the KA model and observations. The probability of Annual ET being less than 1250 mm is 12.3% and 11.5% for the observations and KA model respectively, and 5.6% and 5.0% for the ZL and OMC models. Similarly, the probability of Annual ET exceeding 1750 mm is 12.3%, 9.6%, 3.8%, and 3.9% for the observations, KA model, OMC model, and ZL model, respectively.

tions from each of the three models (ZL, OMC, and KA) by design. By subtracting the daily mean values we obtain an anomaly time series for the daily precipitation observations. The important differences between these anomaly time series will in the degree and quality of nonstationarity – the observations will show complex autocorrelation structure (and not simply that represented by a Guassian AR model) pertaining to storm-track processes and multi-day land-atmosphere feedbacks. The ZL model will be statistically stationary; the OMC model will capture the nonstationarity due to occurence triggering processes, and he kernel autoregressive model will capture both occurrence and intensity autocovariance. Similar analysis can be performed on seasonal or annual totals without removing the daily climatological mean, and often is for the sake of obtaining more normal residual distributions. The simple least-squares regression line through the 57 years of precipitation data has a slope of +21.2 mm/year per decade. t-distribution based confidence intervals using standard assumptions of residual normality suggest that this is significantly different from zero using $\alpha = 0.05$ (in either direction). It is clear, of course, that when looking at daily precipitation values, the individual observations are not normally distributed around the mean trend line, are heavily skewed towards zero-values, and are likely not independent samples, suggesting that this may not be a robust test of trend significance.

Alternatively, since the KA, OMC and ZL models have no representation of climate variability or trends, we can use simulations from these climatological null models to assess the significance of the observed trend. For each model, we simulate 1,000 ensemble members of 57 years of daily data, determine 1,000 corresponding linear trends in the anomaly time series, and determine a distribution of possible stochastically-generated slope magnitudes. If the observed slope is sufficiently far in the tails (below the 5th or above the 95th quantile), then the slope is significant. Using the ZL and OMC models, the slope appears to be significant (ZL confidence interval: [-19.2, 19.9] mm/year per decade; OMC confidence interval [-20.7, 19.9] mm/year per decade). Using the KA model, the trend is *not* significant ([-22.6 and 21.3] mm/year per decade), suggesting that the climatological null used to determine significance of trends can have an impact in terms of climate signal detection. For an observed trend to be marked as likely due to climate change or climate variability, we want to ensure that it is not the product of stochastically-probable strings of auto-correlated anomalies. The KA model, by better representing the combined memory in short-term precipitation occurrence and intensity provides a more appropriate (in this case more stringent) test of expected stochastic trends than advanced WGMs using occurrence alone.

We can then apply this approach to determine the significance of trends across the Contiguous United States (ConUS). Figure 6 shows the magnitude of trends in precipitation in aggregated annual and seasonal precipitation. Larger markers show statistical significance using the distribution of slopes generated by 1000 "weather only" realizations from the kernel autoregressive model ($\alpha = 0.05$ increasing or decreasing). Most significant trends in annual precipitation are for increases in the eastern ConUS (notably not in the Southeast), with relatively few decreases, mostly located in the Cascades. Winter (DJF) trends appear to follow the El Niño/Southern Oscillation (ENSO) precipitation pattern, which corresponds to a slight increase in DJF ENSO indices over that period. The autumn (SON) shows the most and strongest significant trends, with wetting over much of the area east of the Rockies, particularly along the southern Mississippi drainage.

3.3 Drivers of interannual variability

Well-tuned weather generator models can also allow us to partition the sources of variability in observed data. We can think of precipitation as having (1) some secular trend (as in Figure 6) which contributes to differences year-to-year from the mean, (2) interannual variability driven by interannual-to-decadal climate modes (ENSO, etc.), and (3) varying because of year-to-year aggregated differences in daily weather phenomena

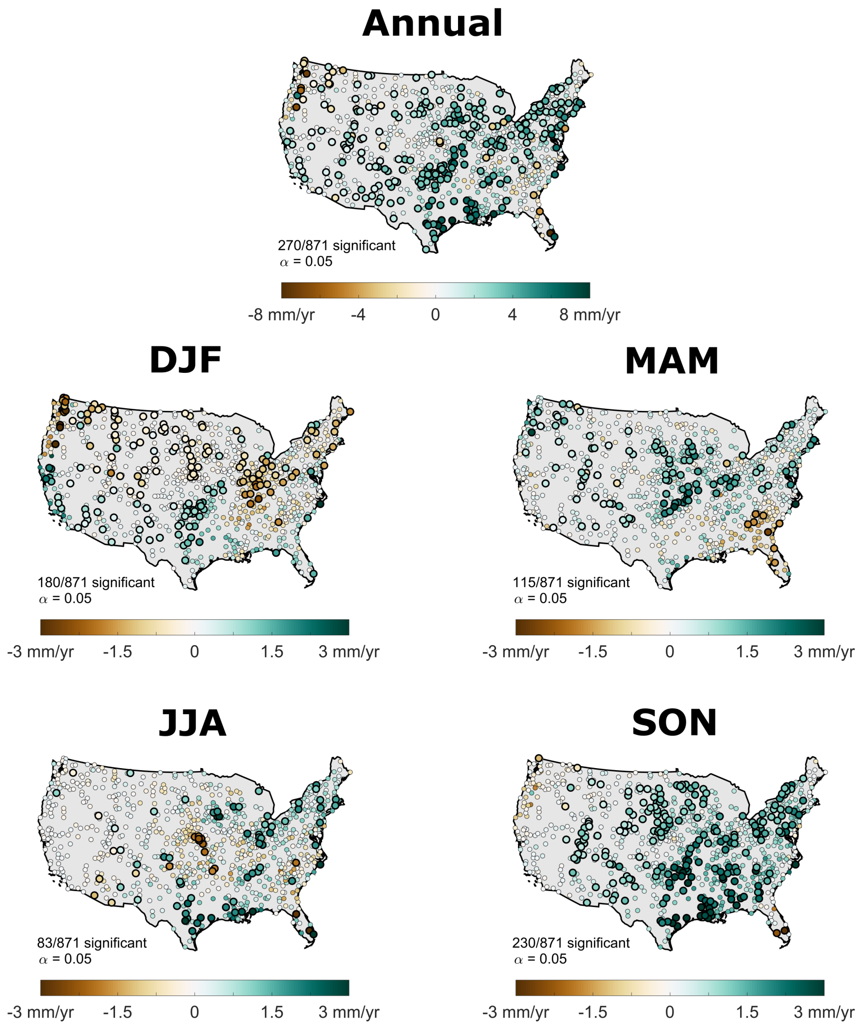

Figure 6. Trends in annual and seasonal precipitation from daily GHCN data (1948–2004). Larger markers indicate significance at the $\alpha = 0.05$ level (in either direction) relative to the distribution of slopes generated by 1000 "weather only" realizations from the kernel autoregressive model. Ratios show the fraction of stations with significant trends.

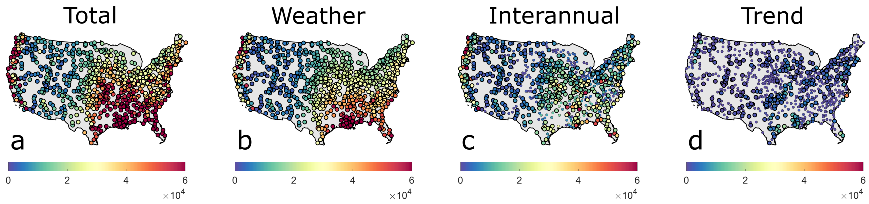

Figure 7. Interannual variance in annual precipitation from daily GHCN data (1948–2004) [(mm/yr)^2]. a) Total interannual variance of annual precipitation. b) Variance of weather-scale processes from 1000 annually-stationary kernel autoregressive weather generator simulations. c) Non-trend interannual variability, calculated as the variance of the detrended observations minus the variance of the detrended weather simulations. Larger markers indicate significantly different from zero at the $\alpha = 0.05$ level using the distribution of 1000 weather simulations. d) Interannual variance due to linear trend in observations. Larger markers denote a significant ($\alpha = 0.05$) trend as in Figure 6.

which are not driven by the persistence of climate modes. Figure 7a shows the total interannual variance of precipitation, which to large degree scales with the mean. Simulations from the kernel autoregressive model capture much of this variability (Figure 7b, typically 50–95%, with larger values in the East). The KA model, however, by design does not represent the interannual variability due to trends, nor does it represent interannual variability from climate modes.

The trend contribution is calculated from the linear regression as the explained sum of squares divided by the number of observations (also equivalent to the total variance times the coefficient of determination; Figure 7d, with significance determined as in Fig. 6). This contribution is typically small, on the order of 15% of the total variance or less. The remaining interannual variability (Figure 7c) is that variance which is neither determined by the trend, nor is it able to be captured by the aggregated non-stationary anomalies of weather-scale processes as represented in the KA model. To estimate this contribution, the observed annual data are detrended, as are each of the 1000 weather-only simulations from the KA model. The remaining interannual variance of each simulation is subtracted from the remaining interannual variance of the observations, which gives a 1000-member distribution of unexplained variances for each location. the mean of these (always positive) is shown in Figure 7c, and significance is determined by whether zero is below the 0.05 quantile. Where significant, this climate variability is typically 30–50% of the the total variance. See Supplementary Figure 1 for these data in a normalized “fraction of variance explained” format.

To see if certain portions of the year are subject to differing drivers of variability, we can perform the same decomposition using seasonal accumulated totals. Figure 8 shows the “Total,” “Weather”-scale, “Interannual” (climate mode), and “Trend” components of the variance for each of DJF, MAM, JJA, and SON. The “Weather,” “Interannual,” and “Trend” components are shown normalized by the total variance to give the fraction of variance explained. As with the annual totals, weather-scale processes—that is, those easily captured by weather generator models with only a few days of memory—are the primary source of interannual variability in seasonal precipitation, typically explaining more than 50% of the observed variability. Interannual processes, which we hypothesize can be largely explained by known modes of annual-to-decadal climate variability, explain something around 1/3 to 1/2 of interannual variability in large swaths of the ConUS, but more in the winter, and less in SON. The trend components explain a relatively small portion of the total variability in precipitation over this time period,

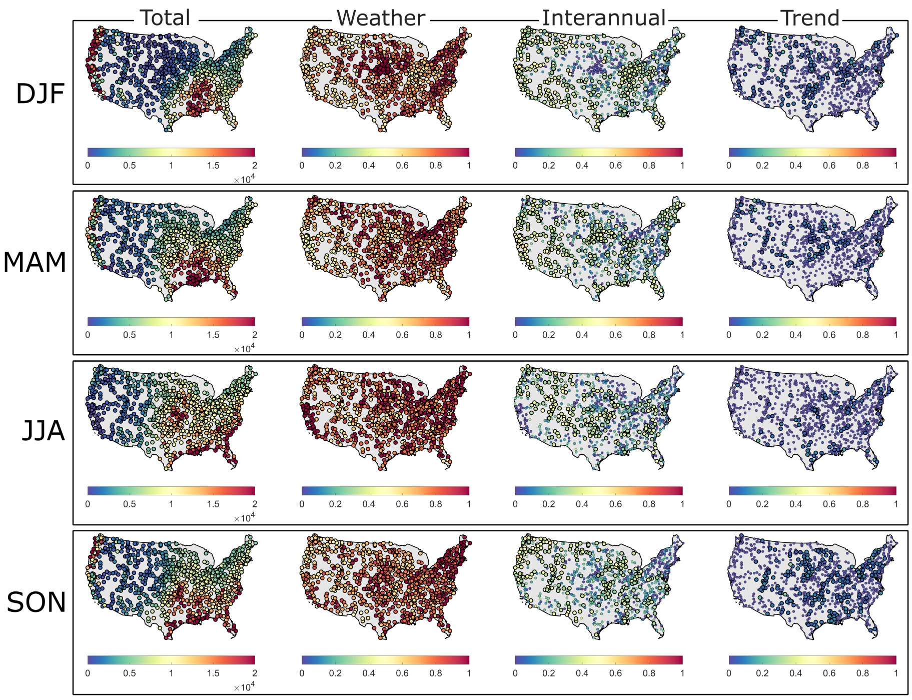

Figure 8. Interannual variance in seasonal precipitation from daily GHCN data (1948–2004). "Total" column shows total interannual variance of seasonal precipitation $[(\mathrm{mm/yr})^2]$ for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). "Weather" column shows variance of weather-scale processes from 1000 annually-stationary kernel autoregressive weather generator simulations as in Fig. 7b, but normalized by the "Total" column (i.e., fraction of variance explained by weather-scale phenomena). "Interanual" column shows non-trend interannual variability, calculated as the variance of the detrended observations minus the variance of the detrended weather simulations as in Fig. 7c, but normalized by "Total." Larger markers indicate significantly different from zero at the $\alpha = 0.05$ level using the distribution of 1000 weather simulations. "Trend" column shows interannual variance due to linear trend in observations as in Fig. 7d, but normalized by "Total" (i.e., the R^2 metric of the time-series regression). Larger markers denote a significant ($\alpha = 0.05$) trend as in Figure 6. See Supplementary Material for un-normalized maps.

nearly always less than 20% (although variations of these magnitudes can of course have major impacts on regional water balance).

4 Discussion and Conclusions

The kernel-auto-regressive model is better able to capture the variability of accumulated precipitation than an advanced occurrence-chain-based model (OMC). Even when the OMC model was able to separately condition intensity on previous occurrence patterns to provide additional memory structure, the added benefits were almost never justifiable from an information criterion perspective for any day at hundreds of U.S. locations (see Short Gianotti et al., 2014). And while most of the memory in station data is in the occurrence signal (Short Gianotti et al., 2014), once we look at larger scales local occurrence information is lost, variability is reduced, and methods that represent occurrence and intensity separately will under-represent the daily correlation structure of precipitation.

Beyond its uses in establishing variances and potential predictability estimates for precipitation at varying scales, or as a stochastic weather generator model, the kernel auto-regressive model can be used to represent conditional probabilities and empirical probability distributions for any regularly-sampled variable — particularly those with some degree of auto-correlation or non-stationarity that may be poorly represented by a multivariate Gaussian correlation structure. Examples include weather generators for meteorologic variables (temperature, wind speed, pressure levels, etc.); conditional distributions of one earth system variable on another and/or their lags (evaporation given wind speed, convective available potential energy given net surface radiation, etc.); and representations of auto-correlation in non-linear biological and ecological processes (vegetation transpiration rates, population dynamics, cellular metabolic processes, etc.). The kernel auto-regressive method also provides (1) a means of representing complex auto-correlation in advanced modeling situations, e.g., empirical distributions for use in mixed process/data models in machine learning (the KA method with Gaussian kernels is a class of Gaussian Process Models); (2) a means of assessing the degree of (non-)stationarity in time series analysis; and (3) a means of reducing assumptions inherent in common auto-regressive models (e.g., providing a means of properly dealing with all members of "Anscombe's Quartet" and similarly devious statistical relationships).

We demonstrate the use of our improved WGM for better confidence intervals for precipitation trends (Fig. 6). These confidence intervals are specifically designed to define as significant those trends which are not likely to be due to weather fluctuations. We show that although the trend component of precipitation variability is small (Figs. 7–8), many regions see significant trends that are detectable beyond the "noise" of stochastic weather variability, particularly driven by increased autumn rainfall. Interannual variability not due to trends (e.g., due to climate variability) is almost uniformly of intermediate magnitude (25%) between the (dominant 70%) weather-driven variability and (smaller 5%) trend components.

This innovation in representation of memory processes in precipitation has impacts for stochastic simulation of precipitation (and other variables) that drive land surface processes, as shown for soil moisture and evapotranspiration in Figure 5. This is of particular relevance for variables such as soil moisture which integrate and smooth precipitation on time-scales on the order of weeks (McColl et al., 2017), thereby enhancing the effects of daily-scale correlation structure, as well as downstream variables such as surface water and energy fluxes and ecological variables.

Appendix A Variance of a Kernel Density Estimate

Although using kernel density methods to estimate empirical probability densities is fairly common practice, there are important caveats for their use, specifically regarding variance in this context. Kernel-based probability density functions are essentially mixture model distributions, using one mixture for each observed data point, and are not appropriate for either maximum-likelihood or method-of-moments fitting techniques (both of which will be optimized with a bandwidth of zero, equivalent to bootstrapping). Because of this, cross-validation (or some approximation thereof) is typically employed, but this does not preserve the variance of the sample to which the kernel density was fit, nor is it tied to asymptotic estimators for the variance of the population from which the data sample was drawn. While the bias in the variance is typically small, when variance is a key feature of your model, this needs to be addressed (and is the reason why simulated samples in this paper are re-standardized prior to transformation back to the data domain). It is worth noting that when using an axially-symmetric kernel (as is typical), the sample mean is always preserved.

A1 Kernel Density

For a set of N observed points, $\{y_i\}$, the univariate probability density using Gaussian kernels with bandwidth (standard deviation in this case), h, is

$$f_K(x; \{y_i\}, h) = \frac{1}{N} \sum_{i=1}^{N} \phi(x; y_i, h), \tag{A1}$$

where $\phi(x; y_i, h)$ is the normal density function with mean y_i and standard deviation h.

A2 Mean of Kernel Density

Using $E\{\cdot\}$ to denote expectation and $\mathcal{N}$ to denote the normal (Gaussian) distribution, the mean of the Gaussian kernel density is

$$\begin{aligned}
E\{x\} &= \int x f_K(x)\,\mathrm{d}x && \text{(A2)}\\
&= \int \frac{x}{N} \sum_{i=1}^{N} \phi(x; y_i, h)\,\mathrm{d}x && \text{(A3)}\\
&= \frac{1}{N} \sum_{i=1}^{N} \int x\phi(x; y_i, h)\,\mathrm{d}x && \text{(A4)}\\
&= \frac{1}{N} \sum_{i=1}^{N} E\{\mathcal{N}(y_i, h)\} && \text{(A5)}\\
&= \frac{1}{N} \sum_{i=1}^{N} y_i && \text{(A6)}\\
&= \overline{y_i}, && \text{(A7)}
\end{aligned}$$

which is just the sample mean of the points $\{y_i\}$ used to define the kernel densities.

A3 Variance of Kernel Density

The variance of the kernel density function is

$$\begin{aligned}
\mathrm{Var}\{x\} &= E\{x^2\} - (E\{x\})^2 && \text{(A8)}\\
&= \int x^2 f_K(x)\,\mathrm{d}x - (\overline{y_i})^2 && \text{(A9)}\\
&= \int \frac{x^2}{N} \sum_{i=1}^{N} \phi(x; y_i, h)\,\mathrm{d}x - (\overline{y_i})^2 && \text{(A10)}\\
&= \frac{1}{N} \sum_{i=1}^{N} \int x^2\phi(x; y_i, h)\,\mathrm{d}x - (\overline{y_i})^2 && \text{(A11)}
\end{aligned}$$

But for any individual normal distribution with mean y_i and standard deviation h,

$$E\{(x - y_i)^2\} = E\{x^2\} - (E\{x\})^2 \tag{A12}$$

Rearranging gives

$$\begin{aligned}
E\{x^2\} &= E\{(x - y_i)^2\} + (E\{x\})^2 && \text{(A13)}\\
\int x^2\phi(x; y_i, h)\,\mathrm{d}x &= E\{(x - y_i)^2\} + (E\{x\})^2 && \text{(A14)}\\
&= h^2 + y_i^2, && \text{(A15)}
\end{aligned}$$

since for kernel density centered at y_i the variance is just h^2 and the mean is y_i. Substituting back into A11,

$$\mathrm{Var}\{x\} = \frac{1}{N}\sum_{i=1}^{N}\left(h^2 + y_i^2\right) - \left(\overline{y_i}\right)^2 \quad \text{(A16)}$$

$$= \left[\overline{y_i^2} - \left(\overline{y_i}\right)^2\right] + h^2 \quad \text{(A17)}$$

The term on the left in A17 is just the sample variance, and so the variance of the kernel density is effectively additively inflated by the squared bandwidth, h^2. In this paper, where N for any given daily model is on the order of (31 days)·(57 years) = 1767 data points, and h^2 is on the order of 0.1 (in squared z-score units), the bandwidth variance inflation is larger than the effect of the typical multiplicative "Bessel correction" (i.e., $n/(n-1)$) used for unbiased population variance estimates. If our data did not go through an explicitly relative rank-driven CDF transformation (but rather some absolute mapping from the data domain to the Gaussian CDF-transformed domain and back), very large bandwidth values could lead to biases in the correlation structure of the simulated data. The rank-based transformation, however, eliminates this potential problem, but requires that any simulated data be re-standardized before back-transformation to explicitly preserve the proper variance.

This re-standardization can lead to statistical problems of its own if an insufficient number of simulated data points are used. To prevent the daily simulated variances from matching the observed sample variance exactly, a large number of simulations can be performed, re-standardized, and back-transformed. Then a subsample of the simulated data can be used for analysis. As an example, in this research, 1000 stochastic recreations of the historic record were simulated, and for each day of the year the 57,000 simulated daily values match the observed mean and sample variance, but the means and variances of each individual 57-year simulation can vary stochastically.

Acknowledgments

The authors would like to thank Michael Dietze and Dara Entekhabi for feedback on earlier versions of this manuscript. All data used in this study are available from the United State Historical Climatology Network at ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/v2.5/.

References

Ailliot, P., Allard, D., Monbet, V., Naveau, P., Ailliot, P., Allard, D., … Naveau, P. (2015). Stochastic weather generators: an overview of weather type models Titre: Générateurs stochastiques de condition météorologiques : une revue des modèles à type de temps. *Journal de la Société Française de Statistique*, *156*(1), 101–113. Retrieved from `http://www.sfds.asso.fr/journal`

Akbar, R., Short Gianotti, D. J., Salvucci, G. D., & Entekhabi, D. (2019). Mapped Hydroclimatology of Evapotranspiration and Drainage Runoff Using SMAP Brightness Temperature Observations and Precipitation Information. *Water Resources Research*, *55*, 3391–3413.

Akritas, M. G. (1990). The Rank Transform Method in Some Two-Factor Designs. *Journal of the American Statistical Association*, *85*(409), 73—-78.

Anderson, B. T., Gianotti, D., & Salvucci, G. (2015a). Characterizing the Potential Predictability of Seasonal, Station-Based Heavy Precipitation Accumulations and Extreme Dry Spell Durations*. *Journal of Hydrometeorology*, *16*(2), 843–856. doi: 10.1175/jhm-d-14-0111.1

Anderson, B. T., Gianotti, D. J., Salvucci, G., & Furtado, J. (2016). Dominant time scales of potentially predictable precipitation variations across the continental United States. *Journal of Climate*, *29*(24), 8881–8897. doi: 10.1175/JCLI-D-15-0635.1

Anderson, B. T., Gianotti, D. J., & Salvucci, G. D. (2015b). Detectability of historical trends in station-based precipitation characteristics over the continental united states. *Journal of Geophysical Research*, *120*(10), 4842–4859. doi: 10.1002/2014JD022960

Cai, X., Li, H., & Liu, A. (2016). A marginal rank-based inverse normal transformation approach to comparing multiple clinical trial endpoints. *Statistics in Medicine*, *35*, 3259—-3271.

Chen, M., & Xie, P. (2008). CPC Unified Gauge-based Analysis of Global Daily Precipiation. In *Western pacific geophysics meeting (9 july – 1 august).* Cairns, Australia.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., … Van Zyl, J. (2010, may). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*(5), 704–716. Retrieved from http://ieeexplore.ieee.org/document/5460980/ doi: 10.1109/JPROC.2010.2043918

Gianotti, D., Anderson, B., & Salvucci, G. (2013). What do rain gauges tell us about the limits of precipitation predictability? *Journal of Climate*, *26*(15), 5682—-5688. doi: 10.1175/JCLI-D-12-00718.1

Hurvich, C. M. ., & Tsai, C.-L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, *76*(2), 297–307.

Katz, R. W., & Parlange, M. (1998). Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, *11*(4), 591—-602.

Katz, R. W., & Zheng, X. (1999). Mixture model for overdispersion of precipitation. *Journal of climate*, *12*, 2528—-2537.

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences Journal*, *49*(4), 575–590. doi: 10.1623/hysj.49.4.575.54430

Madden, R. A., Shea, D. J., Katz, R. W., & Kidson, J. W. (1999). The potential long-range predictability of precipitation over New Zealand. *International Journal of Climatology*, *19*(4), 405–421. doi: 10.1002/(SICI)1097-0088(19990330)19:4⟨405::AID-JOC355⟩3.0.CO;2-U

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., … Albers, J. (2020, 05). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608-E625. Retrieved from https://doi.org/10.1175/BAMS-D-18-0326.1 doi: 10.1175/BAMS-D-18-0326.1

McColl, K. A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D. J., Lu, H., … Entekhabi, D. (2017). Global characterization of surface soil moisture drydowns. *Geophysical Research Letters*, *44*(8), 3682–3690. doi: 10.1002/2017GL072819

Menne, M., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., … Houston, T. (2012). *Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.* NOAA National Climate Data Center.

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378–381. Retrieved from http://dx.doi.org/10.1038/nature09763 doi: 10.1038/nature09763

National Acadamies of Science, Engineering, and Medicine. (2020). *Proceedings of a workshop–in brief.* The National Acadamies Press.

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., & Bindlish, R. (2016). *SMAP enhanced L3 radiometer global daily 9 km EASE-grid soil moisture, version 1.* Boulder, Colorado USA. doi: https://doi.org/https://doi.org/10.5067/ZRO7EXJ8O3XI

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* Cambridge MA: MIT Press.

Richardson, C. W. (1981). Dependence Structure of Daily Temperature and Solar

Radiation. *Water Resources Research*, *17*(1), 182—-190. doi: 10.13031/2013.33604

Short Gianotti, D. J. (2016). *Occurrence Markov Chain daily precipitation model.* Retrieved from https://doi.org/10.5281/zenodo.45435 doi: 10.5281/zenodo.45435

Short Gianotti, D. J., Anderson, B. T., & Salvucci, G. D. (2014). The potential predictability of precipitation occurrence, intensity, and seasonal totals over the continental United States. *Journal of Climate*, *27*(18), 6904–6918. doi: 10.1175/JCLI-D-13-00695.1

Short Gianotti, D. J., Akbar, R., Feldman, A. F., Salvucci, G. D., & Entekhabi, D. (2020). Terrestrial evaporation and moisture drainage in a warmer climate. *Geophysical Research Letters*, *47*(5), e2019GL086498. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL086498 (e2019GL086498 2019GL086498) doi: 10.1029/2019GL086498

U.S. Climate Prediction Center. (2015). *CPC REGIONAL US Mexico daily gridded archive.* National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction.

von Storch, H., & Zwiers, F. (2013). Testing ensembles of climate change scenarios for "statistical significance". *Climatic Change*, *117*, 1–9. doi: 10.1007/s10584-012-0551-0

Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography1*, *23*(3), 329—-357.

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *62*(1-3), 189–216. doi: 10.1023/B:CLIM.0000013685.99609.9e

SUPPLEMENTAL INFORMATION FOR Separating weather and climate using a spatially-scalable precipitation model with optimized subseasonal-to-seasonal statistics

Daniel J. Short Gianotti[1*], Guido D. Salvucci[2], and Bruce T. Anderson[2]

[1]Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[2]Department of Earth and Environment, Boston University, Boston, Massachusetts, USA.

*Parsons Laboratory, Massachusetts Institute of Technology, 15 Vassar St., Cambridge, MA 02139, USA

Corresponding author: Daniel J. Short Gianotti, `gianotti@mit.edu`

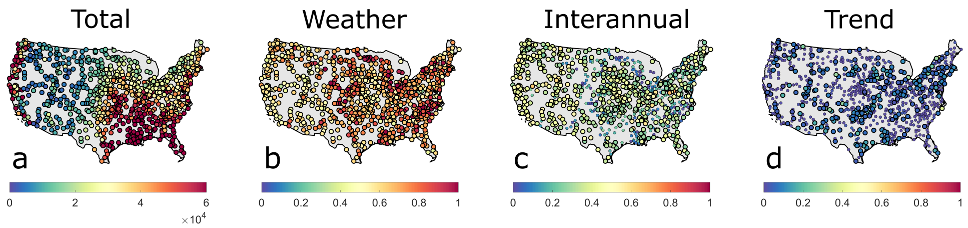

Figure 1. Normalized interannual variance in annual precipitation from daily GHCN data (1948–2004) [$(\mathrm{mm/yr})^2$]. Compare with main text Figure 7. a) Total interannual variance of annual precipitation. b) Fraction of total variance due to weather-scale processes from 1000 annually-stationary kernel autoregressive weather generator simulations. c) Fraction of total variance due to non-trend interannual variability, calculated as the variance of the detrended observations minus the variance of the detrended weather simulations, all divided by (a) to give a variance explained. Larger markers indicate significantly different from zero at the $\alpha = 0.05$ level using the distribution of 1000 weather simulations. d) Fraction of total interannual variance due to linear trend in observations. Larger markers denote a significant ($\alpha = 0.05$) trend as in Figure 6 in the main text. The large majority of interannual variability is driven by weather-scale processes in the Eastern ConUS, equally by weather- and climate-scale variability in the Western ConUS, and to a lesser degree by significant trends across ConUS.

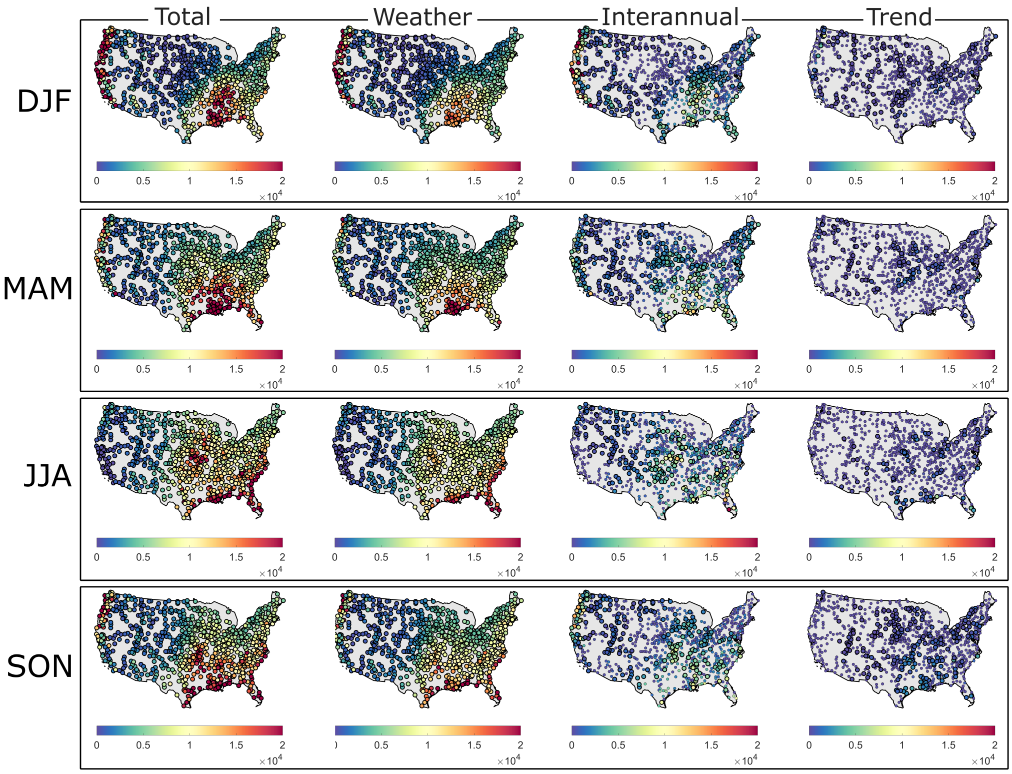

Figure 2. Interannual variance in seasonal precipitation from daily GHCN data (1948–2004). "Total" column shows total interannual variance of seasonal precipitation [$(\text{mm/yr})^2$] for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). "Weather" column shows variance of weather-scale processes from 1000 annually-stationary kernel autoregressive weather generator simulations as in Main Text Fig. 7b. "Interanual" column shows non-trend interannual variability, calculated as the variance of the detrended observations minus the variance of the detrended weather simulations as in Main Text Fig. 7c. Larger markers indicate significantly different from zero at the $\alpha = 0.05$ level using the distribution of 1000 weather simulations. "Trend" column shows interannual variance due to linear trend in observations as in Main Text Fig. 7d. Larger markers denote a significant ($\alpha = 0.05$) trend as in Main Text Figure 6. See Main Text Figure 8 for normalized maps.