A Nonstationary Standardized Precipitation Index (NSPI) Using Bayesian Splines

JAMES H. STAGGE and KYUNGMIN SUNG

Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, Ohio

(Manuscript received 9 November 2021, in final form 2 March 2022)

ABSTRACT: The standardized precipitation index (SPI) measures meteorological drought relative to historical climatology by normalizing accumulated precipitation. Longer record lengths improve parameter estimates, but these longer records may include signals of anthropogenic climate change and multidecadal natural climate fluctuations. Historically, climate nonstationarity has either been ignored or incorporated into the SPI using a quasi-stationary reference period, such as the WMO 30-yr period. This study introduces and evaluates a novel nonstationary SPI model based on Bayesian splines, designed to both improve parameter estimates for stationary climates and to explicitly incorporate nonstationarity. Using synthetically generated precipitation, this study directly compares the proposed Bayesian SPI model with existing SPI approaches based on maximum likelihood estimation for stationary and nonstationary climates. The proposed model not only reproduced the performance of existing SPI models but improved upon them in several key areas: reducing parameter uncertainty and noise, simultaneously modeling the likelihood of zero and positive precipitation, and capturing nonlinear trends and seasonal shifts across all parameters. Further, the fully Bayesian approach ensures all parameters have uncertainty estimates, including zero precipitation likelihood. The study notes that the zero precipitation parameter is too sensitive and could be improved in future iterations. The study concludes with an application of the proposed Bayesian nonstationary SPI model for nine gauges across a range of hydroclimate zones in the United States. Results of this experiment show that the model is stable and reproduces nonstationary patterns identified in prior studies, while also indicating new findings, particularly for the shape and zero precipitation parameters.

SIGNIFICANCE STATEMENT: We typically measure how bad a drought is by comparing it with the historical record. With long-term changes in climate or other factors, however, a typical drought today may not have been typical in the recent past. The purpose of this study is to build a model that measures drought relative to a changing climate. Our results confirm that the model is accurate and captures previously noted climate change patterns—a drier western United States, a wetter eastern United States, earlier summer weather, and more extreme wet seasons. This is significant because this model can improve drought measurement and identify recent changes in drought.

KEYWORDS: Drought; Precipitation; Bayesian methods; Climate change; Climatology; Nonlinear models; Multidecadal variability

1. Introduction

Normalized meteorological drought indices like the standardized precipitation index (SPI) (McKee et al. 1993; Guttman 1999; Lloyd-Hughes and Saunders 2002) are used to measure meteorological drought severity relative to a reference climate, represented by probability distribution estimates. In particular, the SPI transforms precipitation accumulated over a fixed number of prior months to the standard normal distribution, making the index statistically interpretable in terms of probability or frequency (return period) (Guttman 1999). For example, an SPI of −2 implies that accumulated precipitation is two standard deviations drier than a typical year on this date and has a 2.3% chance of occurring in any given year (44-yr return period). Normalization permits comparisons of drought severity across regions with widely varied climatologies and highly nonnormal precipitation distributions. This statistical interpretability combined with simplicity of calculation are some of the reasons the SPI is recommended by the World Meteorological Organization (WMO 2006) and is used operationally by the U.S. Drought Monitor.

The SPI transforms precipitation using a backward-looking rolling window into a normally distributed index, with a mean of 0 and standard deviation of 1, \( N(\mu = 0; \sigma = 1) \) (McKee et al. 1993; Guttman 1999; Lloyd-Hughes and Saunders 2002). The SPI is typically named based on the number of preceding months included in its rolling window. For example, the SPI-3 measures the preceding 3-month window. By adjusting this period to mimic drought propagation, the SPI may approximate drought severity in other parts of the hydrologic cycle, such as soil moisture drought (Lloyd-Hughes and Saunders 2002) or hydrologic drought (Van Loon 2015).

Seasonally accumulated precipitation distributions are often positively skewed, leading to the use of distributions like the two-parameter gamma distribution (Stagge et al. 2015; Guttman 1999; Lloyd-Hughes and Saunders 2002; Giddings et al. 2005), which can approximate positively skewed and Gaussian shapes by adjusting its shape parameter. Other distributions have been used when calculating the SPI (Guttman 1999; Sienz et al. 2012; Vicente-Serrano et al. 2012) but are not considered in this study. The gamma distribution is strictly

---

Denotes content that is immediately available upon publication as open access.

Corresponding author: James H. Stagge, stagge.11@osu.edu

DOI: 10.1175/JAMC-D-21-0244.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
positive, so a piecewise distribution is typically used to account for the probability of zero precipitation:

\[ p(P_n) = \begin{cases} 
\theta + (1 - \theta)F_{\text{gamma}}(\alpha, \beta) & \text{if } P_n > 0 \\
\theta & \text{if } P_n = 0 
\end{cases} \]  

(1)

where \( p(P_n) \) is the cumulative probability of \( P_n \) precipitation accumulated over \( n \) months and \( F_{\text{gamma}}() \) is the gamma cumulative probability density function with shape and scale parameters \( \alpha \) and \( \beta \), respectively. The \( \theta \) parameter represents the probability of zero precipitation, which is important for semiarid regions and short accumulation periods for which zero precipitation is likely. The gamma distribution parameters are typically estimated by fitting 365 or 12 univariate distributions, independent of each other, for daily or monthly resolutions, respectively. The \( \theta \) parameter is regularly estimated separately using an empirical Weibull plotting position function:

\[ \theta = \frac{n_{\text{zero}}}{n + 1}, \]  

(2)

where \( n_{\text{zero}} \) represents the number of zero precipitation observations in the fitting period and \( n \) is the number of observations (Wu et al. 2007; Stagge et al. 2015). When estimated empirically, \( \theta \) is rarely presented with uncertainty estimates. This approach frequently places the SPI value associated with zero precipitation at the upper bound of probability, which can produce consistent overestimates. Stagge et al. (2015) suggests using the centroid of the zero probability mass:

\[ p(P_n) = \begin{cases} 
\theta + (1 - \theta)F_{\text{gamma}}(\alpha, \beta) & \text{if } P_n > 0 \\
\theta/2 & \text{if } P_n = 0 
\end{cases} \]  

(3)

where \( \theta/2 \) could be estimated empirically:

\[ \frac{\theta}{2} = \frac{n_{\text{zero}} + 1}{2(n + 1)}. \]  

(4)

This method places the resultant SPI value at the centroid for zero precipitation probability, rather than at its upper limit, which should produce a less biased SPI estimate (Stagge et al. 2015; Solakova et al. 2014).

Parameter estimates for the SPI are strongly influenced by record length, with increased record length providing more data for calibration, and thus less uncertainty in the parameter estimates (Carbone et al. 2018; Vergni et al. 2017; Wu et al. 2005). This would lead a researcher to attempt to maximize record length; however, calibrating the SPI based on decades or century long precipitation records can introduce other types of errors due to a changing climate over this longer period. The typical SPI fitting approach [Eq. (1)] does not account for a changing climate, instead assuming stationarity, that the parameters underlying each distribution are constant over time. This approach typically fits 365 unique distributions, one for each day of the year. This is represented graphically in Fig. 1a, where 365 unique parameter estimates are fit, along with their unique uncertainty. This creates a “stationary” model with respect to long-term climate variability, changing model parameters with seasonality, but keeping them fixed from year to year (Fig. 1a).

Given modern anthropogenic climate change (Slater et al. 2021; Ganguli and Ganguly 2016) and natural multidecadal climate fluctuations (Seager and Hoerling 2014; Poore et al. 2009; Nigam et al. 2011; Coats et al. 2016; Stahle et al. 2020), considering nonstationarity becomes important when calibrating drought indices from an instrumental record influenced by these effects (Cammalleri et al. 2022). Historically, two approaches have been used when calibrating the SPI under a nonstationary climate: one can either normalize based on a quasi-stationary subset of the calibration period, such as the WMO 30-yr reference period, or normalize based on the full historical time series. The first approach, using a quasi-stationary reference period, for example might fit the parameters of Eq. (3) based on the previous WMO baseline 1961–90. All SPI values would then be normalized relative to this baseline. If precipitation underwent an increasing trend during the twentieth century, the SPI during the baseline period (1961–90) would be centered on zero, but SPI values for an earlier period would be centered on a value less than zero and the later period would be centered around an SPI value greater than zero. By using a 30-yr subset, this approach decreases the record length, thereby increasing parameter uncertainty (Dubrovsky et al. 2009; Heinrich and Gobiet 2012; Russo et al. 2013; Carbone et al. 2018; Vergni et al. 2017; Wu et al. 2005). This approach, however, has the benefit of generating easily comparable results between studies that share a common reference period.

The second approach uses the full record to fit the parameters in Eq. (3), assuming the observations all come from the same distribution, thereby ignoring any long-term trends (nonstationarity). In the previous example with steadily increasing precipitation, this would produce SPI values centered on zero for the middle year of the record length. This approach uses the full record, decreasing parameter uncertainty by using more data, but implicitly ignores climate nonstationarity. For the simple, linear precipitation increase, parameter estimates of the mean would be too high for the early period and too low for the later period. This is further complicated if the trends are more complex. Further, this approach centers the SPI on the middle year of the record. If two studies were to use different record length, for example as new data became available, each would be centered on a different year, making comparisons across studies more challenging.

Recent studies have introduced an alternative for developing a nonstationary SPI (NSPI), designed to permit changes in the underlying distribution parameters over decadal scales (Russo et al. 2013; Shiau 2020). These approaches generally rely on the generalized linear model (GLM) framework, an extension of linear regression that removes the assumption of normality McCullagh and Nelder (1989). For GLMs, linear combinations of explanatory variables are transformed via a link function, which could be the identity (original value) or transforms for the log, inverse log, logit, and others. Further, rather than assuming a normal distribution around the estimate, GLMs often assume a distribution from the exponential family, for example the gamma, exponential, normal, Poisson,
or binomial. In this way, GLMs generalize linear regression to include a more flexible range of links and distributions. Russo et al. (2013) and Shiau (2020) both use a GLM framework with a gamma distribution, whereas this study’s hurdle model in Eq. (5) (see below) includes a gamma distributed GLM for positive precipitation and a binomial GLM for zero precipitation likelihood. To permit greater flexibility in the predictor equation, GLMs can be combined with generalized additive models (GAMs), which use nonlinear functions as predictors (Hastie and Tibshirani 1986). GAMs therefore refer to the explanatory variables and whether they use nonlinear terms, whereas GLMs refer to how the response variable is distributed. Generalized additive models for location scale and shape (GAMLSS) (Rigby and Stasinopoulos 2005) are a special case whereby nonlinear GAM functions can be used to explain the location, scale, and shape parameters of a distribution separately. Using the previous example of precipitation modeled as a two parameter gamma distribution, a GAMLSS model would use unique nonlinear GAM terms to explain the shape and scale parameters, $\alpha$ and $\beta$. We should clarify the terms BAMLSS and BAM, where BAMLSS refers to GAM based location-scale-shape model that use a Bayesian framework (Umlauf et al. 2018) and BAM is a commonly used software approach for fitting GAM models with extremely large datasets (Wood et al. 2015, 2017).

In Russo et al. (2013), the scale parameter $\beta$ of each daily gamma distribution was allowed to change based on the year via a linear trend, while the shape parameter $\alpha$ remained

FIG. 1. Illustrative example showing parameter estimates for (a),(b) stationary climate models and (c),(d) nonstationary climate models. The left column illustrates prior methods for mean parameter estimation, each using 365 unique fits for a stationary model in (a) and a nonstationary model in (c) with a linear trend during the twentieth century. The right column illustrates the methods proposed here for a stationary model in (b) using a single cyclic spline to estimate all days simultaneously and a nonstationary model in (d) with splines for repeating seasonality and long-term trends. (e) Illustration of how the constituent seasonal, annual, and tensor interaction terms sum to create the behavior in (d), producing nonlinear trends for each day while retaining their relative relationships. The tensor interaction and tensor product terms display day on the x axis and year on the y axis, with color representing parameter value.
constant. Because the gamma distribution mean is equal to \( \alpha \times \beta \), this is equivalent to linearly modifying the distribution mean. Using the previously introduced terminology, the Russo et al. (2013) model relied on a GLM to capture a linear trend in a single parameter of the gamma distribution. This is equivalent to Fig. 1c, where the mean parameter varies linearly throughout the twentieth century, but where each day’s model is fit independently. Shiau (2020) expands on this approach, allowing both the shape and scale parameters to change nonlinearity with a spline function, fit using the aforementioned GAMLSS framework. The Shiau (2020) approach adds two features relative to the Russo et al. (2013) approach: allowing nonlinearity and allowing the shape parameter to vary, but still relies on 365 nonstationary distributions fit independently of one another and does not address periods of zero precipitation.

This study proposes a novel Bayesian model based on cyclic splines, designed to better constrain stationary (Fig. 1b) and nonstationary (Fig. 1d) parameter estimates by fitting a single model that simultaneously estimates all parameters from Eq. (1) \((\alpha, \beta, \text{and } \theta)\). We show how a stationary model using this framework can be expanded into an NSPI model by slowly varying each parameter through time using Bayesian splines. The nonstationary term is penalized to mimic a WMO 30-yr reference period, making it compatible with quasi-stationary approaches (Russo et al. 2013; Shiau 2020). This study is organized primarily as a model description and validation, directly comparing the proposed Bayesian spline SPI approach with existing MLE methods for stationary and nonstationary climates. This comparison is performed across three experiments of increasing complexity: 1) synthetic precipitation from a stationary climate with known parameters, 2) synthetic precipitation from a nonstationary climate, and 3) real-world precipitation gauges across a variety of climate regimes within the United States.

2. Model

a. Bayesian NSPI model framework

The Bayesian NSPI model translates the piecewise SPI equation [Eq. (1)] to a Bayesian framework using a hurdle model (Mullahy 1986), which simultaneously models the likelihood of zero precipitation and the cumulative distribution of positive precipitation:

\[
P_{\text{smooth}} \sim \begin{cases} 
\gamma \text{amma}(\rho, \alpha, \beta) & \text{if } \rho > \theta \\
0 & \text{if } \rho \leq \theta 
\end{cases},
\]

where \( \theta \) is modeled as a binomial process with logit link via the \( \rho \) parameter:

\[
\theta \sim \text{Bernoulli}(\rho).
\]

Rather than estimating the shape \( \alpha \) and scale \( \beta \) parameters, the proposed model estimates the mean \( \mu \) and shape \( \alpha \) parameters, with scale derived as \( \beta = \mu/\alpha \). This provides more stable parameter estimates (Wood 2004). The three parameters \( \mu, \alpha, \text{and } \rho \) are then constructed using nonlinear spline functions in a GAM framework:

\[
\log(\mu) = b_{0 \mu} + f_{\mu}(\text{day}),
\]

\[
\frac{1}{\log(\alpha)} = b_{0 \alpha} + f_{\alpha}(\text{day}), \quad \text{and}
\]

\[
\rho = b_{0 \rho} + f_{\rho}(\text{day}).
\]

where \( f() \) is a spline function (Hastie and Tibshirani 1986).

The log link and inverse log link are used for the \( \mu \) and \( \alpha \) parameters, respectively, ensuring strictly positive estimates. The \( \rho \) parameter uses the identity link but is subsequently applied to the Bernoulli distribution to estimate \( \theta \). For the remainder of this paper, we will refer to \( \theta \) to discuss the modeled probability of zero precipitation. In this model, \( f(\text{day}) \) is a spline function of \( 1-365 \), and \( b_0 \) represents the intercept. Subscripts are used to denote the parameter being estimated.

The stationary model allows \( \mu, \alpha, \text{and } \theta \) parameters to vary seasonally, constrained by a spline that repeats annually (Fig. 1b).

In this way, the parameters represent a stationary, cyclic climatology for a location. The nonstationary model retains an annual cyclic pattern but also allows \( \mu, \alpha, \text{and } \theta \) parameters to slowly vary through time (Figs. 1d,e). The model then becomes nonstationary, as long-term climate variability or trends can slowly adjust the climatology of a site.

b. Stationary cyclic model

For the stationary climate model, a cyclic cubic regression spline \( f(t) \) is used to model annually repeating seasonal parameters, smoothly adjusting parameter estimates for each day of the year. This creates a stationary climate model, with the parameters changing seasonally to capture climatology, but repeating annually without changing from year to year, as shown in Fig. 1b. Fitting a single model for all days is distinct from other approaches that fit 365 independent models (Fig. 1a). Use of a spline enforces similarity of parameter values from one day to the next, which follows logically from the use of a moving window in the SPI, where all but the first and last precipitation measurements are shared. Our hypothesis was that using a single common model for all daily parameters increases the number of data points used in regression, thereby improving parameter estimates relative to the use of 365 individual daily parameter fits.

The “cyclic” descriptor refers to an added constraint that enforces continuity between days 365 (31 December) and 1 (1 January). The “cubic regression” spline descriptor indicates the spline is constructed from sections of a cubic polynomial, dictated by values at control points called “knots” and joined together to ensure the spline is continuous up to and including the second derivative (Fig. 2a). The spline magnitude at each knot is controlled by a fitted value, represented in equations here as \( b \) values. The \( b \) values are analogous to pulling or pushing on points along a spring, where the spring’s resistance is defined by a “wiggliness” penalty from the spline’s instantaneous second derivative. This penalty is described in greater detail in section 2c. Estimates at any point along the spline are calculated by multiplying a model basis matrix \( \mathbf{X} \) by the vector of \( b \) values (Hastie and Tibshirani 1986).
A further constraint is added for “identifiability,” ensuring that the smooths sum to zero over the observed values. This centers the spline about the intercept $b_0$. The cyclic and identifiability constraints are absorbed into the model basis matrix $\mathbf{X}$ by reparametrizing $\mathbf{X}$ to account for fewer degrees of freedom (Wood et al. 2017). This reparameterization has the added benefit of decreasing the free parameters to be estimated.

c. Nonstationary tensor product model

For the nonstationary model, an additional dimension is added to each cyclic spline to incorporate interannual trends (Figs. 1d,e). This type of multidimensional spline is known as a tensor product spline (Wood et al. 2013; Pedersen et al. 2019). A tensor product spline can be visualized as a lattice of strips made up of splines running perpendicularly, in this case with day along one dimension and year along the second. This expansion of the stationary model means that a single model can estimate distribution parameters for all days and years, while enforcing a smooth transition in both dimensions, as illustrated in Fig. 1d. This is unique from previous approaches, which use 365 independent nonstationary models and can produce unrelated long-term trends (Fig. 1c).

The resulting tensor product is anisotropic, meaning it permits different scales and wiggliness along each dimension (Wood 2004), as is the case for day and year. This model uses a cyclic spline along the day dimension, as in the stationary model, but a noncyclic cardinal spline along the year dimension. Cardinal splines ensure the annual dimension becomes linear at its earliest and latest knots by requiring the second derivative equal zero at these boundaries. Figure 2b illustrates the anisotropy, showing how higher density knots along the day $x$ axis produce an elongated response and how the knot centered on 1 January 2008 is cyclic in the day dimension. The tensor product in this application follows the tensor product interaction basis, incorporated in the “mgcv” package (Wood et al. 2017, 2016), which allows the tensor product terms to be separated into their constituent parts for greater control (Fig. 1e). The proposed nonstationary model for the mean parameter could then be written as

$$\log(\mu) = b_0 + f_\mu(\text{day}) + f_\mu(\text{year}) + f_\mu(\text{day}, \text{year}).$$  \hspace{1cm} (10)

The first two terms in Eq. (10), $b_0$ and $f(\text{day})$, are similar to the stationary model and represent regular, cyclic seasonality (Fig. 1e). The $f(\text{year})$ term models long-term interannual change in a parameter, common to all days of the year. The final $f(\text{day}, \text{year})$ interaction term represents interannual trends specific to a given season or day of the year. For example, the increasing $f(\text{year})$ term in Fig. 1e indicates an overall wetting trend across all seasons, whereas $f(\text{day}, \text{year})$ adds an exaggerated wetting effect over time centered on day 250 (September), indicated by a trend from blue to yellow moving from 1920 (bottom) to 2020 (top). The reversed pattern around day 40 (February) indicates a weakening of this long wetting term trend for the winter. In this way the first seasonal term is stationary with respect to climate, and the second two are nonstationary, summing to create the tensor product used to model the mean parameter (Fig. 1e). This structure [Eq. (10) and Fig. 1e] is also applied to the shape and $\theta$ parameters, following the framework of Eqs. (7)–(9).

The proposed model follows a Bayesian formulation of the wigginess penalty term (Wood 2016) by assigning a multinormal prior to the $\mathbf{b}$ parameter vector with a mean of zero and variance–covariance matrix defined by a weight $\lambda$ and the penalty matrix $\mathbf{S}$:

$$\mathbf{b}_{\mu, \text{day}} \sim N(0, \lambda_{\mu, \text{day}} \mathbf{S}_{\mu, \text{day}}).$$ \hspace{1cm} (11)

This keeps the $b$ values centered on the intercept $b_0$ while the $\mathbf{S}$ matrix penalizes rapid changes in $b$ values between adjacent knots, decreasing this effect for more distant knots. The prior for each $\lambda$ tuning parameter is gamma distributed to ensure they remain strictly positive. Equation (11) specifies $\mu$ and day subscripts as a reminder that each spline basis is assigned

![Diagram](image-url)
its own $\lambda$ tuning parameter. By adjusting the $\lambda$ prior, the model can replicate a 30-yr WMO period along the year dimension, while protecting against excess wiggliness or overfitting. We utilize a preprocessing step that estimates $\lambda$ using the mgcv package (Wood 2004; Wood et al. 2017) to improve model convergence. The gamma distributed prior for $\lambda$ used a relatively large shape parameter (500), which approximates a normal distribution and produces a relatively uninformative prior, while the scale parameter was selected to center the distribution on the preprocessing estimate.

d. Generating NSPI values

Once the parameters of the nonstationary model are fit, an NSPI time series can be generated in two ways. The more typical approach is to fix a reference period and calculate the NSPI based on this reference period (Cammalleri et al. 2022). Using Figs. 1d and 1e as an example, the parameter estimates from the nonstationary tensor product model in 1976 are representative of the WMO 30-yr climate normal 1961–90. If one would like to generate SPI values based on the 1961–90 climate, you could extract the mean parameter for 1976 by taking a vertical slice through Fig. 1d or a horizontal slice through the resultant tensor product plot for 1976 (Fig. 1e). The shape and $\theta$ parameters could be extracted for 1976 in the same way. Similarly, by extracting parameter estimates for 2006, one could approximate a 1991–2020 reference period without refitting the model. For this example, the annual mean $\{f_d(\text{day, year})\}$ increased between these periods with even greater increases in September (day 250), according to $f_d(\text{day, year})$. So, a historical observation of 4.5 mm day$^{-1}$ might be typical (SPI = 0) relative to 1976 climate, but drier than typical (SPI = −0.5) relative to 2006 climate. By fixing a reference year, the entire SPI time series could be generated relative to any baseline reference you desire. This example focused on the mean, but the shape and $\theta$ parameters could also vary between baseline reference periods, affecting the distribution extremes and zero precipitation, respectively.

A second, less common use for the NSPI model would be to generate SPI values relative to a sliding reference period that adjusts the mean, shape, and $\theta$ estimates along with the time series. With this approach, one could imagine extracting the mean, shape, and theta parameters from Figs. 1d and 1e separately for each year before estimating that year’s SPI. In this example, the reference baseline would not stay fixed, but would change along with the SPI time series. This would produce a very different type of SPI time series, where the resulting values would follow a standard normal distribution (mean of 0; standard deviation of 1), thereby removing nonstationary trends from the SPI time series. This is very different from the first approach, which could produce a trend in SPI values, as the observations differ more from the fixed reference year. Therefore, the second approach with a moving reference year should only be used for time series analyses that require stationarity or normality. It should not be used to discuss long-term climate variability or climate change in meteorological drought.

3. Methods

a. Experimental design and evaluation

Evaluation of the proposed model is performed as comparisons with an MLE model across three experiments, using synthetically generated precipitation from a stationary climate (experiment 1), synthetic precipitation from a nonstationary climate (experiment 2), and instrumental data from nine gauges across the United States (experiment 3). Experiment 1 compares two MLE models fit using a 2-parameter gamma distribution and empirical estimate of $\theta$ with the proposed stationary Bayesian model (section 2b) using synthetically generated precipitation from a predetermined stationary climate. The two MLE models considered for comparison use a 30-yr subset of data (1961–90) and the full 100-yr record (1920–2019). They are referred to as MLE short and long, respectively. These MLE models were chosen to contrast the proposed Bayesian model with typical SPI approaches, either using a WMO 30-yr reference period or all available data. Experiment 2 used the same MLE short (1961–90) and MLE long (1920–2019) model alternatives, but instead compared these with the Bayesian nonstationary NSPI model (section 2c) under a nonstationary synthetically generated climate. Experiment 3 used instrumental precipitation fit with the Bayesian nonstationary NSPI model. As such, it does not include a MLE comparison because the true distribution parameters are unknown.

Quantitative model comparisons for experiments 1 and 2 were based on parameter estimates and the resulting SPI values. Visual comparisons were used to identify patterns of bias or exceptional parameter uncertainty, while summary statistics were used to quantify model accuracy and precision. Accuracy was measured using mean absolute error (MAE), root-mean-squared error (RMSE), mean error (ME), variance explained $R^2$, and Nash–Sutcliffe efficiency (NSE), and precision was measured by the 95% confidence interval (CI) width and the percent of accurate hits within the 95% CI. Each metric is calculated using random draws from the MLE fit with uncertainty or Bayesian posterior draws. In this way, the MAE approximates the probabilistic continuous ranked probability skill score (CRPSS) (Hersbach 2000; Gneiting and Raftery 2007) while still allowing a direct comparison between MLE and Bayesian approaches.

b. Synthetic precipitation data

Synthetic time series were used in experiments 1 and 2 to simulate stationary and nonstationary climates, respectively. Synthetic daily precipitation was generated by first randomly simulating a SPI-3 time series based on a moving average (MA) model with a 3-month rolling window. To accomplish this, an MA(91) model was used to simulate a 92-day moving window, with equal weights and daily innovations of $(92)^{1/2}$ to ensure variance of 1. This MA(91) model was used to generate a 100-yr-long time series, from 1 January 1920 to 31 December 2019 (Figs. 2b,d). Dates were included as placeholders but have no connection with true history. The same underlying SPI-3 time series was used.
for experiments 1 and 2 with stationary and nonstationary parameters, respectively.

To generate synthetic precipitation under a stationary climate (experiment 1), the SPI-3 time series was applied via the piecewise SPI-3 model [Eq. (1)] using predetermined daily values for $\alpha$, $\beta$, and $\theta$ (Fig. 3a). These parameters imply a wet winter and dry summer with subseasonal fluctuations (Fig. 3a), with a constant shape parameter, and a probability of zero precipitation $\theta$ that peaks midyear. Any SPI with a corresponding probability below $\theta$ was assigned zero precipitation, as shown by the shaded red region in Fig. 3c.

For experiment 2, climate was designed to be nonstationary. The mean parameter intensified throughout the century, with wet winters becoming wetter and dry summers became drier (Fig. 3b). The shape parameter remained constant throughout the century, while the probability of zero precipitation $\theta$ was designed to remain constant until 1970 and then gradually decrease, as shown by a slowly decreasing red region in Fig. 3d.

c. Instrumental data

Instrumental precipitation data was based on the Global Historical Climatology Network (GHCN) daily dataset, a database of instrumental climate data from land surface stations that been subjected to quality assurance reviews (Menne et al. 2012). Nine gauges across the United States were chosen to capture a wide range of hydroclimatic regions while also ensuring long available time series (Fig. 4). Of the nine gauges, five [Aberdeen, Washington (ABD); Albany, Georgia (ALB); Marysville, Ohio (MAR); New York, New York (NYC); Waxahachie, Texas (WAX)] are considered temperate climates as based on the Köppen–Geiger classification.
system (Peel et al. 2007), two [Paauilo, Hawaii (PAO), and San Juan, Puerto Rico (SAN)] are tropical, one [Los Angeles, California (USC)] is dry, and one [Morris, Minnesota (MOR)] is a cold winter continental climate. This selection of gauges includes two of the wettest gauges with long records in the United States (ABD and PAO) and two gauges that occasionally experience prolonged zero precipitation periods (USC and WAX). The MOR site was chosen as indicative of a cold region, and the two island gauges (PAO and SAN) were chosen to emphasize their unique tropical hydroclimates and geographic location in the Pacific and Atlantic Oceans, respectively. More gauge details are presented in the appendix.

d. Model fitting

The proposed Bayesian spline model, stationary SPI and nonstationary NSPI, were fit using the spibayes package (Stagge 2021), developed for this study. Bayesian inference was performed using the Stan platform (Carpenter et al. 2017), which uses a Hamiltonian Monte Carlo sampling algorithm and is accessed from R via the “cmdstanr” package (Gabry and Češnovar 2020). All basis and penalty matrices were derived from the mgcv package (Wood et al. 2017).

4. Results

a. Experiment 1: Stationary synthetic precipitation

For a stationary climate, the stationary Bayesian spline model performed much better than the MLE short model and marginally better than the MLE long model based on the full 100-yr reference period (Table 1). This finding is reasonable, because the MLE long and Bayesian spline model both fit 100 years of data, while the MLE short model relies on only 30 years of data. Relative to the more comparable MLE long model, the proposed Bayesian model is slightly more accurate for all three parameters ($\mu$, $\alpha$, and $\theta$), quantified by smaller errors (MAE, RMSE) and higher predictive skill ($R^2$; NSE) (Table 1).

For the mean parameter, the Bayesian spline model and MLE long models both produce small errors (RMSE = 0.335 and 0.341, respectively), good modeling of temporal variance (NSE = 0.806 and 0.799, respectively), and relatively unbiased estimates ($ME = -0.063$ and $-0.043$, respectively) (Table 1). Visual comparisons showed that the Bayesian model produced smoother parameter estimates due to its spline constraints, whereas the MLE long estimates are more erratic due to 365 independent daily fits (Fig. 5). Confidence intervals for the mean parameter were smallest for the Bayesian model (Table 1), particularly along the upper limit (Fig. 5), though this produced a hit rate within the 95% CI, which was lower than the expected 95%. The CI for the estimate (Fig. 5) is based on the 95% range of posterior distribution for each parameter, and the wider, light-gray interval visualizes error around each parameter estimate using an additional error variance modeled around each parameter.

The true shape parameter was purposefully held constant to test whether the Bayesian spline model would produce spurious nonlinearities. Results showed that the Bayesian model was visually similar to the MLE long model (Fig. 5), again with slightly smaller errors, as measured by MAE and RMSE (Table 1). Shape parameter estimates produced by the MLE models were especially noisy (Fig. 5), which may explain the larger RMSE and MAE. When adjusted for their magnitudes, shape parameter estimates for all models were less accurate and more uncertain than mean parameter estimates. Calculation of $R^2$ and NSE was not possible for the shape parameter because the true shape values were constant (zero variance). Confidence intervals for the shape parameter were similar for the Bayesian and MLE long models, although both models produced lower hit rates than the expected 95% (Table 1). For the $\theta$ parameter, the most important difference between the Bayesian and MLE approaches was that the Bayesian model produced confidence intervals, whereas the MLE model could

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Shape</th>
<th>Theta</th>
<th>SPI-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE short</td>
<td>MLE long</td>
<td>Bayesian estimate</td>
<td>MLE short</td>
</tr>
<tr>
<td>MAE</td>
<td>0.822</td>
<td>0.285</td>
<td>0.280</td>
<td>0.368</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.982</td>
<td>0.341</td>
<td>0.335</td>
<td>0.476</td>
</tr>
<tr>
<td>ME</td>
<td>-0.364</td>
<td>-0.046</td>
<td>-0.063</td>
<td>-0.111</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.300</td>
<td>0.836</td>
<td>0.838</td>
<td>—</td>
</tr>
<tr>
<td>NSE</td>
<td>-0.662</td>
<td>0.799</td>
<td>0.806</td>
<td>—</td>
</tr>
<tr>
<td>CI width</td>
<td>3.44</td>
<td>1.57</td>
<td>1.33</td>
<td>1.77</td>
</tr>
<tr>
<td>CI hit rate</td>
<td>0.838</td>
<td>0.967</td>
<td>0.923</td>
<td>0.964</td>
</tr>
</tbody>
</table>

For the mean parameter, the Bayesian spline model and MLE long models both produce small errors (RMSE = 0.335 and 0.341, respectively), good modeling of temporal variance (NSE = 0.806 and 0.799, respectively), and relatively unbiased estimates ($ME = -0.063$ and $-0.043$, respectively) (Table 1). Visual comparisons showed that the Bayesian model produced smoother parameter estimates due to its spline constraints, whereas the MLE long estimates are more erratic due to 365 independent daily fits (Fig. 5). Confidence intervals for the mean parameter were smallest for the Bayesian model (Table 1), particularly along the upper limit (Fig. 5), though this produced a hit rate within the 95% CI, which was lower than the expected 95%. The CI for the estimate (Fig. 5) is based on the 95% range of posterior distribution for each parameter, and the wider, light-gray interval visualizes error around each parameter estimate using an additional error variance modeled around each parameter.

The true shape parameter was purposefully held constant to test whether the Bayesian spline model would produce spurious nonlinearities. Results showed that the Bayesian model was visually similar to the MLE long model (Fig. 5), again with slightly smaller errors, as measured by MAE and RMSE (Table 1). Shape parameter estimates produced by the MLE models were especially noisy (Fig. 5), which may explain the larger RMSE and MAE. When adjusted for their magnitudes, shape parameter estimates for all models were less accurate and more uncertain than mean parameter estimates. Calculation of $R^2$ and NSE was not possible for the shape parameter because the true shape values were constant (zero variance). Confidence intervals for the shape parameter were similar for the Bayesian and MLE long models, although both models produced lower hit rates than the expected 95% (Table 1). For the $\theta$ parameter, the most important difference between the Bayesian and MLE approaches was that the Bayesian model produced confidence intervals, whereas the MLE model could
not because of its use of empirical estimates [Eq. (2)]. All models show poor predictive skill for \( \theta \), with \( R^2 \) near 0 and NSE less than 0, though the Bayesian model performs slightly better than the MLE, particularly for the NSE (Table 1). Another benefit of the Bayesian model was noted when there are no zero precipitation observations in the reference period, as around day 80 (21 March). The MLE \( \theta \) estimate is zero, whereas the Bayesian estimate remains an extremely small, but still positive value. If the MLE estimate were used to calculate SPI based on new data that contained zero precipitation on this date, the resulting SPI would be undetermined because the fitted distribution did not allow for this possibility. The Bayesian approach avoids this issue entirely. While the Bayesian model produces a CI around the true value shown in the lower-right corner. Seasonality, along the x axis, was captured well by the Bayesian NSPI and MLE long models, while the MLE short model produced more noisy estimates that remained constant, indicated by vertical constantly shaded lines. The Bayesian NSPI model captured some of the intensification from 1920 (Fig. 8, bottom) to 2019 (Fig. 8, top), although with some exaggeration (Fig. 8). This model also produced a mild pause in intensification around 1980, which may be related to random oversampling of zero precipitation \( \theta \) during this period (Fig. 8b). Of further note is the slight “washing out” of the highest frequency patterns in the Bayesian NSPI model, in which adjacent patterns were merged.

Estimates of the underlying shape parameter, which remained constant throughout the sample period (Fig. 3b), showed that the Bayesian NSPI model produced estimates that were better than the MLE short model but worse than the MLE long model for most metrics (Table 2). The exceptions were absolute bias (ME) and CI width, for which the Bayesian model performed worst. An inspection of errors showed that localized over and undersampling errors propagated both seasonally and interannually, along both dimensions, due to the tensor product constraint.

For the probability of zero precipitation, \( \theta \), the Bayesian NSPI model performs better than the MLE short model, though slightly worse than the MLE long model (Table 2). Parameter estimates of \( \theta \) for all models were subject to random sampling errors, with high estimates in October for all models of zero precipitation. The Bayesian model produces a smooth seasonal estimate of SPI-3 during these periods, for example in 1980 (Fig. 6), and both MLE models produce erratic estimates, dictated by the \( \theta \) parameter. Though MLE models could not produce a true SPI CI during periods of zero precipitation, intervals between \(-3\) and \(\theta\) were added for visualization. The stepped nature of MLE estimates was more apparent when the CDF was zoomed in to show the lowest SPI values (Fig. 7).

**b. Experiment 2: Nonstationary synthetic precipitation**

For the nonstationary synthetic precipitation, the proposed Bayesian NSPI model (section 2c) was generally more accurate than the MLE short model with a quasi-stationary 30-yr reference period and slightly less accurate than the MLE long model, which ignores nonstationarity (Table 2). Comparisons with both models are presented; however, the comparison with MLE short is more apt if one is to assume a nonstationary climate.

Relative to the MLE short model, the Bayesian NSPI mean parameter estimates showed slightly smaller errors (RMSE, MAE), with less bias (ME), better explanation of temporal variance (\( R^2 \) and NSE) and smaller CIs (Table 2). Figure 8 shows how each model captured the mean parameter, with the true value shown in the lower-right corner. Seasonality, along the x axis, was captured well by the Bayesian NSPI and MLE long models, while the MLE short model produced poorer seasonal estimates, particularly around the extremes in April and October. The Bayesian NSPI model produced a smooth seasonality and long-term trend, constrained by the tensor product, unlike both MLE models, which produced more noisy estimates that remained constant, indicated by vertical constantly shaded lines. The Bayesian NSPI model captured some of the intensification from 1920 (Fig. 8, bottom) to 2019 (Fig. 8, top), although with some exaggeration (Fig. 8). This model also produced a mild pause in intensification around 1980, which may be related to random oversampling of zero precipitation \( \theta \) during this period (Fig. 8b). Of further note is the slight “washing out” of the highest frequency patterns in the Bayesian NSPI model, in which adjacent patterns were merged.

Estimates of the underlying shape parameter, which remained constant throughout the sample period (Fig. 3b), showed that the Bayesian NSPI model produced estimates that were better than the MLE short model but worse than the MLE long model for most metrics (Table 2). The exceptions were absolute bias (ME) and CI width, for which the Bayesian model performed worst. An inspection of errors showed that localized over and undersampling errors propagated both seasonally and interannually, along both dimensions, due to the tensor product constraint.
(Fig. 8b) despite this being a relatively low period for the true value (Fig. 3b). This effect was most noticeable for the MLE short model and slightly lessened in the Bayesian NSPI model. The Bayesian NSPI model was impacted by edge effects, or biased behavior at the edges of the data (Miller et al. 2013), due to a high-leverage outlier period in the late 2010s that caused overestimates in this latter period. Studies have shown that extreme outliers at the edges of the available data, in this case year, can cause a fitted surface to increase unrealistically for these regions (Miller et al. 2013). These edge effects greatly diminished by moving 15–30 years, or two knots, from the tensor product edges. The Bayesian NSPI did not adequately capture the nonstationary decrease in $\theta$ through the century, whereas the MLE models were also not able to capture this trend, but due to their assumption of stationarity (Fig. 8b). It appears that the Bayesian NSPI model’s theta estimate was more sensitive to random sampling errors than other parameters.

Estimates of the underlying SPI-3 time series using a nonstationary climate reinforced the Bayesian NSPI models’ improvement over the MLE short model and nearly equal performance to the MLE long model (Table 2). The good performance of the MLE long model, despite its incorrect assumption of stationarity, suggested that the magnitude and pattern of nonstationarity was not severe enough to produce large nonstationary biases over time. This may not always be true for real-world climate change and therefore future studies should evaluate the degree of climate trend necessary to warrant use of a nonstationary SPI model. The Bayesian NSPI model therefore had the benefit of capturing nonstationarity, unlike the MLE long model, but produced SPI estimates that were more accurate than the MLE short model, of the type typically used to represent nonstationarity.

c. Experiment 3: Instrumental precipitation

The final experiment, using instrumental precipitation, was designed to demonstrate the potential for the Bayesian NSPI model following validation in experiments 1 and 2. One of the previously noted features of the Bayesian NSPI model is its simultaneous estimation of all three parameters: $\mu$, $\alpha$, and $\theta$. The MOR gauge demonstrated this benefit (Fig. 9). Figure 9 displays the seasonal and long-term marginals for the mean parameter, $f_{\mu}(\text{day})$ and $f_{\mu}(\text{year})$, respectively (Figs. 9a,b), while the full tensor product for the mean parameter is shown in Fig. 9c. Figure 9d visualizes the fitted gamma distribution on the first day of each month ($x$ axis) and for several slices through time (individual colored boxplots). This location underwent a relatively steady increase in overall precipitation (Fig. 9b) that was strongest during the July–December wet period (Figs. 9c,d) with a temporary regime change or pause between 1930 and 1960 (Figs. 9b,c). A second, less obvious pattern occurred in April through June, when the mean remained constant but the shape parameter increased, producing a distribution with less skew, fewer extremes, and more Gaussian shape (Fig. 9d). These simultaneous and seasonally specific nonstationary trends in the mean and shape parameters would be of value to water resources planners, while the fitted model could estimate the NSPI relative to any reference period.

The proposed Bayesian NSPI model was also designed to capture seasonal shifts, best demonstrated by the WAX gauge. Historically, this region had a bimodal seasonality, with two wet periods centered on June–July and December–January (Fig. 10). During the last century, the wet season peak shifted earlier by approximately one month, from 28 June 1900 (day 179) to 4 June 2010 (day 155), as shown in red (Fig. 10). The majority of this shift happened during the latter part of the century (1960–2010), suggesting an
The USC gauge has the largest proportion of zero precipitation and so was used to check the Bayesian NSPI zero precipitation model when applied to real-world data. The cyclic seasonal predictor, \( f_d(\text{day}) \), had a sharp increase from near zero probability to 53\% on day 255 (13 September), representing Southern California’s dry period from mid-June until mid-September (Fig. 11a). The interannual marginal, \( f_d(\text{year}) \), showed a gradual increase, punctuated by some extreme decades and edge effects during the first and last 10 years (Fig. 11b). Estimates of the \( \theta \) parameter using the full tensor product appeared overly sensitive to periodic fluctuations in zero precipitation, as indicated by more decadal variation than was originally intended by the 30-yr spline penalties (Fig. 11c).

There were three periods of abnormally high zero precipitation likelihood: 1912–20, 1955–78, and 1999–2018. Additionally, the seasonal period of zero precipitation expanded during the last century by approximately 45 days, with the approximate onset shifting from August to July (day 210 to 180) and cessation shifting later from mid-October to November (day 290 to 305) (Fig. 11c).

The fitted nonstationary models showed consistent regional patterns (Fig. 12), though we must stress that nine gauges are not sufficient to draw any statistically significant conclusions. Rather, the spatiotemporal patterns are presented to illustrate the potential for the NSPI approach and to note where findings are consistent with prior studies of precipitation trends. Figure 12 shows the 3-month mean precipitation associated with SPI equal to 0 as a solid blue line and precipitation associated with an SPI of \( \pm 2 \) and \( \pm 1 \) as differently shaded regions. This figure isolates only the \( f(\text{year}) \) distribution, meaning only annually representative trends are shown, ignoring seasonal-specific trends. A consistent feature found was a wetting trend for the eastern part of the country beginning around 1960. This pattern is evident in gauges ALB, MAR, MOR, NYC, SAN, and WAX, which all shown an increase in SPI = 0 precipitation (blue line) relative to the WMO 1961–90 baseline (dashed line). For gauges like NYC and WAX, the period prior to this increase in 1960 was relatively stationary. If a nonstationary model using linear trends, such as in Russo et al. (2013), were to model these gauges, it would interpolate

\[ \text{Table 2. Nonstationary model goodness of fit for selected parameters and the resulting SPI-3. The Bayesian estimate is based on a moving reference period. MLE short is based on the 1961–90 reference period, and MLE long is based on the full period of 1920–2019.} \]

<table>
<thead>
<tr>
<th>Mean</th>
<th>MLE short</th>
<th>MLE long</th>
<th>Bayesian estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.819</td>
<td>0.350</td>
<td>0.488</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.987</td>
<td>0.420</td>
<td>0.610</td>
</tr>
<tr>
<td>ME</td>
<td>-0.369</td>
<td>-0.0498</td>
<td>-0.0863</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.292</td>
<td>0.750</td>
<td>0.644</td>
</tr>
<tr>
<td>NSE</td>
<td>-0.716</td>
<td>0.690</td>
<td>0.345</td>
</tr>
<tr>
<td>CI width</td>
<td>3.40</td>
<td>1.56</td>
<td>1.63</td>
</tr>
<tr>
<td>CI hit rate</td>
<td>0.838</td>
<td>0.840</td>
<td>0.909</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shape</th>
<th>MLE short</th>
<th>MLE long</th>
<th>Bayesian estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.390</td>
<td>0.213</td>
<td>0.320</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.502</td>
<td>0.274</td>
<td>0.447</td>
</tr>
<tr>
<td>ME</td>
<td>-0.096</td>
<td>0.122</td>
<td>0.287</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.000</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>NSE</td>
<td>-3.84</td>
<td>-0.908</td>
<td>-2.36</td>
</tr>
<tr>
<td>CI width</td>
<td>1.80</td>
<td>0.99</td>
<td>1.91</td>
</tr>
<tr>
<td>CI hit rate</td>
<td>0.956</td>
<td>0.975</td>
<td>0.957</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theta</th>
<th>MLE short</th>
<th>MLE long</th>
<th>Bayesian estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.0427</td>
<td>0.0278</td>
<td>0.0341</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0538</td>
<td>0.0338</td>
<td>0.0448</td>
</tr>
<tr>
<td>ME</td>
<td>0.00729</td>
<td>-0.0089</td>
<td>-0.0116</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.000</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>NSE</td>
<td>-3.84</td>
<td>-0.908</td>
<td>-2.36</td>
</tr>
<tr>
<td>CI width</td>
<td>1.80</td>
<td>0.99</td>
<td>1.91</td>
</tr>
<tr>
<td>CI hit rate</td>
<td>0.956</td>
<td>0.975</td>
<td>0.957</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPI-3</th>
<th>MLE short</th>
<th>MLE long</th>
<th>Bayesian estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.276</td>
<td>0.137</td>
<td>0.182</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.370</td>
<td>0.188</td>
<td>0.248</td>
</tr>
<tr>
<td>ME</td>
<td>0.141</td>
<td>0.003</td>
<td>-0.001</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.884</td>
<td>0.963</td>
<td>0.937</td>
</tr>
<tr>
<td>NSE</td>
<td>0.854</td>
<td>0.962</td>
<td>0.934</td>
</tr>
<tr>
<td>CI width</td>
<td>0.937</td>
<td>0.558</td>
<td>0.711</td>
</tr>
<tr>
<td>CI hit rate</td>
<td>0.794</td>
<td>0.855</td>
<td>0.856</td>
</tr>
</tbody>
</table>
onset. The USC gauge in the semiarid west shows the opposite trend, drying during the modern period for typical years and the extreme tails of the distributions (Fig. 12). Because only the $f(\text{year})$ predictor was shown in Fig. 12, precipitation remained positive; however, if seasonality were included, the distribution would show zero precipitation for SPI ≤0 during late summer when $\theta$ values were greater than 50%. The other West Coast gauge (ABD), the wettest gauge in the continental United States, showed negligible changes for median precipitation (SPI = 0) but an expansion of both the wet and dry extremes as controlled by the shape parameter. A shape parameter trend was also detected for the WAX gauge, where increasing trends for the wet extremes (SPI = 2) exceeded trends for the median (SPI = 0). Capturing such changes in interannual variance would be important for decision makers who are often most concerned with wet and dry extremes.

5. Discussion

a. Benefits and limitations

The Bayesian NSPI model introduced in this paper was purpose-built to address several limitations of traditional SPI and nonstationary SPI fitting. A first benefit confirmed by testing was a reduction in parameter uncertainty and an increase in estimate stability by constraining estimates from day-to-day (stationary model) and over multidecadal periods (nonstationary model). Results showed that the Bayesian model performed slightly better than the most appropriate MLE methods for the stationary (MLE long) and nonstationary comparisons (MLE short). The mechanism for improvement in a stationary climate is presumed to be the temporal spline constraint on the parameters, allowing surrounding days to provide valuable information, rather than assuming independent fits. The mechanism for nonstationary estimate improvement is presumed to be the use of information from the full 100-yr dataset, while still mimicking a 30-yr reference period. While the Bayesian NSPI model performed slightly worse than the MLE long model for nonstationary climates (experiment 2), the MLE long model is a stationary model using the entire record, which does not account for climate trends and makes direct interstudy comparisons more challenging, as described later in this section.

A second benefit is the simultaneous modeling of the mean, shape, and theta parameters within a single model, rather than multiple independent daily or monthly models. Some prior NSPI methods have focused on changes in the mean parameter while keeping the shape parameter fixed (Russo et al. 2013). The GAMLSS approach of Shiau (2020) modeled mean and shape parameters, but fit separate daily models and did not consider the $\theta$ parameter. The added flexibility of simultaneously modeling mean and shape is a significant benefit for modeling realistic nonstationarity, as some climate change projections show more significant trends in extreme precipitation values (Wuebbles et al. 2014; Donat et al. 2013; Bao et al. 2017; Papalexiou and Montanari 2019). This was notable for...
the ABD and WAX annual trends (Fig. 11), which both showed increasing interannual variance captured by the $\alpha$ parameter and the MOR gauge (Fig. 9), which showed a decrease in interannual variance during spring. The importance of modeling the $\theta$ parameter alongside parameters of the gamma distribution was most evident in experiment 1 (Fig. 8), while the importance of estimating uncertainty around zero precipitation was highlighted throughout the study. As noted in experiment 1, most MLE methods use empirical estimates, which do not provide uncertainty estimates and have the potential for infeasible SPI values.

A third benefit of the proposed model is the flexibility to model nonlinear trends, capturing features like pauses (Fig. 9b), step changes, multidecadal cyclicity (Fig. 12, PAO gauge), and delayed onset or acceleration of trends (Fig. 12, NYC or USC gauges). Prior nonstationary methods that use linear predictors (Russo et al. 2013; Pryor et al. 2009) would limit or distort some of these patterns. The ability to model delayed onset or acceleration is particularly important when models incorporate precipitation measurements that extend far into the past, for example with global climate simulations, recovered historical records, or proxy-based reconstructions.

**Fig. 9.** Fitted Bayesian nonstationary model for the MOR gauge. Marginal plots for the mean parameter show the effect of (a) day and (b) year in isolation, and the (c) tensor product shows the combined effect of these two marginals and the full tensor model. For (a) and (b), the 95% CI is shown in light gray, and the 95% interval for the estimate is shown in dark gray, as in Fig. 5. For (a) in particular, the dark-gray region is so small that it may appear indistinct from the estimate. (d) A box-and-whisker plot showing the combined effect of mean and shape parameters on the precipitation distribution, taken at month ($x$ axis) and year time slices (colors).
FIG. 10. Fitted Bayesian nonstationary model of the mean parameter for the WAX gauge. As in prior tensor product plots, axes represent time intervals and the color scheme represents the parameter estimate. The red line shows the date of the annual maximum mean parameter through time.

By modeling preindustrial data alongside modern observations, it becomes useful to identify the onset of deviations from long-term patterns. Nonlinear flexibility is also valuable for capturing seasonal shifts, as in the WAX or USC gauges (Figs. 10 and 11c). The expansion or contraction of wet/dry seasons is another mode of change implied by climate change that is captured by this model. For example, the USC gauge (Fig. 11c) appears to show an expansion of the dry season in the later twentieth century, beginning several weeks earlier and ending several weeks later.

The proposed Bayesian NSPI model does have several limitations. First, although model flexibility is a strength, it comes at the cost of potential overfitting and excessive wiggliness. Despite attempts to control for overfitting through Bayesian priors that approximate a 30-yr WMO period, some overfitting occurred in this study, primarily for the $\theta$ parameter (Figs. 5 and 11). The $\theta$ parameter appears particularly susceptible to overfitting because most gauges have relatively few zero precipitation periods, making the logistic regression model overly sensitive to isolated periods of zero precipitation. Future iterations of the Bayesian model could build better constraints to minimize this overfitting. This opportunity for improvement is contrasted with the MLE empirical approach that also overfit $\theta$, but could not be modified in the future. A second noted limitation is the presence of edge effects, which can induce errors in the earliest and latest time periods. While cardinal splines were used to prevent extreme behavior at the edges, we still recommend users avoid relying on parameter estimates from within half a WMO period (15 years) of the edges. Also, the proposed model is limited by computational speed. While a fit using Stan’s non-Bayesian method (penalized MLE) required 2 or 4 min for the stationary and nonstationary models, respectively, a full Bayesian fit with 2000 posterior samples required 15 and 20 h for the stationary and nonstationary models, respectively. Fitting independent, stationary MLE models is significantly faster than this, requiring only 40 s. We anticipate speed will increase with improvements in code efficiency, computing power, or if users are willing to simplify portions of the model framework. For example, the approaches by Wood et al. (2017) could provide quicker model fitting but are not suited to hurdle models needed to model $\theta$ simultaneously [Eq. (6)]. Similarly, if the user is willing to use Stan’s approximate Bayesian automatic differentiation variational inference algorithm (Kucukelbir et al. 2015), times are much closer to MLE, requiring 2 or 18 min for the stationary and nonstationary models, respectively.

b. Spatiotemporal patterns in instrumental gauges

While the gauges in experiment 3 represent a small subset of the United States, it is useful to compare findings with prior studies of historical trends and climate modeling to identify commonalities. The nine instrumental gauges in this study tend to cluster into broad spatial patterns, which agree with several previous studies. The most consistent pattern occurred across much of the eastern and central United States, where precipitation increased during the twentieth century for most seasons (Fig. 11), punctuated by a pause during the 1950s and 1960s and frequently followed by acceleration until the present. Previous studies show similar wetting trends across a more complete set of instrumental gauges for the central and eastern United States (Garbretch and Rossel 2002; Mallakpour and Villarini 2015; Pryor et al. 2009; Schilling and Libra 2003). This observed wetting trend is further supported by modeled precipitation trends for eastern North American under the influence of anthropogenic greenhouse gases (Cook et al. 2018; Almazroui et al. 2021).

The temporary interruption between 1960 and 1970 of the century-long wetting trend for the eastern United States (Fig. 12) has been noted previously (Bishop et al. 2019; Andreadis and Lettenmaier 2006; Sarkar and Maity 2021; Reid et al. 2016; Meehl et al. 2009; Groisman et al. 2004). Some of this pause has been attributed to a climatic regime change in the Indian and Pacific Oceans (Strong et al. 2020; Sarkar and Maity 2021; Reid et al. 2016; Meehl et al. 2009; Groisman et al. 2004), a phase change for the Atlantic multidecadal oscillation (Enfield et al. 2001), and a change in moisture delivery from the Gulf of Mexico (Bishop et al. 2019). The resultant precipitation signal would be similar to that presented here, with natural climate variability overlaid on anthropogenic driven long-term wetting trends. The NSPI model captures the resultant nonlinear pause, followed by accelerating wetting trends, but does not disaggregate the anthropogenic effect from natural climate variability.

In opposition to the wetting eastern United States, analysis of the western continental gauges (USC and ABD) indicated a drying southwest (Fig. 12) and relatively unchanged median
behavior for the northwest, coupled with intensifying extremes (Fig. 11). A drying southwest, represented here by the USC gauge is both a feature of historical analyses of precipitation (Griffin and Anchukaitis 2014) and projections of climate change (Cook et al. 2018). The periods of excessively high \( \theta \) values for the USC gauge correspond to periods 1910-30, 1960-70, and 2000-present, each of which are documented periods of multiyear droughts (Griffin and Anchukaitis 2014). These periods also correspond to consistently negative (La Niña) periods of the multivariate ENSO index (MEI) (Wolter and Timlin 2011), which aligns with current understanding of Southern California hydroclimatic drivers (Cook et al. 2017). The negligible trend for median precipitation in the northwest agrees with prior studies (Booth et al. 2012; Mote 2003); however, the finding of intensifying extremes and interannual variability may be of greater importance to water managers and has not been previously noted. The lack of long-term trend for the PAO gauge agrees with a lack of significant trends found previously for the northern shore of Hawaii’s “Big Island” (Doty 1982; Chu et al. 2010).

Seasonal specific trends found via the Bayesian NSPI also corroborate prior findings. For example, the autumn- and winter-specific precipitation increase for the southeast (ALB and WAX) and Puerto Rico (SAN) has been documented (Mote 2003; Bishop et al. 2019). The seasonal shift toward earlier precipitation for the WAX gauge (Fig. 10) was noted for other gauges near the Gulf of Mexico (Pal et al. 2013; Coopersmith et al. 2014). Each of these findings provide greater confidence in the validity of the proposed nonstationary NSPI method.

6. Conclusions

This study was designed to introduce and evaluate a Bayesian nonstationary SPI (NSPI) model using tensor product splines to represent parameters that smoothly change both seasonally and over multidecadal scales. All comparisons were made relative to SPI approaches using MLE for predetermined stationary and nonstationary climates. This comparison showed the proposed Bayesian NSPI model cannot only reproduce the performance of MLE models but improves upon them in several key areas. Notably, for a stationary climate, the proposed model was better than its direct comparison model for accurately reproducing distribution parameters and the resulting SPI. This improvement was attributed to the seasonal spline, which permits sharing of information between nearby days, rather than the MLE approach, which assumes independent distributions for each day. For nonstationary climates, the proposed Bayesian NSPI model performed better than the existing MLE alternative using a quasi-stationary WMO 30-yr period (MLE short), with these benefits attributed to use of the entire dataset, made possible by imposing a tensor product spline that mimics the 30-yr period. The MLE long model, performed better with respect to several metrics, but did not capture any of the true, underlying nonstationarity. This leads to a
recommendation for future studies to evaluate the degree of nonstationarity needed to warrant a nonstationary SPI model. The Bayesian NSPI model was shown to have several other benefits: capturing nonlinear trends and seasonal shifts, permitting trends in the higher moment parameters (shape and \( \theta \)), and providing uncertainty bounds for the \( \theta \) parameter. For semiarid climates or for short accumulation periods (\( \leq 3 \) months), the improved handling of zero precipitation \( \theta \) would produce better SPI estimates, even for the stationary model. These benefits were further demonstrated when the model was applied to instrumental gauges. Nonstationary model findings for instrumental gauges broadly agreed with current best understanding of long-term precipitation trends across the United States during the last century.

The model presented here is designed as a flexible framework, rather than a single, static model. For example, even if users do not want to consider nonstationarity, their calibration could benefit from using the stationary model because of the findings described above. Also, the existing models could be improved to incorporate better constraints, for example on the \( \theta \) parameter or merging multiple data sources. This framework also could be extended to other normalized indices, such as the standardized precipitation–evapotranspiration index (Vicente-Serrano et al. 2010) or the standardized runoff index (Shukla and Wood 2008), which share a method with the SPI but normalize other measures of the hydrologic cycle.

The benefits of this Bayesian NSPI model will be important for water managers and drought agencies, which must estimate the severity of meteorological or hydrologic drought relative to agreed-upon historical baselines while using instrumental data subject to a nonstationary climate. The challenges of climate change mean that nonstationarity must be incorporated into relative drought estimates and the proposed Bayesian NSPI model permits this calculation, with better accuracy than a quasi-stationary reference period and with more flexibility with regard to differing reference periods.

Acknowledgments. This work was supported primarily by the Paleo Perspectives on Climate Change (P2C2) Program of the National Science Foundation under NSF Project 2002539: “Leveraging Bayesian Approaches to Link Reconstructed, Observed, and Projected Meteorological Drought while Accounting for Inherent Data Biases.” Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation. The work was also supported by the Byrd Polar and Climate Research Center at The Ohio State University.

Data availability statement. All code, data, and instructions are available via an open-access repository (https://doi.org/10.5281/zenodo.5642734). This code has been tested to generate all figures and tables presented in this paper.

APPENDIX

Gauge Details

Table A1 shows details for the gauges included in the NSPI analysis.
### Table A1. Details for gauges included in the NSPI analysis.

<table>
<thead>
<tr>
<th>Gauge</th>
<th>Site</th>
<th>Tavg (C)</th>
<th>Pavg (mm)</th>
<th>Snow (mm)</th>
<th>Elev (m)</th>
<th>Köppen–Geiger</th>
<th>Climate group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABD</td>
<td>Aberdeen, WA</td>
<td>10.3</td>
<td>2110</td>
<td>120</td>
<td>3</td>
<td>Cfb</td>
<td>Temperate</td>
</tr>
<tr>
<td>ALB</td>
<td>Albany, GA</td>
<td>19.3</td>
<td>1270</td>
<td>0</td>
<td>55</td>
<td>Cfa</td>
<td>Temperate</td>
</tr>
<tr>
<td>MAR</td>
<td>Marysville, OH</td>
<td>10.7</td>
<td>960</td>
<td>640</td>
<td>302</td>
<td>Cfa</td>
<td>Temperate</td>
</tr>
<tr>
<td>MOR</td>
<td>Morris, MN</td>
<td>5.6</td>
<td>630</td>
<td>1120</td>
<td>348</td>
<td>Dfb</td>
<td>Continental</td>
</tr>
<tr>
<td>NYC</td>
<td>New York, NY</td>
<td>12.3</td>
<td>1160</td>
<td>700</td>
<td>40</td>
<td>Cfa</td>
<td>Temperate</td>
</tr>
<tr>
<td>PAO</td>
<td>Paauilo, HI</td>
<td>21.3</td>
<td>2460</td>
<td>0</td>
<td>70</td>
<td>Af</td>
<td>Tropical</td>
</tr>
<tr>
<td>SAN</td>
<td>San Juan, PR</td>
<td>26.9</td>
<td>1430</td>
<td>0</td>
<td>3</td>
<td>Af</td>
<td>Tropical</td>
</tr>
<tr>
<td>USC</td>
<td>Los Angeles, CA</td>
<td>18.5</td>
<td>360</td>
<td>0</td>
<td>70</td>
<td>BSk</td>
<td>Dry</td>
</tr>
<tr>
<td>WAX</td>
<td>Waxahachie, TX</td>
<td>18.9</td>
<td>920</td>
<td>20</td>
<td>191</td>
<td>Cfa</td>
<td>Temperate</td>
</tr>
</tbody>
</table>

### REFERENCES


Heinrich, G., and A. Gobiet, 2012: The future of dry and wet spells in Europe: A comprehensive study based on the


Stahle, D. W., and Coauthors, 2020: Dynamics, variability, and change in seasonal precipitation reconstructions for North


