सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

Auto-Regressive Standard Precipitation Index: A Bayesian Approach for Drought Characterization

Sujay Mukhoti

Soham Ghosh

Pritee Sharma

WP/03/2024-25/OM&QT
April 2025

Auto-Regressive Standard Precipitation Index: A Bayesian Approach for Drought Characterization

Soham Ghosh[a], Sujay Mukhoti[b], Pritee Sharma[a]

[a]*Department of Economics, School of Humanities and Social Sciences, Indian Institute of Technology Indore, Indore, 453552, Madhya Pradesh, India*
[b]*Operations Management and Quantitative Techniques Area, Indian Institute of Management Indore, Indore, 453556, Madhya Pradesh, India*

Abstract

This study proposes Auto-Regressive Standardized Precipitation Index (ARSPI) as a novel alternative to the traditional Standardized Precipitation Index (SPI) for measuring drought by relaxing the assumption of independent and identical rainfall distribution over time. ARSPI utilizes an auto-regressive framework to tackle the auto-correlated characteristics of precipitation, providing a more precise depiction of drought dynamics. The proposed model integrates a spike-and-slab log-normal distribution for zero rainfall seasons. The Bayesian Monte Carlo Markov Chain (MCMC) approach simplifies the SPI computation using the non-parametric predictive density estimation of total rainfall across various time windows from simulated samples. The MCMC simulations further ensure robust estimation of severity, duration, peak and return period with greater precision. This study also provides a comparison between the performances of ARSPI and SPI using the precipitation data from the Colorado River Basin (1893–1991). ARSPI emerges to be more efficient than the benchmark SPI in terms of model fit. ARSPI shows enhanced sensitivity to climatic extremes, making it a valuable tool for hydrological research and water resource management.

Keywords: Standardized Precipitation Index, Auto-correlated Rainfall, Bayesian Estimation, Markov Chain Monte-Carlo Simulation, Predictive Density.

1. Introduction

Drought is recognized as a significant natural disaster that affects every region of the world (Wilhite et al., 2007). Its effects are felt widely across agriculture, socio-economics, and the environment. Existing drought indicators assume that the precipitation data are independent over time, although there is substantial amount of literature indicating existence of weak-stationarity in precipitation data. In this work, we develop a drought index accounting for the weak stationarity using a Bayesian auto-regressive model and study the key characteristics of drought based on the proposed index.

A key challenge in studying drought is the lack of a universally accepted definition, as experts from various disciplines define it based on their specific perspectives. Accordingly, drought is categorized into four types: Meteorological, Agricultural, Hydrological, and Socio-Economic (Yihdego et al., 2017). Meteorological drought refers to a deficiency in precipitation over a given time period. Agricultural drought is linked to prolonged insufficient crop yields and inadequate soil moisture. Socio-Economic drought occurs when water resource systems are unable to meet societal water needs, while hydrological drought is characterized by a lack of surface and subsurface water resources. Efforts to refine drought definitions continue, focusing on its impacts on different societal and environmental aspects.

Numerous drought indices have been developed so far to monitor and assess various types of drought. One of the most popular drought indices is the Standardized Precipitation Index (SPI), which is used for measuring meteorological drought across the globe

(Ntale and Gan, 2003, Tsakiris et al., 2007, Zhao et al., 2020, Yildirim and Rahman, 2022, Ullah et al., 2023). SPI provides valuable insights including a spectrum from arid to drought conditions, which can be challenging to assess otherwise (Mukhawana et al., 2024). The World Meteorological Organization (WMO) recommends SPI as a primary indicator of drought (WMO, 2006).

The computation of the Standardized Precipitation Index (SPI) is based on the premise that the moving totals of rainfall (MTR) data over a window of time are independent and identically distributed (i.i.d.) realizations of Gamma random variables (Mckee et al., 1993). Numerous studies have shown that rainfall series from various regions and time windows tend to exhibit positive skewness (Saha et al., 2018, Haktanir et al., 2012, Sankarasubramanian and Srinivasan, 1999, YU and NAGHAVI, 1994). The use of the Gamma distribution helps to address the non-negativity as well as the skewness in precipitation data. However, the Gamma distribution also presents certain limitations. For instance, it has a thin tail, meaning it underestimates the likelihood of extreme rainfall events, making it less suitable for modeling rainfall series that include frequent heavy rainfalls. Additionally, the Gamma distribution may not fit well with data that have extremely low variance, potentially leading to poor parameter estimation due to over-dispersion.

A key challenge in constructing the SPI, which has not been thoroughly addressed in the literature, is how to model MTR data that do not satisfy the i.i.d. assumption. The i.i.d assumption reflects strict stationarity of the MTR series, which in many real-world situations, including precipitation series, is often violated. Das et al. (2020) developed a drought index for non-stationary precipitation data and demonstrated its efficiency in capturing different drought characteristics using a generalized additive model. However, weak-stationarity (or time independent auto-correlation) in precipitation data is most commonly reported in the literature. For example, Soltani et al. (2007) reported significant variation in the temporal patterns of monthly rainfall series across much of Iran. Obeysekera et al. (1987) also observed auto-correlation in hourly rainfall data in northeastern Colorado, while Delleur and Kavvas (1978) demonstrated similar findings for monthly rainfall series in 15 river basins in Indiana, Kentucky, and Illinois. Additionally, Labadie et al. (1981) identified significant short-term auto-correlation in rainfall data for northern San Francisco, and Mahmud et al. (2017) documented auto-correlation in monthly rainfall series with a 12-month lead time across 30 meteorological stations in Bangladesh.

Another notable limitation of the Gamma distribution is its complexity in modeling auto-correlated MTR data. In contrast, the log-normal distribution distribution has been recommended as a viable alternative (see Angelidis et al., 2012, Lloyd-Hughes and Saunders, 2002, Lana and Burgueño, 2000, Stagge et al., 2015, Shoji and Kitaura, 2006, Moccia et al., 2022, for example). It provides a more parsimonious model to accommodate non-negative support, positive skewness and auto-correlation. The conditional mean of log-transformed MTR data can be effectively represented as a stationary Gaussian auto-regressive (AR) process under the assumption of log-normality (Wilks, 2011). A key challenge in constructing the Standardized Precipitation Index (SPI) is accurately accounting for the occurrence of zero-rainfall days, which vary with seasonal changes. Addressing this zero-inflation alongside auto-correlation proves to be difficult. In this paper, we propose a spike-and-slab log-Normal AR model for MTR data, where MTR is modeled in two parts: a probability spike at zero and the remainder distributed as a log-Normal variable with AR conditional mean (see Lempers, 1971, Mitchell and Beauchamp, 1987, for review of spike and slab). During the construction of the Standardized Precipitation Index (SPI), it's necessary to assess the cumulative distribution function (CDF) at the de-seasonalised MTR values and create a Gaussian mapping based on the resulting CDFs. It's worth noting that parameter and CDF estimation in such a complex model can be challenging. In this paper we use Bayesian method for parameter estimation and

simulate predictive densities of MTR for CDF estimation.

Recent advances in computational techniques, such as Markov Chain Monte Carlo (MCMC) methods, have contributed to the widespread adoption of Bayesian inference in hydrology (Tang et al., 2016, Smith et al., 2015). The Bayesian approach incorporates uncertainty in model parameters by assigning known prior probability distributions. Bayesian inference seeks to update these priors by deriving the posterior distributions of the parameters based on observed data. However, deriving the exact posterior can be challenging for complex models. This challenge is typically addressed by using MCMC methods, like the Gibbs sampler or Metropolis-Hastings algorithms, to generate samples from the posterior distribution (see Chib and Greenberg (1995), for a review). In practice, the posterior mean is often reported as the Bayesian estimate, aligning with the frequentist perspective.

In this study, the proposed spike-and-slab log-Normal AR model is structured as a Bayesian network. Here, the nodes represent MTR distributions, which are either degenerate at zero on dry days or log-Normal with an auto-regressive (AR) mean function on rainy days. The prior for these two states is encoded by the edges, modeled as a Bernoulli distribution with the probability of success representing the likelihood of no rainfall. The weak stationarity of the MTR distribution is accounted for by including a Uniform prior on the lagged log-MTR coefficient within the mean function of the log-transformed MTR data, referred to as a log-stationary process. Other parameters are assigned non-informative priors due to a lack of particular prior knowledge.

The model's predictive density is obtained by averaging the model's density over the posterior distribution of the parameters. In this Bayesian framework, simulated MTR approximates the posterior predictive distribution, allowing for the empirical estimation of the cumulative distribution function (CDF) at observed precipitation values. Furthermore, uncertainty in drought characteristics, such as severity, duration, and peak, can be assessed using independent and identically distributed (i.i.d.) predictive samples.

SPI was invented by a group of scientists at Colorado state University and applied as an experimental tool for drought monitoring in Colorado river basin during its inception. Climate condition of the basin is extremely vulnerable and it faces dual challenges of extreme drought and periodic flooding conditions. In this paper we use the same data, *viz.* the monthly precipitation dataset for the time period 1893–1991, for calculating and comparing the performances of the SPI and the proposed one.

The rest of the paper is organized as follows: Section 2 introduces the development of ARSPI. The index accounts for the auto-correlated structure of the precipitation time series. Section 3 elaborates a Bayesian MCMC approach for estimating the probabilistic structure of rainfall, leveraging Bayesian methods to impose necessary constraints. Section 4 demonstrates the real application of ARSPI using precipitation data from the Colorado River Basin (1893-1991). Finally, the paper concludes with a discussion of the key findings and their implications.

2. Precipitation Index for Auto-Correlated Rainfall Data

2.1. Conventional SPI

The Standardized Precipitation Index (SPI), introduced by Mckee et al. (1993), is based on monthly rainfall data. The construction approach can be described as follows: let $X_1, X_2, \ldots, X_T$ represent i.i.d. monthly rainfall observations collected over T consecutive months (where $T \geq 360$, as per Mckee et al. (1993)). A moving total over a window of ζ months is computed as

$$r_t = \sum_{i=t-\zeta+1}^{t} X_i, \quad \zeta = 1, 3, 6, \ldots, 72.$$

The resulting series r_t is referred to as moving total rainfall (MTR) from here onward. The non-zero MTR, denoted as

$$Y_t = I(r_t > 0)r_t,$$

where $I(\cdot)$ is the indicator function, is then fitted to a two-parameter Gamma distribution. The probability density function (pdf) of the Gamma distribution with shape parameter k and scale parameter τ is given as follows:

$$f_{k,\tau}(y) = \frac{1}{\Gamma(k)\tau^k}y^{k-1}e^{-\frac{y}{\tau}}, \quad y, k, \tau > 0, \tag{1}$$

where $\Gamma(k)$ is the Gamma function, defined as $\Gamma(k) = \int_0^\infty u^{k-1}e^{-u}\, du$. The shape and scale parameters are estimated using the method of moments. The estimates are given by the following equations:

$$\hat{\alpha} = \frac{1}{4B}\left(1 + \sqrt{1 + \frac{4B}{3}}\right), \quad \hat{k} = \frac{\bar{y}}{\hat{\tau}},$$

where $B = \ln(\bar{y}) - \frac{\sum_{j=1}^m \ln(y_j)}{m}$, $y_1, y_2, \ldots, y_m$ are samples from the MTR series, and $\bar{y}$ is the sample mean. Using these estimates, the Cumulative Density Function (CDF) is computed for each observation $y_1, y_2, \ldots, y_m$ as:

$$\hat{F}(y_t) = \int_0^{y_t} f_{k,\tau}(x)\, dx\bigg|_{k=\hat{k},\tau=\hat{\tau}}, \quad t = 1, 2, \ldots, m, \tag{2}$$

where $f_{k,\tau}(x)$ represents the Gamma probability density function with shape parameter k and scale parameter τ. The probability of a zero-rainfall window is estimated as

$$\pi = \frac{v}{T},$$

where v represents the number of zero MTR values, and T is the total number of observations. Thus the modified CDF of the MTR data is given by:

$$G(y_t) = \begin{cases} \pi, & \text{if } y_t = 0 \\ \pi + (1-\pi)F(y_t), & \text{if } y_t > 0, \quad t = 1, 2, \ldots, m \end{cases} \tag{3}$$

Finally, the CDF values are transformed into a standard Normal variate via the inverse Gaussian transformation, $SPI_t = \Phi^{-1}(G(y_t))$, $t = 1, 2, \ldots m$.

2.2. Precipitation Index for Auto-Correlated Rainfall

Lloyd-Hughes and Saunders (2002) demonstrated the effectiveness of the Standardized Precipitation Index (SPI) for drought classification following normalization. Table 1 presents the index values corresponding to both wet and drought conditions. The SPI is calculated using a moving window across various time scales. While shorter time scales tend to produce frequent fluctuations around zero, longer time scales yield a smoother SPI response to changes in rainfall. However, in regions with low seasonal precipitation, shorter time scales can inaccurately indicate severe drought through highly negative SPI values.

In this study, we relax the traditional assumption of independent and identically distributed (i.i.d.) observations commonly used in SPI computation. We introduce a more realistic framework for calculating the Standardized Precipitation Index (SPI) by modeling the logarithm of the mean total rainfall (MTR) during rainy periods, $\log y_t$, as a weakly stationary Gaussian auto-regressive (AR) process, resulting in a log-normal distribution for non-zero MTR values. To account for seasons with zero rainfall, we

represent the MTR as a spike-and-slab process with a time-varying probability of zero rainfall ($r_t = 0$) within a given window. The MTR at time t (r_t) is modeled as follows:

$$
\begin{aligned}
Z_t &\sim \text{Bernoulli}(\pi_t), \quad 0 < \pi_t < 1 \\
r_t &= Z_t + (1 - Z_t)y_t \\
y_t \mid r_{t-1} &\sim \text{log-Normal}(\mu_t, \sigma_t^2) \\
logit(\pi_t) &= \alpha + \phi \, logit(\pi_{t-1}) \\
\mu_t &= e^{\beta_1 + \frac{\sigma_t^2}{2}} r_{t-1}^{\beta_2} \\
\sigma_t &= \sigma(1 - Z_t), \quad t = 1, 2, \ldots, T
\end{aligned}
\tag{4}
$$

This approach allows for a more general and data-driven representation of rainfall dynamics, particularly in regions where dry periods are frequent.

The proposed model suggests that, following a dry period, rainfall events are driven by a log-normal shock with a constant variance σ^2. During wet periods, the MTR at each time point depends on the preceding rainfall amount. To ensure the expected MTR remains stable and does not diverge, we impose the condition $|\beta_2| < 1$, thereby guaranteeing that $\log \mu_t = \log \mathbb{E}[Y_t \mid r_{t-1}]$ is weakly stationary. The Auto-Regressive Standardized Precipitation Index (ARSPI) at time point t is obtained by applying the inverse Gaussian transformation to $F(r_t) = P[r_t \leq r_t]$, similar to the conventional SPI, ensuring consistent interpretation. However, calculating $F(r_t)$ based on the proposed model is inherently complex, in addition to the challenges involved in estimating the unknown model parameters.

3. Bayesian Estimation of ARSPI

We employ a Bayesian Monte Carlo Markov Chain (MCMC) approach for estimation (Carlin and Louis, 2008). The Bayesian estimation method offers several advantages in this context. First, Bayesian methods facilitate the straightforward imposition of constraints required for ensuring the stationarity of auto-correlated rainfall and the time-dependent rainfall probabilities through the selection of appropriate priors, a task that can be more complex in a frequentist framework. Another advantage of the Bayesian approach is that it simplifies the estimation process, reducing the computational complexity compared to frequentist methods.

In the initial stage of Bayesian analysis, the probabilistic model underlying the MTR data is defined through the likelihood function (joint density) $L(\mathbf{r} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \phi, \beta_1, \beta_2, \sigma) \in \Theta$. In the Bayesian framework, $\boldsymbol{\theta}$ is treated as a random vector with associated probability distribution $\pi(\boldsymbol{\theta})$, referred to as the prior distribution. Priors can be subjective, derived from expert knowledge relevant to the data context. Alternatively, they may be non-informative, conveying minimal prior information, and are often considered objective in nature. The objective of Bayesian inference is to determine the posterior distribution of $\boldsymbol{\theta}$ given the observed data $\mathbf{r}$, which is obtained using Bayes' theorem:

$$
\pi(\boldsymbol{\theta} \mid \mathbf{r}) = \frac{\pi(\boldsymbol{\theta})L(\mathbf{r} \mid \boldsymbol{\theta})}{\int_\Theta \pi(\boldsymbol{\theta})L(\mathbf{r} \mid \boldsymbol{\theta})d\boldsymbol{\theta}}.
\tag{5}
$$

Bayesian estimates of the model parameters are typically represented by the posterior mean. For the computation of the Standardized Precipitation Index (SPI), it is necessary to evaluate the cumulative distribution function (CDF) at specified MTR values. In the Bayesian framework, this computation is simplified by deriving the empirical CDF from simulated observations of the predictive distribution of the MTR data. The predictive density of MTR at i^{th} time point ($\tilde{r}_i$) is defined as:

$$
f(\tilde{r}_i \mid \boldsymbol{r}) = \int_\Theta f(\tilde{r}_i \mid \boldsymbol{r}, \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \boldsymbol{r})d\boldsymbol{\theta}
$$

where Θ denotes the parameter space. The cumulative distribution function (CDF) $F(r_t) = P[r \leq r_t \mid \boldsymbol{r}]$ is estimated empirically from a large sample $\{\tilde{r}_t^{(i)}, i = 1, 2, \ldots, M\}$ drawn from the predictive density $f(\tilde{r}_t \mid \boldsymbol{r})$. The empirical CDF is given by:

$$\hat{F}(r_t) = \frac{1}{M} \sum_{i=1}^{M} I(\tilde{r}_t^{(i)} \leq r_t)$$

where $I(\cdot)$ is the indicator function. Finally, the auto-regressive SPI (ARSPI) is obtained by applying the inverse Gaussian transformation $\Phi^{-1}\left(\hat{F}(r_t)\right)$. This newly derived index have the interpretation as outlined in Table 1.

3.1. MCMC Convergence

Evaluating the convergence of MCMC chains is a critical step in Bayesian analysis. To assess convergence, we employ graphical approaches. The trace plots, generated after discarding the burn-in period and applying thinning (or selecting every "thin"-th value), visualize the sampled values and provide insights into the behavior of the MCMC sequence. Evidence of non-convergence would be indicated by discernible trends or persistent correlations in the trace plots.

3.2. Model Adequacy Measure

To assess the fit of the model, we use the Deviance Information Criterion (DIC) as introduced by Spiegelhalter et al. (2002). The DIC consists of two main components: the deviance ($\bar{D}$) and the model complexity (p_D). The deviance is defined as:

$$\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}} \left[-2 \log f(\mathbf{y} \mid \boldsymbol{\theta}) \right] \tag{6}$$

and serves as a measure of how well the model fits the observed data. The complexity of the model (p_D) is determined by the effective number of parameters and is given by:

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}) = \bar{D} + 2 \log f(\mathbf{y} \mid \bar{\boldsymbol{\theta}}) \tag{7}$$

where $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta} \mid \mathbf{y}]$ denotes the posterior mean of the model parameters.

4. Colorado Basin Rainfall Data Analysis

The rainfall data comprises monthly measurements from 1893 to 1991 in the Colorado River basin. Below, we briefly highlight the significance of this dataset. The Colorado River stretches approximately 2,300 kilometers, originating in the Rocky Mountains of Colorado and Wyoming and flowing to the Gulf of California. The geographical coordinates of the basin range from 31° N to 42° N and 105.5° W to 118° W (see Figure 1 (Harris, 2012)). Water resources within the basin are shared among the growing populations of seven U.S. states and parts of Mexico. The basin is home to two of the largest reservoirs, Lake Mead and Lake Powell. In recent decades, the Colorado River Basin has faced substantial drought conditions, marking one of the driest periods in its history. This prolonged drought is exacerbated by rising temperatures and diminishing snowpacks, which are critical for water supply in the region. The ongoing drought has the potential to significantly impact agricultural practices and energy production for a large portion of the population.

4.1. Time-series Statistics of the Colorado Rainfall Data

Figure 2 illustrates the monthly precipitation series for Colorado River basin. The series demonstrates the highly variable nature of monthly rainfall in Colorado, with values ranging from 0 to 9.27 inches. The precipitation series in the basin has fluctuated without any long-term trend over nearly a 100-year period. Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) are widely used statistical techniques to determine the relationship between observations in a time series with one another. Plotting the ACF and PACF values against subsequent time lags is a statistical method for choosing appropriate time series model. Figure 3 depicts ACF plots across different accumulation windows (3,6,12,24 months). The ACF values across all windows gradually decrease with increasing lags and eventually becoming statistically insignificant after some point. Significant spikes in the initial lags indicate short-time dependencies in the data. Lower-order auto-regressive processes are suitable for modeling these gradual decaying patterns in the data. The PACF plots across all accumulation windows reflect a similar pattern (see Figure 4). We observe presence of significant spikes at lag 1 and spikes at subsequent lags dropping quickly. This pattern in the time series across all windows indicates that the Auto-regressive process of order 1 (AR(1)) is the best model.

To illustrate the practical application of ARSPI using real-world data, we utilize the rainfall dataset from Mckee et al. (1993) and apply the proposed index. Additionally, we compare the performance of ARSPI with two other competing models, as outlined below.

Model: ARSPI with π_t

The ARSPI model, as defined in Equation 4, is utilized in this analysis. We adopt a hierarchical Bayesian framework for the prior distributions, which are specified as follows:

$$
\begin{aligned}
\beta_1 &\sim N(0, \sigma_\beta^2), \\
\beta_2 &\sim \text{Uniform}(-1, 1), \\
\alpha &\sim N(0, 0.25), \\
\phi &\sim \text{Uniform}(-1, 1), \\
\sigma &\sim \text{Inverse-Gamma}(\nu/2, \nu/2) \\
\nu &\sim \text{Exponential}(0.1) \\
\sigma_\beta &\sim \text{Inverse-Gamma}(\nu_1, \nu_2), \\
\nu_1 &\sim \text{Exponential}(0.1), \\
\nu_2 &\sim \text{Exponential}(0.1).
\end{aligned}
$$

4.2. MCMC Results

We conducted an MCMC simulation with a total of 450,000 iterations on a dedicated 11th Gen Intel®Core^{TM} i5-1135G7 processor 8.00 GB RAM with 2667 MHz RAM speed. This involved running three MCMC chains, each consisting of 150,000 iterations, with a thinning interval of 10 and a burn-in period of 5,000, using Just Another Gibbs Sampler (JAGS-Parallel) within R.

Outcomes: The convergence of the Markov Chain Monte Carlo (MCMC) chains is verified through trace plots shown in Figure 5 to 8. The Deviance Information Criterion (DIC) for this model is presented in Table 2. Initially, the DIC decreases with increasing MTRs, but a significant increase is observed as the MTR rises from 12 to 24.

Parameter Estimates: Table 3 presents the Bayesian estimates and standard errors in terms of posterior mean and standard deviation of the related parameters for all the three models. It may be noted that the model precision (*i.e.* $\frac{1}{\sigma}$) is high for the ARSPI model indicating higher accuracy of the parameter estimates. Further, $| \beta_2 | \leq 1$ across all accumulation windows. High values of the β_2 values confirms the presence of auto-regression in MTR series. The probability of zero rainfall windows also remains stationary as the corresponding auto-regression coefficient ϕ is bounded in magnitude by unity. In other words, the probability of drought does not explode to one in the longer time horizon.

ARSPI Computation: The computation of ARSPI is carried out similarly to the conventional SPI. To achieve this, the cumulative distribution function (CDF) must be estimated at each observed MTR data point. This estimation is performed using the predictive density. Let $\tilde{\boldsymbol{\theta}}^{(i)}$ represent the ith draw from the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{r})$, where $i = 1, 2, \ldots, 45000$. A predictive sample of size 45000 is then generated at a specific time point t by drawing one sample from the density $f(r_t \mid \tilde{\boldsymbol{\theta}}^{(i)})$ for each i. The empirical CDF of this sample is then used to estimate the CDF at time t.

4.3. Comparative Analysis of SPI and ARSPI in the Colorado River Basin

This study presents the ARSPI as a flexible alternative to traditional SPI, leveraging Hierarchical Bayesian structure to better estimate meteorological drought conditions. The foundational SPI was developed by Mckee et al. (1993) using monthly precipitation data from the Colorado River Basin between January 1889 and December 1991.

For this analysis, we compare ARSPI with SPI using almost same historical precipitation dataset spanning 1188 months from January 1893 onward (PSL, 2024). Both indices are computed over 3-, 6-, 12-, and 24-month accumulation windows. Shorter time windows (3 and 6 months) are useful for assessing agricultural impacts, whereas longer windows (12 and 24 months) are more relevant for water resource management decisions.

The first plot in Figure 9 shows SPI-3 and ARSPI-3 values for the Colorado River Basin. Both indices exhibit considerable variability, with SPI-3 values ranging from -3.6447 to 2.9582, while ARSPI-3 spans a wider range of -3.9798 to 3.4658. The oscillatory nature of the time series reflects climatic fluctuations with frequent transitions between dry and wet periods. The broader range of ARSPI suggests its enhanced ability to capture precipitation anomalies compared to SPI.

Similarly, the second plot of Figure 9 illustrates SPI-6 and ARSPI-6. While SPI-6 ranges between -3.1774 and 2.9838, ARSPI-6 shows sharper peaks and troughs, with a range from -4.0752 to 3.8118, emphasizing extreme climatic conditions more effectively.

The third plot highlights SPI-12, which exhibits smoother oscillations over a range of -2.8892 to 2.4220, capturing long-term trends and annual precipitation anomalies. In contrast, ARSPI-12 demonstrates a higher amplitude, ranging from -3.8567 to 3.9109, making it more sensitive to short-term anomalies within annual cycles.

Finally, the fourth plot compares SPI-24 and ARSPI-24. SPI-24 ranges from -2.7586 to 2.2796, suitable for analyzing prolonged droughts. However, ARSPI-24 shows more pronounced peaks and troughs, with a range of -4.0752 to 3.9109, indicating greater sensitivity to short-term variations within the 24-month window.

Mismatch Points in SPI Comparison

Discrepancies between ARSPI and SPI are identified when the indices classify drought conditions differently. Two primary types of mismatches are defined as follows. *Type 1* mismatch occurs ARSPI indicates wet conditions ($ARSPI > 0$), while SPI suggests severe or extreme drought ($SPI < -1.50$). *Type 2* mismatch is defined when ARSPI classifies conditions as severe or extreme drought ($ARSPI < -1.50$), whereas SPI indicates wet conditions ($SPI > 0$).

In the present data set total 4.42% is the mismatch rate, of which 4.59% is of type 1 (4.65%,3.99%,5.14%,4.61% for 3,6,12,24 month accumulation windows) and 4.25% is of type 2 (4.23%,4.58%,4.54%,3.65% for 3,6,12,24 month accumulation windows). Figure 10 illustrates these mismatches for accumulation windows of 3, 6, 12, and 24 months. Type 1 mismatches are shown by the blue dashed line, and Type 2 by the red solid line.

3-Month Accumulation Window: Clusters of mismatches are observed in the early 1900s, 1940s, and 1960s. Type 2 events show higher rainfall variability, while Type 1 captures moderate to low rainfall values.

6-Month Accumulation Window: Mismatches occur irregularly, with notable clustering in the 1920s and 1960s. Type 2 events are more frequent and associated with

higher rainfall magnitudes.

12-Month Accumulation Window: The 12-month window shows a dominance of Type 2 mismatches, often linked to significant rainfall anomalies, particularly in the early and mid-20th century.

24-Month Accumulation Window: The 24-month window also demonstrates a dominance of Type 2 mismatches, often associated with significant rainfall anomalies, particularly in the early and mid-20th century.

5. Conclusion

In this study, we introduce the AutoRegressive Standardized Precipitation Index (AR-SPI), a novel drought index that incorporates the auto-correlated nature of precipitation series. By capturing evolving precipitation patterns under dynamic environmental conditions, ARSPI addresses limitations in traditional drought assessment methods.

The key contribution of this work is the construction of ARSPI, and a comparative analysis of ARSPI and the conventional SPI across the Colorado River Basin. Utilizing Bayesian techniques, ARSPI provides a more flexible and adaptive framework for quantifying precipitation anomalies. Our evaluation across multiple accumulation windows (3, 6, 12, and 24 months) demonstrates that ARSPI outperforms SPI in identifying extreme drought and wet conditions. The broader range and pronounced oscillations observed in ARSPI, characterized by sharper peaks and troughs, enhance its sensitivity to climatic extremes, offering a more refined depiction of drought variability.

Mismatch analyses between ARSPI and SPI reveal significant discrepancies in drought classification, particularly under extreme conditions. These differences highlight the nuanced response of each index to precipitation variability. The temporal clustering of mismatches underscores the heightened sensitivity of ARSPI to historical climatic patterns, consistently capturing extreme rainfall anomalies with greater accuracy.

ARSPI's flexibility, enhanced sensitivity to precipitation extremes, and ability to capture nuanced drought characteristics make it a valuable tool for hydrological and climatic studies. Future research should explore its application in other regions and evaluate its integration into decision-making frameworks for drought mitigation and water resource management.

References

Angelidis, P., Maris, F., Kotsovinos, N., Hrissanthou, V., 2012. Computation of drought index spi with alternative distribution functions. Water Resour Manage 26, 2453–2473. `https://doi.org/10.1007/s11269-012-0026-0`.

Bonacci, O., 1993. Hydrological identification of drought. Hydrological Processes 7, 249–262. `https://doi.org/10.1002/hyp.3360070303`.

Carlin, B.P., Louis, T.A., 2008. Bayesian Methods for Data Analysis, 3rd ed. CRC press, New York.

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. The American Statistician 49, 327–335. `https://doi.org/10.1080/00031305.1995.10476177`.

Das, J., Jha, S., Goyal, M.K., 2020. Non-stationary and copula-based approach to assess the drought characteristics encompassing climate indices over the himalayan states in india. Journal of Hydrology 580, 124356. `https://doi.org/10.1016/j.jhydrol.2019.124356`.

Delleur, J.W., Kavvas, M.L., 1978. Stochastic models for monthly rainfall forecasting and synthetic generation. J. Appl. Meteor. 17, 1528–1536. https://doi.org/10.1175/1520-0450(1978)017<1528:smfmrf>2.0.co;2.

Ganguli, P., Reddy, M.J., 2012. Risk assessment of droughts in gujarat using bivariate copulas. Water Resour Manage 26, 3301–3327. https://doi.org/10.1007/s11269-012-0073-6.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7. https://doi.org/10.1214/ss/1177011136.

Haktanir, T., Cobaner, M., Kisi, O., 2012. Frequency analyses of annual extreme rainfall series from 5 min to 24 h. Hydrological Processes 24, 3574–3588. https://doi.org/10.1002/hyp.7759.

Harris, C. , 2012. Map of colorado river basin. URL https://gcdamp.com/index.php/File:Colorado_River_Basin-_MAP-_CRBC-_Chris_Harris.jpg.

Kim, T.W., Valdés, J.B., Yoo, C., 2003. Nonparametric approach for estimating return periods of droughts in arid regions. J. Hydrol. Eng. 8, 237–246. https://doi.org/10.1061/(asce)1084-0699(2003)8:5(237).

Labadie, J.W., Lazaro, R.C., Morrow, D.M., 1981. Worth of short-term rainfall forecasting for combined sewer overflow control. Water Resources Research 17, 1489–1497. https://doi.org/10.1029/wr017i005p01489.

Lana, X., Burgueño, A., 2000. Some statistical characteristics of monthly and annual pluviometric irregularity for the spanish mediterranean coast. Theoretical and Applied Climatology 65, 79–97. https://doi.org/10.1007/s007040050006.

Lempers, F.B., 1971. Posterior probabilities of alternative linear models: Some theoretical considerations and empirical experiments. Universitaire Pers Rotterdam, Rotterdam.

Lloyd-Hughes, B., Saunders, M.A., 2002. A drought climatology for europe. Intl Journal of Climatology 22, 1571–1592. https://doi.org/10.1002/joc.846.

Mahmud, I., Bari, S.H., Rahman, M.T.U., 2017. Monthly rainfall forecast of bangladesh using autoregressive integrated moving average method. Environmental Engineering Research 22, 162–168. https://doi.org/10.4491/eer.2016.075.

Mckee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales, California. pp. 179–183.

Mishra, A.K., Singh, V.P., Desai, V.R., 2009. Drought characterization: a probabilistic approach. Stoch Environ Res Risk Assess 23, 41–55. https://doi.org/10.1007/s00477-007-0194-2.

Mishra, V., Aadhar, S., Asoka, A., Pai, S., Kumar, R., 2016. On the frequency of the 2015 monsoon season drought in the indo-gangetic plain. Geophysical Research Letters 43. https://doi.org/10.1002/2016gl071407.

Mitchell, T.J., Beauchamp, J.J., 1987. Bayesian variable selection in regression. Journal of the american statistical association 83, 1023–1032. https://doi.org/10.2172/6842309.

Moccia, B., Mineo, C., Ridolfi, E., Russo, F., Napolitano, F., 2022. Spi-based drought classification in italy: Influence of different probability distribution functions. Water 14, 3668. https://doi.org/10.3390/w14223668.

Mukhawana, M.B., Kanyerere, T., Kahler, D., Masilela, N.S., Lalumbe, L., Umunezero, A.A., 2024. Hydrological drought assessment using the standardized groundwater index and the standardized precipitation index in the berg river catchment, south africa. Journal of Hydrology: Regional Studies 53, 101779. `https://doi.org/10.1016/j.ejrh.2024.101779`.

Ntale, H.K., Gan, T.Y., 2003. Drought indices and their application to east africa. Intl Journal of Climatology 23, 1335–1357. `https://doi.org/10.1002/joc.931`.

Obeysekera, J.T.B., Tabios, G.Q., Salas, J.D., 1987. On parameter estimation of temporal rainfall models. Water Resources Research 23, 1837–1850. `https://doi.org/10.1029/wr023i010p01837`.

PSL, 2024. Boulder colorado monthly precipitation 1893-present: NOAA physical sciences laboratory. URL `https://psl.noaa.gov/boulder/Boulder.mm.precip.html`.

Saha, S., Chakraborty, D., Paul, R.K., Samanta, S., Singh, S.B., 2018. Disparity in rainfall trend and patterns among different regions: analysis of 158 years' time series of rainfall dataset across india. Theor Appl Climatol 134, 381–395. `https://doi.org/10.1007/s00704-017-2280-9`.

Sankarasubramanian, A., Srinivasan, K., 1999. Investigation and comparison of sampling properties of l-moments and conventional moments. Journal of Hydrology 218, 13–34. `https://doi.org/10.1016/s0022-1694(99)00018-9`.

Shiau, J., Feng, S., Nadarajah, S., 2007. Assessment of hydrological droughts for the yellow river, china, using copulas. Hydrological Processes 21, 2157–2163. `https://doi.org/10.1002/hyp.6400`.

Shoji, T., Kitaura, H., 2006. Statistical and geostatistical analysis of rainfall in central japan. Computers and Geosciences 32, 1007–1024. `https://doi.org/10.1016/j.cageo.2004.12.012`.

Smith, T., Marshall, L., Sharma, A., 2015. Modeling residual hydrologic errors with bayesian inference. Journal of Hydrology 528, 29–37. `https://doi.org/10.1016/j.jhydrol.2015.05.051`.

Soltani, S., Modarres, R., Eslamian, S.S., 2007. The use of time series modeling for the determination of rainfall climates of iran. Intl Journal of Climatology 27, 819–829. `https://doi.org/10.1002/joc.1427`.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit 64, 583–639. `https://doi.org/10.1111/1467-9868.00353`.

Stagge, J.H., Tallaksen, L.M., Gudmundsson, L., Van Loon, A.F., Stahl, K., 2015. Candidate distributions for climatological drought indices (spi and spei). Intl Journal of Climatology 35, 4027–4040. `https://doi.org/10.1002/joc.4267`.

Tang, Y., Marshall, L., Sharma, A., Smith, T., 2016. Tools for investigating the prior distribution in bayesian hydrology. Journal of Hydrology 538, 551–562. `https://doi.org/10.1016/j.jhydrol.2016.04.032`.

Tsakiris, G., Pangalou, D., Vangelis, H., 2007. Regional drought assessment based on the reconnaissance drought index (rdi). Water Resour Manage 21, 821–833. `https://doi.org/10.1007/s11269-006-9105-4`.

Ullah, I., Ma, X., Yin, J., Omer, A., Habtemicheal, B.A., Saleem, F., Iyakaremye, V., Syed, S., Arshad, M., Liu, M., 2023. Spatiotemporal characteristics of meteorological drought variability and trends (1981–2020) over south asia and the associated large-scale circulation patterns. Clim Dyn 60, 2261–2284. `https://doi.org/10.1007/s00382-022-06443-6`.

Wilhite, D.A., Svoboda, M.D., Hayes, M.J., 2007. Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. Water Resour Manage 21, 763–774. `https://doi.org/10.1007/s11269-006-9076-5`.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed. Elsevier.

World Meteorological Organization: Geneva, Switzerland, 2006. "Concepts, progress and future challenges, wmo-no. 1006".

Yevjevich, V.M., 1969. An objective approach to definitions and investigations of continental hydrologic droughts. Journal of Hydrology 7, 353. `https://doi.org/10.1016/0022-1694(69)90110-3`.

Yihdego, Y., Webb, J., Vaheddoost, B., 2017. Highlighting the role of groundwater in lake– aquifer interaction to reduce vulnerability and enhance resilience to climate change. Hydrology 4, 10. `https://doi.org/10.3390/hydrology4010010`.

Yildirim, G., Rahman, A., 2022. Spatiotemporal meteorological drought assessment: a case study in south-east australia. Nat Hazards 111, 305–332. `https://doi.org/10.1007/s11069-021-05055-3`.

YU, F.X., NAGHAVI, B., 1994. Estimating distribution parameters using optimization techniques. Hydrological Sciences Journal 39, 391–403. `https://doi.org/10.1080/02626669409492757`.

Zhao, C., Brissette, F., Chen, J., Martel, J.L., 2020. Frequency change of future extreme summer meteorological and hydrological droughts over north america. Journal of Hydrology 584, 124316. `https://doi.org/10.1016/j.jhydrol.2019.124316`.

Appendix: Tables and Figures

Tables

Table 1: Drought Classification using ARSPI

ARSPI	Category	Probability
ARSPI $\geq$ 2.00	Extreme Wet	0.023
1.50 < ARSPI < 1.99	Severe Wet	0.044
1.00 < ARSPI < 1.49	Moderate Wet	0.092
0.00 < ARSPI < 0.99	Mild Wet	0.341
−0.99 < ARSPI < 0.00	Mild Drought	0.341
−1.49 < ARSPI < −1.00	Moderate Drought	0.092
−1.99 < ARSPI < −1.50	Severe Drought	0.044
ARSPI $\leq$ −2.00	Extreme Drought	0.023

Table 2: DIC values for the ARSPI model

Time	*ARSPI Model*
3 MTR	4632.639
6 MTR	4526.911
12 MTR	3726.612
24 MTR	3636.744

Table 3: Parameter estimates for different Models.

Model	Parameters	3 MTR	6 MTR	12 MTR	24 MTR
ARSPI Model	β_1	0.3756(0.0299)	0.2465(0.0291)	0.1388(0.0245)	0.0885(0.0205)
	β_2	0.7220(0.0200)	0.8854(0.0133)	0.9523(0.0084)	0.9755(0.0057)
	σ	0.4411(0.0091)	0.1916(0.0039)	0.0650(0.0013)	0.0313(0.0006)
	α	-6.8149(0.7851)	-6.8025(0.7835)	-6.8023(0.7827)	-6.7914(0.7847)
	ϕ	0.0026(0.5724)	0.0076(0.5716)	0.0012(0.5707)	0.0012(0.5698)

Figures

Figure 1: Colorado River Basin.

Colorado Rainfall data (1893–1991)
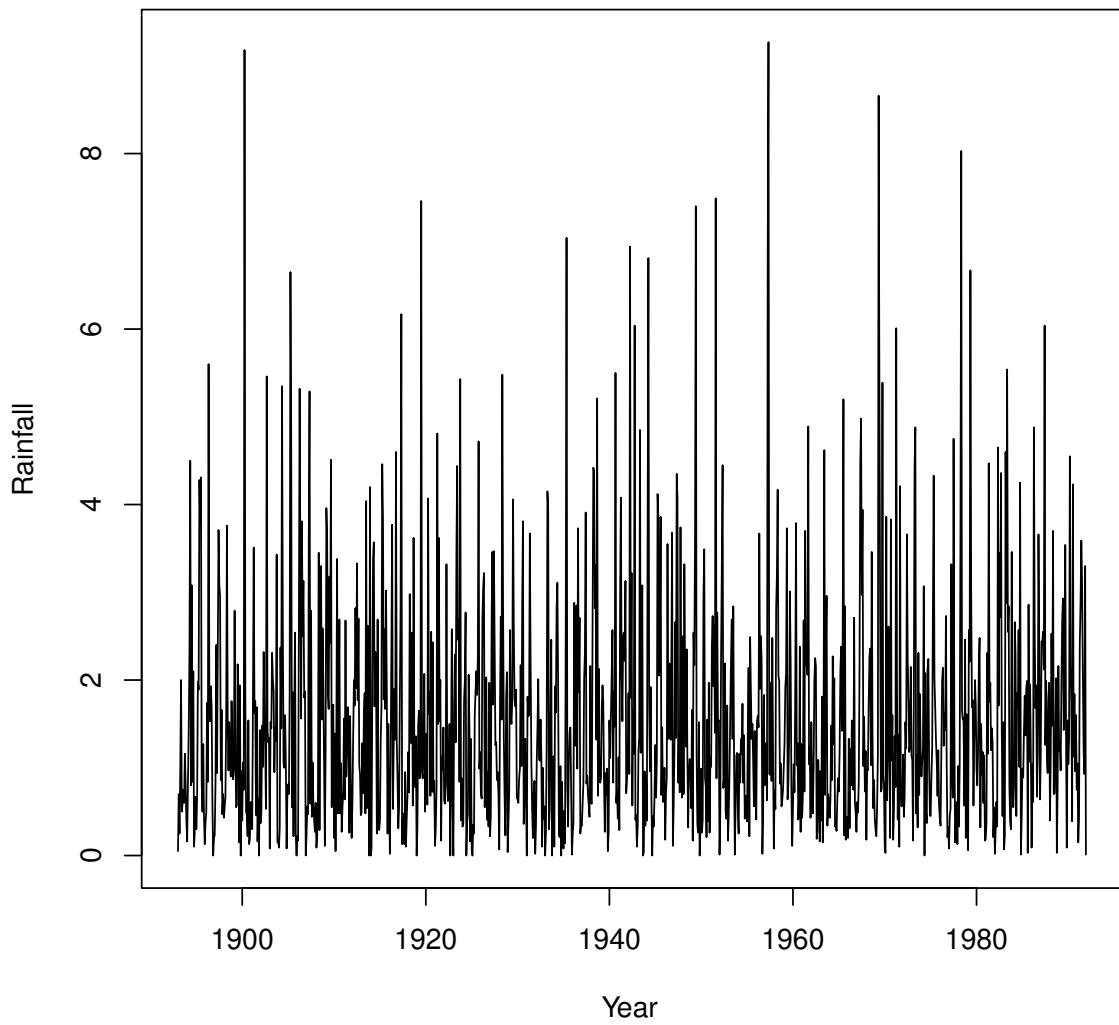
Figure 2: Rainfall of Colorado River Basin.
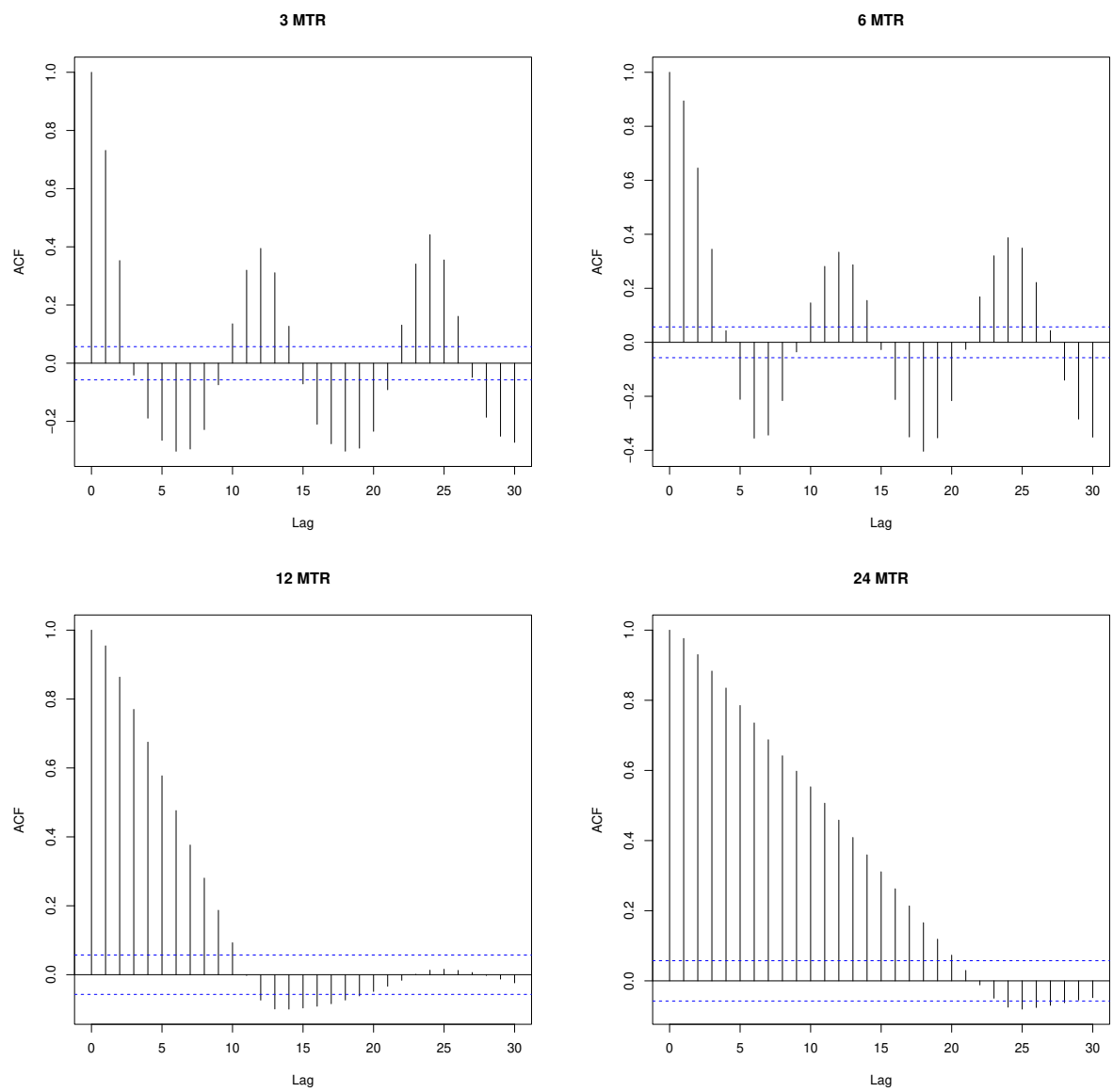
Figure 3: ACF plots of across different accumulation windows.
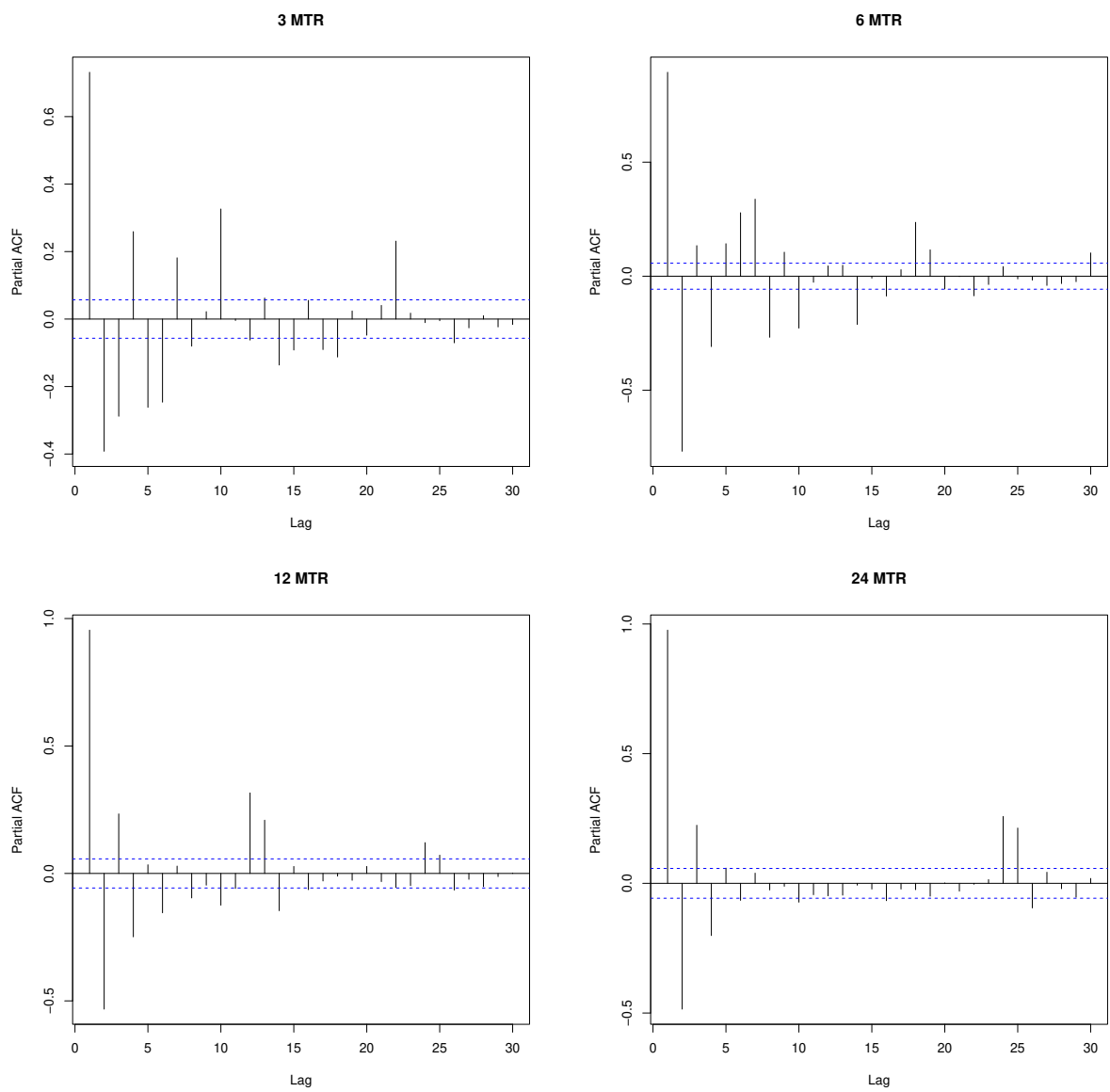
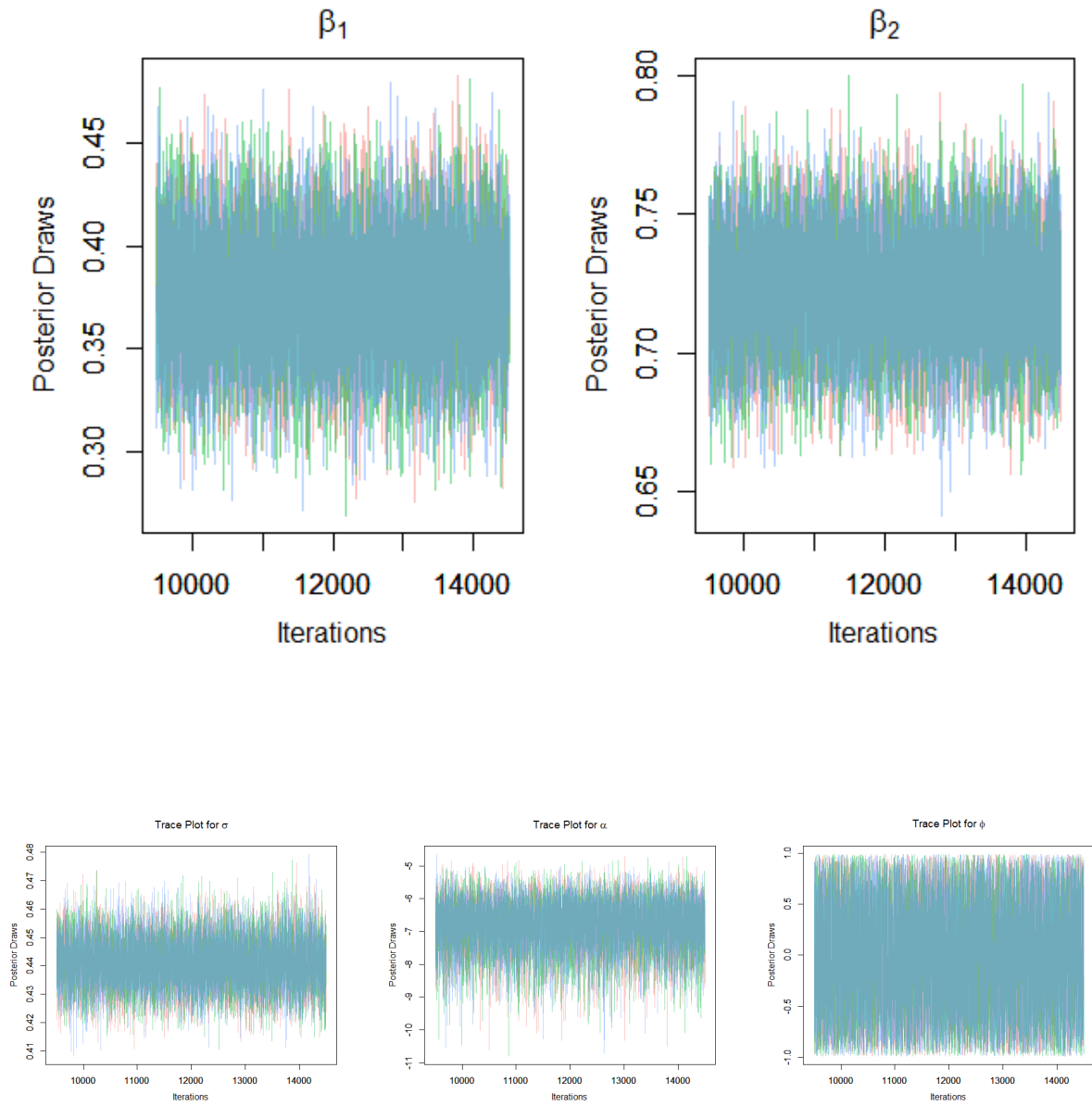Figure 4: Partial ACF plots across different accumulation windows.

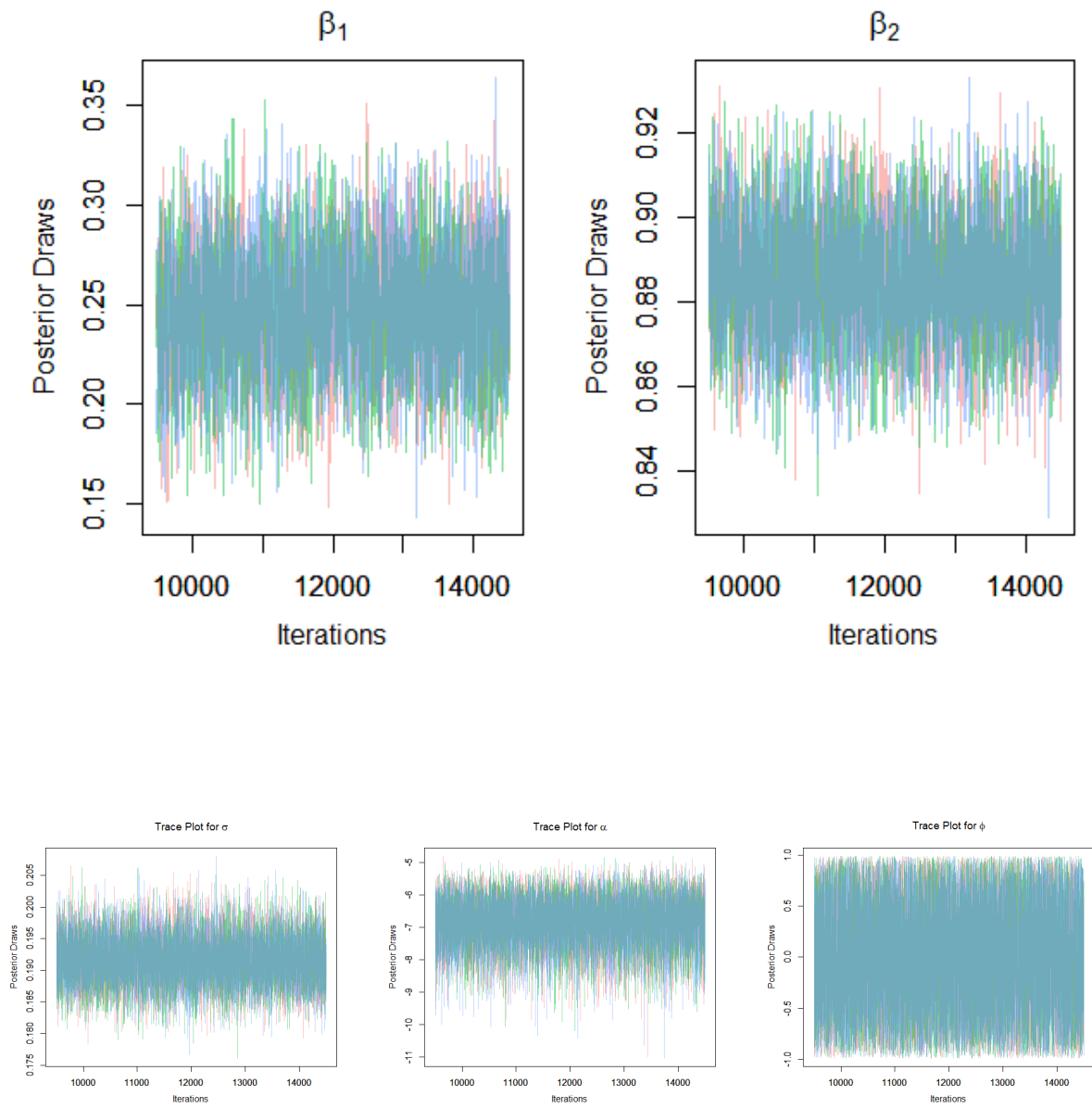Figure 5: Trace plots for parameters for *Model III* (3 month accumulation window).

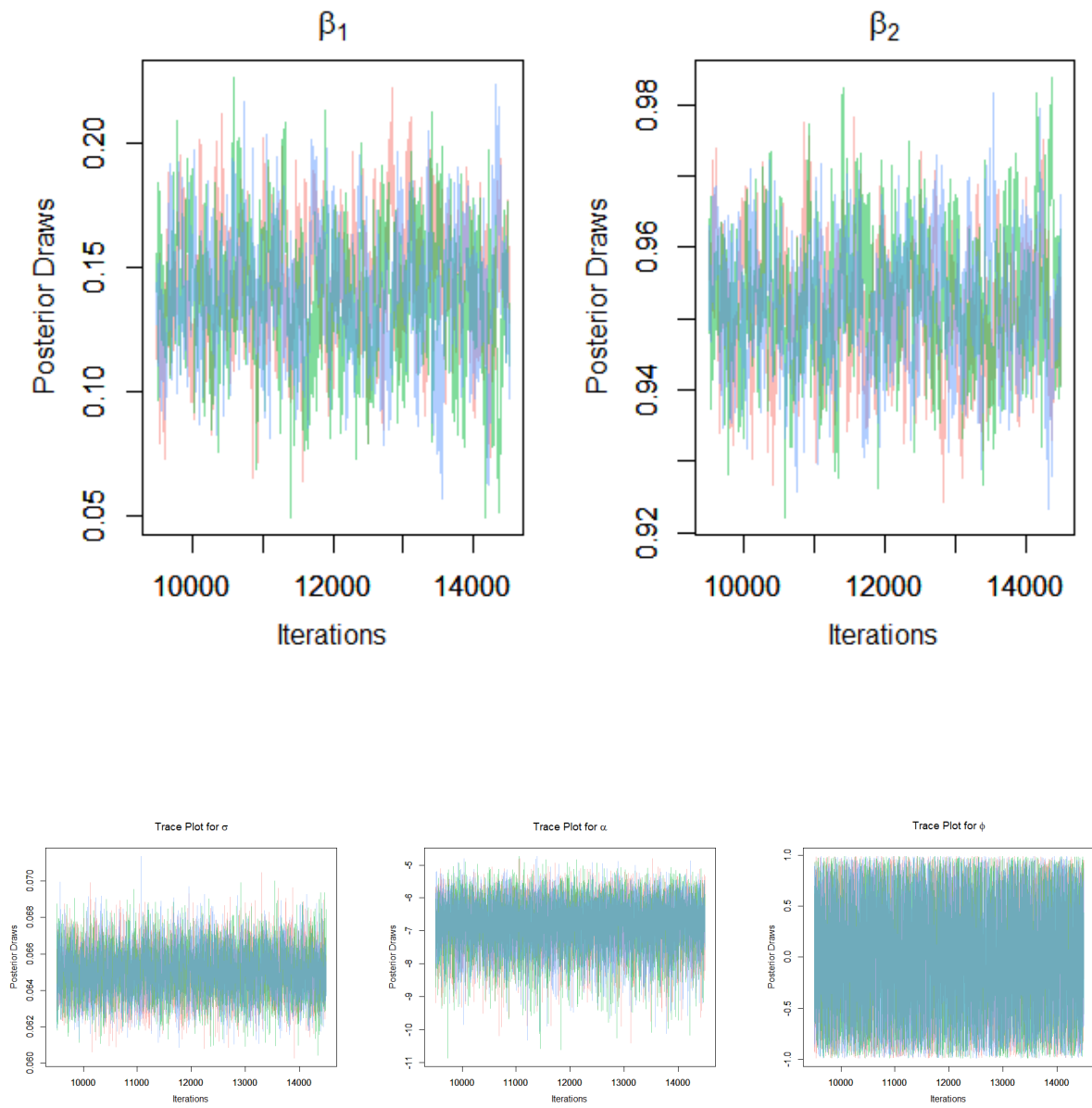Figure 6: Trace plots for parameters for *Model III* (6 month accumulation window).

Figure 7: Trace plots for parameters for *Model III* (12 month accumulation window).
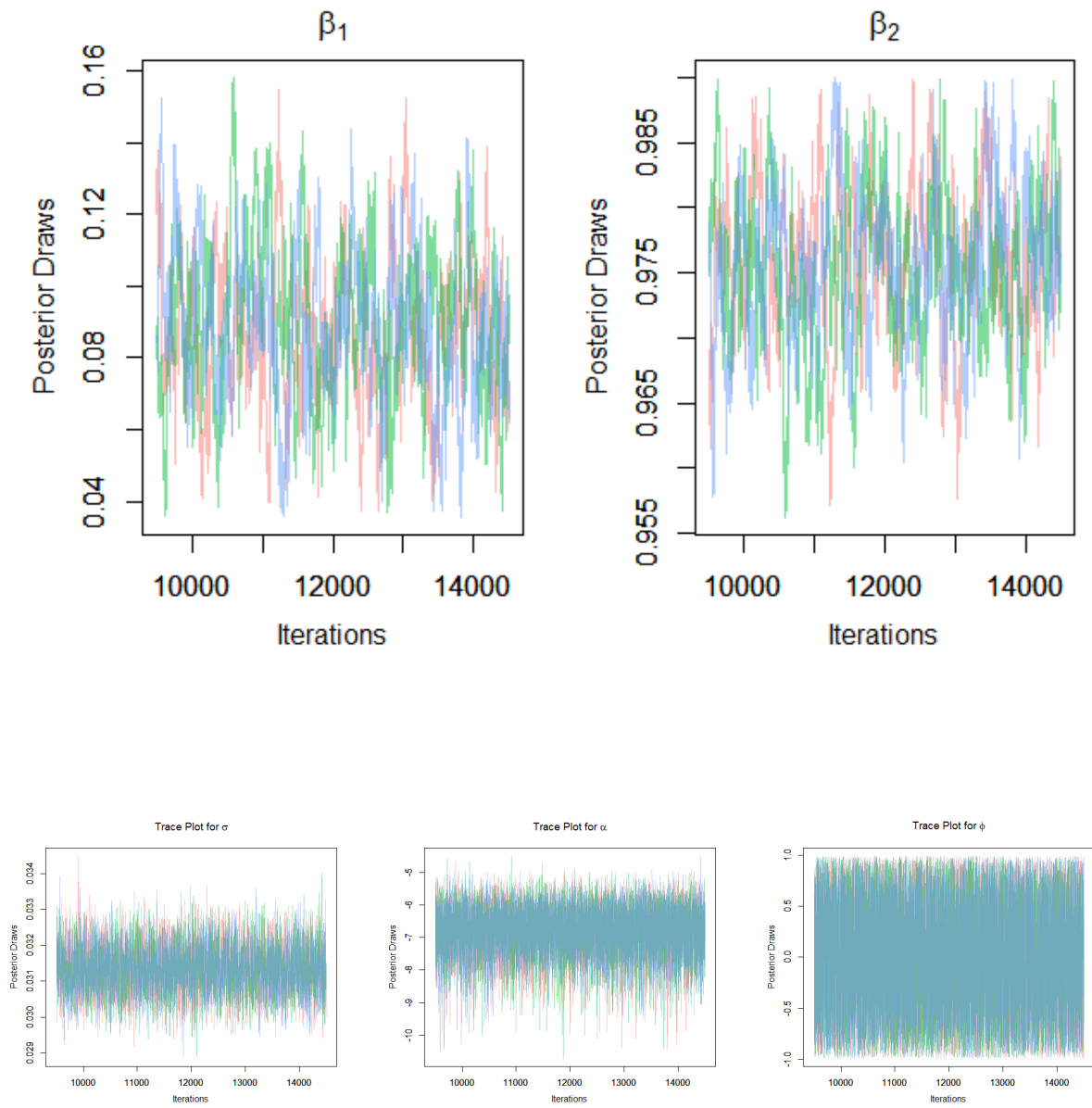
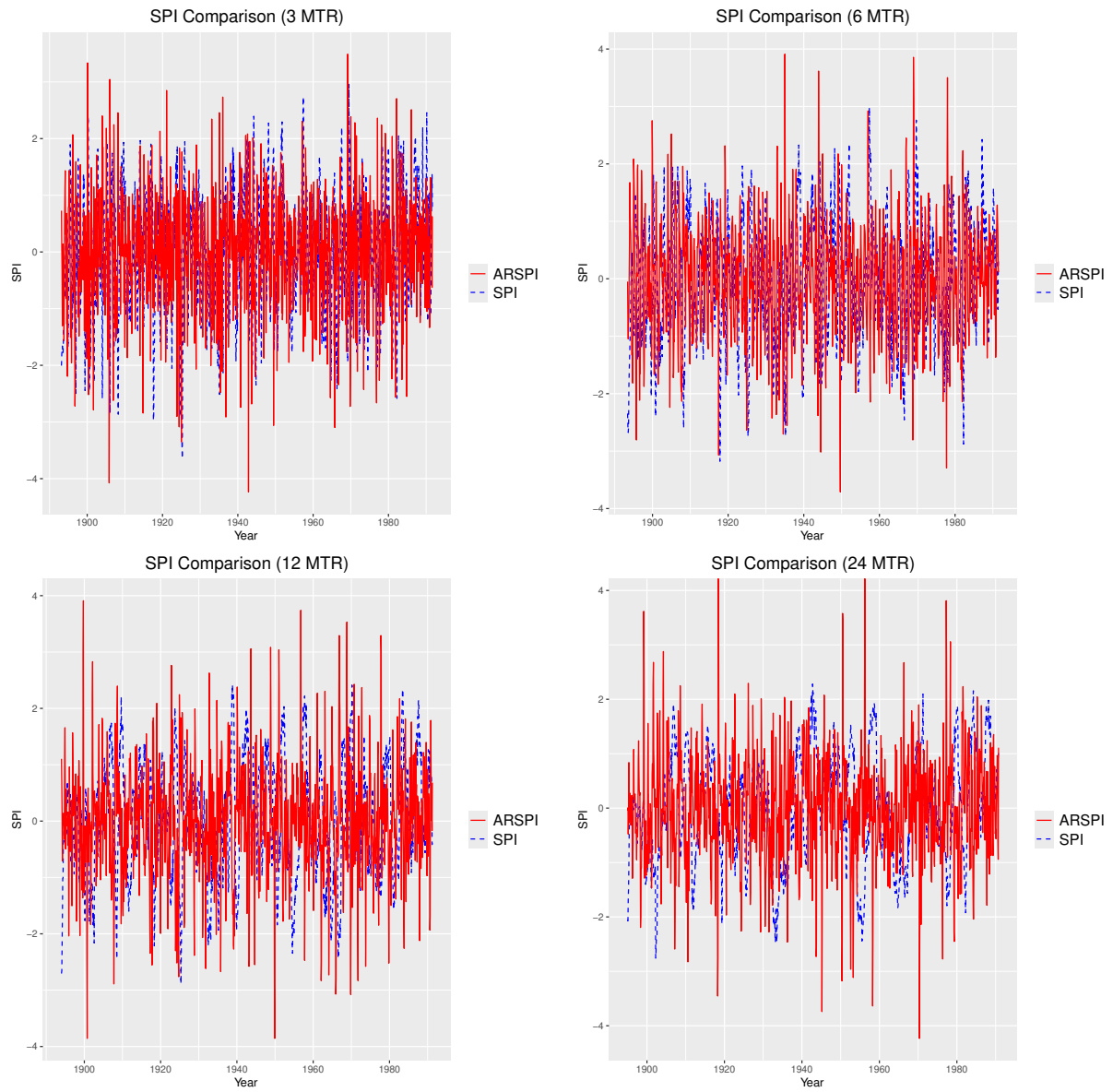Figure 8: Trace plots for parameters for *Model III* (24 month accumulation window).

Figure 9: Comparison of SPI and ARSPI for different accumulation windows (3,6,12,24 months).
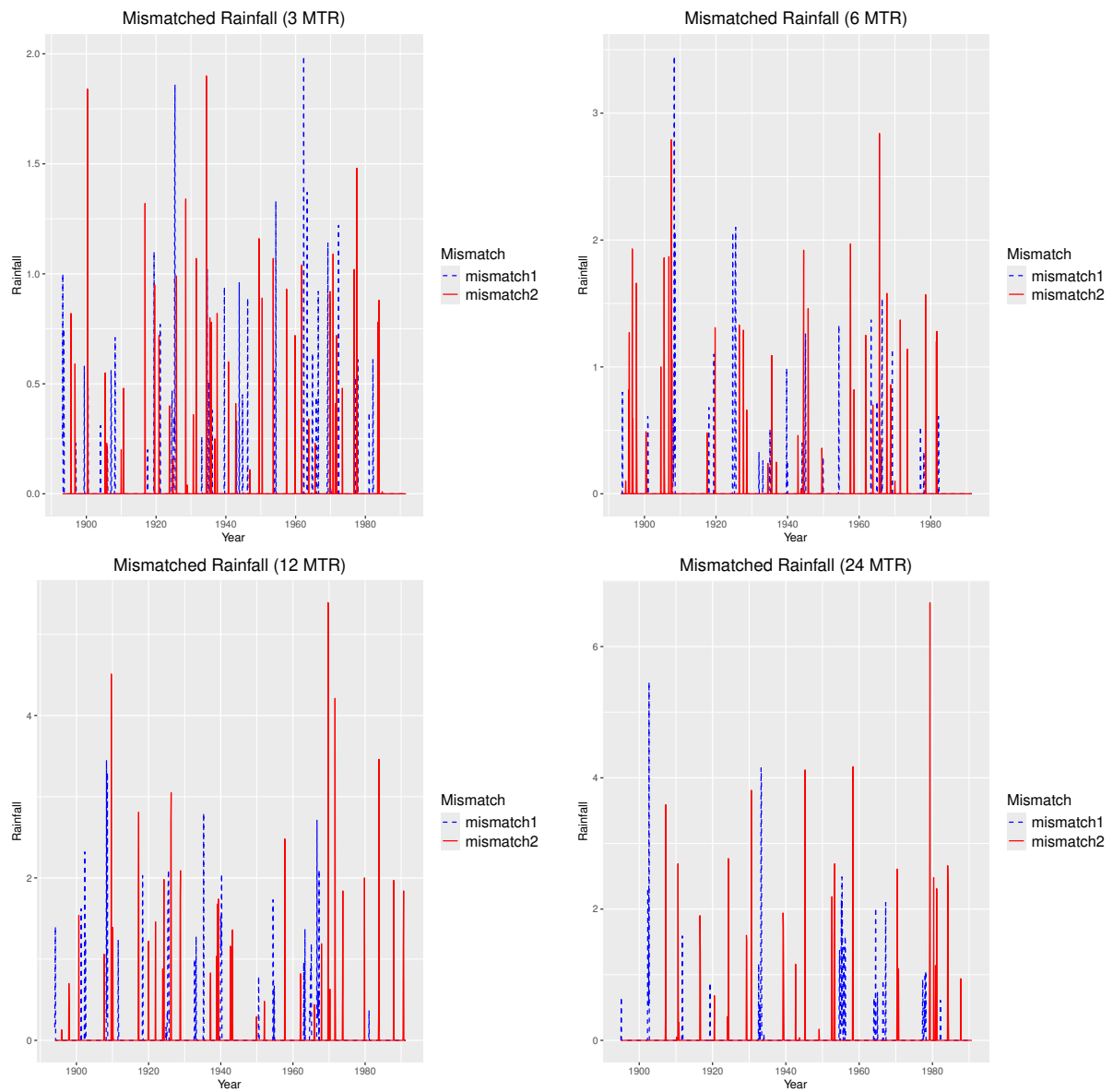
Figure 10: Plots representing the mismatched rainfall for different accumulation windows (3,6,12,24 months).