A multisite Stochastic Watershed Model (SWM) with intermittency for regional low flow and flood risk analysis

Zach Brodeur[1], Rohini Gupta[2], Scott Steinschneider[3]
Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY

1. Postdoctoral Associate, 111 Wing Drive, Riley-Robb Hall, Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, 14853. Email: zpb4@cornell.edu, Phone: 607-255-2155 (Corresponding Author).

2. Graduate Research Assistant, 111 Wing Drive, Riley-Robb Hall, Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, 14853. Email: rg727@cornell.edu, Phone: 607-255-2155.

3. Assistant Professor, 111 Wing Drive, Riley-Robb Hall, Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, 14853. Email: ss3378@cornell.edu, Phone: 607-255-2155.

Key Points:

- We develop a Stochastic Watershed Model (SWM) that simulates multisite streamflow ensembles and captures spatial patterns in model error
- The SWM also reproduces multisite and Markovian properties of flow intermittency
- We show that capturing multisite error properties and intermittency is critical for reproducing regional high and low flow design events

Abstract

Stochastic Watershed Models (SWMs) are an important innovation in hydrologic modeling that propagate uncertainty into model predictions by adding samples of model error to deterministic simulations. A growing body of work shows that univariate SWMs effectively reduce bias in hydrologic simulations, especially at the upper and lower flow quantiles. This has important implications for short term forecasting and the estimation of design events for long term planning. However, the application of SWMs in a regional context across many sites is underexplored. Streamflow across nearby sites is highly correlated, and so too are hydrologic model errors. Further, in arid and semi-arid regions streamflow can be intermittent, but SWMs rarely model zero flows at one site, let alone correlated intermittency across sites. In this technical note, we contribute a multisite SWM that captures univariate attributes of model error (heteroscedasticity, autocorrelation, non-normality, conditional bias), as well as multisite attributes of model error (cross-correlated error magnitude and persistence). The SWM also incorporates a multisite, auto-logistic regression model to account for multisite persistence in streamflow intermittency. The model is applied and tested in a case study that spans 14 watersheds in the Sacramento, San Joaquin, and Tulare basins in California. We find that the multisite SWM is able to better reproduce regional low and high flow events and design statistics as compared to a single-site SWM applied independently to all locations.

1. Introduction

Stochastic watershed models (SWM) are a recent innovation in hydrologic prediction that enable the generation of streamflow ensembles for water resources planning and management (Shabestanipour et al., 2023; Vogel, 2017). SWMs build from stochastic streamflow models (SSM) (Maass et al., 1962; Teegavarapu et al., 2019; Vogel, 2017), which are statistical models fit directly to observed streamflow and enable ensemble simulation of synthetic streamflow traces with plausible extremes that extend beyond the historical record. SSMs work well under an assumption of stationarity but are challenging to implement without that assumption (Vogel, 2017). In contrast, SWMs produce simulations using the output from a deterministic watershed model (DWM) coupled with simulations of DWM error drawn from the model's predictive uncertainty distribution (i.e., the distribution of errors between the DWM simulation and the observations). The stochasticity of SWMs is critical to ensure that hydrologic model simulations are unbiased around high and low extreme events (Farmer & Vogel, 2016), which is important for both short-term prediction (e.g., flood forecasting; Troin et al., 2021; Vannitsem, 2018; Zha et al., 2020) and long-term planning (e.g., design event estimation; Shabestanipour et al., 2023). Moreover, the incorporation of process-oriented predictions from the DWM allows for non-stationary simulations that can capture the hydrologic response to climate or land use change (Steinschneider et al., 2015).

This technical note focuses on the development of a multisite SWM, which to date has been understudied but is needed to capture joint hydrologic risks. Joint behaviors in streamflow extremes across locations can create spatially compounding events (Zscheischler et al., 2020) that produce far greater risks than those events considered at individual locations (Serinaldi &

Kilsby, 2018; Simpson et al., 2021; Zscheischler, 2020). Joint hydrologic risks extend both to regional floods and to extreme low flow events, the latter which threaten human and environmental water needs (Hanak, 2011; Loucks & Van Beek, 2017).

Over the last decade, copulas have been widely employed to capture joint behaviors in the observational data directly, allowing both risk quantification and stochastic simulation for joint hydrologic risk assessments (Chen et al., 2015; Chen & Guo, 2018; Favre et al., 2004; Teegavarapu et al., 2019). However, capturing joint risk in multisite DWM simulations is a subtly different problem. This endeavor requires accounting for the multisite dependencies in DWM predictive errors, not the observations directly. These dependencies result when attributes of DWM errors at individual sites (e.g., underprediction bias, autocorrelation; Vogel, 2017) are correlated in space and time. DWM predictive errors are difficult to model, as they exhibit heteroscedasticity, non-normality, autocorrelation, and conditional bias, especially when the model operates on short (e.g., daily, hourly) timescales (McInerney et al., 2017; Schoups & Vrugt, 2010; Vogel, 2017). Intermittency in the observed streamflow data further complicates SWM development (Ye et al., 2021). The most commonly employed deterministic hydrologic models use an exponential decay to simulate baseflow, making them incapable of producing zero flows (Shabestanipour et al., 2023). For a SWM to capture intermittency, simulated errors must produce periods of zero flow that occur across sites with the correct spatial correlation and persistence.

A number of recent studies have explored different SWM approaches (Farmer & Vogel, 2016; Hah et al., 2022; Koutsoyiannis & Montanari, 2022; McInerney et al., 2017; Sikorska et al., 2015; Vogel, 2017; Shabestanipour et al., 2023), but they all have targeted SWM simulations at individual sites and without consideration of streamflow intermittency. In contrast, much recent work has been devoted to multisite, intermittent SSMs (Efstratiadis et al., 2014; Haktanir et al., 2022; Saad et al., 2015; Papalexiou, 2018; Papalexiou & Serinaldi, 2020; Tsoukalas et al., 2019, 2020). Many of these studies use 'Nataf-based' approaches that rely on a framework of copulas, multivariate autoregressive models, and flexible distributional forms to capture a wide range of spatiotemporally correlated stochastic behavior. Intermittency is accounted through the use of censored distributions (Papalexiou, 2018; Papalexiou & Serinaldi, 2020; Wang & Robertson, 2011), mixture distributions (Ye et al., 2021), or other techniques (e.g., truncation, Markovian models, and non-parametric methods; Efstratiadis et al., 2014; Nowak et al., 2010).

This technical note contributes for the first time a multisite, intermittent SWM by adapting these recent advances in SSM to the case of process-oriented hydrologic model error simulation. We develop and test the multisite, intermittent SWM in a case study of 14 watersheds in California across the Sacramento, San Joaquin, and Tulare basins that feed the agriculturally and ecologically important Central Valley and San Francisco Bay-Delta. We utilize the adaptable framework developed in previous work (Brodeur & Steinschneider, 2021) and related stochastic simulation studies (Efstratiadis et al., 2014; Papalexiou, 2018; Papalexiou & Serinaldi, 2020) to account for complex properties of DWM predictive errors across watersheds, and introduce an auto-logistic model to account for multisite intermittency. We compare the proposed model against a univariate SWM benchmark to evaluate the importance of spatiotemporally correlated errors and intermittency for the simulation and estimation of joint high and low flow events relevant to water resources planning.

2. Data

This study spans 14 watersheds in California that make up the Sacramento, San Joaquin, and Tulare basins and collectively drain into the San Francisco Bay at the Sacramento-San Joaquin Delta (Figure 1a). These watersheds range from mostly perennial, snowmelt dominated catchments in the north to smaller, rain-fed catchments with high intermittency in the south. All 14 watersheds exhibit some degree of intermittency, which is not uncommon in U.S. watersheds (Levick et al., 2008; Ye et al., 2021). Observed flows between water years (WY) 1988-2013 were collected for each watershed from the full natural flow dataset from the California Data Exchange Center (CDEC). The Sacramento Soil Moisture Accounting (SAC-SMA) model (i.e., the DWM) was calibrated to each watershed, as detailed in Wi & Steinschneider (2022). The SAC-SMA model was forced with $1/16^{o}$ meteorological data (Livneh et al. 2013) and calibrated by maximizing the Nash-Sutcliffe Efficiency (NSE) using a genetic algorithm. We direct readers to Wi & Steinschneider (2022) for further details on the hydrologic model setup, calibration, and validation. Multisite correlations are prevalent in the DWM errors (Figure 1b), motivating the need for additional treatment of multisite correlation in a SWM.

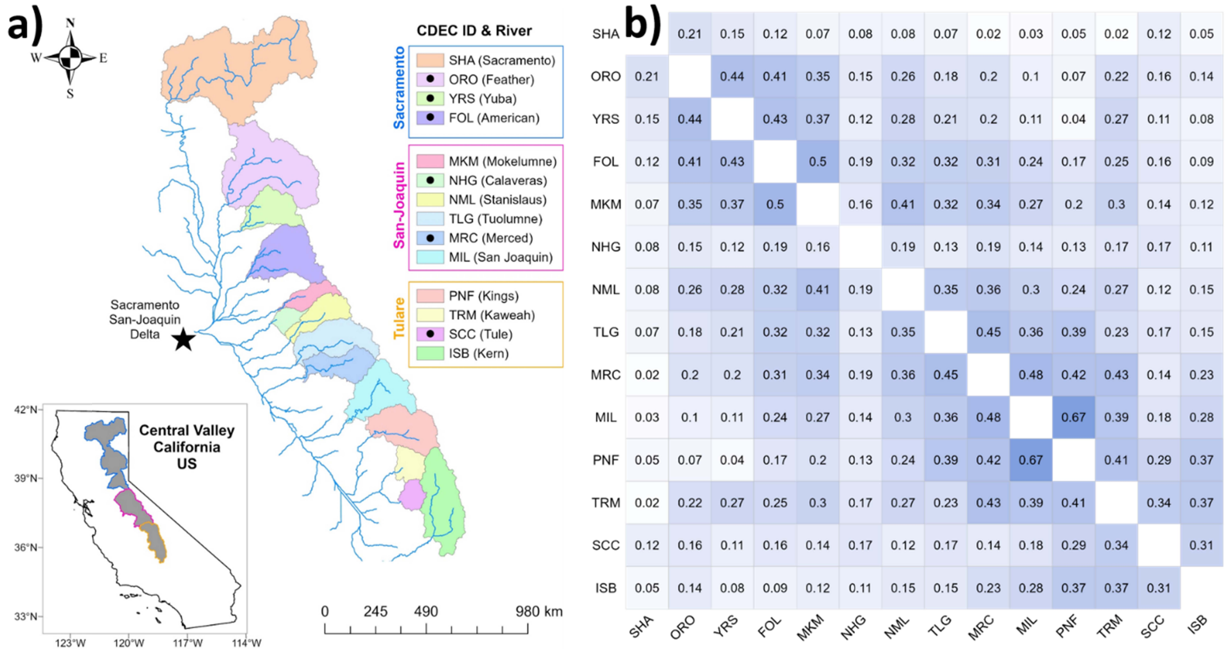

Figure 1. *(a) Map of 14 watersheds modeled in this work. Watersheds marked with a black dot are evaluated in detail in Figures 3 & 4. (b) Spearman correlations between errors of the SAC-SMA model across sites (error correlations shown after conditional debiasing; see Section 3.1).*

3. Methods

We construct our multisite SWM utilizing an aggregated approach first forwarded in Montanari & Brath (2004):

$$Q_{t,k} = F(X_{t,k}, \pi_k) + e_{t,k} \qquad \text{(Eq. 1)}$$

$F(X_{t,k}, \pi_k)$ is the process based DWM simulation of the observed flows $Q_{t,k}$ for site k based on inputs $X_{t,k}$ and parameters π_k. In the SWM, $F(X_{t,k}, \pi_k)$ is first calibrated to $Q_{t,k}$, yielding errors $e_{t,k}$ that are assumed to aggregate all uncertainties (Montanari & Koutsoyiannis, 2012). The proposed SWM is based on a multivariate stochastic model for these errors (see Figure 2), which we summarize here and then describe in more detail below.

The multivariate error model (Figure 2a; described in Section 3.1) is adapted from Brodeur & Steinschneider (2021). This model captures the cross-correlation and autocorrelation of $e_{t,k}$ while faithfully preserving distributional attributes like heteroscedasticity and conditional bias. After simulation of new synthetic errors via this model, we post-process streamflow simulations to incorporate intermittency with a novel, multisite auto-logistic model, which we compare to a simpler truncation approach (Figure 2b; described in Section 3.2).

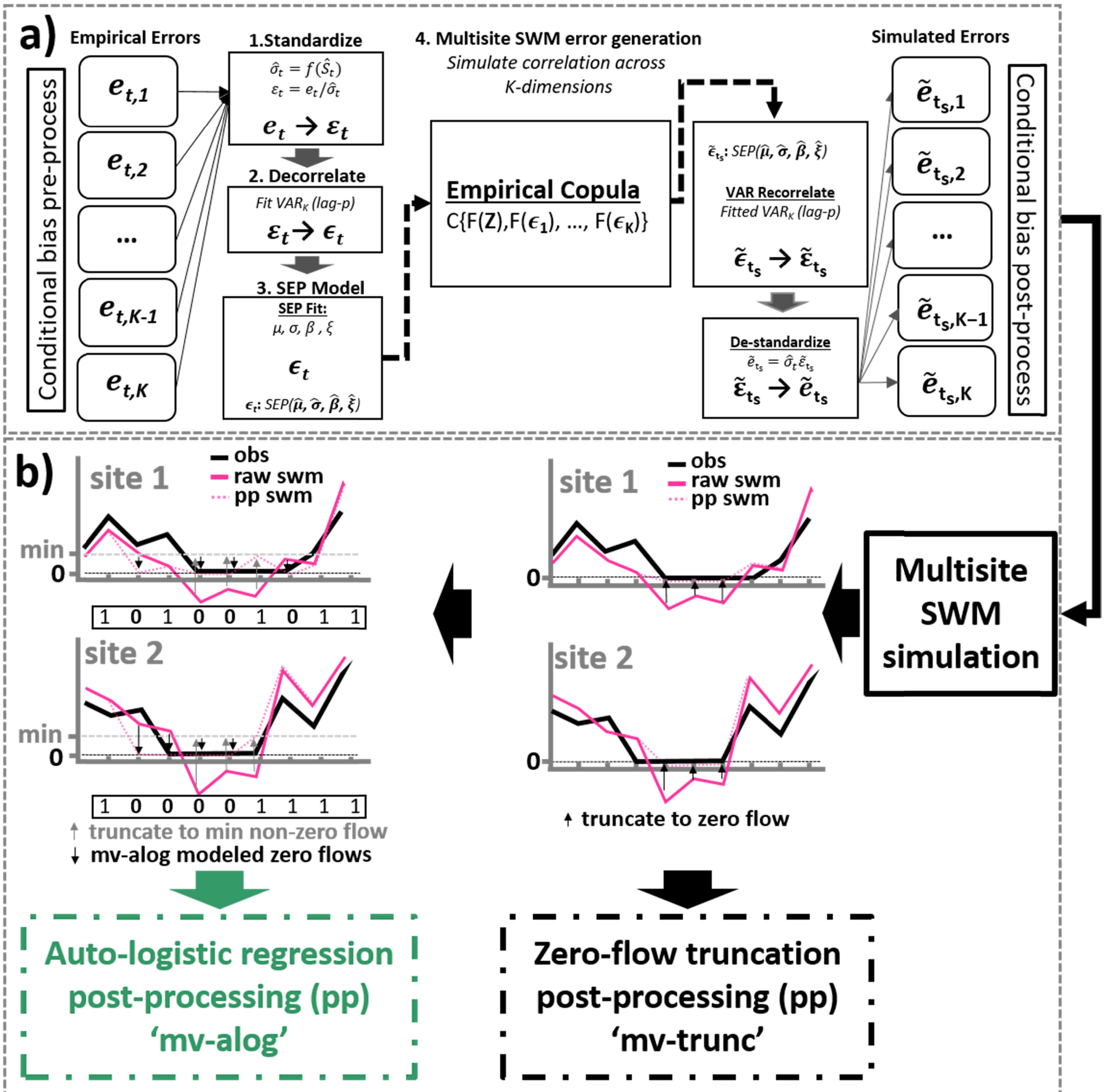

Figure 2. *Conceptual diagram describing: a) the multivariate error model, and b) the two intermittency post-processing strategies.*

3.1. Multisite SWM Error Model

The first step in the multisite error model is to remove conditional bias in the empirical DWM errors. DWM biases are often conditional, in that they change depending on the prevailing hydrologic regime, e.g., DWM estimates that consistently overestimate low flows and underestimate high flows (Farmer & Vogel, 2016). Conditional bias can lead to unstable estimation of the autoregressive models employed later in the modeling process, and so need to be removed beforehand. We estimate conditional bias by fitting a locally weighted polynomial regression (LOESS) for each site between the raw DWM simulations $\left(S_{t,k} = F\left(X_{t,k}, \pi_k\right)\right)$ and the observations $(Q_{t,k})$. Application of this LOESS model yields a conditionally debiased DWM estimate of $Q_{t,k}$, which we refer to as $\hat{S}_{t,k}$. To reduce edge effects in the sparse upper tail of the data, we linearly extrapolate $\hat{S}_{t,k}$ from the monotonic portion of the LOESS model into the upper tail. Replacing $F\left(X_{t,k}, \pi_k\right)$ in Eq. 1 with $\hat{S}_{t,k}$ leaves $e_{t,k}$ as the conditionally debiased errors.

We then account for heteroskedasticity in $e_{t,k}$ by fitting a model between $\hat{S}_{t,k}$ and $|e_{t,k}|$, where $|e_{t,k}|$ serves as a proxy for the standard deviation of the errors at time *t* and site *k*. We again use a LOESS model to estimate the conditional standard deviation $\hat{\sigma}_{t,k}$, which is then used to estimate standardized errors $\left(\varepsilon_{t,k}\right)$ for each site:

$$\varepsilon_{t,k} = \frac{e_{t,k}}{\hat{\sigma}_{t,k}} \qquad \text{(Eq. 2)}$$

We then fit a vector autoregressive (VAR) model to the multisite vector of standardized errors $(\boldsymbol{\varepsilon_t})$ using a robust, multivariate least trimmed squares estimator (Croux & Joossens, 2008; Galanos, 2022). We used a lag-3 VAR model in line with our previous work (Brodeur & Steinschneider, 2021). Application of the VAR model yields a vector of decorrelated and standardized residuals $(\boldsymbol{\epsilon_t})$. We then model these residuals at each site with the skew exponential power distribution (SEP; Schoups & Vrugt, 2010), alternately called the skew generalized error distribution (SGED; Wuertz et al., 2022), which is well suited for non-Gaussian, fat tailed, and skewed distributions.

Samples of $\epsilon_{t,k}$ (denoted $\tilde{\epsilon}_{t,k}$) from the fitted SEP distributions form the basis for the synthetic generation of new model errors. However, even after the VAR model, the residual vector $\boldsymbol{\epsilon_t}$ may still exhibit multisite correlations, and so independent, site-by-site samples of $\tilde{\epsilon}_{t,k}$ from the SEP distribution may lose important multisite patterns of correlation. To address this issue, we employ the empirical copula and kNN sampling procedure developed in Brodeur & Steinschneider (2021). In short, the approach randomly generates new sequences of residuals ($\tilde{\epsilon}_{t,k}$) by sampling from the SEP distribution for each site, and then reorders the samples via the Schaake Shuffle (Clark et al. 2004) to emulate the rank correlation structure of the empirical residuals. For each time step in the simulation, kNN sampling is then used to sample a vector of sampled residuals ($\tilde{\boldsymbol{\epsilon}}_k$) conditional on the bias-corrected DWM simulations at that time ($\widehat{\boldsymbol{S}}_t$), which ensures that any correlation between residuals and DWM simulations is preserved.

As depicted in Figure 2, after the generation of a new residual vector, the remainder of the steps are inverted to produce stochastic hydrologic model ensembles. First, multisite autocorrelation is reintroduced via sequential VAR simulation to produce $\tilde{\varepsilon}_{t,k}$. Then, the heteroscedasticity is reintroduced via inversion of Eq. 2 to produce $\tilde{e}_{t,k}$. Finally, the conditional bias is reintroduced by adding the resultant errors to the DWM conditional bias estimator $\hat{S}_{t,k}$, producing SWM simulations $\tilde{Q}_{t,k}$:

$$\tilde{Q}_{t,k} = \hat{S}_{t,k} + \tilde{e}_{t,k} \qquad \text{(Eq. 3)}$$

We note that the multivariate error model described above is fit separately for each month, since the properties of DWM errors can vary depending on prevailing hydrologic regimes (e.g., snow vs. rain dominated runoff response) that vary across the year.

3.2. Multisite Intermittency

After generating a SWM simulation, we post-process the data to simulate streamflow intermittency. A simple approach is to truncate any negative SWM simulations to zero. We term this the mv-trunc approach, which serves as a benchmark method. However, the mv-trunc approach is not designed to preserve the persistence of zero flow events or cross-correlation of zero flows, and so an alternative approach based on an auto-logistic regression model is also forwarded that is designed to preserve these properties. This approach (termed mv-alog) relies on a logistic regression to estimate the Bernoulli probability p of a zero-flow event based on a set of predictor variables (x_1, x_2,…, x_m):

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m)}} \qquad \text{(Eq. 4)}$$

To implement this model we use a sequential fitting procedure (see Figure S1 for a graphical depiction). At site 1, the auto-logistic regression model is estimated with the lag-1 binary timeseries from site 1 observations ($Q_{t-1,k}^{bin};\ 0 \text{ for zero flow}, 1 \text{ for non} - \text{zero flow}$) and the entire vector of DWM simulations across sites ($\boldsymbol{S_t}$),. That is, $x_1, \dots, x_m = \{Q_{t-1,1}^{bin}, S_{t,1:k}\}$. Site 2 includes the same covariates but adds the concurrent binary timeseries for site 1, i.e., $x_1, \dots, x_m = \{Q_{t-1,2}^{bin}, S_{t,1:k}, Q_{t,1}^{bin}\}$. Site 3 includes the same covariates as site 1 but adds the concurrent binary timeseries for site 1 and 2, i.e., $x_1, \dots, x_m = \{Q_{t-1,3}^{bin}, S_{t,1:k}, Q_{t,1:2}^{bin}\}$. This sequential fitting proceeds through to the final site k.

Simulation proceeds in the same order, requiring only the specification of a random binary starting value for site 1. That is, the binary value generated for time $t = 1$ is a Bernoulli draw based on the probability from Eq. 4, using as covariates the DWM simulation values across sites at $t = 1$ $\left(S_{t=1,1:k}\right)$ and a random binary value for t=0 $(Q_{t=0,1}^{bin})$. The remainder of binary values at site 1 are generated sequentially through time, using the estimated binary values from the previous time step as a covariate. Once the binary simulation for site 1 is complete, the binary sequence for site 2 is simulated using the generated binary sequence for site 1 as an additional covariate. The remainder of the sites are generated sequentially in this manner. This novel procedure enables the generation of random binary sequences that preserve multisite correlations and persistence in zero flows, as well as dependence between DWM simulations and observed zero flows.

To postprocess SWM simulations using mv-alog, we first need to remove negative flows generated via the baseline SWM algorithm. We employ a rudimentary procedure to do this by setting negative flows generated by the SWM to the minimum non-zero observation or minimum simulation value, whichever is smaller. We also note that the auto-logistic intermittency model is fit to the entire dataset, rather than by month, as zero flows were mostly isolated to the summer season.

4. Results

To assess model performance, we generate 1000 samples from the multisite SWM model and employ the two post-processing techniques (mv-trunc and mv-alog) to the resulting ensemble. We also generate an independent SWM benchmark (termed 'ind') using a model very similar to the multisite SWM but with independent, univariate replacements for the VAR and copula models (i.e., replacing the VAR(3) with AR(3) and empirical copula with random, independent residual generation). SWM simulations are generated for all 14 watersheds shown in Figure 1, but we focus on a subset of 6 watersheds (ORO, YRS, FOL, NHG, MRC, SCC) when presenting results for the purposes of illustration. Additional verification results for all 14 sites are shown in Supporting Information S2.

We first verify that the multisite SWM can replicate the statistical attributes of the data to which it was trained (Stedinger & Taylor, 1982; Shabestanipour et al., 2023). In Figure 3a, we highlight univariate distributional properties of the VAR residuals ($\boldsymbol{\epsilon_t}$) and the fitted SEP distributions for a selection of sites and months. Overall, the residual distributions are centered around 0 and have standard deviations near unity ($\mu \approx 0, \sigma \approx 1$), suggesting that the conditional bias correction and heteroscedasticity models function properly. This is true for two geographically separate sites (FOL in the north that is snowmelt driven and NHG in the south that is rainfed and highly intermittent) and two separate months (cold/wet December and warm/dry June). The distributions are relatively symmetric, albeit with a slight tendency towards right skew ($\xi > 1$) in some months (i.e., larger underpredictions). The distributions across sites and months differ the most based on their shape parameter (β), where the NHG site in December exhibits a near-Gaussian shape ($\beta = 0$), while other site/month combinations exhibit fat-tailed, Laplace-like distributions ($\beta = 1$).

For streamflow intermittency at individual sites, we compare the 0-0 transition probabilities (zero-flow persistence) between the truncated (mv-trunc) and auto-logistic (mv-alog) approaches across the subset of 6 watersheds (Figure 3b). At all selected sites, the mv-alog approach reproduces zero-flow persistence well and with limited sampling variability. In contrast, the mv-trunc approach performs well at certain sites but with high sampling variability, and it underpredicts persistence at NHG and MRC. We also examine the frequency of non-zero flows across the two methods (Figure 3c). Again, we find that the mv-alog approach reproduces the fraction of non-zero flows well across all sites, whereas the mv-trunc approach tends to overestimate non-zero flow days for NHG and to a lesser extent MRC. These findings suggest that simple truncation can work well at sites with moderate intermittency, but may underestimate zero-flow behavior at sites with higher intermittency (NHG).

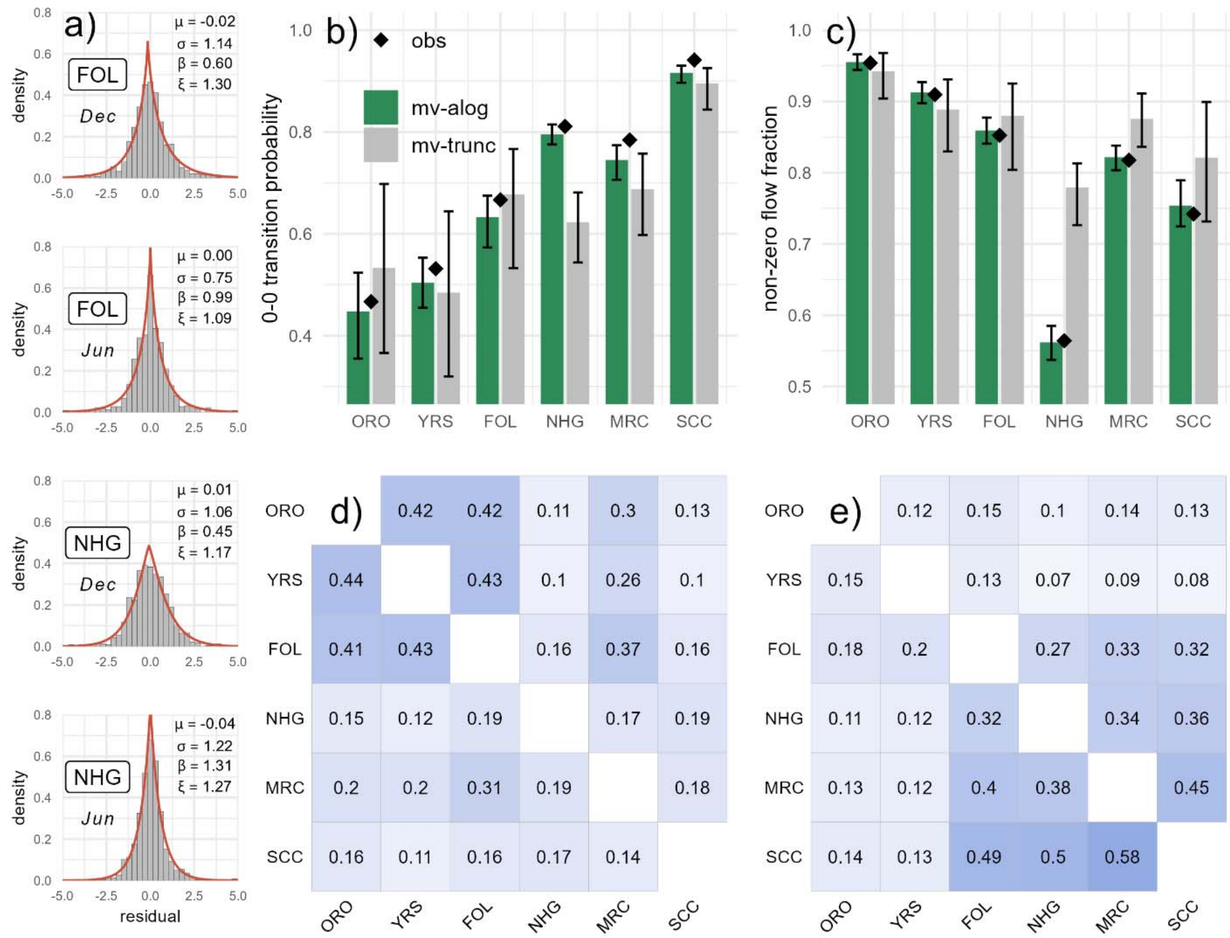

__Figure 3__. a) Empirical residuals (histogram) and fitted SGED pdf (red line) for two locations (FOL and NHG) and two months (December and June). b) 0-0 Markov transition probability (zero flow persistence) across six locations in the observations (diamonds) and for the SWM simulations using both the mv-alog and mv-trunc approaches. The bars show the median values and the whiskers show the full range of values across the 1000 SWM simulations. c) Same as in (b) but for the fraction of days with non-zero flows. d) Spearman correlations in empirical debiased errors (lower left triangle) versus the median correlation across 1000 simulated samples (upper right triangle). Results here are only shown for the mv-alog approach. e) Same as in (d) but for Pearson correlations on binary flow outcomes (0/1 = zero-flow/non-zero flow).

Across the 6 selected sites, we examine Spearman rank based correlations in empirical and SWM simulated errors in Figure 3d, while Figure 3e shows Pearson correlations for both empirical and SWM simulated binary series (0 = zero-flow, 1 = non-zero-flow). Correlations from the empirical errors are shown in the lower left portion of each matrix, while simulation-based correlations are shown in the upper-right. Here, results are only shown for the SWM with mv-alog. Broadly, the model replicates well the general pattern of correlations in error magnitude and intermittency. There is a tendency towards overestimating correlations between certain sites in error magnitude (i.e., MRC against ORO, YRS, and FOL), while intermittency correlations are generally underestimated (e.g., see SCC versus FOL, NHG, and MRC). However, these biases are relatively small.

The results in Figure 3 show that the SWM correctly captures many multisite statistical properties of DWM errors. However, we are most interested in whether the SWM ensemble is 'fit for purpose' in hydrologic risk analysis (Stedinger & Taylor, 1982; Shabestanipour et al., 2023), which in this case involves capturing the attributes of multisite extremes. For demonstration, we choose three northern sites (ORO, YRS, FOL) and focus on both multisite flooding and low flow events. These three locations are near one another in snowmelt dominated catchments, and so have correlated floods that are often driven by rain-on-snow events. In addition, these locations have important environmental low flow requirements driven by Chinook salmon and Steelhead spawning requirements.

For flooding, we first focus on the largest observed event in the record (January 1, 1997). Figure 4a shows daily flows from the observations, the DWM simulation, a single trace from the SWM ensemble, and the 90% bounds from the SWM ensemble, all for 15 days prior to and after the event. The DWM simulations are all biased below the observations at the peak of the event, while the SWM ensembles correct for this low bias and encapsulate the observations. Focusing on the single SWM trace, one can also see deviations both above and below the observations that are correlated across the three sites.

Next, we focus on flood metrics that are commonly used in planning studies, such as design events. Figure 4b shows the 10-year flood for flows summed across the three sites (ORO, YRS, FOL), which provides one measure of joint flood risk. The 10-year flood was estimated from the observations and the DWM simulation by fitting a GEV distribution to the annual maxima of the combined flows, and are shown as triangles in Figure 4b. A similar approach was taken for the 1000 multisite SWM simulations (green density), as well as for flows simulated for each location separately from the independent SWM benchmark (orange density). Figure 4c shows the same results as Figure 4b, but for the 100-year event.

For both the 10-year and the 100-year events, the DWM results are biased low compared to the observations. The median of both the multisite and independent SWM flood event distributions are closer to the observed flood event estimates than the estimate from the DWM. However, the multisite SWM exhibits more probability density in the upper tails and brings the median of the SWM distribution closer to the observed estimate, especially for the 10-year event. This finding shows that preserving multisite correlations has important ramifications for SWM estimation of combined extreme outflows from multiple watersheds, and that independent simulations from a SWM at multiple sites can underestimate combined flood flows.

A similar result is seen in Figure 4d, which shows the likelihood (expressed as a return period) of the SWM ensemble producing 3-day summed flows (Q_t^{*3d}) at each of the three locations that exceed the January 1, 1997 observed 3-day summed flow (i.e., the probability that $\tilde{Q}_t^{ORO3d} \geq Q_{1-1-97}^{ORO3d}\ and\ \tilde{Q}_t^{YRS3d} \geq Q_{1-1-97}^{YRS3d}\ and\ \tilde{Q}_t^{FOL3d} \geq Q_{1-1-97}^{FOL3d}$). We use 3-day summed flows to highlight longer duration flow dynamics important for water systems design and that are characteristic of the most intense storms in the region (Lamjiri et al., 2017; Ralph et al., 2019). We show these results using simulations from the independent SWM and the multisite SWM using both the mv-alog and mv-trunc approaches. Both multisite versions of the SWM estimate a substantially lower return period (~ 85-year event) for the January 1, 1997 flood across the three sites compared to the single site SWM (~ 125-year event), again showing how the multisite version of the model produces joint extremes across sites with much higher likelihood than a SWM applied independently to multiple sites.

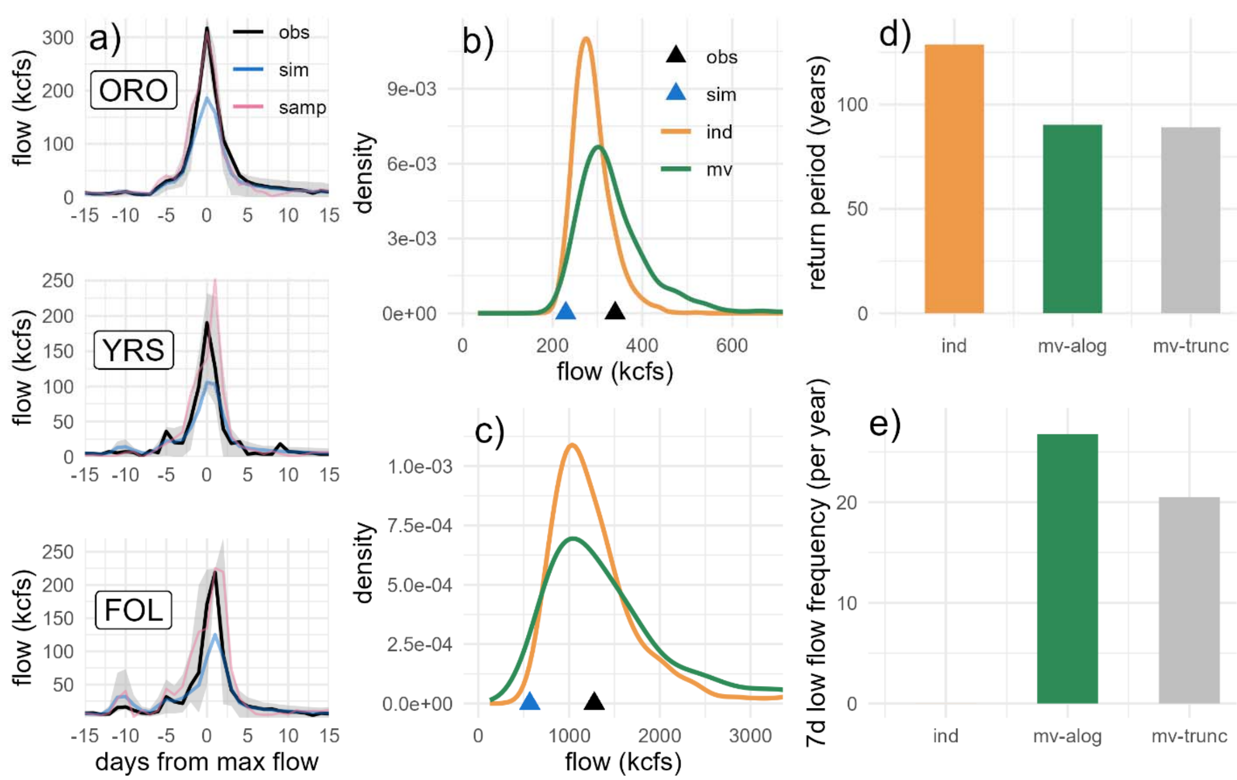

Figure 4. *a) Timeseries of daily flows in the ORO, YRS, and FOL watersheds around the flood of Jan 1, 1997, where flows are shown for 15 days prior and after the maximum flow on January 1. Observed flows (black), simulated flows from the DWM (blue), and a single flow simulation from the SWM (pink) are shown, along with the 90th percentile bounds from the SWM ensemble (grey). b) Point estimates of the 10-year flood event for flows summed across ORO, YRS, and FOL, shown for both the observations (black triangle) and the DWM simulation (blue triangle). Also shown are the distributions of the 10-year flood event across the ensemble of SWM simulations from the multisite model (green density) and the independent SWM benchmark (orange density). c) Same as in (b) but for the 100-year event. d) A trivariate return period estimate of the largest observed joint flow event (Jan 1, 1997) for 3-day summed flows from 1000 concatenated samples of the independent SWM (ind) and two versions of the multisite SWM (mv-alog, mv-trunc). e) Joint low flow frequency of 7-day average flows below environmental flow minimums for the three sites, shown for the independent SWM and two versions of the multisite SWM.*

Finally, Figure 4e shows a similar analysis to that in Figure 4d but for low flow extremes relevant to environmental flow requirements. We focus on the period of October-January when fall-run Chinook migrate upstream to their spawning grounds. We define environmental thresholds of 700 cfs, 700 cfs, and 500 cfs for ORO, YRS, and FOL, respectively, based on applicable local environmental flow regulations (Cain & Monohan, 2008; Lauer & McClurg, 2009; USACE, 2017; Yuba Water Agency, 2023), and then determine how often 7-day average low flows simulated by the SWM are below these environmental low flow thresholds simultaneously across all three sites. The joint occurrence of these low flow events is important because they would stress the regional ecology and could require joint releases from all three reservoirs to support environmental flows, with implications for water supply later in the season.

The results in Figure 4e show that the independent SWM never produces events that are jointly below the environmental thresholds at all three sites. In contrast, the two multisite SWM ensembles produce more than 20 occurrences per year on average. In the observations, these joint low flow events occur 16.5 times per year on average. The mv-alog model produces about

25% more occurrences per year compared to the mv-trunc approach, showing a moderate effect from employing an explicit zero-flow model in the SWM simulations.

5. Conclusion

In this study, we contribute a multisite SWM that captures correlated behavior in DWM simulations across sites, leveraging recent advances in SSM (Papalexiou, 2018; Papalexiou & Serinaldi, 2020; Tsoukalas et al., 2019, 2020) and tailoring them for the SWM context. We also developed a multisite auto-logistic regression to account for streamflow intermittency and its spatial and Markovian structure. We demonstrate that the multisite SWM replicates multivariate statistical attributes of DWM errors, and that the multisite auto-logistic regression helps improve the representation of zero-flow behavior over a simpler truncation method. We further investigated the importance of multisite modeling in the context of operationally relevant design statistics, and found that the multisite version of the SWM estimated joint flood and low flow events across sites with a much greater likelihood than a comparable SWM applied independently to each site. These results show that single-site applications of SWMs can significantly underestimate joint hydrologic risks.

Future work should consider the application of multisite SWMs with different transforms of the predictive uncertainty (e.g. logarithm, logarithmic ratio, Box-Cox) and the application of Nataf-based multivariate designs employed in recent stochastic simulation studies (Papalexiou & Serinaldi, 2020; Tsoukalas et al., 2020). In addition, intermittency modeling featuring mixture models or censored distributions could be considered (Ye et al., 2021). The application of machine learning based hydrologic prediction and uncertainty estimation techniques, especially regionalized approaches, offers an exciting area of exploration for SWMs (Frame et al., 2021; Klotz et al., 2022; Nearing et al., 2020). Finally, the need to understand SWM predictive uncertainty under non-stationarity is critical to its use for water resources planning purposes and is an important area of future work.

Data Availability Statement

All code and data are available in public GitHub (https://github.com/zpb4/multivariate_swm) and Zenodo repositories (https://doi.org/10.5281/zenodo.8155751) respectively and cited in the references.

Acknowledgements

This project was supported through funds provided under U.S. National Science Foundation award no. 2205239

Reference

Brodeur, Z. P., & Steinschneider, S. (2021). A Multivariate Approach to Generate Synthetic Short-To-Medium Range Hydro-Meteorological Forecasts Across Locations, Variables, and Lead Times. *Water Resources Research*, *57*(6). https://doi.org/10.1029/2020WR029453

Cain, J., & Monohan, C. (2008). *Estimating Ecologically Based Flow Targets for the Sacramento and Feather Rivers* (DWR 1126; p. 75). National Heritage Institute (NHI). https://n-h-i.org/wp-content/uploads/2017/01/Sacramento-and-Feather-Env-Flows-Doc_April2008.pdf

Chen, L., & Guo, S. (2018). *Copulas and its application in hydrology and water resources*. Springer Berlin Heidelberg.

Chen, L., Singh, V. P., Guo, S., Zhou, J., & Zhang, J. (2015). Copula-based method for multisite monthly and daily streamflow simulation. *Journal of Hydrology*, *528*, 369–384. https://doi.org/10.1016/j.jhydrol.2015.05.018

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004). The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields. *Journal of Hydrometeorology*, *5*(1), 243–262. https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2

Croux, C., & Joossens, K. (2008). Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares Procedure. In P. Brito (Ed.), *COMPSTAT 2008* (pp. 489–501). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2084-3_40

Efstratiadis, A., Dialynas, Y. G., Kozanis, S., & Koutsoyiannis, D. (2014). A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environmental Modelling and Software*, *62*, 139–152. https://doi.org/10.1016/j.envsoft.2014.08.017

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, *52*(7), 5619–5633. https://doi.org/10.1002/2016WR019129

Favre, A.-C., Adlouni, S. E., Perreault, L., Thiémonge, N., Bobée, B., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, *40*(1). https://doi.org/10.1029/2003wr002456

Frame, J., Frame, J., Kratzert, F., Klotz, D., Klotz, D., Gauch, M., Shelev, G., Gilon, O., Qualls, L. M., Gupta, H. V., Nearing, G., & Grey Nearing. (2021). Deep learning rainfall-runoff predictions of extreme events. *Hydrology and Earth System Sciences Discussions*, 1–20. https://doi.org/10.5194/hess-2021-423

Galanos, A. (2022). *Multivariate GARCH Models* (1.3-9) [R]. https://cran.r-project.org/web/packages/rmgarch/rmgarch.pdf

Hah, D., Quilty, J. M., & Sikorska-Senoner, A. E. (2022). Ensemble and stochastic conceptual data-driven approaches for improving streamflow simulations: Exploring different hydrological and data-driven models and a diagnostic tool. *Environmental Modelling & Software*, *157*, 105474. https://doi.org/10.1016/j.envsoft.2022.105474

Haktanir, T., Kara, M. B., & Acanal, N. (2022). A multiseries stochastic model for synthetic monthly flows. *Hydrological Sciences Journal*, *67*(5), 741–758. https://doi.org/10.1080/02626667.2022.2039662

Hanak, E. (Ed.). (2011). *Managing California's water: From conflict to reconciliation*. Public Policy Institute of California. http://www.ppic.org/content/pubs/report/R_211EHR.pdf

Klotz, D., Kratzert, F., Gauch, M., Sampson, A. K., Johannes Brandstetter, Johannes Brandstetter, Günter Klambauer, Günter Klambauer, Sepp Hochreiter, Sepp Hochreiter, Grey Nearing, & Grey Nearing. (2022). Uncertainty estimation with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences*, *26*(6), 1673–1693. https://doi.org/10.5194/hess-26-1673-2022

Koutsoyiannis, D., & Montanari, A. (2022). Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions. *Water Resources Research*, *58*(1). https://doi.org/10.1029/2021WR031215

Lamjiri, M. A., Dettinger, M. D., Ralph, F. M., & Guan, B. (2017). Hourly storm characteristics along the U.S. West Coast: Role of atmospheric rivers in extreme precipitation: Storm Characteristics in U.S. West Coast. *Geophysical Research Letters*, *44*(13), 7020–7028. https://doi.org/10.1002/2017GL074193

Lauer, S., & McClurg, S. (2009). *The Lower Yuba River Accord: From Controversy to Consensus* (ISBN: 1-893246-88-4; p. 24). Water Education Foundation. https://www.yubawater.org/DocumentCenter/View/84/Lower-Yuba-River-Accord-Overview-PDF

Levick, L. R., Goodrich, D. C., Hernandez, M., Fonseca, J., Semmens, D. J., Stromberg, J., Tluczek, M., Leidy, R. A., Scianni, M., Guertin, D. P., & Kepner, W. G. (2008). *The*

Ecological and Hydrological Significance of Ephemeral and Intermittent Streams in the Arid and Semi-arid American Southwest (EPA/600/R-08/134, ARS/233046; p. 116). U.S. Environmental Protection Agency and USDA/ARS Southwest Watershed Research Center.

Loucks, D. P., & Van Beek, E. (2017). *Water Resource Systems Planning and Management*. Springer International Publishing. https://doi.org/10.1007/978-3-319-44234-1

Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, H. A., Marglin, S. A., & Fair, G. M. (1962). *Design of Water-Resources Systems*. Harvard University Press.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*(3), 2199–2239. https://doi.org/10.1002/2016WR019168

Montanari, A., & Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations: RAINFALL-RUNOFF UNCERTAINTY. *Water Resources Research*, *40*(1). https://doi.org/10.1029/2003WR002540

Montanari, A., & Koutsoyiannis, D. (2012). A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, *48*(9), 2011WR011412. https://doi.org/10.1029/2011WR011412

Nearing, G., Grey Nearing, Grey Nearing, Kratzert, F., Sampson, A. K., Craig S. Pelissier, Pelissier, C., Klotz, D., Klotz, D., Frame, J. M., Frame, J., Cristina Prieto, Cristina Prieto, Prieto, C., & Gupta, H. V. (2020). What Role Does Hydrological Science Play in the Age of Machine Learning. *Water Resources Research*, *57*(3). https://doi.org/10.1029/2020wr028091

Nowak, K., Prairie, J., Rajagopalan, B., & Lall, U. (2010). A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow: NON-PARAMETRIC DAILY DISAGGREGATION. *Water Resources Research*, *46*(8). https://doi.org/10.1029/2009WR008530

Papalexiou, S. M. (2018). Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources*, *115*, 234–252. https://doi.org/10.1016/j.advwatres.2018.02.013

Papalexiou, S. M., & Serinaldi, F. (2020). Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity. *Water Resources Research*, *56*(2). https://doi.org/10.1029/2019WR026331

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., Schick, L. J., & Smallcomb, C. (2019). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269–289. https://doi.org/10.1175/BAMS-D-18-0023.1

Saad, C., Adlouni, S. E., St-Hilaire, A., & Gachon, P. (2015). A nested multivariate copula approach to hydrometeorological simulations of spring floods: The case of the Richelieu River (Québec, Canada) record flood. *Stochastic Environmental Research and Risk Assessment*, *29*(1), 275–294. https://doi.org/10.1007/s00477-014-0971-7

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, *46*(10), 2009WR008933. https://doi.org/10.1029/2009WR008933

Serinaldi, F., & Kilsby, C. G. (2018). Unsurprising Surprises: The Frequency of Record-breaking and Overthreshold Hydrological Extremes Under Spatial and Temporal Dependence. *Water Resources Research*, *54*(9), 6460–6487. https://doi.org/10.1029/2018WR023055

Shabestanipour, G., Brodeur, Z., Farmer, W. H., Steinschneider, S., Vogel, R. M., & Lamontagne, J. R. (2023). Stochastic Watershed Model Ensembles for Long-Range Planning: Verification and Validation. *Water Resources Research*, *59*(2), e2022WR032201. https://doi.org/10.1029/2022WR032201

Sikorska, A. E., Montanari, A., & Koutsoyiannis, D. (2015). Estimating the Uncertainty of Hydrological Predictions through Data-Driven Resampling Techniques. *Journal of Hydrologic Engineering*, *20*(1), A4014009. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000926

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., Lawrence, J., Lempert, R. J., Muccione, V., Mackey, B., New, M. G., O'Neill, B., Otto, F., Pörtner, H.-O., Reisinger, A., Roberts, D., Schmidt, D. N., Seneviratne, S., Strongin, S., … Trisos, C. H. (2021). A framework for complex climate change risk assessment. *One Earth*, *4*(4), 489–501. https://doi.org/10.1016/j.oneear.2021.03.005

Stedinger, J. R., & Taylor, M. R. (1982). Synthetic streamflow generation: 1. Model verification and validation. *Water Resources Research*, *18*(4), 909–918. https://doi.org/10.1029/WR018i004p00909

Steinschneider, S., Wi, S., & Brown, C. (2015). The integrated effects of climate and hydrologic uncertainty on future flood risk assessments: FLOOD RISK UNDER HYDROLOGIC AND CLIMATE UNCERTAINTY. *Hydrological Processes*, *29*(12), 2823–2839. https://doi.org/10.1002/hyp.10409

Teegavarapu, R. S. V., Salas, J. D., Stedinger, J. R., American Society of Civil Engineers, & Environmental and Water Resources Institute (U.S.) (Eds.). (2019). *Statistical analysis of hydrologic variables: Methods and applications*. American Society of Civil Engineers.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. (2021). Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research*, *57*(7). https://doi.org/10.1029/2020WR028392

Tsoukalas, I., Efstratiadis, A., & Makropoulos, C. (2019). Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *Journal of Hydrology*, *575*, 354–380. https://doi.org/10.1016/j.jhydrol.2019.05.017

Tsoukalas, I., Kossieris, P., & Makropoulos, C. (2020). Simulation of Non-Gaussian Correlated Random Variables, Stochastic Processes and Random Fields: Introducing the anySim R-Package for Environmental Applications and Beyond. *Water*, *12*(6), 1645. https://doi.org/10.3390/w12061645

USACE. (2017). *Folsom Dam and Lake, American River, California: Water Control Manual* (December 1987, Revised September 2017). USACE. https://www.spk.usace.army.mil/Portals/12/documents/civil_works/JFP/Water%20Control%20Manual%20Update/DSEAandAppendices2017/DraftWCMUpdate_06062017.pdf?ver=2017-06-06-170628-783

Vannitsem, S. (Ed.). (2018). *Statistical postprocessing of ensemble forecasts*. Elsevier.

Vogel, R. M. (2017). Stochastic watershed models for hydrologic risk management. *Water Security*, *1*, 28–35. https://doi.org/10.1016/j.wasec.2017.06.001

Wang, Q. J., & Robertson, D. E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences: MULTISITE PROBABILISTIC FORECASTING OF STREAMS. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009333

Wi, S., & Steinschneider, S. (2022). Assessing the Physical Realism of Deep Learning Hydrologic Model Projections Under Climate Change. *Water Resources Research*, *58*(9). https://doi.org/10.1029/2022WR032123

Wuertz, D., Chalabi, Y., Setz, T., Maechler, M., Boudt, C., Chausse, P., Miklovac, M., & Boshnakov, G. N. (2022). *Rmetrics—Autoregressive Conditional Heteroskedastic Modelling* (4022.89) [R]. https://www.rmetrics.org

Ye, L., Gu, X., Wang, D., & Vogel, R. M. (2021). An unbiased estimator of coefficient of variation of streamflow. *Journal of Hydrology*, *594*, 125954. https://doi.org/10.1016/j.jhydrol.2021.125954

Yuba Water Agency. (2023). *Lower Yuba River Accord Flow Schedules, Marysville Gauge*. https://www.yubawater.org/397/4193/Lower-Yuba-River-Accord-Flow-Schedules

Zha, X., Xiong, L., Guo, S., Kim, J.-S., & Liu, D. (2020). AR-GARCH with Exogenous Variables as a Postprocessing Model for Improving Streamflow Forecasts. *Journal of Hydrologic Engineering*, *25*(8), 04020036. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001955

Zscheischler, J. (2020). Moving beyond isolated events. *Nature Climate Change*, *10*(7), 583–583. https://doi.org/10.1038/s41558-020-0846-5

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., Van Den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., & Vignotto, E. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, *1*(7), 333–347. https://doi.org/10.1038/s43017-020-0060-z