Copula Theory as a Generalized Framework for Flow-Duration Curve Based Streamflow Estimates in Ungaged and Partially Gaged Catchments

Scott. C. Worland1,2, Scott Steinschneider3, William Farmer1, William Asquith4, and Rodney Knight1

1U.S. Geological Survey, Lower Mississippi-Gulf Water Science Center, Nashville, TN, USA, 2Cornell University Department of Biological and Environmental Engineering, Ithaca, NY, USA, 3U.S. Geological Survey, Denver, CO, USA, 4U.S. Geological Survey, Texas Water Science Center, Lubbock, TX, USA

Abstract  Flow-duration curve (FDC) based streamflow estimation methods involve estimating an FDC at an unaged or partially gaged location and using the time series of nonexceedance probabilities estimated from donor streamgage sites to generate estimates of streamflow. We develop a mathematical framework to illustrate the connection between copulas and prior FDC-based approaches. The performance of copula methods is compared to several other streamflow estimation methods using a decade of daily streamflow data from 74 sites located within two river basins in the southeast United States with different climate characteristics and physiographic properties. We show that copula approaches: (1) outperform other methods in the limiting case of perfect information with regard to the rank-based correlation structure and FDCs across the gaging network; (2) provide a hedge against poor performance when donor information becomes sparser and less informative; (3) outperform other methods when used for partially gaged sites with several years of available data; and (4) remain a competitive albeit nondominating method for unaged sites and partially gaged sites with limited data when realistic error is introduced in the estimation of FDCs and correlations across the gaging network.

1. Introduction

Less than 1% of streams in the United States are instrumented with a streamgage (National Academies of Sciences, Engineering, and Medicine, 2018; U.S. Environmental Protection Agency, 2013). Further, many streams that have been gaged have limited data records and require extension or gap filling for water-resource applications, either because the gages were installed in recent decades, were discontinued due to budgetary constraints, or failed over an extended period of time. Hydrologists rely on process or statistical models to estimate streamflow in unaged or partially gaged catchments (Blöschl, 2013, 2016; Booker & Snelder, 2012). Process-based models estimate streamflow by approximating the various physical processes of the hydrologic cycle using a conceptual understanding of the governing relations between climate, landscape properties, and streamflow. Statistical models use patterns discovered in observational data from gaged catchments to estimate streamflow in unaged or partially gaged catchments. Both approaches have distinct benefits and drawbacks depending on the modeling context. For process models, a parameter set for a target site of interest (particularly an unaged site) needs to be estimated by transferring calibrated parameter values from gaged (donor) sites (Hrachowitz et al., 2013), for instance, based on spatial proximity or physical similarity (Steinschneider et al., 2015). This approach enables direct estimates of future streamflow for the target site, as would be needed under projections of climate change (Beck et al., 2016). However, structural and parametric uncertainties (Arsenault & Brissette, 2014, 2016), coupled with uncertain climate forcing data (Bisselink et al., 2016), are major challenges for process model regionalization. Statistical models have the benefit of leveraging observed flow at gaged sites, which already contains information on the integration of all underlying hydrologic processes that control streamflow. Thus, if the goal of the regionalization process is restricted specifically to the reconstruction of historical streamflow at unaged or partially gaged sites in a region with moderate gaging density, statistical methods tend to result in more accurate retrospective estimates that better capture the distributional properties of historical streamflow (Farmer et al., 2015a).
Of the suite of statistical approaches that are available (Razavi & Coulibaly, 2012), flow-duration curve (FDC) based methods are considered one of the most reliable (Archfield & Vogel, 2010; Booker & Snelder, 2012; Farmer et al., 2015b; Lorenz & Ziegeweid, 2016; Mohamoud, 2008; Zhang et al., 2015). This is in part because they separately model the shape of the FDC and the timing of streamflow events at the target site, allowing for the use of information tailored for each of those components. In particular, the FDC shape can be determined by the physical attributes of the catchment (a relation learned using many other gaged sites), while streamflow timing can be modeled based on the timing of events only at nearby gaged sites. FDC-based methods for estimating streamflow were concurrently developed in the mid-1990s in the United States by Fennessey (1994) and in South Africa by Hughes and Smakhtin (1996). Although FDC-based methods have been referred to by several names (e.g., spatial nonlinear interpolation (Archfield & Vogel, 2010; Mohamoud, 2008; Hughes & Smakhtin, 1996; Farmer et al., 2015b)), we use “QPPQ”—the name given in Fennessey (1994). QPPQ generally involves the following four steps: (1) estimating an FDC for the target catchment of interest; (2) choosing one or multiple donor sites for each target catchment; (3) transferring the time series of nonexceedance probabilities (Ps) from the donor site(s) to the target catchment; (4) and using the estimated FDC for the target catchment to map the donated Ps back to streamflow. Steps 3 and 4 involve converting streamflow (Q) to P for the donor and converting P back to Q for the target, hence the name “QPPQ.” If the target site is partially gaged, estimating the FDC (Step 1) can be based directly on the available record. If the target site is unengaged or partially gaged with a very short record, FDCs can be estimated using regression models to either regionalize the parameters or moments of theoretical distributions (e.g., Log-Pearson Type III) that approximate the FDC (Atieh et al., 2017; Blum et al., 2017) or to regionalize a discrete set of quantiles that are interpolated to create a continuous FDC (Farmer et al., 2015b; Schnier & Cai, 2014). Although generating accurate FDC estimates for target locations is an important step in QPPQ, this paper primarily focuses on how the donor sites are chosen (step 2) and how the Ps are transferred between sites (Steps 3 and 4).

In its most basic form (i.e., one donor and one target site), the P to P component of QPPQ assumes that the Ps at the target site are exactly equal to the Ps at the donor site. This probability equivalence assumption is almost certainly inaccurate. Rather, it is more realistic to assume that the value of P at the target site will be similar but not equal to that of the donor site. In addition, there will be uncertainty around the value of P estimated for the target site, and this uncertainty will influence the uncertainty in the final streamflow estimate. Previous studies have recognized the need to improve the estimate of P at the target by relaxing the 1:1 correspondence with P at a donor site, often by combining information across many donors. For example, Smakhtin (1999) used a weighted mean of Ps from multiple donors, whereas Farmer (2015) used a kriging method to regionalize daily standard normal deviates (i.e., z scores) based on the values of P from many donors passed through the quantile function of a standard normal distribution, which were then back transformed using the standard normal cumulative distribution function (CDF) to a value of P at the target site. Recently, Farmer and Levin (2018) used another FDC-based regionalization approach to estimate uncertainty in daily streamflow at the target site, although this was based on cross-validated estimates of error for streamflow predictions at donor sites rather than uncertainty in P at the target site.

Many of the approaches mentioned above can be viewed as models of the joint distribution of Ps between the target and donor sites, even if they are not presented in such probabilistic terms. Under this framing, a key step in the QPPQ method becomes the estimation of the conditional distribution of P at the target given known values of Ps at the donors. If this conditional distribution is available, then it becomes straightforward to develop a point estimate of P at the target site, as well as a formal quantification of its uncertainty. Uncertainty in P can then be directly propagated into an estimate of uncertainty in the associated daily streamflow estimate for the target site.

In this study, we argue that copula theory provides a unifying framework for the estimation of the joint and conditional distributions between Ps at the target and donor sites. Copulas have become increasingly popular in the field of hydrology, with applications in flood-frequency analysis, drought analysis, forecasting, and multisite streamflow simulation (see Chen & Guo, 2019 for a recent review). Copulas describe the family of multivariate distributions for uniform random variables and are commonly used in a two-step procedure to model the joint distribution of random variables with arbitrary marginal distributions. In this procedure, the original multivariate data (e.g., streamflow at different sites) are first passed through their marginal distributions (e.g., FDCs estimated at each site) to produce estimated Ps. These Ps are thus uniformly distributed on
The interval $(0, 1)$, and their joint distribution is then modeled using a copula. This procedure perfectly mirrors that of QPPQ, and in fact, one approach described in (Farmer, 2015) effectively used a Gaussian copula, although this connection was not stated. The introduction of copula theory to the QPPQ problem presents an opportunity to guide future model development and more formally characterize the joint distribution of the $P$s across sites. In particular, many copula models provide closed-form expressions for the conditional distributions of $P$ at the target site given known values of $P$ from each donor site, providing a direct way to quantify uncertainty at the target. In addition, there are a range of copula models to account for complexity in the joint distribution of $P$s across sites, such as tail dependence and asymmetry (Salvadori & De Michele, 2004).

Our results show that QPPQ-type methods can be described within the framework of copulas and demonstrate how this framing can be used to estimate streamflow in practice for ungaged and partially gaged sites. We begin by providing a brief introduction to copula theory and its relation to QPPQ. We then build six different streamflow-transfer models, including both copula- and non-copula-based methods, using data from 2000–2009 for 74 sites in two regions in the southeast United States with distinct hydrologic regimes. To isolate the performance of the different streamflow-transfer methods, we first build the models using the empirical FDCs and observed correlations between streamflow at the different sites over the entire 10-year period. This provides an upper bound of performance for the different methods. We then use estimated FDCs and correlations for both partially gaged and completely ungaged sites under cross-validation to compare the streamflow-transfer methods when there is additional error introduced from each modeling step in the QPPQ framework.

### 2. QPPQ and Copulas

Let $p_{1,t}, \ldots, p_{k,t}$ be the nonexceedance probabilities from $k$ donor sites at time $t$, $p_{k+1,t}$ be the nonexceedance probability at the target site at time $t$, $F_{k+1}$ be the FDC at the target site, and $q_{k+1,t}$ is the estimated streamflow at the target site. In its simplest form, $f(p_{1,t}, \ldots, p_{k,t})$ is a function that relates the nonexceedance probabilities from the donor sites to the target site. In its simplest form, $f(p_{1,t}, \ldots, p_{k,t})$ is the nonexceedance probability at time $t$ from the $i$th site that has the smallest geographic distance, $\|s_i - s_{k+1}\|$, from the target site, where $s$ is a vector of geographic coordinates. This approach is commonly referred to as the “nearest-neighbor method” (e.g., Farmer et al., 2015b). Another option is to use a weighted contribution ($o_i$) of each $p_{1,t}, \ldots, p_{k,t}$ (e.g., Hughes & Smakhtin, 1996). For example, if we calculate $o_i$ as the squared inverse geographic distance between each donor and the target, $1/\|s_i - s_{k+1}\|^2$, then an inverse distance-weighted (IDW) estimate of $p_{k+1,t}$ can be calculated as:

$$f(p_{1,t}, \ldots, p_{k,t}) = \sum_{i=1}^{k} o_i p_{i,t} / \sum_{i=1}^{k} o_i$$

Various other methods, such as using physiographic rather than geographic distance to create $o_i$, have been explored in the literature (e.g., see Castellarin et al., 2018 and Pugliese et al., 2014). Many methods used to transfer the probabilities can be recast in the form of equation (2). For example, using the nearest-neighbor approach would result in a vector of weights where every weight was zero except the nearest neighbor, which would have a weight of one.

#### 2.1. The Gaussian Copula

Copulas offer a formal framework to model the dependence structure between $p_{1,t}, \ldots, p_{k,t}$ and $p_{k+1,t}$. For ease of exposition, we first demonstrate this using a particular copula model, the Gaussian copula, which can be written as:

$$C(p_{1,t}, \ldots, p_{k+1,t}) = \Phi_{k+1} (\phi^{-1}(p_{1,t}), \ldots, \phi^{-1}(p_{k+1,t})|\Sigma) .$$
Here, \( C(p_{1:t}, \ldots, p_{k+1,t}) \) denotes a generic joint, multivariate CDF for the uniformly distributed nonexceedance probabilities (i.e., a copula). In equation (3), the joint CDF is specified using the functional form of the Gaussian copula, where \( \Phi_{k+1} \) is the \((k + 1)\) dimensional Gaussian CDF, \( \phi^{-1} \) is the one-dimensional standard normal inverse CDF, and \( \Sigma \) is a \((k + 1) \times (k + 1)\) correlation matrix. \( \Sigma \) can be partitioned as:

\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}
\]

where \( \Sigma_{22} = 1 \) represents the unit correlation of the target site with itself, \( \Sigma_{11} \) represents the correlation matrix for the donor sites, and \( \Sigma_{12} \) and \( \Sigma_{21} \) represent correlations between the target and donor sites (the estimation of which is described later). If we let \( Z_{k+1,t} = \phi^{-1}(p_{k+1,t}) \), then the conditional distribution of \( Z_{k+1,t} \) with known values for \( Z_{1:t} \), \( \ldots \), \( Z_{k,t} \) is normal and given by:

\[
Z_{k+1,t} | Z_{1:t} \sim \mathcal{N}(\mu_{k+1,t} | Z_{1:t}, \sigma_{k+1,t}^2 | Z_{1:t}),
\]

where \( \mu_{k+1,t} | Z_{1:t} \) is the conditional mean, \( \sigma_{k+1,t}^2 | Z_{1:t} \) is the conditional variance, and \( z_{1:t} \) are the real-valued \( z \) scores for each of the donor sites, and \( \mathcal{N} \) represents that Gaussian distribution. For the bivariate case and \( i^{th} \) donor site, equation (5) reduces to:

\[
Z_{k+1,t} | Z_{1:t} \sim \mathcal{N}(\rho \times z_{it}, (1 - \rho^2)),
\]

where the conditional mean estimate of \( Z_{k+1,t} \) only depends on the value of \( Z_{it} \) and the correlation coefficient \( \rho \) between the two sites, and the conditional variance only depends on the correlation coefficient. The Gaussian copula relies on the correlation between donors and the target to weight daily \( z \) scores across donors before using them to adjust the unconditional mean of the target, which is zero in \( z \) space and the median flow in the original space of streamflow. Unless the correlation values are equal to unity, this weighting scheme effectively shrinks the estimate at the target toward its median, which will help reduce the likelihood of over- or under-estimations at the target.

To produce a point estimate of streamflow at the target site, the conditional mean of \( Z_{k+1,t} \) can be mapped back to a nonexceedance probability using the standard normal CDF (equation (9)), and then back to streamflow using the inverse FDC at the target site (equation (10)):

\[
\hat{p}_{k+1,t} = \phi(\mu_{k+1,t} | Z_{1:t}) \quad \text{(9)}
\]

\[
\hat{q}_{k+1,t} = F_{k+1}^{-1}(\hat{p}_{k+1,t}) \quad \text{(10)}
\]

Similarly, the full conditional distribution of \( q_{k+1,t} \) given known values for \( q_{1:t}, \ldots, q_{k,t} \) can be simulated by generating random samples of \( Z_{k+1,t} \) from its conditional distribution (equation (5)) and then passing those samples through the standard normal CDF and then the inverse FDC of the target site. This provides a straightforward way of estimating the full conditional distribution of streamflow at the target site based on measured streamflow at nearby donor sites.

### 2.2. Alternative Copula Models

A Gaussian copula is often an appropriate model when the nonexceedance probabilities between sites are correlated in the body of the distribution but are not asymptotically dependent in the tails (Figure 2a). There are situations, however, where streamflow data do not meet this assumption. When data in the tails of the distribution across sites are correlated, this is referred to as tail dependence and can be defined separately for lower tail dependence (\( \lambda_L \)) and upper tail dependence (\( \lambda_U \)) (Schmid & Schmidt, 2007):

\[
\lambda_L = \lim_{p \to 0} P\{Q_1 \leq F_{Q_1}^{-1}(p) | Q_2 \leq F_{Q_1}^{-1}(p)\}
\]

\[
\lambda_U = \lim_{p \to 1} P\{Q_1 > F_{Q_1}^{-1}(p) | Q_2 > F_{Q_1}^{-1}(p)\}
\]
Figure 2. Simulations from five copula models with normally distributed marginals. The normal copula (a) exhibits zero-tail dependence, the Student’s \(t\) (b) copula exhibits symmetric tail dependence, the Gumbel copula (c) exhibits asymmetric upper-tail dependence, the Clayton copula (d) exhibits asymmetric lower-tail dependence, and the Frank copula (e) exhibits diffuse dependence throughout.

The dependence in the tails becomes more prominent as the \(\lambda\) values approach unity. If there is near-equal and nonzero correlation between streamflow at both the extreme low and high end of the hydrograph between sites, then the tail dependence is considered symmetrical. This situation is possible when sites are located in a relatively homogeneous region where, for instance, the low flows are predominantly controlled by either a connection to an aquifer or upstream regulation, and the high flows are controlled by storm events that similarly affect each site. If these conditions hold and an elliptical distribution is still appropriate in the body of the distribution, then the Student’s \(t\) copula can be used to model the joint distribution between the nonexceedance probabilities across sites (Figure 2b, Demarta & McNeil, 2005):

\[
C(p_{1:t}, \ldots, p_{k+1:t}) = \Psi_{k+1}(\psi^{-1}(p_{1:t}), \ldots, \psi^{-1}(p_{k+1:t}), \Sigma, \nu)
\]

Here, \(\Psi_{k+1}\) is the \(k + 1\)-dimensional Student’s \(t\) CDF, \(\psi^{-1}\) is the one-dimensional Student’s \(t\) inverse CDF, \(\Sigma\) is a \((k + 1) \times (k + 1)\) scale matrix, and \(\nu\) is the degrees of freedom. Letting \(T_{1:t} = \psi^{-1}(p_{1:t})\), the conditional distribution of \(T_{k+1:t}\) given values for \(T_{1:k:t}\) is Student’s \(t\), with a conditional mean similar to the Gaussian formulation in equation (6), but a conditional variance that depends on the values of \(T_{1:k}\) and the degrees of freedom, \(\nu\), in addition to the scale matrix \(\Sigma\) (Käärik et al., 2011):

\[
\mu_{k+1:t|T_{1:k:t}} = \Sigma_{21} \Sigma_{11}^{-1} \times T_{1:k:t}
\]

\[
\sigma_{k+1:t|T_{1:k:t}}^2 = \frac{\nu + \Sigma_{12} \Sigma_{22}^{-1} T_{1:k:t}^T}{\nu + k - 1} (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})
\]

Similar to the Gaussian copula, samples of \(T_{k+1:t}\) from this conditional distribution can be passed through the Student’s \(t\) CDF and then through the inverse FDC of the target site to generate a conditional distribution of \(q_{k+1:t}\) given values of \(q_{1:k:t}\).

Streamflow across sites might also exhibit asymmetric tail dependence, where there is correlation in one tail and not the other. For example, high flows at two locations may be controlled by storm events that similarly affect each basin, but the low flows could be disconnected if one stream is regulated and the other is not, or if the two sites have distinct underlying geologies that exert different controls over low-flow behavior. In this situation, there would be asymmetric upper-tail dependence, which can be modeled using a Gumbel copula (Figure 2c) (Salvadori & De Michele, 2004). Asymmetric lower tail dependence is also possible, for instance, if baseflow for two sites is supplied from the same aquifer system but high flows are influenced by local convective systems that often impact one or the other site but not both. A Clayton copula would be appropriate in this situation (Figure 2d). Conversely, the distribution of nonexceedance probabilities across sites could exhibit no tail dependence but also could be more diffuse in the body of the distribution than can be modeled using a Gaussian copula. In this case, a Frank copula may be appropriate (Figure 2e). These commonly used copulas are just some of the models that are available.
The Gumbel, Clayton, and Frank models are all members of a class of Archimedean copulas (Genest & Rivest, 1993), commonly used for bivariate modeling and written here generically for the target site \((p_{k+1,t})\) and the \(i^{th}\) donor site \((p_{i,t})\):

\[
C(p_{i,t}, p_{k+1,t}) = \pi^{-1}(\pi(p_{i,t} | \theta) + \pi(p_{k+1,t} | \theta) | \theta)
\]  
(16)

where \(\pi\) is referred to as the generator function. Each Archimedean copula (Gumbel, Clayton, Frank) has its own generator function, which has a single parameter \(\theta\). Conditional simulation of \(p_{k+1,t}\) for the target site given a known value of \(p_{i,t}\) for the donor site can be based on the conditional distribution function, shown here for a generic bivariate copula:

\[
C(p_{k+1,t} | p_{i,t}) = \frac{\partial}{\partial p_{i,t}} C(p_{i,t}, p_{k+1,t})
\]  
(17)

Aas et al. (2009) provide expressions for the conditional CDF of several bivariate copula models. Conditional simulations of \(p_{k+1,t}\) can then be passed through the inverse FDC of the target site to develop the conditional distribution of \(q_{k+1,t}\).

### 2.3. Estimation of Copula Parameters

The modeling framework above requires estimates for a few key terms, including parameters for each copula and \(F_{k+1}\). We discuss the estimation of \(F_{k+1}\) separately in section 3.4. For the Gaussian copula, an estimate of the correlation matrix \(\Sigma\) is needed and must include not only correlations across donor sites but also correlations between donors and the target site. If the target site is partially gaged and its streamflow record overlaps with the donor sites, then empirical correlations can be used. However, if the target site is unaged or is partially gaged but its streamflow record does not overlap that of the donors, a maximum-likelihood-based approach is not feasible because there are no data that can be used in the likelihood function. Instead, a correlogram-based model can be used to estimate correlations as a function of the Euclidean distance between sites (Oliver, 2010):

\[
\hat{\rho}(i, j) = \gamma(||s_i - s_j||, \beta)
\]  
(18)

where \(\gamma\) is the correlogram model, ||s_i - s_j|| is the distance between sites \(i\) and \(j\), and \(\beta\) are the parameters of the correlogram. \(\beta\) can be estimated based on known correlations between donor sites, and then equation (18) can be used to estimate correlations between the target site and each donor site. Many correlogram models are possible; a description of the correlogram used in this study is described in section 3.3.

Spearman \((\rho_s)\) or Kendall \((\rho_k)\) rank-based correlation coefficients are invariant under monotonic transformation and are often preferred over Pearson correlation coefficients \((\rho_p)\). Either of these rank-based correlation coefficients can be modeled with the correlogram and then \(\hat{\Sigma}(i, j)\) can be recovered as either (Aas, 2004):

\[
\hat{\Sigma}(i, j) = 2 \times \sin \left( \frac{\pi}{6} \times \hat{\rho}(i, j) \right)
\]  
(19)

\[
\hat{\Sigma}(i, j) = \sin \left( \frac{\pi}{2} \times \hat{\rho}(i, j) \right)
\]  
(20)

The final \(\hat{\Sigma}\) in equations (19) and (20) are based on the empirical correlations between donor sites and either empirical correlations between donors and partially gaged target sites or correlogram-estimated correlations between donors and the target. If the resulting correlation matrix is not positive-definite, adjustments are available to achieve this property (see Rousseau & Molenberghs, 1993). These adjustments require the specification of a tolerance for near positive definiteness, which we treat as a hyperparameter of our modeling framework and calibrate to maximize predictive skill.

The conditional Student’s \(t\) copula (Demarta & McNeil, 2005) requires estimation of both the scale matrix \(\Sigma\) and the degrees of freedom \(\nu\). These parameters can be estimated in stages using data at the donor sites and both method-of-moments and pseudo-maximum likelihood approaches (Aas, 2004). First, empirical \(\rho_s\) correlations or a correlogram for \(\rho_s\), and equation (20) can be used to estimate the scale matrix, \(\hat{\Sigma}\). Then, \(\nu\) can be estimated by maximizing the likelihood function for the Student’s \(t\) copula across donor sites given fixed \(\Sigma\).

The parameter \(\theta\) for the Archimedean copulas (Gumbel, Clayton, Frank) can be estimated based on its relation to \(\rho_s\), between sites (Aas, 2004):

\[
\rho_s = 1 - \theta^{-1}_{\text{Gumbel}}
\]  
(21)
Table 1

The set of six P-transfer methods used in this study

| P-transfer method               | Description                                                                 |
|--------------------------------|-----------------------------------------------------------------------------|
| multivar_norm_cop*             | Multivariate Gaussian copula                                                |
| bivarcop_rho*                  | Bivariate copula estimate between the target and the donor site with the highest $\rho$ |
| QPPQ_highest_rho              | Traditional QPPQ estimate using the target and the site with the highest gaged (obs FDC) relation as the donor |
| QPPQ>IDW                      | Inverse distance-weighted estimate of P from every donor used in QPPQ transfer |
| IDW_log_runoff                 | Inverse distance-weighted estimate using log runoff                          |
| rho_weight_log_runoff          | Correlation-weighted combination of log runoff values from every available donor for each time step |

* indicate copula-based methods

$$\rho_\tau = \frac{\theta_{\text{Clayton}}}{\theta_{\text{Clayton}} + 2}$$

$$\rho_\tau = 1 + \frac{4 \times (D(\theta_{\text{Frank}}) - 1)}{\theta_{\text{Frank}}}$$

where $D()$ is the Debye function (Abramowitz & Stegun, 1965). To estimate the parameter for the Gumbel and Clayton copulas, equation (21) and equation (22) can be directly inverted using an estimated value for $\rho_\tau$, while for the Frank copula, equation 23 needs to be numerically inverted.

3. Application of Copulas to Estimate Streamflow in the Southeast United States

We use 74 sites across two regions of the southeast United States to demonstrate the application of copulas to estimate streamflow in ungaged and partially gaged locations. We first select bivariate copula models for each region and present the specific correlogram used for correlation estimates to ungaged sites. We then present six P-transfer methods, including copula- and noncopula-based approaches (Table 1), that are compared under three assessment frameworks. The first assessment, termed the “fully gaged” scenario, provides an upper bound on method performance by using the observed correlations between target and donor sites and the empirical FDC at the target site for the entire 10-year record. In this assessment, we consider how the six methods perform as the donor network is thinned and its relation to the target degraded. We then present cross-validated assessments of model skill for two more scenarios, one for partially gaged sites and one for completed ungaged sites, where both FDCs and correlations for the target site are estimated (i.e., based on information other than that available in the full 10-year record at the target site). In these cross-validated assessments, the six models are also built using two intermediary scenarios: (1) the observed (i.e., 10-year) correlations between sites and estimated FDCs at the target; and (2) estimated correlations between sites and observed (i.e., 10-year) empirical FDCs at the target. The nested structure of these two intermediary scenarios is designed to determine which set of estimated terms (FDCs vs. correlations) most reduces predictive skill in a real-world QPPQ application. After the intermodel comparison, we decompose model performance into bias, variance, and timing components and also assess the quantification of prediction uncertainty. All analyses were completed using the R programming language (R Core Team, 2019) and the copula R package (Hofert et al., 2017), VineCopula R package (Schepsmeier et al., 2018), and the copBasic R package (Asquith, 2019).

3.1. Data Description

Ten years (1 January 2000 to 31 December 2009) of continuous, daily streamflow data were downloaded from the U.S. Geological Survey (USGS) National Water Information System (NWIS) web interface (U.S Geological Survey NWIS, 2018) for 37 sites in the Mobile-Tombigbee River Basin (MTB) in Alabama and Mississippi and 37 sites in the Galveston-Trinity River Basins (GTB) in Texas and Louisiana (Figure 3). The sites include both regulated and unregulated streams and only nonzero streamflow values were included in the
These two basins are located in roughly the same latitude yet have distinct physiographic regimes. Streams in the MTB have relatively small basin areas (median 371 km²), receive 1,431 mm of precipitation annually, and are characterized by high baseflow (31–57%), minimal regulation, and forested landcover (67% of landcover is forest and 11% is grasslands/pasture). Streams in the GTB have large basin areas (median 1,036 km²), receive 1,170 mm of precipitation annually, and are characterized by low baseflow (9–27%), high regulation, and balanced pastured and forested landcover (33% of landcover is grasslands/pasture and 33% is forested) (Crowley-Ornelas et al., 2019a).

### 3.2. Bivariate Copula Model Selection

Before developing a final set of copula-based QPPQ models and assessing prediction skill, we first diagnose which bivariate copula model is appropriate for each basin. The outcome of this step will support the choice of a single bivariate copula model to be considered in the final assessment framework (i.e., the bivar-cop-rho method in Table 1). We focus on selecting a bivariate copula to have a copula-based method that is directly comparable to the conventional QPPQ method that uses a single donor for streamflow estimation. In addition to the final bivariate models, a multivariate Gaussian copula is also used for both basin applications.

Three characteristics of bivariate copula models are considered to inform the choice of a final model: tail dependence, model fit, and uncertainty quantification. For each basin, upper and lower tail dependence is examined for all pairwise combinations of sites. This assessment is used to help determine whether symmetric (Gaussian, Student’s *t*, Frank) or asymmetric (e.g., Gumbel, Clayton) bivariate copulas are more appropriate, on average, for pairwise modeling in each basin. We also pair each site with the site to which it is most correlated and examine the distribution of Akaike Information Criterion (AIC) values for bivariate copula models fit to those pairs. In addition to the AIC, we also conduct goodness-of-fit hypothesis tests using the Cramer-von Mises statistic described in (Genest et al., 2009). Finally, we compare the ability of each bivariate copula to generate confidence intervals (CIs) that correctly bound predictions. We calculate 95% CIs for each model, which are equal to the 2.5th and 97.5th percentiles of the conditional distribution of target site flows given known values for the donor site (see sections 2.1 and 2.2). Coverage probabilities are the proportion of the time that observations fall into these CIs and are calculated by pooling the data over all target sites for each model and basin. Coverage probabilities are calculated separately for different ranges of flow to determine if uncertainty is being correctly estimated across low-, medium-, and high-flow ranges.

### 3.3. Correlogram Model for Estimated Correlations to Ungaged Sites

When copulas are used to estimate streamflow at an ungaged target site, correlations need to be estimated between the target and donor sites using a correlogram model (introduced in section 2.3). The correlogram model...
model used in this study was fit using the weighted least squares algorithm provided by Ribeiro Jr and Diggle (2018) in their “geoR” package of the R programming language (R Core Team, 2019). For a particular target site, a unique correlogram was fit for each donor site in the region after removing the target site from the data. For instance, for one target site in the MTB (with 37 sites in total), this resulted in 36 unique correlogram models, with each being built on the correlations of a single donor site with the 35 other donors. These correlations were used to build a semivariogram cloud by looking at intersite differences (semivariances) in relation to distance. This cloud was then binned into 10 identically sized distance bins, summarizing each bin as the central distance and mean semivariance. A spherical correlogram was then fit to these 10 points (see Supporting Information Figure S1). Using this correlogram model, it is then possible to use kriging to predict the correlation between an arbitrary point (the left-out target site) and the base donor site for which the correlogram was developed. Figure S2 shows the resulting surface of correlations based on the correlogram model defined in Figure S1 (see supporting information).

3.4. Model Intercomparison Framework

After initially testing 16 different models (see supporting information Figure S3 and Table S1), we select a final set of six approaches for comparison (Table 1): (1) multivar-norm-cop, (2) bivar-cop-rho, (3) QPPQ-highest-rho, (4) QPPQ-IDW, (5) IDW-log-runoff, and (6) rho-weight-log-runoff. These six models are chosen based on their performance and to support a nested experimental design. Specifically, we choose a model that does not rely on either estimated correlations or estimated FDCs (IDW-log-runoff), a model that relies only on estimated correlations (rho-weight-log-runoff), a model that relies only on estimated FDCs (QPPQ-IDW), and three models that rely on both the estimated FDCs and estimated correlations (multivar-norm-cop, bivar-cop-rho, QPPQ-highest-rho). As stated earlier, the bivar-cop-rho model provides the closest comparison to the conventional, single-donor QPPQ method, while the multivar-norm-cop model provides a comparable method to QPPQ models that pool nonexceedance probabilities across multiple donors (e.g., QPPQ-IDW). Three separate assessments are conducted on the set of six models:

3.4.1. Scenario 1: Fully Gaged Target Sites

To isolate the theoretical benefits of copulas, we first consider target sites to be “fully gaged.” Here, we use the observed correlations between target and donor sites and the empirical FDC for the target site over the entire 10-year period. This scenario allows us to ask the question, “What is the best way to transfer nonexceedance probabilities between donors and the target with perfect information about correlation and FDC structure?” For the fully gaged scenario, we also use a pairwise, iterative procedure that assumes perfect information about correlation and FDC structure at the target site, but compares intermodel performance as correlations between the target and available donor sites are systematically reduced. In the first iteration of this procedure, all models are built using donor sites with the highest pairwise correlation to the target site. Methods that use multiple sites (e.g., multivar-norm-cop and QPPQ-IDW) are built using all donors, including the one with the highungaged (obs FDC)elation to the target. In the second iteration, models are refit, but to a database in which the donor site with the highungaged (obs FDC)elation to the target has been removed. The third iteration removes the two donor sites with the first and second highungaged (obs FDC)elations to the target, and this is repeated until only one donor site remains with the lowungaged (obs FDC)elation to the target. The full iterative procedure is repeated for all sites, using a different site as the target each time. Predictive skill across the six models for each iteration and target site is then compared to the maximum pairwise correlation for each iteration. This provides an assessment of how the models perform as information from the donor network becomes sparser and degrades in quality.

3.4.2. Scenario 2: Partially Gaged Target Sites

In the second scenario, the performance of the six models is tested for partially gaged sites in a leave-one-out cross-validation framework. Here, we randomly select m years of data to be assumed known at the target site and assume the remaining 10- m years of data are missing and require estimation. Correlations to donors and the empirical FDC are estimated based on the m years of known data. All six models are then used to develop out-of-sample estimates for the remaining 10- m years. We consider two cases where m equals one and five.

To supplement the partially gaged assessment, predictions are also produced using two intermediary scenarios: (1) observed correlations (i.e., based on all 10 years) and estimated FDCs from the partial record and (2) estimated correlations from the partial record and 10-year empirical FDCs. These intermediary scenarios are used to better isolate how additional error from the estimated correlations and FDCs for the partially gaged target sites influence prediction skill.
3.4.3. Scenario 3: Completely Ungaged Target Sites

Model performance under leave-one-out cross-validation is also tested for ungaged sites. Here, we assume correlations to donors and the FDC are completely unknown at the target site. The correlations between sites are estimated using a spherical correlogram model fit to the available donor sites, as described in section 3.3. The FDC for each target site is estimated using a cross-validated prediction from a multioutput neural network (Worland et al., 2019a, 2019b). This model predicts 15 monotonically increasing quantiles in the output layer of a neural network based on a large set of basin characteristics (including characteristics on upstream dams), which are then interpolated to create a continuous FDC. The result of the ungaged scenario is an out-of-sample estimate of 10 years of streamflow at each site in both basins.

Similar to the partially gaged scenario, two intermediary scenarios are used to supplement the ungaged assessment and better isolate how additional error from the estimated correlations and FDCs influence prediction skill: (1) predictions based on observed (i.e., 10 years) correlations and estimated FDCs from the neural network and (2) estimated correlations from the correlogram and observed (i.e., 10 years) empirical FDCs.

3.5. Nash–Sutcliffe Efficiency Explanation and Decomposition

The Nash–Sutcliffe model efficiency (NSE) is utilized to measure prediction skill across models (Nash & Sutcliffe, 1970):

\[
\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (\hat{q}_i - \bar{q}_i)^2}{\sum_{i=1}^{n} (q_i - \bar{q}_i)^2} = 1 - \frac{\text{Mean Squared Error}}{\text{Var}(q)},
\]

where \(\hat{q}\) is the vector of estimated streamflow values, \(q\) is the vector of observed streamflow values, and \(\bar{q}\) is the mean of the observed streamflow values. A limitation of the NSE is that the squared error results in overestimation of model performance for high streamflow values and underestimation for low streamflow values (Krause et al., 2005). Taking the natural log of both the observed and estimated values prior to calculating the NSE provides one way to partially mitigate this issue:

\[
\log \text{NSE} = 1 - \frac{\sum_{i=1}^{n} (\log(\hat{q}_i) - \log(q_i))^2}{\sum_{i=1}^{n} (\log(q_i) - \log(q_i))^2}.
\]

An additional limitation is that, due to the normalization by the variance in equation (24), models must perform better in catchments with low variance to obtain similar NSE values as worse performing models in catchments with higher variance. Therefore, it is also helpful to decompose the mean squared error in the numerator of equation (24) into three components and examine each component individually:

\[
1 - \text{NSE} = \frac{(\mu_{\hat{q}} - \mu_q)^2}{\sigma_q^2} + \frac{(\sigma_{\hat{q}} - \sigma_q)^2}{\sigma_q^2} + \frac{2\sigma_{\hat{q}}\sigma_q(1 - \rho)}{\sigma_q^2},
\]

where \((\mu_{\hat{q}} - \mu_q)^2\) provides a measure of how well the mean of the estimated data matches the mean of the observed data, \((\sigma_{\hat{q}} - \sigma_q)^2\) is a measure of how well the variance of the estimated data matches the variance of the observed data, and \(2\sigma_{\hat{q}}\sigma_q(1 - \rho)\) provides an indication of the correlation between the estimated and observed streamflow and is commonly referred to as the “timing” component. All three terms are standardized by the variance of the observed streamflow.

4. Results

4.1. Bivariate Copula Model Selection

A pairwise analysis of sites in the MTB and GTB reveals a range of asymmetric tail dependence across the two basins (Figure 6a). To help illustrate the patterns of streamflow under these various tail dependencies, the pairs of sites in both the GTB and MTB with the highest \(\lambda_L\) and \(\lambda_U\) are identified (highlighted in Figure 6a) and are presented in terms of their joint distribution of nonexceedance probabilities (Figure 6b) and raw streamflow (Figure 6c). In the GTB, several site-pairs show strong upper tail dependence, but very few show lower-tail dependence, indicating that streamflow variability tends to be more spatially coherent in the GTB under high-flow conditions. Conversely, site-pairs in the MTB exhibit upper- and lower-tail dependence in roughly equal proportion. However, these tail dependencies tend to vary across space, that is, many site-pairs exhibit either upper- or lower-tail dependence but not both.
Figure 4. Density of normalized, pairwise negative Akaike Information Criterion (AIC) across sites for five different bivariate copula models fit using donors that have the highest pairwise correlation coefficient to each target site. The AIC is normalized by adjusting the AIC values for each pair of sites so that the best AIC value is equal to zero.

The different types of tail dependence present in both basins indicate that multiple copulas could be considered for modeling the joint distribution of streamflow. Figure 4 shows the distribution of normalized AIC values across all sites and different bivariate copula models fit to sites with the highest pairwise correlation coefficients. The Clayton copula performs poorly for both basins, but particularly in the GTB, which shows almost no lower-tail dependence between sites (Figure 6). Due to its poor performance, the Clayton copula is not used for further analysis. The normal copula assumes zero tail dependence between pairs and performs well in both basins, likely because there are many site-pairs that are correlated but exhibit weak tail dependence (Figures 4 and 6). The Frank, Student's $t$, and Gumbel copulas all result in the best AIC values for both basins. The Gumbel model supports upper-tail dependence, which is common across the basins (Figure 6), while the Student's $t$ copula supports both upper- and lower-tail dependence. The Frank copula does not support tail dependence but provides a wide and uniform correlation structure throughout the joint distribution (see Figure 2). These features can improve mean prediction of target site flows and can lead to better uncertainty characterization, both of which can improve AIC values. The $p$ values resulting from the goodness-of-fit tests (Genest et al., 2009) suggest that all of the copulas considered should be rejected as potential models for the majority of site-pairs. However, when dealing with large samples (i.e., 10 years of daily data), small discrepancies between the data and proposed distributions often result in a statistically significant lack of fit, even if the departure from the specified distribution is small (Johnson & Wichern, 1992). Therefore, we look at the distribution of test statistics, which suggests that the Gumbel copula is the best model for the GTB, while the Student's $t$ or normal copula provide the best fit for the MTB (see supporting information, Figure S4).

Figure 5 more clearly shows how these models characterize prediction uncertainty. Figure 5a shows coverage probabilities for a 95% CI across different ranges of flow, expressed as nonexceedance probability ranges, and aggregated across all target sites. Ideally, these CIs would contain 95% of the observed values for the entire data set, regardless of flow range. The coverage probabilities for the Frank, Gumbel, Normal, and Student's $t$ copula range from 90.5–93.5% when considering all the data without separation by flow magnitude. However, the coverage probabilities vary appreciably depending on the location along the CDF and the basin. For example, in the MTB, CIs under the Frank copula capture almost 90% of the lowest and highest streamflow but are somewhat under-dispersed around the median streamflow relative to the other models. The Student's $t$ and normal copula result in very similar CIs for both basins, and generally achieve 95% coverage except for the largest and smallest flows. The Gumbel copula returns CIs that perform the worst for high streamflow, likely because the Gumbel assumes upper-tail dependence and is less dispersed in its conditional estimates for high streamflow values. This difference is apparent when examining the 95% CI for a particular target site (Figure 5b). For streamflow values over 30 $m^3s^{-1}$, the Gumbel results in estimates...
Figure 5. (a) Coverage probabilities from bivariate normal, Student’s $t$, Gumbel, and Frank copulas for both basins across nonexceedance probabilities. The nonexceedance probabilities were separated into 20 bins that contained the same number of streamflow values (~6,700) and the $x$ axis is the mean nonexceedance probability for each bin. The black dashed line indicates 95%, which is the expected coverage probability across all nonexceedance probabilities if the data-generating process is accurately captured by the chosen model. B. 95% CIs for each copula for an example time period (1 June 2008 to 31 December 2008) for a particular site in the GTB (USGS site number 08075400). The site was chosen because each method has roughly the same NSE, but the coverage probabilities vary depending on the properties of each copula.

with a much smaller conditional variance compared to the other copulas, especially the Frank. Importantly though, there is tradeoff between the size of the CIs and the ability to consistently capture the true value. A CI that is too large may contain the observed values but is not useful in practice because of the wide range of estimated values.

Based on relations shown in Figures 4, 5, 6, and S4, we select the Gumbel copula as an appropriate bivariate copula model in the GTB, because upper-tail dependence is relatively common in that basin and this model has the best overall performance (as estimated by the AIC and goodness-of-fit test statistic). In the MTB, we select the normal copula as an appropriate bivariate model, given its adequate performance and the inconsistency in tail dependence across site-pairs. We also could have selected the Student’s $t$ copula, which exhibited the best AIC and goodness-of-fit statistics, but the Gaussian and Student’s $t$ copulas produce the same expected value estimate and the Student’s $t$ copula requires an additional estimated parameter ($\nu$). These models are used as the bivariate copula methods in the experiments below. However, we note that if uncertainty estimates of target site flows were of high interest, the Student’s $t$ or Frank bivariate copulas would likely have been selected.

4.2. Model Intercomparison of Predictive Skill
4.2.1. Streamflow Estimation for Fully Gaged Target Sites

We first examine the log-NSE values associated with the six different models built assuming the target sites are fully gaged, that is, using known correlations and empirical FDCs for each target site based on the full 10-year record (Figure 7). Under this scenario, the multivariate normal copula results in the highest log-NSE values for both basins compared to the other regionalization methods. The bivariate copula approach that uses the most correlated donor site to the target provides the second best performance, while a direct application of QPPQ with the most correlated donor is the next best model. Notably, methods that utilize the general QPPQ framework perform better than those that do not, given known correlation and FDC structure at the target sites.
Figure 6. (a) Pairwise upper and lower tail dependence (parameters $\lambda$ : $\lambda_L$ or $\lambda_U$) for site-pairs in the Galveston-Trinity (left) and the Mobile-Tombigbee (right) basins. The upper triangular component of the matrix contains $\lambda$ values for upper-tail dependence and the lower triangular component of the matrix contains $\lambda$ values for the lower-tail dependence. (b) Scatter plots of nonexceedance probabilities for select sites, which are indicated by red and green site numbers and black boxes in a, that either exhibit upper-tail dependence or lower-tail dependence. (c) Example time period (1 April 2000–1 November 2000) of streamflow for the sites described in b.

The copula models also outperform other approaches when the information in the donor network is thinned and degraded (Figure 8). The six methods result in median log-NSE values between 0.63–0.92 in the MTB and 0.39–0.86 in the GTB when all of the available donors are available in the network (far left of Figure 8). However, the log-NSE values of the noncopula models decrease at a much faster rate than the copula models when correlated donors are removed. For example, when the maximum pairwise correlation coefficient is around 0.4, the noncopula models median log-NSE is −0.14 (MTB) and −0.227 (GTB), whereas the log-NSE for the mutlivar-norm-cop is 0.194 (MTB) and 0.33 (GTB) and the bivar-cop-rho is 0.18 (MTB) and 0.07

Figure 7. Density of log Nash-Sutcliffe efficiency for six different streamflow estimation methods. If a method relied on correlations or FDCs, the known correlations between sites and empirical flow-duration curves at the target site were selected. For the “bivar_cop_rho” models, the Gumbel bivariate copula was selected in the Galveston-Trinity River Basin and the normal bivariate copula was selected in the Mobile-Tombigbee River Basin.
Figure 8. Performance of the six models with different levels of correlations to donor sites. The correlations on the x axis are binned by the mean maximum correlation in the set. For example, on the far left of the plot, the models use all of the sites, so the x axis simply indicates the highest pairwise correlation for each region. Moving to the second column on the far left, the pairwise correlations are calculated between each target, and all possible remaining donors after the donor site with the highest correlation to each region is dropped. This continues until every model uses only one site that has the highest correlation to the target. For the “bivar_cop_rho” models, the Gumbel bivariate copula was selected in the Galveston–Trinity River Basin, and the normal bivariate copula was selected in the Mobile–Tombigbee River Basin. (GTB). Although log-NSE values between 0.07–0.33 are not indicative of a good model, they do reveal that copula models provide a hedge in near worst-case scenarios. This is most evident in the limit of the experiment, when only one donor site remains that is highly uncorrelated to the target (far right of Figure 8). In this scenario, the conditional expectation of the copula models collapses to approximately the median of the target because the correlation-weighted transfer between the donor and the target is negligible. The other models do not account for this degradation of information in their predictive schemes.

4.2.2. Streamflow Estimation for Partially Gaged Target Sites

This section focuses on predictive skill for partially gaged target sites. Figure 9 shows the NSE and log-NSE for the six models where both the FDC and the correlations are estimated based on the available data in the partial record (either 1 or 5 retained years, labeled “partial 1 year” and “partial 5 year,” respectively). The results from the fully gaged assessment in 4.2.1 (Figure 7) are also shown for comparison (labeled “fully gaged”). Other intermediary scenarios for the partially gaged sites where either correlations or FDCs were based on the full 10-year record are shown in the supporting information (Figures S5 and S6).

When 5 years of data are available, predictions for the out-of-sample 5 years are quite promising for the multivar-norm-cop model in both basins. The degradation in performance that does occur is primarily caused by the estimated correlation values based on a shorter record (Figure S5). However, when the partial record is only 1 year long, the copula-based methods exhibit equal or worse performance compared to the noncopula based approaches in both basins. Sampling variability in the correlation estimates from the short record are the largest cause of this degradation, although the estimated FDC based on 1 year of data also leads to similar performance loss (Figure S6). We note that correlation sampling variability is quite high when only 1–3 years of data are available in the partial record, but becomes relatively negligible after the partial record grows to more than 7 years (Figure S7). We also note that predictions for partially gaged sites do not improve substantially if multioutput neural network FDC estimates are selected instead of the empirical estimates (not shown).
4.2.3. Streamflow Estimation for Completely Ungaged Target Sites

Figure 9 also shows the NSE and log-NSE for completely ungaged target sites, where the FDC and the correlations are estimated based on the neural network model and correlogram, respectively (labeled “ungaged”). The figure also shows intermediary scenarios where either the FDC or the correlations are estimated without observed streamflow data at the target site, but the other is based on 10 years of observed target site flows (labeled “ungaged (obs corr)” and “ungaged (obs FDC)”, respectively).

The results show that if the correlations between every site in the network are relatively high (i.e., $|\rho| \gg 0$), then the relative performance of the six models is fairly consistent when the estimated information is used in lieu of the observed information. This can be seen in the MTB (minimum $|\rho| = 0.22$), where the multivar-norm-cop outperforms the other methods for both metrics and across scenarios (right panel, Figure 9). The only exception is for the raw NSE and the “ungaged” scenario, where the multivar-norm-cop (median NSE of 0.59) performs similarly to the IDW-log-runoff (median NSE of 0.63). It is also noteworthy that in the MTB, estimated correlations cause the most degradation for the log-NSE, while estimated FDCs lead to a larger loss of performance when measured by the NSE. In the GTB, the minimum $|\rho|$ equals 0.006, and estimated correlations and FDCs cause significant degradation in the multivar-norm-cop, bivar-cop-rho, and QPPQ-highest-rho models that are clearly superior when built using observed information (left panel, Figure 9). Much of the error in the copula models is introduced by the estimated correlations compared to the estimated FDCs. Notably, the IDW-log-runoff and QPPQ-IDW approaches result in NSE values that are equal to or better than the multivar-norm-cop in the “ungaged” scenario for both log-NSE and raw NSE in the GTB.

4.3. NSE Decomposition and Uncertainty

Figure 10 shows the decomposition of the NSE and log-NSE into their component parts (bias, variance, and timing errors) for the completely ungaged and fully gaged scenarios in both basins. When considering the log-NSE, prediction errors for all models and both basins are primarily due to timing errors. For models based on correlation (multivar-norm-cop, bivar-cop-rho, and QPPQ-highest-rho), these timing errors grow substantially when using estimated (i.e., unaged) instead of observed (i.e., fully gaged) correlations. The error components for the multivar-norm-cop and the QPPQ-IDW methods in the MTB and the unaged scenario
Figure 10. Decomposition of the log-Nash-Sutcliffe efficiency (a) and NSE (b) into bias, variance, and timing components for both basins and the completely “ungaged” and “fully gaged” scenarios. For the “bivar_cop_rho” models, the Gumbel bivariate copula was selected in the Gavleston-Trinity River Basin and the normal bivariate copula was selected in the Mobile-Tombigbee River Basin.

are almost identical, indicating that for low-to-medium flows, there is not much added benefit when making the calculations in \( z \) space (multivar-norm-cop) versus nonexceedance probability space (QPPQ-IDW) when correlations are estimated. Models based directly on runoff (IDW-log-runoff, rho-weight-log-runoff) also exhibit some degree of mean bias, which grows in these models and emerges in the other models when using estimated FDCs.

For the raw NSE, timing errors still dominate the error metric and grow substantially when using estimated instead of observed correlation values, particularly in the GTB. However, variance errors also play a large role in reduced performance, particularly for QPPQ-IDW, IDW-log-runoff, and rho-weight-log-runoff. This suggests that the range of raw streamflow values, including the largest values, are more likely to be captured if nonexceedance probabilities are selected directly from the most correlated site (QPPQ-highest-rho) or are transferred to the target site using a copula (bivar-cop-rho, multivar-norm-cop).

Finally, Figure 11 shows the out-of-sample coverage probabilities for a 95% CI using the multivar-norm-cop model under the completely ungaged and fully gaged scenarios for both basins. Similar to before, coverage probabilities are shown across different ranges of flow, expressed as nonexceedance probability ranges, and

Figure 11. Coverage probabilities of the multivariate normal copula for both basins and the completely “ungaged” and “fully gaged” scenarios.
aggregated across all target sites. When correlations and FDCs are known, the coverage probabilities for the multivar-norm-cop model perform reasonably well, achieving near 95% coverage except for the largest and especially the smallest flows. The coverage errors for the largest flows actually decline slightly when the correlations and FDCs are estimated, but low-flow coverage gets markedly worse, particularly in the MTB. Overall, when examining the data in aggregate without distinguishing by flow range, the coverage probabilities with known correlations and FDCs are 0.93 (MTB) and 0.94 (GTB), which change to to 0.88 (MTB) and 0.94 (GTB) when correlations and FDCs are estimated.

5. Discussion

5.1. Diagnosing and Selecting Copula Models
The range of tail dependencies seen in Figure 6 indicates that the mechanisms dominating the high- and low-flow regimes vary across the two basins. In both basins, large $\lambda_U$ values indicate that high flows are often responding to similar storm events. Given the general proximity of the sites, this is unsurprising. The lack of lower-tail dependence in the GTB reveals that the processes governing low streamflow for each site, such as the baseflow contribution or upstream regulation, are disconnected. For example, USGS site numbers 08066500 (Trinity River at Romayor, TX) and 08066300 (Menard Creek near Rye, TX) are only 9.14 km apart and share almost exactly the same physiographic characteristics (Figure 3). Yet the $\lambda_L$ between the sites is $\sim$0, indicating no correlation for extremely low streamflow values. In many cases, this disconnect can be most easily explained by upstream reservoir operations. Site 08066500 is 50 km downstream from the Lake Livingston dam, a water supply dam operated by the Trinity River Authority, whereas 08066300 is located in a mostly unregulated basin. The operating decisions for the Lake Livingston dam are dictated by priorities that do not coincide with natural low-flow conditions, resulting in release polices that alter the timing and volumes of low flows downstream that differ greatly from those of a natural flowing river. Conversely, many sites in the MTB show a strong lower-tail dependence. This is likely because low flows at a majority of sites in the MTB are maintained by groundwater (as indicated by high baseflow index values), and additionally, there is less regulation in the MTB relative to the GTB.

While both basins exhibit a tendency toward certain patterns of tail dependence, there is a substantial amount of variability in the dependence across site-pairs. This complicates the choice of a single bivariate copula model. Arguably, sites need to be grouped into more homogeneous regions that limit the range of variability in tail dependence. However, upstream regulation across sites can vary substantially, even for sites in close proximity, and distance may not be a sufficient criterion to determine homogeneous regions. This can add a significant barrier to the choice of copula model, particularly for ungaged sites. Thus, an important avenue for future work is to determine the controls (e.g., catchment characteristics or dominant climate processes Szolgay et al., 2015) over tail dependence between sites, so that appropriate copula models can be chosen for a target site in the absence of streamflow data to verify this choice. Importantly, any regionalization of tail dependence would have to contend with a high degree of sampling variability in the associated estimates.

5.2. Benefits and Limitations to Prediction Using Copulas
When the correlation and FDC structure at the target sites are known, the multivariate Gaussian copula significantly outperforms all other methods tested in this study (Figures 7 and 9). This indicates that this approach efficiently leverages available information in its prediction scheme compared to many other methods available in the literature. On average, the bivariate copula approaches do not perform significantly better than a direct QPPQ approach based on the most correlated site. The Archimedean copulas tested in this work are conventionally used as bivariate copulas, limiting their application to a single donor. However, vine copulas (Pereira et al., 2017; Vernieuwe et al., 2015) can be used to extend the application of bivariate Archimedean copulas to multiple donors. Given the asymmetric tail dependence shown across the two case study basins and the potential improvements demonstrated with a multivariate Gaussian copula, this presents a promising avenue for future research. However, given the number of parameters that require estimation, this may be more feasible in the case of partially gaged sites. We also note that both the multivariate and bivariate copulas do provide a strong hedge against poor performance as the correlation structure in the donor network is thinned and degraded (Figure 8). This property could have particularly important benefits when working in data-scarce regions, especially when performing gap-filling or record extension on partially gaged sites where correlation and FDC estimates will tend to be more accurate. However, for some
degree of information loss from donor sites, process models will likely provide equal or better streamflow estimates than statistical methods.

For ungaged catchments, the performance of the copula models declines substantially once they are based on estimated rather than observed information, and no longer provide consistent improvements over other, simpler methods (IDW-log-runoff, QPPQ-IDW). Much of this performance loss is linked to the estimated correlations, rather than the FDCs. This is unsurprising considering that the modeled FDCs were the result of an extensive study (Worland et al., 2019a, 2019b), whereas the correlations were estimated using a simple kriging model. This result suggests that considerable care needs to be given to the estimate of correlations across the gaging network if the copula-based approach is going to provide sufficient benefits for prediction in ungaged catchments to justify its use over simpler methods. Similar to the argument in section 5.1, the selection of homogeneous regions could play an important role in the improvement of correlation estimates. For instance, while the sites in the MTB are generally clustered in the northern portion of that basin and have a minimum $|\rho|$ of 0.22, the sites within the GTB have a minimum $|\rho|$ of 0.006. These sites are spatially clustered in two distinct regions that are separated by the 31.5° N latitude line (see Figure 3). If the correlation matrix is independently calculated for only the sites above this latitude and the sites below it, the minimum $|\rho|$ increases to $\sim 0.1$. Beyond the selection of homogeneous regions, improved correlogram models could also be considered that utilize information beyond geographic distance, such as physiographic distance (e.g., accounting for differences between regulated and unregulated sites). This exploration is left for future work.

For gap filling of missing data or record extension in partially gaged catchments, the copula models show substantial promise, especially if there are sufficient data at the partially gaged site to estimate FDCs and correlations with reasonable accuracy. This result is consistent with recent work on this topic (Aissia et al., 2017). However, if the record at the partially gaged site is too short, direct empirical estimates of both FDCs and correlations to donors can suffer from a substantial amount of sampling uncertainty. This uncertainty can substantively degrade copula-based estimates. However, alternative approaches are possible to help reduce sampling uncertainty in both FDC and correlation estimates. For instance, FDC estimates from regression-based models (like the multioutput neural network used in this work), the empirical FDC at the partially gaged site, and even an FDC estimated from a process based model could be statistically combined to make the final estimate more robust. A similar approach could be taken for correlation estimates from the correlogram model and the empirical correlation between the partially gaged site and donor sites. These approaches should be tested in future work. Further, we limited our assessment to 10 years of data across the gaging network, and it is also likely that many of these sampling uncertainties (including those for tail dependence) would be resolved if longer records were used to support the estimation of all terms. However, potential nonstationarity in FDCs and correlations would have to be considered due to the influences of land-use and climate changes.

In addition to mean prediction skill, results suggest that copulas provide a statistically sound and straightforward method for calculating uncertainty at target sites. In most applications considered here, 95% CIs based on fitted copula models provided adequate coverage of the observations, except in the tails of the distribution (Figures 5 and 11). It is possible that some of these tail uncertainty estimates could be improved if uncertainty from other sources was propagated forward, for instance uncertainty in the FDC estimate (see Worland et al., 2019a, 2019b) or even measurement error at the donor sites. In addition, when the NSE values were decomposed for each model, copula-based approaches tended to have less variance error in the NSE (Figure 10). Therefore, the copula methods provide an important advantage in recreating the full range of flows at the target site, which might be particularly relevant for certain applications (e.g., ecohydrologic studies). Further work could examine the ability of these approaches to preserve key ecohydrologic signatures (Poff et al., 2010).

6. Conclusion

This study contributes a formal statistical framework based in copula theory to improve traditional QPPQ approaches for transferring streamflow from donor to target sites. Several benefits are possible in this framework, including the ability to account for asymmetric tail dependence between sites, a formal statistical approach for estimating uncertainty, and the ability to weight the contributions of multiple donor sites based on their correlation coefficient—even with differing marginal distributions of the contributing sites. The
benefits of a copula-based approach can be confounded by errors when correlations and FDCs at the target site are estimated. This presents the largest obstacle to the application of this approach for streamflow estimation in ungaged sites or partially gaged sites with a very limited record. When the correlations and FDCs are calculated from longer (>5 year) partial records, then the copula methods may prove to be far superior to other methods for purposes like gap filling or extending missing streamflow records. Overall, the utility of copula approaches, combined with their straightforward implementation, makes them a useful extension of QPPQ worthy of additional research.

Acknowledgments
The data release and model archive associated with this manuscript are available online at Crowley-Ornelas et al., 2019b, (2019b, 2019a); Worland et al. (2019c). Funding for this research was provided by the Gulf Coast Ecosystem Restoration Council. We would like to thank three anonymous reviewers for their constructive feedback, which served to considerably improve this work. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References
Aas, K. (2004). Modelling the dependence structure of financial assets: A survey of four copulas. *Samba, 22*(4), 18.
Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics, 44*(2), 182–198.
Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions: With formulas, graphs, and mathematical Tables* (Vol. 55). Courier Corporation.
Alessa, M.-A. B., Chebana, F., & Ouarda, T. B. (2017). Multivariate missing data in hydrology–review and applications. *Advances in Water Resources, 110*, 299–309.
Archfield, S. A., & Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research, 46*, W10513. https://doi.org/10.1029/2009WR008481
Arsenault, R., & Brisette, F. P. (2014). Continuous streamflow prediction in ungaged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research, 50*, 6135–6153. https://doi.org/10.1002/2013WR014898
Arsenault, R., & Brisette, F. (2016). Multi-model averaging for continuous streamflow prediction in ungaged basins. *Hydrological Sciences Journal, 61*(13), 2443–2454.
Asquith, W. (2019). copBasic—general bivariate copula theory and many utility functions. R package version 2.1.2.
Atieh, M., Taylor, G., Sattar, A. M., & Gharabaghi, B. (2017). Prediction of flow duration curves for ungaged sites. *Journal of Hydrology, 545*, 383–394.
Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research, 52*, 3599–3622. https://doi.org/10.1002/2015WR018247
Bisselinik, B., Zambrano-Bigiarini, M., Burek, P., & De Roo, A. (2016). Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. *Journal of Hydrology: Regional Studies, 8*, 112–129.
Bölschli, G. (2013). *Runoff prediction in ungaged basins: Synthesis across processes, places and scales*. New York City, NY: Cambridge University Press.
Bölschli, G. (2016). Predictions in ungaged basins—Where do we stand? *Proceedings of the International Association of Hydrological Sciences, 373*, 57–60.
Blum, A. G., Archfield, S. A., & Vogel, R. M. (2017). On the probability distribution of daily streamflow in the United States. *Hydrology and Earth System Sciences, 21*(6), 3093–3103.
Booker, D., & Snelder, T. (2012). Comparing methods for estimating flow duration curves at ungaged sites. *Journal of Hydrology, 434*, 78–94.
Castellarin, A., Persiano, S., Pugliese, A., Aloe, A., Skoien, J., & Pistocchi, A. (2018). Prediction of streamflow regimes over large geographical areas: Interpolated flow–duration curves for the Danube region. *Hydrological Sciences Journal, 63*(6), 845–861.
Chen, L., & Guo, S. (2019). Copulas and its application in hydrology and water resources. Singapore: Springer.
Crowley-Ornelas, E., Asquith, W. H., Worland, S. C., & Knight, R. R. (2019b). Summary of streamflow statistics for USGS streamgages in the southeastern United States: 1950 - 2010. In *U.S. Geological Survey Data Release. https://doi.org/10.5066/P91DOYCI*
Crowley-Ornelas, E., Worland, S. C., Wiecezorck, M., Asquith, W. H., & Knight, R. R. (2019a). Summary of basin characteristics for National Hydrography Dataset, version 2 catchments in the southeastern United States. In *1950 – 2010 U.S. Geological Survey Data Release. https://doi.org/10.5066/P9KXTDU4*
Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review, 73*(1), 111–129.
Farmer, W. (2015). *Estimating records of daily streamflow at ungaged locations in the southeast United States PhD dissertation*. USA: Tufts University MA.
Farmer, W. H., Archfield, S. A., Over, T. M., Hay, L. E., LaFontaine, J. H., & Kiang, J. E. (2015b). A comparison of methods to predict historical daily streamflow time series in the southeastern United States: U.S. Geological Survey Scientific Investigations Report. https://doi.org/10.3133/sir20145231
Farmer, W. H., Knight, R. R., Eash, D. A., Hutchinson, K. J., Linhart, S. M., Christiansen, D. E., et al. (2015a). Evaluation of statistical and rainfall-runoff models for predicting historical daily streamflow time series in the Des Moines and Iowa River watersheds: US Geological Survey Scientific Investigations Report. https://doi.org/10.3133/sir20155089
Farmer, W. H., & Levin, S. (2018). Characterizing uncertainty in daily streamflow estimates at ungaged locations for the Massachusetts sustainable yield estimator. *JAWRA Journal of the American Water Resources Association, 54*(1), 198–210.
Fennessey, N. (1994). *A hydro-climatological model of daily streamflow for the northeast United States (Ph.D. thesis)*, Medford MA Tufts University.
Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics, 44*(2), 199–213.
Genest, C., & Ríovest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American statistical Association, 88*(423), 1034–1043.
Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2017). copula: Multivariate dependence with copulas. r package version 0.999-18.
Hrachowitz, M., Savenije, H., Bölschli, G., McDonnell, J., Sivapalan, M., Pomeroy, J., et al. (2013). A decade of predictions in ungaged basins (PUB)—A review. *Hydrological Sciences Journal, 58*(6), 1198–1255.
Hughes, D., & Smakhtin, V. (1996). Daily flow time series patching or extension: A spatial interpolation approach based on flow duration curves. *Hydrological Sciences Journal, 41*(6), 851–871.
Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (3rd ed.). New Jersey: Prentice-Hall.

Käärik, M., Selart, A., Käärik, E., & Livi, J. (2011). The use of copulas to model conditional expectation for multivariate data, in *ISI Proc. 58th World Statistical Congress* , (pp. 5533–5538): International Statistical Institute.

Krause, P., Boyle, D., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97.

Lorenz, D. L., & Ziegeweid, J. R. (2016). Methods to estimate historical daily streamflow for ungauged stream locations in Minnesota: U.S. Geological Survey Scientific Investigations Report. https://doi.org/10.3133/sir20155181

Mohamoud, Y. M. (2008). Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. *Hydrological Sciences Journal*, 53(4), 706–724.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I–A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.

National Academies of Sciences, Engineering, and Medicine (2018). *Future water priorities for the nation: Directions for the U.S. Geological Survey Water Mission Area*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25134

Oliver, M. A. (2010). The variogram and kriging, in *Handbook of applied spatial analysis* pp. 319–352: Springer.

Pereira, G. A., Veiga, Á., Erhardt, T., & Czado, C. (2017). A periodic spatial vine copula model for multi-site streamflow simulation. *Electric Power Systems Research*, 152, 9–17.

Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., et al. (2010). The ecological limits of hydrologic alteration (ELoHA): A new framework for developing regional environmental flow standards. *Freshwater Biology*, 55(1), 147–170.

Pugliese, A., Castellarin, A., & Brath, A. (2014). Geostatistical prediction of flow–duration curves in an index–flow framework. *Hydrology and Earth System Sciences*, 18(9), 3801–3816.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Razavi, T., & Coulibaly, P. (2012). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering*, 18(8), 958–975.

Ribeiro Jr, P. J., & Diggle, P. J. (2018). *geor: Analysis of geostatistical data*. r package version 1.7-5.2.1.

Rousseeuw, P. J., & Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics–Theory and Methods*, 22(4), 965–984.

Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40, W12511. https://doi.org/10.1029/2004WR003133

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., & Erhardt, T. (2018). *VineCopula: Statistical inference of vine copulas*. r package version 2.1.8.

Schmid, F., & Schmidt, R. (2007). Multivariate conditional versions of Spearman’s rho and related measures of tail dependence. *Journal of Multivariate Analysis*, 98(6), 1120–1140.

Schmier, S., & Cai, X. (2014). Prediction of regional streamflow frequency using model tree ensembles. *Journal of Hydrology*, 517, 298–309.

Smakhtin, V. Y. (1999). Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique. *Regulated Rivers: Research & Management: An International Journal Devoted to River Research and Management*, 15(4), 311–323.

Steinschneider, S., Yang, Y.-C. E., & Brown, C. (2015). Combining regression and spatial proximity for catchment model regionalization: A comparative study. *Hydrological Sciences Journal*, 60(6), 1026–1043.

Szolgay, J., Gaal, L., Kohnova, S., Hlavcova, K., Vyleta, R., Bacigal, T., & Blöschl, G. (2015). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97.

Vernieuwe, H., Vandenberghe, S., De Baets, B., & Verhoest, N. (2015). A continuous rainfall model based on vine copulas. *Advances in Geosciences*, 5066/F7P55KJN.

Vöröskövesdi, M. (2015). A comparative study of copula types for peak-volume flood relationships, in *Proceedings of the International Association of Hydrological Sciences (IAHS)*. https://doi.org/10.5194/piase-370-183-2015

Worland, S. C., Knight, R. R., & Asquith, W. H. (2019a). Estimated quantiles for the pour points of 9,203 level-12 hydrologic unit codes in the southeastern United States. *U.S. Geological Survey Data Release*. https://doi.org/10.5066/F9P55KJN

Worland, S. C., Knight, R. R., & Asquith, W. H. (2019b). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97.

Worland, S. C., Knight, R. R., & Asquith, W. H. (2019c). Observed and modeled daily streamflow values for 74 U.S. Geological Survey Data Release. https://doi.org/10.5066/P9YGKZZY

Worland, S. C., Knight, R. R., & Asquith, W. H. (2019d). Observed and modeled daily streamflow values for 74 U.S. Geological Survey Data Release. https://doi.org/10.5066/P92F1ROU

Worland, S. C., Steinschneider, S., Asquith, W. K., Knight, R., & Wieczorek, M. (2019b). Prediction and inference of flow-duration curves using multi-output neural networks. *Water Resources Research*, 55, 6850–6868. https://doi.org/10.1029/2019WR024463

Zhang, Y., Vaze, J., Chiew, F. H., & Li, M. (2015). Comparing flow duration curve and rainfall–runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology*, 525, 72–86.