TEST
https://doi.org/10.1007/s11749-023-00895-6

ORIGINAL PAPER

Bayesian joint quantile autoregression

Jorge Castillo-Mateo[1] · **Alan E. Gelfand**[2] · **Jesús Asín**[1] · **Ana C. Cebrián**[1] · **Jesús Abaurrea**[1]

Received: 7 May 2023 / Accepted: 6 October 2023

Abstract
Quantile regression continues to increase in usage, providing a useful alternative to customary mean regression. Primary implementation takes the form of so-called *multiple* quantile regression, creating a separate regression for each quantile of interest. However, recently, advances have been made in *joint* quantile regression, supplying a quantile function which avoids crossing of the regression across quantiles. Here, we turn to quantile autoregression (QAR), offering a fully Bayesian version. We extend the initial quantile regression work of Koenker and Xiao (J Am Stat Assoc 101(475):980–990, 2006. https://doi.org/10.1198/016214506000000672) in the spirit of Tokdar and Kadane (Bayesian Anal 7(1):51–72, 2012. https://doi.org/10.1214/12-BA702). We offer a directly interpretable parametric model specification for QAR. Further, we offer a pth-order QAR(p) version, a multivariate QAR(1) version, and a spatial QAR(1) version. We illustrate with simulation as well as a temperature dataset collected in Aragón, Spain.

Keywords Copula model · Gaussian process · Joint quantile model · Markov chain Monte Carlo · Spatial quantile autoregression

✉ Jorge Castillo-Mateo
jorgecm@unizar.es

Alan E. Gelfand
alan@stat.duke.edu

Jesús Asín
jasin@unizar.es

Ana C. Cebrián
acebrian@unizar.es

Jesús Abaurrea
abaurrea@unizar.es

[1] Department of Statistical Methods, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain

[2] Department of Statistical Science, Duke University, 415 Chapel Dr, Durham, NC 27705, USA

Published online: 12 November 2023

Mathematics Subject Classification 62F15 · 62G08 · 62H05 · 62M10 · 62M30

1 Introduction

For time series data, autoregressive (AR) modeling is perhaps the most common approach. A lag one, AR(1), takes the form $Y_t = \mu + \rho(Y_{t-1} - \mu) + \epsilon_t$, with ϵ_t following a suitable zero-mean error distribution; a conditional mean is provided. By analogy, quantile autoregression (QAR) considers conditional quantiles.

An issue with quantile regression (QR) is the so-called quantile crossing problem. Modeling quantiles individually enables rich modeling for a given quantile but allows for crossing of quantiles across quantile level τ. For arbitrary values of the regressors, $\mathbf{X}$, we can not ensure that the resulting modeled quantiles will increase in τ. Such modeling is referred to as *multiple* QR. Inference typically proceeds by minimizing a check loss function or, more formally, assuming an asymmetric Laplace (AL) error term. Examples of multiple QR with AL errors appear in Yu and Moyeed (2001), and Kozumi and Kobayashi (2011) present a Gibbs sampler model fitting implementation. Following those ideas, Peng et al. (2023) deals with variable selection in the context of QAR models with AL errors. Lum and Gelfand (2012) work in the context of spatially referenced data and extend the AL model to a spatial process. Castillo-Mateo et al. (2023) propose a very flexible spatial AL mixed effects QAR model.

Recent effort has focused on a *joint* QR modeling to avoid quantile crossing. Adopting restricted support for the regressors, $\mathbf{X}$, the τ-quantile will increase monotonically over $\tau \in (0, 1)$. Bondell et al. (2010) offer a non-crossing approach for a fixed set of quantiles of interest. Foundational work appears in Tokdar and Kadane (2012) using Gaussian processes (GPs) with follow on work in Das and Ghosal (2017) using splines. Reich et al. (2011) developed a spatial joint QR model through spatially varying regression coefficients using Bernstein polynomials. Yang and Tokdar (2017) propose a novel parameterization that characterizes any collection of non-crossing quantile planes over arbitrarily shaped convex predictor domains. This parameterization was extended to spatial data by Chen and Tokdar (2021) through a copula process but a non-spatially varying quantile function results. Joint modeling imposes strong restrictions on the class of permissible specifications; models outside of this class may be preferred.

Motivation for joint or non-crossing quantile modeling appears in, e.g., Bondell et al. (2010) who highlight a problem that appears when modeling a wind speed dataset given climatological regressors; the upper quantiles cross not far from the mean. As a further example in this regard, accurate quantile predictions across quantile levels are essential in forecasting of wind power generation (Cui et al. 2023). Formal joint modeling is necessary in applications where coherent estimates of several quantiles or a generative model are of interest; otherwise, quantile crossing leads to an invalid distribution for the response. Also, multiple modeling fails to do justice to the full potential of the model. Joint modeling helps to avoid the lack of data which emerges when attempting to fit individual quantile curves (Tokdar and Kadane 2012). Joint modeling in the context of QAR is relevant in risk management for estimating value-at-risk, as well as in demand forecasting, where understanding the complete demand distribution

is crucial for effective production planning and supply chain management. Further, these models for daily temperatures could improve operational prediction accuracy with forecast intervals, and high quantile simulated series behavior (Thrasher et al. 2012).

Koenker and Xiao (2006) offered an initial version of a joint QAR(p) model. Illustrating with $p = 1$, they consider the generative model

$$Y_t = \theta_0(U_t) + \theta_1(U_t)Y_{t-1}, \tag{1}$$

where U_t is a sequence of independent and identically distributed (i.i.d.) standard uniform random variables. The θ functions, from $[0, 1] \to \mathbb{R}$, need to be estimated. Provided that the right side of expression (1) is monotone increasing in U_t, the τ conditional quantile function of Y_t given y_{t-1} increases in τ and is

$$Q_{Y_t}(\tau \mid y_{t-1}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1}. \tag{2}$$

Koenker and Xiao (2006) required both θ_0 and θ_1 to be strictly increasing functions (referred to as co-monotonicity). Their suggested choices were $\theta_0(\tau) = \sigma\Phi^{-1}(\tau)$ with Φ the cumulative distribution function (cdf) of a standard normal distribution and $\theta_1(\tau) = \min\{\gamma_0 + \gamma_1\tau, 1\}$ for $\gamma_0 \in (0, 1)$ and $\gamma_1 > 0$. If $y_{t-1} \geq 0$, co-monotonicity ensures that $Q_{Y_t}(\tau \mid y_{t-1})$ will not cross as τ increases but under the restrictive assumption that the autoregression coefficient strictly increases in τ.

Our contribution is to reconsider the work of Koenker and Xiao (2006) in the context of Tokdar and Kadane (2012), providing flexible joint QAR modeling in a Bayesian framework. We characterize non-crossing QAR(1) also using two monotone curves, through a convenient class of cdf's. We note extension to the QAR(p) model. We consider bivariate QAR, capturing dependence through a copula process. Then, for spatially referenced time series, we introduce spatial dependence in the realizations and obtain spatially varying QARs through spatially varying coefficients.

QAR models arise when time series are observed to display asymmetric dynamics; such data often appears in economic applications. Koenker and Xiao (2006) show empirical applications of the QAR model to the USA unemployment rate and gasoline prices. Further examples in the literature consider dynamic additive quantile models, QR with cointegrated time series, and conditional quantiles with GARCH models. Applications include stock returns, house price returns, and gold prices. See, e.g., Li et al. (2015) or Yang et al. (2023) and references therein. QR is also popular for climate data (see Gao and Franzke 2017 for an extensive review). Yang et al. (2018) propose a semiparametric QAR model including lagged data to develop quantile-based temperature extreme indices. Zhang et al. (2022) use QR models conditional on the state of the previous observation time to predict short-term wind speed or velocity. Castillo-Mateo et al. (2023) use a rich QAR model to compare the effects of climate change in daily maximum temperature.

The outline of the paper is as follows. Section 2 provides a model characterization for the QAR(1) case. Further, it offers explicit parametric model specifications, the resulting likelihood for Bayesian model fitting, some criteria for model assessment, and a simulation study. Section 3 looks at the QAR(p) case. Section 4 considers the

bivariate QAR(1) setting. Section 5 develops a fully spatial version through the use of a Gaussian copula. Section 6 employs time series of daily temperature data from 18 spatial locations to illustrate the previous four sections. Finally, Sect. 7 presents a brief summary and possibilities for future work.

2 The QAR(1) case

2.1 The support of the data

For a non-crossing *linear* QAR specification we need to restrict the support of the time series data, $\{y_t^* : t = 1, \ldots, T\}$, to a bounded interval on the real line.[1] We take this interval to be [0, 1] and implement this by making a transformation of the data,

$$y_t = \frac{y_t^* - m}{M - m}, \tag{3}$$

where $m < \min y_t^*$ and $M > \max y_t^*$. In fact, m and M are chosen such that $\min y_t$ is close to but above 0 and $\max y_t$ is close to but below 1. This enables the most flexibility for the quantile function under our proposed QAR modeling and we offer an automatic selection approach below.

Two points are important to note. First, we can not take $m = \min y_t^*$ and $M = \max y_t^*$. The data must be in the interior of the unit interval in order to enable distinct quantiles as τ varies across (0, 1). Second, choosing m and M is merely a device for working on the unit interval. There is no connection between these values and the potential practical support of the y_t^*'s. Imposing bounding on the support is unavoidable for a valid linear specification of $Q_{Y_t}(\tau \mid y_{t-1})$ of the form $\theta_0(\tau) + \theta_1(\tau) y_{t-1}$ because the only non-intersecting lines under unbounded support are parallel lines.

A convenient "automatic" strategy for selecting m and M is as follows. The idea recalls basic results from the theory of order statistics. If we have T independent observations from a uniform distribution on (m, M), $\{y_t^* : t = 1, \ldots, T\}$, then $[E(Y_{(1)}^*) - m]/(M - m) = 1/(T + 1)$ and $[E(Y_{(T)}^*) - m]/(M - m) = T/(T + 1)$. So we can say $y_{(1)}^* \approx (mT + M)/(T + 1)$ and $y_{(T)}^* \approx (m + TM)/(T + 1)$. This gives two equations in two unknowns to solve for m and M. We obtain

$$m = \frac{T y_{(1)}^* - y_{(T)}^*}{T - 1} \quad \text{and} \quad M = \frac{T y_{(T)}^* - y_{(1)}^*}{T - 1}. \tag{4}$$

Of course, the Y_t^*'s are not independent, they do not come from a distribution on a bounded interval, and marginally, we would not expect them to follow a uniform distribution on (m, M). We only implement a simple automatic bounding strategy.

[1] This is the analogue of the restriction over the predictor domain in Yang and Tokdar (2017).

2.2 The model

A straightforward characterization of the required monotonicity of the QAR lines is offered by the following result, inspired from Tokdar and Kadane (2012).

Theorem 1 *An autoregressive specification of the form of* (2) *with* $\theta_1(\tau) \in [-1, 1]$ *for* $\tau \in [0, 1]$ *is monotonically increasing in* τ *for* Y_t *taking values in* $[0, 1]$ *and* $y_{t-1} \in [0, 1]$ *if and only if*

$$Q_{Y_t}(\tau \mid y_{t-1}) = y_{t-1}\eta_1(\tau) + (1 - y_{t-1})\eta_2(\tau) \tag{5}$$

where $\eta_1, \eta_2 : [0, 1] \longrightarrow [0, 1]$ *are monotonically increasing.*

Proof Any monotonicity obeying $Q_{Y_t}(\tau \mid y_{t-1})$ given by (2) can be expressed as (5) by taking $\eta_1(\tau) = \theta_0(\tau) + \theta_1(\tau) = Q_{Y_t}(\tau \mid 1)$ and $\eta_2(\tau) = \theta_0(\tau) = Q_{Y_t}(\tau \mid 0)$. For the converse, if $Q_{Y_t}(\tau \mid y_{t-1})$ is given by (5) then it must be monotonically increasing in τ for every $y_{t-1} \in [0, 1]$ for which both y_{t-1} and $1 - y_{t-1}$ are nonnegative. One can express such a $Q_{Y_t}(\tau \mid y_{t-1})$ by defining $\theta_0(\tau) = \eta_2(\tau)$ and $\theta_1(\tau) = \eta_1(\tau) - \eta_2(\tau) \in [-1, 1]$. □

If we focus on (5), a model for functions η_1 and η_2, each from $[0, 1] \to [0, 1]$, induces a QAR(1) model over all valid QAR(1) specifications of $Q_{Y_t}(\tau \mid y_{t-1})$, provided the boundary conditions $Q_{Y_t}(0 \mid y_{t-1}) = 0$ and $Q_{Y_t}(1 \mid y_{t-1}) = 1$ for all $y_{t-1} \in [0, 1]$ are satisfied. The above condition can be rewritten as $\eta_j(0) = 0$ and $\eta_j(1) = 1$ for $j = 1, 2$. Next we show how to specify these two monotone functions.

2.2.1 Specification for the two monotone curves

Specifically, both $\eta_1(\cdot)$ and $\eta_2(\cdot)$ again must be strictly monotone from $[0, 1] \to [0, 1]$. A convenient class to work with are cdf's for continuous random variables with support $[0, 1]$. In fact, a rich class would arise as probabilistic mixtures of such cdf's, leading to the general form

$$\eta(\tau) = \sum_{k=1}^{K} \lambda_k F(\tau \mid \mathbf{\Omega}_k) \tag{6}$$

such that $\lambda_k \geq 0$, $\sum_k \lambda_k = 1$ and $F : [0, 1] \to [0, 1]$ is strictly increasing for any parameters $\mathbf{\Omega}_k$.

A convenient class of F's to work with are the cdf's of the two parameter Kumaraswamy (1980) distribution (also known as the minimax distribution, Jones 2009). Specifically, the probability density function (pdf) and cdf are

$$f(x \mid a, b) = abx^{a-1}(1 - x^a)^{b-1} \quad \text{and} \quad F(x \mid a, b) = 1 - (1 - x^a)^b, \tag{7}$$

where $x \in [0, 1]$ and $a, b > 0$. The Kumaraswamy distributions are a family with behavior similar to the beta distribution. However, for our purposes, they are much simpler to use especially in the context of simulation since the cdf and quantile function

can be expressed in closed form, i.e., $Q(\tau \mid a, b) = (1 - (1 - \tau)^{1/b})^{1/a}$ where $\tau \in [0, 1]$. The flexibility of the Kumaraswamy distributions is shown in Section S1 of the Supplementary Information (SI) employing different combinations of parameters (a, b).

To work with the mixture form for $\eta(\tau)$, we investigated two mixture strategies. The first lets K be small but assumes the a's and b's are unknowns. The second lets K be larger but adopts a fixed set of a's and b's, in the spirit of basis function forms. Specifically, we consider K Kumaraswamy distributions with medians $k/(K + 1)$, respectively. In the former, with $K = 2$ we have a total of five parameters (two a's, two b's, and a λ) while in the latter, with $K = 6$ again we have five parameters (five λ's). Increasing the number of "basis" components in the specification of the η's need not provide better model performance. From considerable simulation experience, model performance is very sensitive to the choice of parameters in the mixture components. So, in the sequel, we work with $K = 1$ or 2 (QAR1K1 and QAR1K2, hereafter) and fit the a's and b's. As for priors, with $K = 1$, we consider $\log a_1, \log b_1 \sim N(0, \sigma_{ab}^2)$ with $\sigma_{ab} = 3$, which gives a weak prior on the log scale. With $K = 2$, we consider $\lambda_1 \sim U(0, 1/2)$ and $\log a_1, \log a_2, \log b_1, \log b_2 \sim N(0, \sigma_{ab}^2)$ with $\sigma_{ab} = 1.5$. Restricting λ_1 to $(0, 1/2)$ avoids identification issues, while σ_{ab} is taken smaller than in the $K = 1$ case to penalize values of a's and b's too small or large. Values of a's and b's that are close to zero or very large can cause negligible numerical errors in the root-finder to generate a numerical overflow in the likelihood (see Eqs. 8 and 9) and thus degeneracy.

2.3 Likelihood evaluation and model fitting

An important feature of a valid joint specification of $Q_{Y_t}(\tau \mid y_{t-1})$ for all $\tau \in (0, 1)$, following Tokdar and Kadane (2012), is that it uniquely defines the conditional response density given $y_{t-1} \in [0, 1]$. Specifically, this density is given by

$$f_{Y_t}(y_t \mid y_{t-1}) = \left. \frac{1}{\frac{\mathrm{d}}{\mathrm{d}\tau} Q_{Y_t}(\tau \mid y_{t-1})} \right|_{\tau = \tau_{y_{t-1}}(y_t)}, \tag{8}$$

where $\tau_{y_{t-1}}(y_t)$ solves $y_t = y_{t-1}\eta_1(\tau) + (1 - y_{t-1})\eta_2(\tau)$ in τ and is numerically approximated to arbitrary precision via a one-dimensional root-finder. We implement the hybrid root-finding algorithm combining the bisection method, the secant method, and inverse quadratic interpolation, so-called Brent's method (Brent 1973). Consequently, given the data at $t = 1$, y_1, we can write a valid log-likelihood score

$$\begin{aligned} \ell(\boldsymbol{\Omega} \mid \mathbf{y}) &= \sum_{t=2}^{T} \log f_{Y_t}(y_t \mid y_{t-1}) \\ &= -\sum_{t=2}^{T} \log \left\{ y_{t-1}\dot{\eta}_1(u_t) + (1 - y_{t-1})\dot{\eta}_2(u_t) \right\}, \end{aligned} \tag{9}$$

where $u_t = \tau_{y_{t-1}}(y_t)$, $\mathbf{y}^\top = (y_1, \ldots, y_T)$ are all of the observed data, $\boldsymbol{\Omega}$ are the model parameters, and the $\dot{\eta}$'s are the derivatives of the η's.

We implement a block Metropolis sampler algorithm with an adaptive period (Haario et al. 2001) during warm-up to obtain Markov chain Monte Carlo (MCMC) samples from the posterior distribution of the parameters and to summarize the posterior distribution of the conditional quantile function. Furthermore, with a posterior realization of the model parameters and a given value of y_{t-1}, we can use (8) with discretization, to obtain a posterior realization of the density function that is driving the joint quantiles. Averaging over these realizations provides the posterior mean of the density.

Expression (8) reveals an important difference between our QAR approach and other nonlinear joint modeling versions in the literature. For example, the nonlinear QAR model in Chen et al. (2009) specifies a joint distribution for (Y_t, Y_{t-1}) using a copula. It yields a conditional distribution for $Y_t \mid Y_{t-1}$ which has a nonlinear quantile function that is monotone in τ. What we do is the reverse. We specify a non-crossing quantile function and obtain the induced conditional distribution for $Y_t \mid Y_{t-1}$. Our quantile function is also nonlinear as a function of Y_{t-1}. Their quantile function depends upon the choice of copula and the copula parameters. Our quantile function depends on the Kumaraswamy distribution and the associated parameters. Their Gaussian version has a conditional quantile function which is linear in Y_{t-1}, which may be restrictive. Their t-copula version yields a quantile function which has a perhaps unattractive form as the square root of a function of the square of Y_{t-1}. If the goal is to model the quantile function directly as nonlinear and flexible, rather than seeing what is induced by a copula, our approach yields a simple form and may be more attractive.

2.4 Model comparison and simulation study

Working within our parametric Bayesian framework, for any τ, posterior samples of the model parameters, $\{\boldsymbol{\Omega}_b^* : b = 1, \ldots, B\}$, produce posterior samples of the conditional quantile function for Y_t, $Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega}_b^*)$. Essentially, for each Y_t (with associated y_{t-1}) and any τ, we obtain the posterior distribution of $Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega})$. We use these posterior distributions along with the dataset, $\mathbf{y}$, to offer model assessment.

We propose two novel approaches. First, for any y, consider $\mathbf{1}(y < Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega}))$ where $\mathbf{1}$ denotes the indicator function. Then, let $p_t(\tau) \equiv E[\mathbf{1}(y_t < Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega})) \mid \mathbf{y}]$, i.e., the posterior probability that $Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega})$ exceeds y_t. Suppose we compute $p(\tau) \equiv \sum_{t=2}^{T} p_t(\tau)/(T-1)$. We note that for any τ and Y regardless of its distribution, $E[\mathbf{1}(Y < Q_Y(\tau))] = \tau$ and $Var[\mathbf{1}(Y < Q_Y(\tau))] = \tau(1-\tau)$. If we let $v \geq 1$ be a real number, then

$$\tilde{p}_v \equiv \sqrt[v]{\int_0^1 \left| \frac{p(\tau) - \tau}{\sqrt{\tau(1-\tau)/(T-1)}} \right|^v \mathrm{d}\tau} \tag{10}$$

provides a standardized deviation form as a dimensionless measure of how well the quantile function under the model is capturing conditional quantiles for the given time series. We propose this as a (*global*) measure of model accuracy. With a minimum value

of zero, a smaller $\tilde{p}_v$ indicates better accuracy of the model. We would approximate the integral by discretizing τ, in particular, we consider $\tau \in \{0.01, 0.02, \ldots, 0.99\}$.

As a second measure, we turn to the check loss function, usually employed as an optimality function to obtain the τ empirical quantile (Koenker and Bassett 1978). Here, we adopt $\delta_\tau(u) = u(\tau - \mathbf{1}(u < 0))$, the check loss function associated with the AL distribution. Again, from the posterior distribution of $Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega})$, for any Y_t (with associated y_{t-1}) and τ, we can obtain $\Delta_t(\tau) \equiv \delta_\tau(y_t - E[Q_{Y_t}(\tau \mid y_{t-1}; \boldsymbol{\Omega}) \mid \mathbf{y}])$. As above, suppose we compute $\Delta(\tau) \equiv \sum_{t=2}^{T} \Delta_t(\tau)/(T-1)$. Then, for a given τ, $\Delta(\tau)$ provides an average discrepancy for the τ quantile function. The smaller the value, the better the model performance. Then, we propose to weight $\Delta(\tau)$,

$$\tilde{\Delta} \equiv \int_0^1 \omega(\tau)\Delta(\tau)\,\mathrm{d}\tau \tag{11}$$

to provide a global measure of model performance. We propose this as a *relative* measure of model performance in making model comparison. The weighting function, $\omega(\tau)$, compensates for the variation in mean of $\Delta(\tau)$ across τ. Again, we would approximate the integral by discretizing τ.

For the weight function, we consider

$$\omega(\tau \mid \mathbf{y}) = \frac{1}{\sum_{t=2}^{T} \delta_\tau(y_t - Q_Y^{emp}(\tau))/(T-1)}. \tag{12}$$

This choice leads to a measure that is closely related to the $R^1(\tau)$ metric by Koenker and Machado (1999). The $R^1(\tau)$ measure is essentially, $1 - \omega(\tau \mid \mathbf{y})\Delta(\tau)$. This measure is viewed as an analogue of R^2 for the classical residual sum of squares, i.e., the check loss function for quantiles replaces the least-squares loss function and the τ empirical marginal quantile $Q_Y^{emp}(\tau)$ replaces the sample mean. With a maximum value of 1, the best model performance is reached at this maximum. Then,

$$\bar{R^1} \equiv \int_0^1 R^1(\tau)\,\mathrm{d}\tau = 1 - \int_0^1 \omega(\tau \mid \mathbf{y})\Delta(\tau)\,\mathrm{d}\tau = 1 - \tilde{\Delta}, \tag{13}$$

provides a dimensionless global measure of model performance which can be used for model comparison.

In Section S2 of the SI, we present the results of a brief simulation study where the goals were (i) to illustrate parameter recovery under fitting for several models, (ii) to investigate model flexibility, i.e., performance when the sampling model is not the same as the fitting model, and (iii) to consider the effect of sample size with regard to (i) and (ii).

3 The QAR(*p*) case

We provide a straightforward extension of our joint QAR(1) model to the lag p case. It is not a characterization of the QAR(p) function of Y_t but offers a flexible specification.

In this regard, we obtain a form with some restrictions on the autoregressive coefficients but no constraints on the y_t's beyond the bounded interval support. By interpreting $\eta_1(\tau)$ and $\eta_2(\tau)$ in (5) as the conditional quantiles of Y_t at $y_{t-1} \in \{0, 1\}$, we build a similar construction for an autoregressive process of order p as follows. Define

$$Q_{Y_t}(\tau \mid y_{t-1}, \ldots, y_{t-p}) = (\eta_1(\tau), \ldots, \eta_{p+1}(\tau)) \begin{pmatrix} 0 & \pi_1 & 0 & \cdots & 0 \\ 0 & 0 & \pi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \pi_p \\ 1 & -\pi_1 & -\pi_2 & \cdots & -\pi_p \end{pmatrix} \begin{pmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix}, \tag{14}$$

where the functions $\eta_1, \ldots, \eta_{p+1} : [0, 1] \to [0, 1]$ are monotonically increasing and the weights $\pi_1, \ldots, \pi_p$ are such that $\pi_j \geq 0$ and $\sum_j \pi_j = 1$. It is easy to see that such $Q_{Y_t}(\tau \mid y_{t-1}, \ldots, y_{t-p})$ is monotonically increasing in $\tau \in [0, 1]$ for every $y_{t-1}, \ldots, y_{t-p} \in [0, 1]$.

In particular, for QAR(2), let $\tau, \pi \in [0, 1]$. Then, define

$$\begin{aligned} Q_{Y_t}(\tau \mid y_{t-1}, y_{t-2}) = \pi y_{t-1}\eta_1(\tau) + (1-\pi)y_{t-2}\eta_2(\tau) \\ +(1-\pi y_{t-1} - (1-\pi)y_{t-2})\eta_3(\tau) \end{aligned} \tag{15}$$

where the three η functions are all strictly increasing, using forms as above. Rewriting the expression as

$$Q_{Y_t}(\tau \mid y_{t-1}, y_{t-2}) = \eta_3(\tau) + \pi(\eta_1(\tau) - \eta_3(\tau))y_{t-1} + (1-\pi)(\eta_2(\tau) - \eta_3(\tau))y_{t-2}, \tag{16}$$

both autoregressive coefficients belong to $[-1, 1]$ and need not be increasing in τ. We fit this QAR(2) model to our real data in Sect. 6.3. In fact, we only attempt this with $K = 1$ mixture components (seven parameters) to keep the model simple. Further, the second autoregressive term results are not influential for our data. Also, we choose $\log a$'s and $\log b$'s to follow a $N(0, 1.5^2)$ prior and $\pi \sim U(0, 1)$ as a non-informative prior for π.

4 Multivariate QAR(1)

Often a collection of dependent times series is gathered over a common time window. For instance, our illustration below considers the dependent pairs $\{(y_t^{\max}, y_t^{\min}) : t = 1, \ldots, T\}$, the daily maximum and minimum temperature for day t at a site. In fact, the collection of time series might be spatially referenced (leading to a spatial copula model construction, as developed in the next section). What we have is the quantile analogue of usual multivariate AR for time series. Implementation using the class of joint QAR(1) models we have proposed has not appeared in the literature. Our interest is in the quantile function for each time series. We are asking about the amount of dependence between quantile levels regarding the marginal quantile functions.

Here, we illustrate with the bivariate case where we have two models each defined as in (1), introducing dependence in the two time series by making the associated U_t's dependent through $T-1$ i.i.d. two-dimensional Gaussian copulas. This specification captures the acknowledged dependence between the pair of time series. We postpone to Sect. 5 the details of modeling using copulas; in particular, that section develops the form of the general n-dimensional joint density. The only detail that we advance here is that the correlation matrix associated with the copulas contains 1's on the diagonal and ρ on the off-diagonal, where $\rho \sim U(-1, 1)$ measures the correlation between series.

Apart from introducing dependence through $U_t^{\max}$ and $U_t^{\min}$, we could introduce dependence in the η's. For instance, using Kumaraswamy cdf's, under the $K = 1$ case, we consider the pairs $\log a_j^{\max}$ and $\log a_j^{\min}$ and the pairs $\log b_j^{\max}$ and $\log b_j^{\min}$ ($j = 1, 2$) to be bivariate normal. In our data, we found little or no correlation between the parameters of the two time series, so in subsequent analyzes we will consider them independent. We do not pursue this case further here except to note the analogy with dependent responses in linear regression models. Introducing dependence through the U_t's is analogous to introducing dependence through the errors in the linear regression while introducing dependence through the η's is analogous to introducing dependence in the mean structure through shared parameters.

An example is presented in Sect. 6.4. Again, with $K = 1$, this yields four η's, i.e., four independent $\log a$'s and four independent $\log b$'s, each following a weak, say $N(0, 3^2)$ prior, as well as the copula parameter. As a by-product, we show the induced bivariate conditional pdf (arising from the bivariate analogue of Eq. 8) for $(Y_t^{\max}, Y_t^{\min})$ with some choices for the y_{t-1}'s.

5 Spatial QAR(1)

In the spatial setting, we consider spatial point-referenced time series data. Here, $Y_t(\mathbf{s})$ denotes the observation for time $t = 1, \ldots, T$ at location $\mathbf{s} \in D$, where $D \subset \mathbb{R}^2$ is the study region. We have a time series at each of the locations, $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, say, the locations of the monitoring stations. The joint spatial QAR model is given by

$$Y_t(\mathbf{s}) = \theta_0(U_t(\mathbf{s}); \mathbf{s}) + \theta_1(U_t(\mathbf{s}); \mathbf{s}) Y_{t-1}(\mathbf{s}), \tag{17}$$

where the θ functions are quantile and spatially varying. Chen and Tokdar (2021) propose to model the spatial dependence of the realizations in a QR model using a spatial copula process. Generalizing it to our model, the vectors $(U_t(\mathbf{s}_1), \ldots, U_t(\mathbf{s}_n))^\top$ follow an independent copula distribution for every t. Supplementing Chen and Tokdar (2021), in (17) we introduce spatially varying coefficients rather than global coefficients. As a consequence, we have dependence in the time series realizations as well as spatially varying quantile functions.

5.1 Modeling spatial dependence

Spatial dependence is captured through spatially varying quantiles which are analogous to introducing spatially varying coefficients in spatial linear regression, and dependent quantile levels which are analogous to introducing dependence through the errors in the linear regression.

5.1.1 Spatially varying quantiles

For the spatially varying coefficients, we consider only one Kumaraswamy cdf for each $\eta(\tau;\mathbf{s})$. In fact, at location $\mathbf{s}$, let assume $\eta_j(\tau;\mathbf{s}) = 1-(1-\tau^{a_j(\mathbf{s})})^{b_j(\mathbf{s})}$ with parameters $a_j(\mathbf{s})$ and $b_j(\mathbf{s})$ for $j = 1, 2$. Then, we introduce four independent GPs for the a's and b's on the log scale. In particular, we model $\log a_j(\mathbf{s}) \sim GP(a_j, \sigma_{a_j}^2\rho(\mathbf{s},\mathbf{s}';\phi_{a_j}))$ and $\log b_j(\mathbf{s}) \sim GP(b_j, \sigma_{b_j}^2\rho(\mathbf{s},\mathbf{s}';\phi_{b_j}))$ where the $\rho(\mathbf{s},\mathbf{s}';\phi)$'s are exponential correlation functions with ϕ's as corresponding decay parameters.

We take the ϕ's to be fixed values, according to the spatial scale, because it is usually difficult to estimate them from the data and typically interest focuses on the σ^2's, the spatial uncertainties (Banerjee et al. 2014). Specifically, we fix $\phi = 3/d_{\max}$, with $d_{\max}$ the maximum distance between any pair of spatial locations. Thus, the spatial GPs are only indexed by a mean and a variance parameter. We choose the priors $a_j, b_j, \log\sigma_{a_j}^2, \log\sigma_{b_j}^2 \sim N(0, 3^2)$ $(j = 1, 2)$.

5.1.2 The spatial copula process

A copula is a multivariate cdf for which the marginal distribution of each variable is $U(0, 1)$. Copulas are used to model the dependence between random variables. Particularly, Sklar's theorem (Sklar 1959) states that any multivariate joint pdf can be written in terms of univariate marginal pdf's and a copula which describes the dependence structure between the variables.

Gaussian spatial copulas enable computational advantages, e.g., ease of parameter estimation and scalability with sample size. For a given correlation matrix R, the n-dimensional Gaussian copula function with parameter matrix R becomes

$$C_\Phi(\mathbf{u} \mid R) = \Phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)), \tag{18}$$

where $\mathbf{u}^\top = (u_1, \ldots, u_n) \in [0, 1]^n$, Φ_R is the joint cdf of a multivariate normal distribution with zero-mean vector and covariance matrix R. According to Xue-Kun Song (2000), the associated copula density is

$$c_\Phi(\mathbf{u} \mid R) = |R|^{-1/2} \exp\left\{\frac{1}{2}\mathbf{q}^\top(\mathbf{I}_n - R^{-1})\mathbf{q}\right\}, \tag{19}$$

with $\mathbf{q}^\top = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))$.

With regard to the copula model for (17), we take the processes $U_t(\mathbf{s})$'s to follow a Gaussian copula for each t, induced by a spatial GP. In the spirit of Chen and Tokdar (2021), we define

$$\begin{aligned} U_t(\mathbf{s}) = \Phi(Z_t(\mathbf{s})), \quad Z_t(\mathbf{s}) = W_t(\mathbf{s}) + \epsilon_t(\mathbf{s}), \\ W_t(\mathbf{s}) \sim GP(0, \gamma\rho(\mathbf{s}, \mathbf{s}'; \phi)), \quad \epsilon_t(\mathbf{s}) \sim \text{i.i.d. } N(0, 1-\gamma). \end{aligned} \tag{20}$$

The process $W_t(\mathbf{s})$ captures spatial dependence while $\epsilon_t(\mathbf{s})$ is independent pure error. The parameter $\gamma \in [0, 1]$ determines the proportion of spatial and independent variation. When $\gamma = 1$, the specification for $Z_t(\mathbf{s})$ is purely spatial. When $\gamma = 0$, we have an independent noise model. With this approach, the Gaussian copula density has correlation matrix $R \equiv \gamma R(\phi) + (1-\gamma)\mathbf{I}_n$ where $R(\phi)$ is the $n \times n$ correlation matrix induced by $\rho(\mathbf{s}, \mathbf{s}'; \phi)$. To address the final copula piece of our model, we fix ϕ as above, and adopt $\gamma \sim U(0, 1)$ as a non-informative prior for γ.

5.2 Likelihood evaluation

We are interested in the likelihood under model (17) using (19) and (20). It is convenient to first obtain the joint distribution for $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_T^\top)$ where $\mathbf{Y}_t^\top = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))$, $t = 1, \ldots, T$. That is, each $\mathbf{Y}_t$ is $n \times 1$ and $\mathbf{Y}$ is $Tn \times 1$. By Sklar's theorem, the joint conditional density of responses, $\mathbf{Y}$, given the data at the initial time, $\mathbf{y}_1$, can be partitioned into a marginal part and a copula part,

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y} \mid \mathbf{y}_1) = \prod_{t=2}^{T} \Bigg[\prod_{i=1}^{n} f_{Y_t(\mathbf{s}_i)}\big(y_t(\mathbf{s}_i) \mid y_{t-1}(\mathbf{s}_i)\big) \\ \times c_\Phi\big(F_{Y_t(\mathbf{s}_1)}(y_t(\mathbf{s}_1) \mid y_{t-1}(\mathbf{s}_1)), \ldots, F_{Y_t(\mathbf{s}_n)}(y_t(\mathbf{s}_n) \mid y_{t-1}(\mathbf{s}_n))\big) \Bigg], \end{aligned} \tag{21}$$

where the cdf $F_{Y_t(\mathbf{s}_i)}$ corresponds to the pdf $f_{Y_t(\mathbf{s}_i)}$ and c_Φ is the Gaussian copula density in (19). As in Sect. 2.3, we evaluate $f_{Y_t(\mathbf{s}_i)}$ and $F_{Y_t(\mathbf{s}_i)}$ using:

$$\begin{aligned} f_{Y_t(\mathbf{s}_i)}(y_t(\mathbf{s}_i) \mid y_{t-1}(\mathbf{s}_i)) &= \left. \frac{1}{\frac{\mathrm{d}}{\mathrm{d}\tau} Q_{Y_t(\mathbf{s}_i)}(\tau \mid y_{t-1}(\mathbf{s}_i))} \right|_{\tau = \tau_{y_{t-1}(\mathbf{s}_i)}(y_t(\mathbf{s}_i))}, \\ F_{Y_t(\mathbf{s}_i)}(y_t(\mathbf{s}_i) \mid y_{t-1}(\mathbf{s}_i)) &= \tau_{y_{t-1}(\mathbf{s}_i)}(y_t(\mathbf{s}_i)), \end{aligned} \tag{22}$$

where $\tau_{y_{t-1}(\mathbf{s}_i)}(y_t(\mathbf{s}_i))$ solves $y_t(\mathbf{s}_i) = y_{t-1}(\mathbf{s}_i)\eta_1(\tau; \mathbf{s}_i) + (1 - y_{t-1}(\mathbf{s}_i))\eta_2(\tau; \mathbf{s}_i)$ in τ. Then, the log-likelihood score of the model parameters $\boldsymbol{\Omega}$ can be expressed by

$$\begin{aligned} \ell(\boldsymbol{\Omega} \mid \mathbf{y}) = \sum_{t=2}^{T} \Bigg[- \sum_{i=1}^{n} \log\left\{ y_{t-1}(\mathbf{s}_i)\dot{\eta}_1(u_t(\mathbf{s}_i); \mathbf{s}_i) + (1 - y_{t-1}(\mathbf{s}_i))\dot{\eta}_2(u_t(\mathbf{s}_i); \mathbf{s}_i) \right\} \\ + \log c_\Phi\left(u_t(\mathbf{s}_1), \ldots, u_t(\mathbf{s}_n) \mid R\right) \Bigg], \end{aligned} \tag{23}$$

with $u_t(\mathbf{s}_i) = \tau_{y_{t-1}(\mathbf{s}_i)}(y_t(\mathbf{s}_i))$. Finally, note that, for the calculation of the log-likelihood, the value of the $u_t(\mathbf{s}_i)$'s must be solved for, so the number of root-finders

needed at each iteration of the MCMC is $n(T-1)$. As a result, likelihood evaluation is expensive, leading to long MCMC run times.

5.3 Spatial interpolation

The quantile $Q_{Y_t(\mathbf{s})}(\tau \mid y_{t-1}(\mathbf{s}))$ is a function of process realizations. Posterior samples for the hyperparameters are available from the model fitting. Posterior samples for the GPs are available, using posterior samples of the hyperparameters, through usual Bayesian kriging (Banerjee et al. 2014). This yields prediction of $a_j(\mathbf{s}_0)$ and $b_j(\mathbf{s}_0)$ $(j = 1, 2)$ at a new $\mathbf{s}_0 \in D$, enabling spatially varying quantile functions. Therefore, we can interpolate conditional quantiles to any desired location in the study region given any proposed or reference value for the previous day's temperature at that location. If we do this over a sufficiently spatially resolved grid, we can obtain the posterior mean at each point and show the posterior τ conditional quantile surface for the given day.

6 Application to temperature data

6.1 The data

We illustrate the proposed modeling methods with analyses of persistence in point-referenced daily maximum temperatures (°C) from $n = 18$ locations in Aragón, Spain. We bring in daily minimum temperatures for the bivariate QAR(1) case. The data is provided by the State Meteorological Agency (AEMET, in its Spanish acronym). Castillo-Mateo et al. (2022) provide exploratory analysis and spatial hierarchical modeling for this dataset. We analyze responses for 2015, an interesting year because the summer was especially hot in Europe (Dong et al. 2016). There were numerous locations with record-breaking temperatures in July 2015 and the heat was maintained over time. The monthly average value of temperatures was a record in July 2015 for 6 of the 18 locations and in the entire region it was among the 10 hottest monthly averages. We restrict analysis to observations from May, June, July, August, and September (denoted as MJJAS), i.e., the hottest months of the year, resulting in $T = 153$ days. The location of the 18 observatories is shown in Fig. 1 and their time series in Figure S4 of the SI.

We begin with a model comparison using QAR(1) and QAR(2) models for all locations. Then, we analyze two illustrative locations within the region, Pamplona and Zaragoza, separately. Subsequently, we implement the bivariate QAR(1) model to the daily maximum and minimum temperature series in Zaragoza. Finally, we implement the general model for spatial QAR(1) with all the locations. Before model fitting, we scale each of the temperature time series to (0, 1) using the transformation in (3) with m and M in (4). We adopt site-level values for m and M.

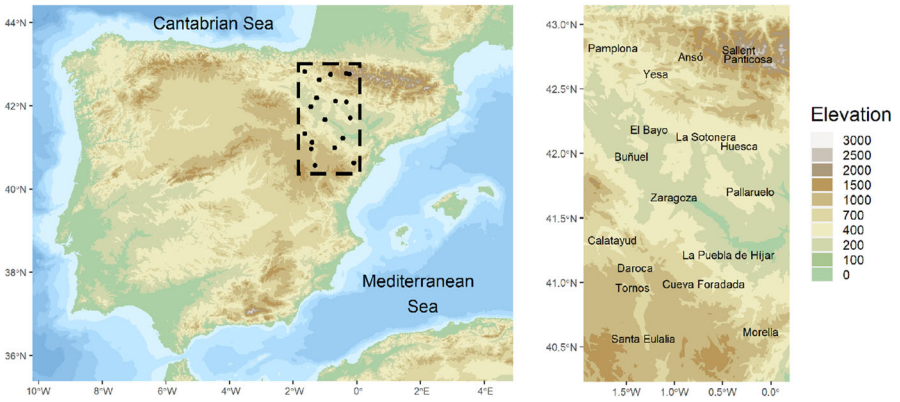

Fig. 1 Location of the 18 sites around Aragón, northeastern Spain

Table 1 Adequacy and comparison metrics $\tilde{p}_2$ in (10) and $\bar{R}^1$ in (13) averaged across locations for QAR1K1, QAR1K2, QAR2K1, and KX2006 models

Model	Description	$\tilde{p}_2$	$\bar{R}^1$
QAR1K1	QAR(1) with $K = 1$ in (6)	0.633	0.365
QAR1K2	QAR(1) with $K = 2$ in (6)	0.402	0.365
QAR2K1	QAR(2) with $K = 1$ in (6)	0.542	0.365
KX2006	Koenker and Xiao (2006)	0.683	0.339

6.2 The QAR(1) case

Table 1 shows, averaged across locations, the metrics of model adequacy $\tilde{p}_2$ and model comparison $\bar{R}^1$ defined in Sect. 2.4 for the QAR1K1 and QAR1K2 models, and the model from Koenker and Xiao (2006) fitted under our Bayesian framework using the density in (8). Table S6 in the SI shows the metrics for each location. For this latter model, denoted as KX2006, we also consider a location parameter μ in the intercept, i.e., $\theta_0(\tau) = \mu + \sigma\,\Phi^{-1}(\tau)$ and $\theta_1(\tau) = \min\{\gamma_0 + \gamma_1\tau, 1\}$ for $\gamma_0 \in (0, 1)$ and $\gamma_1 > 0$. With KX2006 we work on the original scale of the data since they are all positive. We choose the priors $\mu \sim N(0, 10^2)$, $\log\sigma$, $\log\gamma_1 \sim N(0, 3^2)$, and $\gamma_0 \sim U(0, 1)$. The $\bar{R}^1$ does not discriminate much between the proposed models, i.e., the autoregressive term explains much more variability than the difference in specification between the models. However, our proposed models have a slightly higher performance, 0.365, than the KX2006 model, around 0.34. Also, the $\tilde{p}_2$ directly measures how well the quantiles are captured, and its discriminative capacity is much higher. While QAR1K1 obtains a value of 0.633, adding a second component to the mixing improves this measure to 0.402. For its part, the KX2006 model obtains the worst value, 0.683, indicating an overall poorer fitting of the quantiles.

Figure 2 shows the posterior mean of the functions θ_0 and θ_1 in Pamplona and Zaragoza for the models QAR1K1 (dashed) and QAR1K2 (solid). Note that we could recover the intercepts on the original scale as $\theta_0^*(\tau) = m(1-\eta_1(\tau)) + M\eta_2(\tau)$ and the autoregressive coefficients remain invariant. Further, θ_1 is not monotonic; this aspect of temperature dependence with respect to the previous day's temperature was also

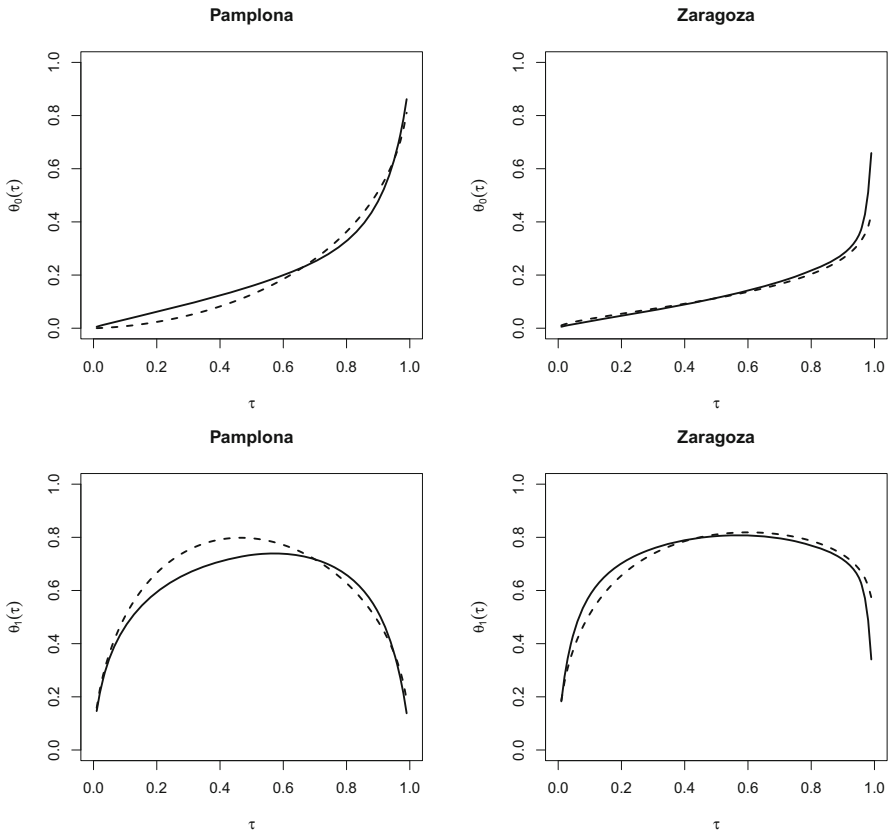

Fig. 2 Posterior mean of $\theta_0(\tau)$ (above) and $\theta_1(\tau)$ (below) vs. τ for QAR1K1 (dashed) and QAR1K2 (solid). Pamplona (left) and Zaragoza (right), MJJAS, 2015

observed by Castillo-Mateo et al. (2023). It cannot be reproduced by KX2006. In Pamplona, the QAR1K2 model (the best) estimates a lower autoregressive coefficient than the QAR1K1 for $\tau \in (0.1, 0.7)$. In Zaragoza, similar curves appear for the two values of K, as shown by $\tilde{p}_2$ and $\bar{R}^1$.

Figure 3 shows the posterior mean of the conditional quantile functions $Q_{Y_t}(\tau \mid y)$ for three situations where y is the empirical τ marginal quantile for $\tau = 0.1, 0.5, 0.9$; the legend shows the values that are conditioned on both the original scale and the $(0, 1)$ scale. The smallest values of θ_1 are in extreme τ's, this means that the previous day's temperature is less influential for high quantiles. In fact, the conditional quantiles in Fig. 3 overlap for τ's near 0 or near 1.

Figure 4 shows the posterior mean of the conditional density function in (8) under the same conditions as Fig. 3. Pamplona presents different shapes in $f_{Y_t}(y_t \mid y)$ for different values of y. The distribution is asymmetrical with positive skewness if we condition on a small value for the previous day's temperature, and negative skewness if we condition on a big value. A general pattern is common in the region, the conditional distribution conditional on the 0.9 marginal quantile is more concentrated than those

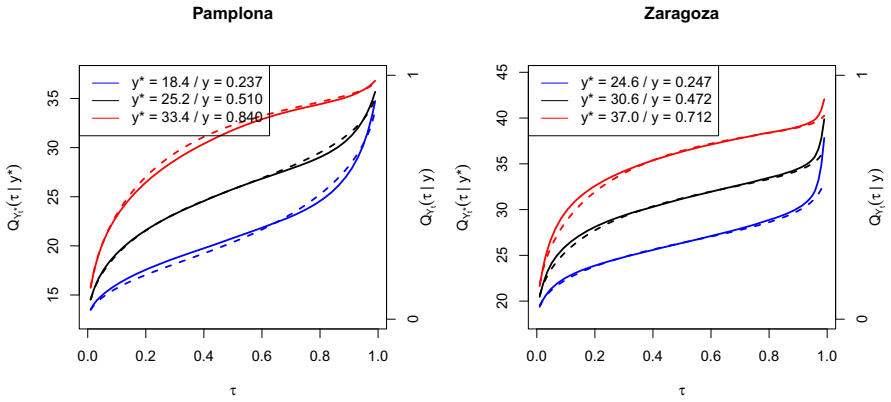

Fig. 3 Posterior mean of the quantile function $Q_{Y_t}(\tau \mid y)$ vs. τ for QAR1K1 (dashed) and QAR1K2 (solid). Here, y is the empirical τ marginal quantile for $\tau = 0.1$ (blue), 0.5 (black), 0.9 (red). Pamplona (left) and Zaragoza (right), MJJAS, 2015 (color figure online)

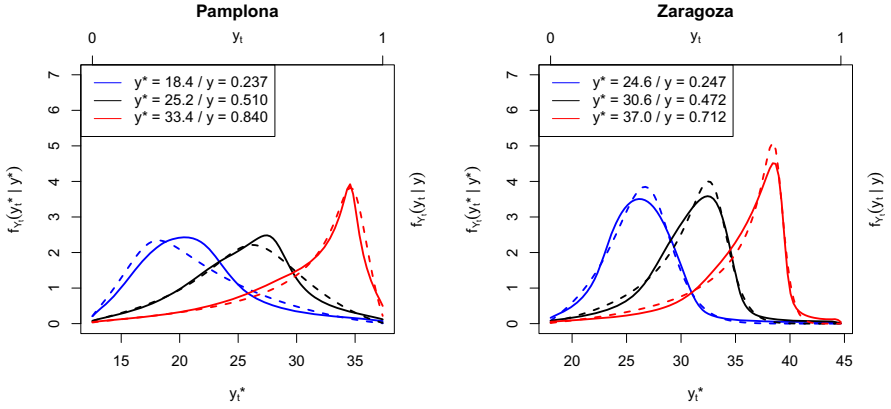

Fig. 4 Posterior mean of the density function $f_{Y_t}(x \mid y)$ for QAR1K1 (dashed) and QAR1K2 (solid). Here, y is the empirical τ marginal quantile for $\tau = 0.1$ (blue), 0.5 (black), 0.9 (red). Pamplona (left) and Zaragoza (right), MJJAS, 2015 (color figure online)

conditional on the 0.1 quantile. Figures S5, S6, S7, and S8 in the SI present the plots for the 18 locations.

6.3 The QAR(2) case

Table 1 uses the criteria $\tilde{p}_2$ and $\bar{R}^1$ for the QAR(2) model with $K = 1$ (QAR2K1). The previous subsection showed that including a first lag improved the performance of the model with respect to an empirical null model. However, including a second lag does not increase the value of $\bar{R}^1$ with respect to a QAR(1) model. On the other hand, the measure of $\tilde{p}_2$ is somewhat better for QAR2K1 than for QAR1K1 but it is still inferior to the QAR1K2 case. Since QAR2K1 does not improve performance, and, as we will see below, there is no evidence that the term $\theta_2(\tau)$ is different from zero for

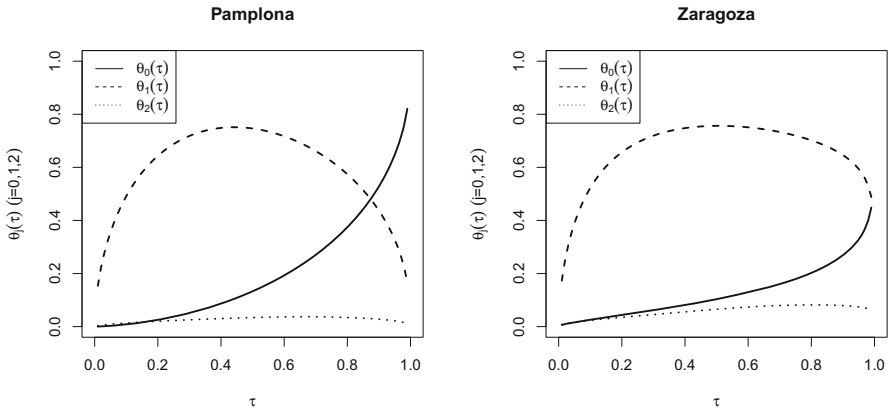

Fig. 5 Posterior mean of $\theta_0(\tau)$ (solid), $\theta_1(\tau)$ (dashed) and $\theta_2(\tau)$ (dotted) vs. τ for QAR2K1. Pamplona (left) and Zaragoza (right), MJJAS, 2015

any τ across most locations, there seems to be no value in exploring a QAR(2) model with $K = 2$.

Figure 5 shows the θ functions in the QAR2K1 model for Pamplona and Zaragoza (see Figure S9 in the SI for all locations). The θ_0 and θ_1 functions have a shape very similar to the QAR1K1 case. The θ_2 functions have values that are essentially centered at zero in most locations, giving more evidence that it is not necessary to introduce a lag of order 2 in the model. However, there are four locations with a coefficient slightly away from zero; Buñuel and La Sotonera have a value of $\theta_2(\tau)$ close to 0.2 for non-extreme τ's while Huesca and La Puebla de Híjar have similar behavior with values around 0.1.

6.4 Multivariate QAR(1)

Here, we fit the multivariate QAR(1) model (MQAR1K1) to the daily maximum and minimum temperature series at Zaragoza, $\{(y_t^{\max}, y_t^{\min}) : t = 1, \ldots, T\}$. The same analyses were developed for Pamplona and Daroca, but with different conclusions. Figure 6 shows the θ functions for the $y_t^{\max}$ (red) and $y_t^{\min}$ (blue) series. We see different patterns for $\theta_1^{\max}$ and $\theta_1^{\min}$; $y_t^{\max}$ shows high autocorrelation for high quantiles while $y_t^{\min}$ has less persistence for those quantiles.

For Zaragoza, the posterior mean of ρ is 0.32 with 95% credible interval $(0.17, 0.45)$, indicating the need to include dependence in the quantile levels of both series. For Pamplona, the posterior mean of ρ is 0.06 with 95% credible interval $(-0.11, 0.23)$. Here, independent models for the conditional quantiles could be adopted. A reasonable explanation is the frequent appearance of fresh wind from the northwest during the night in Pamplona, resulting from proximity to the Cantabrian Sea.

Figure 7 shows level curves of the posterior conditional joint density of the vector $(Y_t^{\max}, Y_t^{\min})$ given the previous day's maximum and minimum temperatures, in

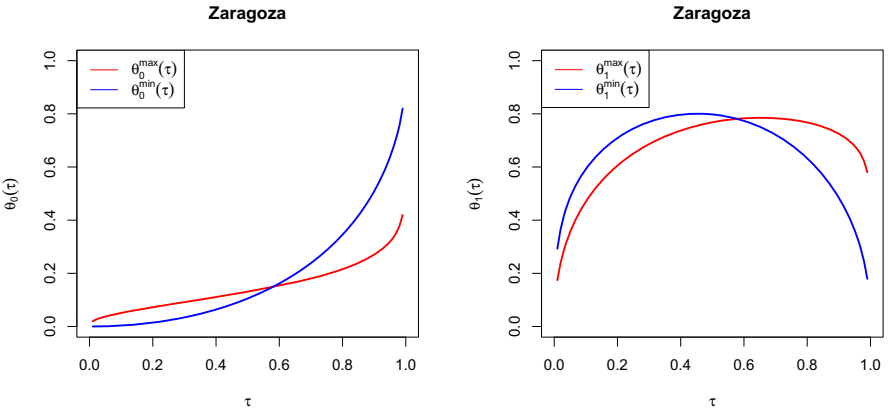

Fig. 6 Posterior mean of $\theta_0^{\max}(\tau)$ and $\theta_0^{\min}(\tau)$ (left), and $\theta_1^{\max}(\tau)$ and $\theta_1^{\min}(\tau)$ (right), vs. τ for MQAR1K1. Zaragoza, MJJAS, 2015

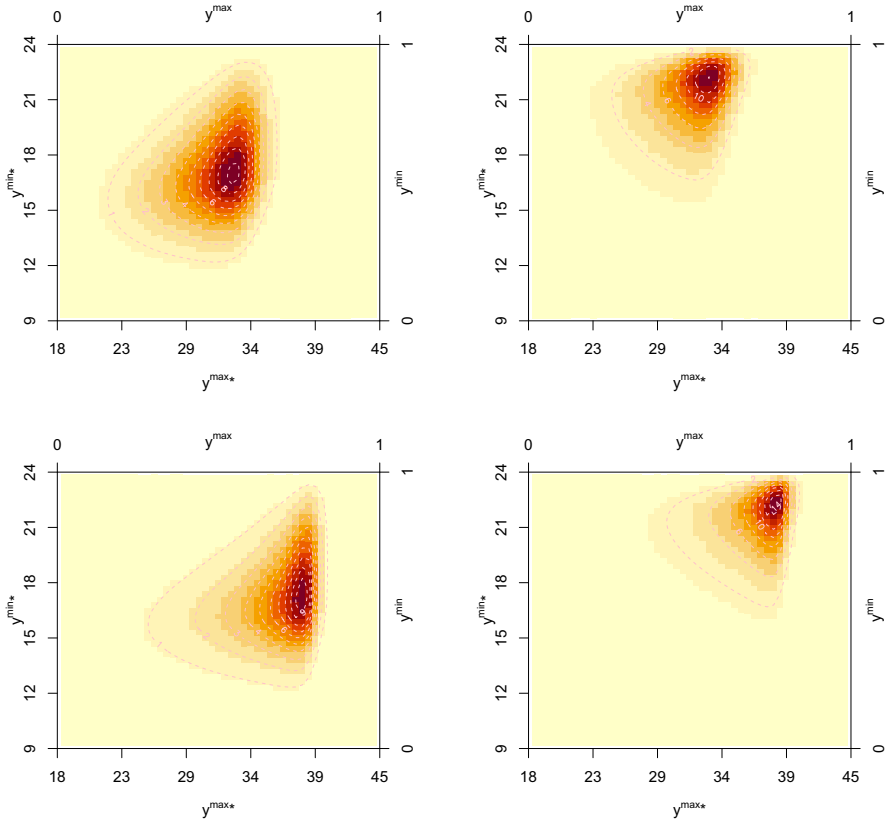

Fig. 7 Posterior mean of the density function of $(Y_t^{\max}, Y_t^{\min})$ conditioned on $(y^{\max}, y^{\min})$ for MQAR1K1. Here, $(y^{\max}, y^{\min})$ is equal to the respective empirical marginal quantiles for $\tau = 0.5$ (above), 0.9 (below) for the maximum; and $\tau = 0.5$ (left), 0.9 (right) for the minimum. Zaragoza, MJJAS, 2015

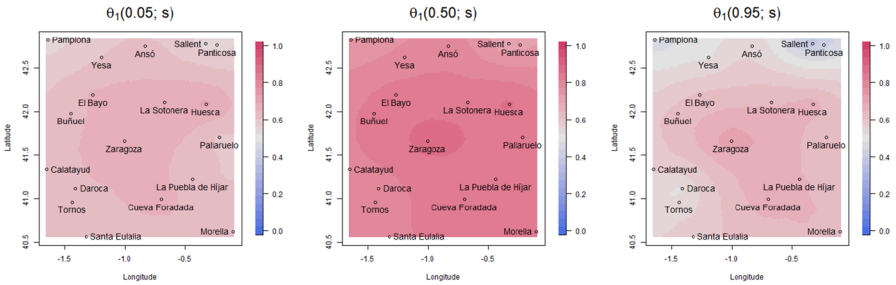

Fig. 8 Maps of the posterior mean of $\theta_1(\tau; \mathbf{s})$ for $\tau = 0.05, 0.50, 0.95$

Zaragoza (see Figures S10 and S11 in the SI for Pamplona and Daroca). The conditioning values are empirical marginal quantiles of $Y_t^{\max}$ and $Y_t^{\min}$. The first row conditions on the quantile $\tau = 0.5$ (30.6°C) and the second row on the quantile $\tau = 0.9$ (37.0°C) of $Y_t^{\max}$, and the same quantiles of $Y_t^{\min}$, for the first (17.2°C) and second (21.8°C) columns. The different patterns observed in the plots reveal a different relation between $Y_t^{\max}$ and $Y_t^{\min}$ depending on the previous day's temperatures. The conditional posterior distribution is not symmetric, with a different mean vector depending on the conditioning temperatures; the variability of the distribution is smaller when it is conditioned on high quantiles.

6.5 Spatial QAR(1)

The spatial QAR model is fitted to the series of MJJAS in 18 locations in Aragón for the year 2015. The posterior mean of γ, the proportion of spatial dependence in (20), is 0.96 with 95% credible interval (0.94, 0.98) indicating very strong spatial dependence in the quantile levels of the temperature series. Figure S12 in the SI provides maps of the posterior mean surface of the model GPs. We notice that $b_1(\mathbf{s})$ and $b_2(\mathbf{s})$ show approximately opposite spatial behavior since $b_1(\mathbf{s})$ has the highest values where $b_2(\mathbf{s})$ has the lowest, in the central and southeastern areas. Figure S13 of the SI shows boxplots of the posterior distribution of the GPs at each observed location; locations are sorted by elevation. The results suggest that the GP of $a_2(\mathbf{s})$ might be not necessary since the boxplots in the 18 locations have very similar ranges. The spatial variability of $a_1(\mathbf{s})$ is higher, and although it is not related to the elevation, it could be related to the distance to the coast.

The posterior distribution of $\theta_1(\tau; \mathbf{s})$, which captures the autoregressive structure, is summarized using the same type of plots. Figure S14 of the SI shows boxplots presenting the posterior distribution of $\theta_1(\tau; \mathbf{s})$ at the observed locations while Fig. 8 the maps of the posterior mean surface of $\theta_1(\tau; \mathbf{s})$, both for $\tau = 0.05, 0.50, 0.95$. The spatial GPs in the parameters of the Kumaraswamy distribution allow the model to fit different spatial patterns in each τ. The results show that the posterior mean of $\theta_1(\tau; \mathbf{s})$ is higher in the central quantiles. The spatial pattern of $\theta_1(\tau; \mathbf{s})$ is not symmetric around $\tau = 0.5$ and, e.g., values of $\theta_1(0.95; \mathbf{s})$ in the Pyrenees and northwestern areas are smaller than $\theta_1(0.05; \mathbf{s})$ in the same areas. Although $\theta_1(\tau; \mathbf{s})$ tends to be lower in

locations with higher elevation, its spatial pattern cannot be explained by elevation alone. Consequently, the spatial GPs cannot be replaced with an elevation fixed effect.

The spatial joint model can also be used to estimate parameters related to the conditional distribution, e.g., conditional quantiles at unobserved locations. As a brief example, Figure S15 of the SI shows this through maps of the posterior mean of $Q_{Y_t(\mathbf{s})}(\tau \mid y)$ for $\tau = 0.05, 0.50, 0.95$, and $y = 0.05, 0.50, 0.95$. If it were desired to obtain the quantiles on the original scale of the data rather than the scale $(0, 1)$, we could consider a kriging of $m(\mathbf{s})$ and $M(\mathbf{s})$. With the same kriging procedure we could condition on values $y(\mathbf{s})$'s relative to a certain empirical marginal quantile for each location.

The spatial modeling here is primarily illustrative. For instance, the assumption of asymptotic tail independence, imposed by the Gaussian copula, may not be suitable. Examination of alternative copulas is beyond the scope of this work.

7 Summary and future work

We have presented consequentially expanded modeling for joint (non-quantile crossing) QAR. In particular, we have characterized the QAR(1) setting in a way that allows for a more flexible autocorrelation structure than the one in the seminal paper by Koenker and Xiao (2006). We have extended this to the QAR(p) case. We have offered a novel multiple time series version using a Gaussian copula. We have elaborated a spatial version, using a GP copula based upon a GP in conjunction with four additional GPs. This model enables spatially varying quantile functions. Our modeling is entirely parametric through the use of the Kumaraswamy distributions. A software implementation of our methods is available as the R-package "QAR" through GitHub: https://github.com/JorgeCastilloMateo/QAR.

We have illustrated the above contributions through time series of daily temperatures from sites in Aragón, Spain. The joint QAR model, with greater flexibility in the modeling of the θ functions, allows us to capture autoregression structure in daily temperature data, which is not strictly increasing in τ, but decreasing in both tails.

A critical challenge in employing this work is model fitting. We can make specifications as rich as needed through the use of probabilistic mixtures of Kumaraswamy cdf's. However, it is well known that model fitting employing MCMC with mixture specifications is often poorly identified. This issue is compounded in our case by the fact that calculation of the likelihood requires constant use of a one-dimensional root-finder. Ongoing work is attempting to address these computational difficulties.

It is important to note that we have not introduced any regressors into our modeling. This adds substantial complication to the joint approach. In order to consider coherent implementation of regressors, conditions have to be imposed on the support for the regressors, seeking to bridge our modeling with the work of Yang and Tokdar (2017). However, we briefly note a simple approximation strategy to incorporate regressors, e.g., seasonality, into our joint QAR approach. Suppose we introduce a regression structure, μ_t into the QAR(1) and estimate by $\hat{\mu}_t$, creating residuals $r_t = Y_t - \hat{\mu}_t$. Then, we could apply the above methodology to obtain the QAR(1) for r_t. Our strategy for selecting m and M can be applied to residuals. More precisely, let $r_t = \theta_{(0)}(U_t) +$

$\theta_{(1)}(U_t)r_{t-1}$. This would yield the conditional quantile function, $Q_{r_t}(\tau \mid r_{t-1}) = \theta_{(0)}(\tau) + \theta_{(1)}(\tau)r_{t-1}$. Solving for the quantile function for Y_t we obtain

$$Q_{Y_t}(\tau \mid Y_{t-1}) = \hat{\mu}_t + \theta_{(0)}(\tau) + \theta_{(1)}(\tau)(Y_{t-1} - \hat{\mu}_{t-1}). \tag{24}$$

We acknowledge that this approximation can be criticized for two reasons: (i) we are creating $\hat{\mu}_t$ as if we were fitting a usual AR(1), and (ii) the resulting quantiles are not coherent since $\hat{\mu}_t$ is a function of $\{Y_t : t = 1, \ldots, T\}$. The QAR(1) is not defined until the end of the observation window.

Sections 4 and 5 could be combined to build a bivariate spatial QAR model for daily maximum and daily minimum temperature. Another challenge for the multivariate and spatial modeling would be to consider alternative copula choices, e.g., t-copulas in order to allow tail dependence for high quantiles.

Supplementary information

SI for "Bayesian joint quantile autoregression" contains details on the Kumaraswamy distribution. Details on the simulation study. More results on the application with temperature series.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-023-00895-6.

Acknowledgements This work has been supported in part by the Grants PID2020-116873GB-I00 and TED2021-130702B-I00 funded by MCIN/AEI/10.13039/501100011033 and Unión Europea NextGenerationEU; and the Research Group E46_20R: Modelos Estocásticos funded by Gobierno de Aragón. J. C.-M. was supported by Gobierno de Aragón under Doctoral Scholarship ORDEN CUS/581/2020 and Mobility Scholarship ORDEN CUS/1668/2022 number MVE_06_23. This work was done in part while J. C.-M. was a Visiting Scholar at the Department of Statistical Science from Duke University. The authors also thank Surya T. Tokdar from Duke University for fruitful discussions on joint QR modeling.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman and Hall/CRC, New York

Bondell HD, Reich BJ, Wang H (2010) Noncrossing quantile regression curve estimation. Biometrika 97(4):825–838. https://doi.org/10.1093/biomet/asq048
Brent RP (1973) Algorithms for minimization without derivatives, 1st edn. Prentice-Hall, Englewood Cliffs
Castillo-Mateo J, Asín J, Cebrián AC, Gelfand AE, Abaurrea J (2023) Spatial quantile autoregression for season within year daily maximum temperature data. Ann Appl Stat 17(3):2305–2325. https://doi.org/10.1214/22-AOAS1719
Castillo-Mateo J, Lafuente M, Asín J, Cebrián AC, Gelfand AE, Abaurrea J (2022) Spatial modeling of day-within-year temperature time series: an examination of daily maximum temperatures in Aragón, Spain. J Agric Biol Environ Stat 27(3):487–505. https://doi.org/10.1007/s13253-022-00493-3
Chen X, Koenker R, Xiao Z (2009) Copula-based nonlinear quantile autoregression. Econom J 12(s1):S50–S67. https://doi.org/10.1111/j.1368-423X.2008.00274.x
Chen X, Tokdar ST (2021) Joint quantile regression for spatial data. J R Stat Soc Ser B (Stat Methodol) 83(4):826–852. https://doi.org/10.1111/rssb.12467
Cui W, Wan C, Song Y (2023) Ensemble deep learning-based non-crossing quantile regression for non-parametric probabilistic forecasting of wind power generation. IEEE Trans Power Syst 38:3163–3178. https://doi.org/10.1109/TPWRS.2022.3202236
Das P, Ghosal S (2017) Bayesian quantile regression using random B-spline series prior. Comput Stat Data Anal 109:121–143. https://doi.org/10.1016/j.csda.2016.11.014
Dong B, Sutton R, Shaffrey L, Wilcox L (2016) The 2015 European heat wave. Bull Am Meteorol Soc 97(12):57–62. https://doi.org/10.1175/BAMS-D-16-0140.1
Gao M, Franzke CLE (2017) Quantile regression-based spatiotemporal analysis of extreme temperature change in China. J Clim 30(24):9897–9914. https://doi.org/10.1175/JCLI-D-17-0356.1
Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. Bernoulli 7(2):223–242
Jones MC (2009) Kumaraswamy's distribution: a beta-type distribution with some tractability advantages. Stat Methodol 6(1):70–81. https://doi.org/10.1016/j.stamet.2008.04.001
Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50. https://doi.org/10.2307/1913643
Koenker R, Machado JAF (1999) Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc 94(448):1296–1310. https://doi.org/10.1080/01621459.1999.10473882
Koenker R, Xiao Z (2006) Quantile autoregression. J Am Stat Assoc 101(475):980–990. https://doi.org/10.1198/016214506000000672
Kozumi H, Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. J Stat Comput Simul 81(11):1565–1578. https://doi.org/10.1080/00949655.2010.496117
Kumaraswamy P (1980) A generalized probability density function for double-bounded random processes. J Hydrol 46(1):79–88. https://doi.org/10.1016/0022-1694(80)90036-0
Li G, Li Y, Tsai CL (2015) Quantile correlations and quantile autoregressive modeling. J Am Stat Assoc 110(509):246–261. https://doi.org/10.1080/01621459.2014.892007
Lum K, Gelfand AE (2012) Spatial quantile multiple regression using the asymmetric Laplace process. Bayesian Anal 7(2):235–258. https://doi.org/10.1214/12-BA708
Peng B, Yang K, Dong X (2023) Variable selection for quantile autoregressive model: Bayesian methods versus classical methods. J Appl Stat. https://doi.org/10.1080/02664763.2023.2178642
Reich BJ, Fuentes M, Dunson DB (2011) Bayesian spatial quantile regression. J Am Stat Assoc 106(493):6–20. https://doi.org/10.1198/jasa.2010.ap09237
Sklar A (1959) Fonctions de réepartition à n dimensions et leurs marges. Publications de l'Institut Statistique de l'Université de Paris 8:229–231
Thrasher B, Maurer EP, McKellar C, Duffy PB (2012) Bias correcting climate model simulated daily temperature extremes with quantile mapping. Hydrol Earth Syst Sci 16(9):3309–3314. https://doi.org/10.5194/hess-16-3309-2012
Tokdar ST, Kadane JB (2012) Simultaneous linear quantile regression: a semiparametric Bayesian approach. Bayesian Anal 7(1):51–72. https://doi.org/10.1214/12-BA702
Xue-Kun Song P (2000) Multivariate dispersion models generated from Gaussian copula. Scand J Stat 27(2):305–320. https://doi.org/10.1111/1467-9469.00191
Yang C, Li L, Xu J (2018) Changing temperature extremes based on CMIP5 output via semi-parametric quantile regression approach. Int J Climatol 38(9):3736–3748. https://doi.org/10.1002/joc.5524
Yang K, Peng B, Dong X (2023) Bayesian inference for quantile autoregressive model with explanatory variables. Commun Stat Theory Methods 52(9):2946–2965. https://doi.org/10.1080/03610926.2021.1964529

Yang Y, Tokdar ST (2017) Joint estimation of quantile planes over arbitrary predictor spaces. J Am Stat Assoc 112(519):1107–1120. https://doi.org/10.1080/01621459.2016.1192545

Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447. https://doi.org/10.1016/S0167-7152(01)00124-9

Zhang C, Ji C, Hua L, Ma H, Nazir MS, Peng T (2022) Evolutionary quantile regression gated recurrent unit network based on variational mode decomposition, improved whale optimization algorithm for probabilistic short-term wind speed prediction. Renew Energy 197:668–682. https://doi.org/10.1016/j.renene.2022.07.123