Statistics and Computing (2026) 36:116
https://doi.org/10.1007/s11222-026-10867-z

ORIGINAL PAPER

Quantile forecast matching with a bayesian quantile gaussian process model

Spencer Wadsworth[1] · **Jarad Niemi**[2]

Received: 9 October 2025 / Accepted: 10 March 2026

Abstract

A set of probabilities along with corresponding quantiles are often used to define predictive distributions or probabilistic forecasts. These quantile predictions offer easily interpreted uncertainty of an event, and quantiles are generally straightforward to estimate using standard statistical and machine learning methods. However, compared to a distribution defined by a probability density or cumulative distribution function, a set of quantiles has less distributional information. When given estimated quantiles, it may be desirable to estimate a fully defined continuous distribution function. Many researchers do so to make evaluation or ensemble modeling simpler. Most existing methods for fitting a distribution to quantiles lack accurate representation of the inherent uncertainty from quantile estimation or are limited in their applications. In this manuscript, we present a Gaussian process model, the quantile Gaussian process –based on established asymptotic results of quantile functions and sample quantiles– to construct a probability distribution given estimated quantiles. In some applications, the form of an unknown distribution function from which sample quantiles are drawn must be estimated, for which case we propose the use of a latent truncated Dirichlet process mixture model for estimation. A Bayesian application of the quantile Gaussian process is evaluated for parameter inference and distribution approximation in simulation studies as well as in a real data analysis of quantile forecasts from the 2023-24 US Centers for Disease Control collaborative flu forecasting initiative. The simulation studies and data analysis show that compared to other existing methods, the quantile Gaussian process leads to accurate inference on model parameters, estimation of a continuous distribution, and uncertainty quantification of sample quantiles.

Keywords Sample quantiles · Quantile regression · Probabilistic forecasting · Disease outbreaks

1 Introduction

The use of quantiles in statistical modeling and in reporting inferential or predictive uncertainty is widespread. Likewise, reporting several quantiles or predictive intervals is common in probabilistic forecasting (Gneiting et al. 2023). Quantiles can be easier to interpret than statistical model parameters and are often used to define confidence or prediction intervals. Thus with multiple quantiles for a particular outcome, one has a measure of uncertainty. Estimating quantiles in the presence of covariates via quantile regression is a common statistical and machine learning strategy, and where parametric models are complicated or nonexistent, quantiles may be easier to estimate via machine learning methods (Martin and Syring 2022; Chung et al. 2021; Koenker 2017; Koenker and Bassett 1978). In estimating multiple quantiles, quantile regression often provides a tradeoff with parametric modeling where for some problems quantile regression is easier to perform, but the predicted quantiles lack the detailed information of a fully defined predictive distribution (Pohle 2020). Perhaps for data privacy reasons, data quantiles are often reported as summaries of the data. Census or medical data, which can be very large or personal to the subjects, may be published as summary or aggregate data including percentiles (quantiles) and medians (Simpson et al. 2023; CDC 2022; Nirwan and Bertschinger 2020). In collaborative forecast hubs, probabilistic forecasts are often submitted

✉ Spencer Wadsworth
spencer.wadsworth@uconn.edu

Jarad Niemi
niemi@iastate.edu

[1] Department of Statistics, University of Connecticut, 215 Glenbrook Rd, Storrs, Connecticut 06269, USA

[2] Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, Iowa 50011, USA

Published online: 04 April 2026

as a set of predictive quantiles (Gneiting et al. 2023; Hong et al. 2016). In recent years, disease outbreak forecast hubs require that all forecasts submitted by outside participants be represented by several predictive intervals or quantiles. This standardized representation allows for straightforward forecast scoring and ensemble building (Mathis et al. 2024; Github 2024; Cramer et al. 2022a, b; Sherratt et al. 2023; Bracher et al. 2021).

A set of quantiles may provide distributional information for an event, but it is not as informative as a distribution defined by a cumulative distribution function (CDF), the inverse-CDF or quantile function (QF), or a probability density/mass function (PDF). Likewise, a set of quantiles provides no information on values below the smallest quantile or above the largest quantile. In other words, it provides no distribution tail information. Another drawback for using quantiles to define a distribution is that many tools for evaluating or scoring continuous distributions require a CDF or PDF (Gneiting and Raftery 2007; Gneiting and Katzfuss 2014). Combining distributions into an ensemble distribution is commonly done by aggregating multiple QFs or CDFs/PDFs, but the latter of these is possible only when a CDF/PDF is available (Gneiting et al. 2005; Wang et al. 2023).

Fitting a distribution given quantiles in order to recover or estimate an unknown distribution is done in many fields, often for the purpose of evaluating forecasts using rules that require a CDF or PDF (Simpson et al. 2023; Gerding et al. 2023). Fitting may also be done to allow for combining of multiple forecasts using aggregation methods that require CDFs or PDFs (Gyamerah et al. 2020; Li et al. 2019; Baran and Lerch 2018; Bogner et al. 2017; He et al. 2016; Gneiting et al. 2005). An example of fitting a distribution given quantiles is given in figure 1. The 12 points in the figure are quantiles estimated from a random sample of size 100 from a standard normal distribution for given probabilities. The QF of the standard normal distribution is represented by the solid grey line in the plot. The dashed grey line is the fitted QF of a normal distribution that was estimated by selecting the mean and standard deviation parameters which minimize the least squares distance between the estimated quantiles and the QF.

Hereafter we refer to estimating a continuous distribution by fitting quantiles as quantile matching (QM). Sgouropoulos et al. performed QM by minimizing the mean square difference between quantiles of a response variable and a linear combination of quantiles of covariates (Sgouropoulos et al. 2015). Related to QM, Feldman and Kowal introduce a margin adjustment for inferring the marginal CDFs of a Gaussian copula model by combining the model with the rank-likelihood in a missing data under multiple data types problem (Feldman and Kowal 2024). Selecting distribution parameters which minimize the mean square error between quantiles and a QF is a common QM method (Dilger et al. 2022; Li et al. 2019; Belgorodski et al. 2017), kernel density estimation and spline interpolation are common nonparametric QM methods (Gerding et al. 2023; Gyamerah et al. 2020; He et al. 2016), and Keelin introduced a semiparametric method based on defining a flexible QF to fit quantiles (Keelin 2016). An issue shared by these methods is that they do not adequately account for the uncertainty inherent from the estimation of quantiles. Nirwan and Bertschinger introduced a QM model based on representing sample quantiles as order statistics and formulated a model likelihood for the quantiles as the joint PDF of multiple order statistics (Nirwan and Bertschinger 2020). This model provides exact inference for quantile uncertainty, but relies on the CDF and PDF of the distribution, making it less practical for fitting quantile defined distributions which have a QF but for which the CDF and PDF either do not exist in closed form or are difficult to evaluate (Perepolkin et al. 2023; Joiner and Rosenblatt 1971; Tukey 1960).

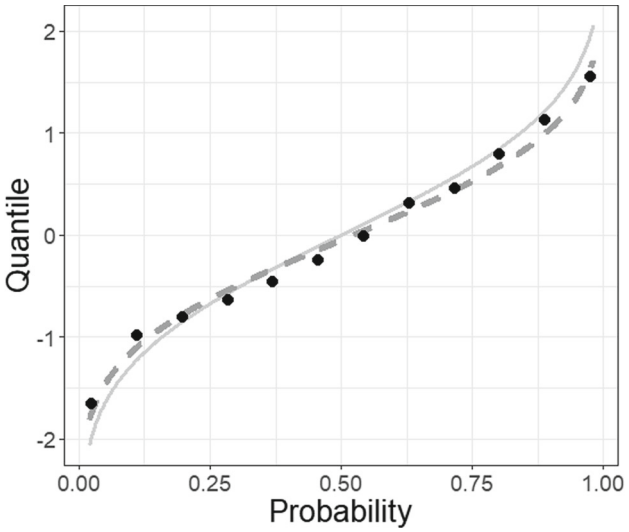

Fig. 1 Points representing 12 quantiles (y-axis) for given probability values (x-axis) estimated from a random sample of size 100 from a standard normal distribution with the true quantile function (QF) (solid grey). The estimated QF (dashed grey) was fit by selecting the mean and standard deviation parameters which minimize the least squares distance between the quantiles and the QF.

In this manuscript we introduce a novel model, the quantile Gaussian process (QGP) which is used specifically for QM. The context for using the QGP is that the available data is a set of quantiles estimated for a given set of probabilities. The QGP relies on established asymptotic theory underlying the relationship between the QF of a continuous distribution and sample quantiles (Parzen 2004; Gilchrist 2000; Hyndman and Fan 1996; Walker 1968; Cramér 1951). The QGP can provide accurate asymptotic inference of the quantile uncertainty, and it works well for QM when modeling quantile defined distributions. The QGP is fit using Bayesian methods and thus allows for using prior information to influence the

QM. The QGP as presented herein requires that either a continuous QF or CDF family be selected a priori. It will often be the case that the true QF which one wishes to estimate is unknown and/or has a complicated shape. To account for this, we use either a finite mixture distribution or a Bayesian nonparametric construction for estimating the unknown function. In section 2 the QF of a continuous random variable and sample quantiles are each defined and important properties and asymptotic theory are reviewed. In section 3 the QGP model for QM is introduced. Section 4 contains simulation studies illustrating the QGP's ability to make parameter inference and match the distribution from which the quantiles were estimated. Section 5 contains an application of QM using the QGP model for quantile forecasts of the 2023-24 United States Centers for Disease Control and Prevention (CDC) flu forecasting competition, also known as FluSight. The analyses in sections 4 and 5 show that the QGP accurately estimates distributions from which the quantiles are estimated and while doing so accurately captures the inherent uncertainty from estimating quantiles. The Bayesian nonparametric construction provides robustness for QM estimation where one might easily overparameterize the model. The QGP outperforms all but one of the competing methods in estimating the unknown distributions and capturing uncertainty, but it also allows for QM in situations where the other methods cannot be used. Section 6 concludes this manuscript with a summary and ideas for further development in QM modeling.

2 Quantile function and sample quantiles

Along with the CDF and PDF, the QF is a defining function of a random variable. The QF, however, often receives less attention in standard statistical training than do the CDF and PDF (Parzen 2004), and the CDF and PDF are used more often for modeling and inference. This section contains the definition of the QF, important properties, the definition of sample quantiles, and key central limit theorems (CLTs) of sample quantiles on which the QGP is based.

2.1 Quantile function definition and properties

The QF is defined in the following.

Definition 1 For a cumulative distribution function $F : \mathbb{R} \to [0, 1]$, the quantile function $Q : [0, 1] \to \mathbb{R}$ is defined as

$$Q(p) = F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in [0, 1]$$

An alternative definition not in terms of the cumulative distribution function is a function $Q : [0, 1] \to \mathbb{R} \cup \{-\infty, \infty\}$, which is nondecreasing and left-continuous.

For most statistical applications, it is useful to define the QF as the inverse of the CDF. However, it is sometimes the case that defining a distribution by a QF rather than a CDF is advantageous. For instance when a distribution is defined only by a QF and has no closed form CDF or when simple QFs may be aggregated to define more complex distributions (Perepolkin et al. 2023; Gasthaus et al. 2019; Alvarez and Orestes 2023). This section continues with important properties of the QF and a related function including the location-scale property, the quantile density function, the probability integral transform, and examples of quantile defined distributions.

2.1.1 Location-scale property

For a random variable with PDF $f_0(x)$ and QF $Q_0(p)$ for $p \in [0, 1]$, a location-scale family takes the PDF form $(1/\sigma) f_0((x - \mu)/\sigma)$ with location parameter $\mu \in (-\infty, \infty)$ and scale parameter $\sigma \in (0, \infty)$ (Casella and Berger 2002). The QF of the location-scale family then takes the form $\mu + \sigma Q_0(p)$ (Parzen 2004). By the addition and multiplication rules from section 3.2 of Gilchrist (2000), this is a valid QF. Gilchrist outlines the standardization rule which states that if the random variable Y with QF $Q_Y(p)$ has a standard distribution –that is the random variable Y is centered (by a mean or median) at 0 and has a scale of 1– then a random variable with quantile function $Q_{T(Y)}(p) = \mu + \sigma Q_Y(p)$ is centered at μ and has scale σ (Gilchrist 2000).

The linear form for a QF is already inherent in some distributions families such as the normal and logistic distribution families, with respective QFs $\mu + \sigma\Phi^{-1}(p)$ and $\mu + \sigma \log(p/(1 - p))$. For the normal distribution, $\Phi^{-1}(p)$ is the QF of a standard normal distribution and μ and σ are the mean and standard deviation of the location-scale transformed normal distribution. For the logistic distribution, $\log(p/(1 - p))$ is the QF of a logistic distribution with mean 0 and scale 1, and μ and σ become the mean and scale parameters of the location-scale transformed logistic distribution. The linear form is convenient for quantile regression modeling and is key to the development of the metalog distribution family which was designed specifically for modeling quantiles (Keelin 2016).

2.1.2 Quantile density function

Besides the CDF, PDF, and QF, a fourth infrequently mentioned defining function of a continuous random variable is the quantile density function (QDF). As the PDF is the derivative of the CDF, the QDF $q : [0, 1] \to \mathbb{R}$ is the derivative of the QF defined as

$$q(p) = \frac{dQ(p)}{dp}.$$

If $Q(p)$ is the inverse of a CDF $F(x)$, a calculus result shows that the reciprocal of the QDF is equal to the PDF evaluated at the QF of p, or $[q(p)]^{-1} = f(Q(p))$ (Perepolkin et al. 2023; Gilchrist 2000). The QDF is important for establishing the CLTs in section 2.2 and for the QGP models introduced in section 3.

2.1.3 Probability integral transform

An important application of the QF is using it to sample from continuous probability distributions via the probability integral transform (PIT). The PIT states that for any continuous random variable Y with CDF $F_Y(y)$, the transformed random variable $X = F_Y(Y)$ is uniformly distributed on (0, 1), or $X \sim \text{Unif}(0, 1)$. Thus if one can sample from a standard uniform distribution, one may also sample from a continuous distribution, provided that the QF of that distribution can be evaluated or reasonably approximated (Wilkinson 2018). The PIT is useful for assessing model fit via generalized residuals (Yang 2024; Cox and Snell 1968) and the calibrating of probabilistic forecasts (Gneiting et al. 2007). Herein the PIT is important for developing one version of the QGP model, and it is used as part of a distance measure for assessing QM.

2.1.4 Examples of quantile defined distributions

An example of a quantile defined distribution family that lacks a CDF in closed form is the generalized lambda distribution (GLD) family (Ramberg and Schmeiser 1974; Perepolkin et al. 2023). A special case of the GLD family is the Tukey lambda distribution (TLD) family, a one parameter distribution family introduced by Tukey (1960); Joiner and Rosenblatt (1971). The QF for the TLD is in (1) and the QDF in (2).

$$Q_\lambda(p) = \begin{cases} \frac{1}{\lambda}\left[p^\lambda - (1-p)^\lambda\right], & \lambda \neq 0 \\ \log\left(\frac{p}{1-p}\right), & \lambda = 0 \end{cases} \tag{1}$$

$$q_\lambda(p) = p^{\lambda-1} + (1-p)^{\lambda-1} \tag{2}$$

Another distribution family defined by its quantile function but without a closed form CDF is the metalog distribution (MLD) family, a generalization of the logistic distribution family (Keelin 2016). One version of the MLD is defined in (3).

$$Q_{\mathbf{a}}(p) = a_1 + a_3(p - 0.5) + a_2 \log\left(\frac{p}{1-p}\right) \tag{3}$$

These two distributions are used in section 4 for comparing different quantile matching methods.

2.2 Sample quantiles

Given a random sample from some distribution, the sample quantile function may be defined in terms of the order statistics. Hyndman and Fan review several definitions of sample quantiles used in statistical packages and state definition 2 as a general definition of the functions reviewed (Hyndman and Fan 1996).

Definition 2 For n independent observations $Y_1, ..., Y_n$ from a distribution with corresponding order statistics $Y_{(1)}, ..., Y_{(n)}$, the sample quantile function $\hat{Q}_n(p)$ may be defined as

$$\hat{Q}_n(p) = (1-\gamma)Y_{(j)} + \gamma Y_{(j+1)} \tag{4}$$

where

$$\frac{j-m}{n} \leq p < \frac{j-m+1}{n}$$

for some $m \in \mathbb{R}$ and $0 \leq \gamma \leq 1$ where γ is some function of $j = \lfloor pn+m \rfloor$, $g = pn+m-j$, and $\lfloor \cdot \rfloor$ is the floor function.

Sample quantiles are used to give a probability summary of a dataset. Sample quantiles of Markov chain Monte Carlo (MCMC) draws from a Bayesian posterior distribution are used to analyze posterior distributions, for example by using quantiles to form credible intervals. Well known asymptotic results of sample quantiles are presented below, and it may be noted that asymptotic results for quantile regression quantiles are similar (Kocherginsky et al. 2005; Koenker and Bassett 1978).

2.2.1 Sample quantile central limit theorem

The asymptotic distribution of a set of sample quantiles is given in (5), and it is the basis for the QGP model for QM in section 3. For independent draws $Y_1, ..., Y_n$ from a continuous random variable with CDF F, PDF f, and QF Q, and with a set of sample quantiles calculated by (4), theorem 1 holds. Note that in (5), $N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ refers to a multivariate normal distribution with K dimensions where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. In (6) the operator $a \wedge b$ is the minimum of a and b, with order of operations such that $a \wedge b - ab = (a \wedge b) - ab$.

Theorem 1 Sample quantile central limit theorem. *Given a vector of length K of probabilities $\boldsymbol{p} = (p_1, ..., p_K)$ and the corresponding quantile vector $\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) = (\hat{Q}_n(p_1), ..., \hat{Q}_n(p_K))$, if F is absolutely continuous for all $y \in \mathcal{Y}$ and is strictly increasing, then*

$$\sqrt{n}(\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) - \boldsymbol{Q}(\boldsymbol{p})) \xrightarrow{D} N_K(0, \boldsymbol{\mathcal{K}}) \tag{5}$$

where $\mathcal{K}$ is a $K \times K$ covariance matrix with the ij^{th} entry

$$\kappa_{ij} = \frac{p_i \wedge p_j - p_i p_j}{f(Q(p_i))f(Q(p_j))}. \tag{6}$$

Theorem 1 has been known since the mid 20th century (Cramér 1951), and a thorough though not unique proof is found in Walker (1968). The entries in the covariance matrix in (6) may equivalently be written as

$$\rho_{ij} = [p_i \wedge p_j - p_i p_j]q(p_i)q(p_j) \tag{7}$$

where q is the QDF. Another CLT in corollary 1 for the transformed set of quantiles $F(\hat{\boldsymbol{Q}}_n(\boldsymbol{p})) = (F(\hat{Q}_n(p_1)), ..., F(\hat{Q}_n(p_k)))$ can be derived from (5) by using the PIT and the Delta method (Parzen 2004).

Corollary 1 PIT transformed sample quantiles. *Under the same conditions as theorem 1, the following holds*

$$\sqrt{n}(F(\hat{\boldsymbol{Q}}_n(\boldsymbol{p})) - \boldsymbol{p}) \xrightarrow{D} N_K(0, \boldsymbol{\Gamma}) \tag{8}$$

Here the covariance matrix Γ has entries $\Gamma_{ij} = p_i \wedge p_j - p_i p_j$, making the asymptotic distribution in (8) a Brownian bridge (Chow 2009). This second result makes QM using the QGP model possible where $Q(p)$ is difficult to evaluate as is shown in the following section.

3 Quantile Gaussian process model

We consider the situation where one is given a set of data including a vector of K probabilities $\boldsymbol{p} = (p_1, ..., p_K)$ and a vector of estimated quantiles at the given probability levels $\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) = (\hat{Q}_n(p_1), ..., \hat{Q}_n(p_K))$. The set of quantiles may provide useful information about a distribution, but the information is limited and one may desire a more complete distribution. The QGP model for QM and estimating a continuous distribution based on the CLT in (5) is in (9).

$$\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) \sim N_K(\boldsymbol{Q}_\theta(\boldsymbol{p}), n^{-1}\mathcal{K}_\theta) \tag{9}$$

Here θ is an unknown parameter vector to be estimated from the data. The covariance matrix $\mathcal{K}_\theta$ has entries from the covariance function $\kappa_\theta(\cdot, \cdot)$ where

$$\kappa_\theta(p, p') = \frac{p \wedge p' - pp'}{f_\theta(Q_\theta(p))f_\theta(Q_\theta(p'))}, \quad p, p' \in (0, 1)$$

The covariance function $\kappa_\theta(\cdot, \cdot)$ initially appears as though it is difficult to evaluate and indeed can be, depending on the functional forms of f_θ and Q_θ. But where these functions belong to distributions in a location-scale family, the covariance function can be greatly simplified. Noting that the covariance is a function of $f_\theta(Q_\theta(p)) = q_\theta(p)$, the QDF, if $\theta = (\mu, \sigma)$ where μ is a location parameter and σ is a scale parameter, then $q_\theta(p) = \sigma q(p)$ where $q(p)$ is the QDF of a standard distribution having location 0 and shape 1. Staudte calls this the location invariant and scale equivariant property (Staudte 2017). For location-scale distribution families, this greatly simplifies the form of $\kappa_\theta(\cdot, \cdot)$ so that it becomes

$$\kappa_\theta(p, p') = \sigma^2[p \wedge p' - pp']q(p)q(p')$$

Thus for estimating θ one only needs to estimate σ^2. A special case of the QGP for QM of a location-scale family is the normal QGP which we define and explore further below.

Often F_θ is not easily invertible making Q_θ a complicated function which can be evaluated only via numerical optimization, which is often computationally expensive and/or inaccurate. A simple solution is to model the PIT transformed quantiles from the data rather than modeling the quantiles directly. The QGP model (9) is then reformulated to be model (10), where $\Gamma_{ij} = p_i \wedge p_j - p_i p_j$ as in (8).

$$F_\theta(\hat{\boldsymbol{Q}}_n(\boldsymbol{p})) \sim N_K(\boldsymbol{p}, n^{-1}\boldsymbol{\Gamma}) \tag{10}$$

By modeling the PIT quantiles, the only function which requires evaluating is the CDF F_θ. Here both $\boldsymbol{p}$ and $\boldsymbol{\Gamma}$ are given, and solving for the transformation $F_{\boldsymbol{\theta}}(\hat{\boldsymbol{Q}}_n(\boldsymbol{p}))$ is unnecessary for our purposes. A Bayesian fit of (10) is straightforward, requiring only that $F_{\boldsymbol{\theta}}(\hat{\boldsymbol{Q}}_n(\boldsymbol{p}))$ be evaluated as part of the acceptance ratio of a MCMC step.

3.1 Normal QGP

If the quantiles in a dataset $(\hat{\boldsymbol{Q}}_n(\boldsymbol{p}), \boldsymbol{p})$ are assumed to be calculated from a normal distribution $N(\mu, \sigma^2)$, and the desire is to estimate the parameters μ and σ by modeling the quantiles according to the QGP in (9), then one may use the model (11) for QM.

$$\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) \sim N_K\left(\mu + \sigma\boldsymbol{\Phi}^{-1}(\boldsymbol{p}), \frac{\sigma^2}{n}\boldsymbol{\Psi}\right) \tag{11}$$

Here $\boldsymbol{\Phi}^{-1}(\boldsymbol{p}) = (\Phi^{-1}(p_1), ..., \Phi^{-1}(p_K))$ is a known vector since $\boldsymbol{p}$ is given, and $\boldsymbol{\Psi}$ is a known $K \times K$ matrix with ij^{th} entry

$$\Psi_{ij} = \frac{2\pi(p_i \wedge p_j - p_i p_j)}{\exp\{-\frac{1}{2}[\Phi^{-1}(p_i)^2 + \Phi^{-1}(p_j)^2]\}}.$$

Model (11) can then be viewed as an atypical normal linear regression model. Two aspects that make this model atypical relative to the standard linear regression model are that the slope parameter σ must be greater than 0 and that σ is both a regression coefficient and part of the variance. Note also that

the sample size n is included as part of the variance. If n is known, then it may be multiplied through $\boldsymbol{\Psi}$ and ignored. If unknown, it can then be accounted for as an unknown part of the variance.

Take for instance the data $\boldsymbol{X} = \left(\mathbf{1}, \Phi^{-1}(\boldsymbol{p})\right)$ and the parameter vector $\beta = (\mu, \sigma)$. Assuming the sample size n is unknown and setting $\sigma^2/n = \gamma^2$, model (11) then becomes

$$\hat{\boldsymbol{Q}}_n(\boldsymbol{p}) \sim N_K\left(X\beta, \gamma^2\boldsymbol{\Psi}\right)$$

In this model, one simplifying assumption may be to treat σ and γ^2 separately and proceed to fit a standard linear regression model. All frequentist and Bayesian results of the linear regression model then apply, including the existence of conditionally conjugate prior distributions. The positive constraint on σ can also be dealt with without adding much complication (see (Gelman et al. 2013) pgs. 377-378). If for a certain problem it is important to make inference on the unknown n, or σ and γ^2 cannot be treated as separate, then one may estimate the two parameters by assigning appropriate prior distributions to σ and n and reverting back to (11). In section 4, we analyze a normal QGP in a simulation study where we assign independent prior distributions to σ and n, and σ is treated as both a regression coefficient and as part of the variance.

For any QGP model where the QF belongs to a location-scale family, the discussed relation to the normal linear regression model applies. The only difference in modeling becomes formulating the matrix $\boldsymbol{\Psi}$ using the proper QDF, and the data $\boldsymbol{X}$ with the proper quantile function.

3.2 Normal mixture distribution QF approximating functions

When the functional form of the true QF which one desires to estimate is unknown or has a complicated shape –such as having several modes– it may be desirous to select a more general Q_θ for approximating the unknown QF. In this section we propose two such options, namely the finite mixture (FM) distribution and the Dirichlet process mixture.

3.2.1 Finite normal mixture

For a continuous random variable distributed according to a FM distribution, the CDF takes the form of (12). Here, there are C component distributions where $c \in \{1, ..., C\}$, F_{ψ_c} is a continuous CDF with parameter ψ_c, $w_c > 0$ is a weight such that $\sum_{c=1}^{C} w_c = 1$, and $\theta = (\psi_1, ..., \psi_C, w_1, ..., w_C)$. When the QF is unknown, one may formulate Q_θ as the inverse of F_θ in (12).

$$F_\theta(x) = \sum_{c=1}^{C} w_c F_{\psi_c}(x) \tag{12}$$

When F_θ is set as a mixture distribution, it is important to consider the functional forms of the component CDFs F_{ψ_c} as well as the total number of components C. A popular decision is to set each F_{ψ_c} to be the CDF of a normal distribution. In this case $F_{\psi_c}(x) = \Phi((x - \mu_c)/\sigma_c))$ where μ_c and σ_c^2 are the mean and variance parameters of a normal distribution, $\psi_c = (\mu_c, \sigma_c)$, and (12) is termed a finite mixture of normals distribution. A common claim is that any continuous distribution may be approximated by a finite mixture of normals with a sufficiently large number of components (Peel and MacLahlan 2000; Nguyen and McLachlan 2019; Nguyen et al. 2020). This claim may be optimistic, but as shown in a simulation study in section 4, setting each F_{ψ_c} to be the CDF of a normal distribution makes for reasonable approximations of non-normal distributions.

When selecting the total number of components, one desires that C be large enough for F_θ to be sufficiently flexible to capture the unknown distribution shape. At the same time, one also desires that C is not so large that overfitting occurs or the computational burden becomes unnecessarily heavy. For a given set of quantile data, one option is to anlayze QM fits for different values of C and select an optimal number of components. A less tedious and often more robust process is to model the QF via a Bayesian nonparametric approach as outlined below.

3.2.2 Dirichlet process mixture

When the form of the QF is unknown, an alternative to estimating Q_θ via a finite mixture of normals is to estimate the function with a Dirichlet process mixture (DPM) model. The DPM as defined in Müller et al. (2015) was introduced by Ferguson (1983) and Lo (1984) and is an infinite extension of the FM distribution in (12). The DPM is defined in (13) where F_ψ is a CDF function and $\mathcal{G}$ is a probability distribution defined on Θ. In the normal mixture case, $\psi = (\mu_\psi, \sigma_\psi)$ where μ_ψ and σ_ψ^2 are the mean and variance parameters respectively.

$$F_{\mathcal{G}}(x) = \int F_\psi(x) d\mathcal{G}(\psi) \tag{13}$$

The DPM is completed by defining the hierarchical prior distributions in (14). Here $\mathrm{DP}(M, G_0)$ denotes the Dirichlet process (DP) prior with total mass parameter M, and G_0 is a base prior distribution.

$$\begin{aligned} \psi_c|\mathcal{G} &\overset{\text{iid}}{\sim} \mathcal{G} \\ \mathcal{G} &\sim \mathrm{DP}(M, G_0) \end{aligned} \tag{14}$$

The DP is an almost surely discrete distribution introduced by Ferguson (1973) as a nonparametric model, and following Sethuraman (1994) can be represented with the stick breaking construction. In this construction, $\mathcal{G}$ is represented by the infinite sum in (15) where the ψ_c are i.i.d draws from the base distribution G_0 and $\delta_a(\cdot)$ is the Dirac measure at a. We set G_0 to be the product of the prior distributions for μ_ψ and σ_ψ. The ω_c are weights defined as $\omega_c = v_c \prod_{l<c}(1 - v_l)$ where $v_c \sim \text{Beta}(1, M)$ and $\{\psi_c\}$ and $\{v_c\}$ are independent.

$$\mathcal{G}(\cdot) = \sum_{c=1}^{\infty} w_h \delta_{\psi_c}(\cdot) \tag{15}$$

Using the stick breaking construction, an equivalent representation of (13) is

$$F_{\mathcal{G}}(x) = \sum_{c=1}^{\infty} \omega_c F_{\psi_c}.$$

The DPM can be approximated by the truncated Dirichlet process mixture (TDPM) by replacing (15) with $\mathcal{G}(\cdot) = \sum_{c=1}^{C} \omega_c \delta_{\psi_c}(\cdot)$ for $C < \infty$. The TDPM was introduced by Ishwaran and Zarepour (2000) and Ishwaran and James (2002) as a close approximation to the DP with major computational advantages. The TDPM stick breaking construction is similar to that defined above with the only difference being that $v_c \sim \text{Beta}(1, M)$ for $c = 1, ..., C-1$ and $v_c = 1$. This ensures that $\sum_{c=1}^{C} w_c = 1$.

In section 4 a simulation study is included where sample quantiles are generated from "unknown" continuous distributions. In a real data analysis in section 5, given quantiles are also from unknown distributions. In these analyses, we analyze the PIT transformed QGP model (10) where F_θ is set as a FM distribution (12) or the TDPM (13) where ∞ is replaced with C. Fitting model (10) proved much faster than fitting (9), and the results were also better. In fact we found that in cases where F_θ was a simpler function, such as a normal or exponential distribution, fitting model (10) was at least as fast and the results were at least as good as fitting (9). Thus when possible, we elected to fit (10). While the methods described in this section provide flexible functions for estimating unknown distributions, we note that any function which meets the technical definition of a CDF would be appropriate to use for F_θ, see (Gasthaus et al. 2019) for an example.

3.3 Competing quantile matching methods

A number of methods already exist for QM. Here we briefly review four methods, and in section 4 we compare performance between these and the QGP. The four methods include spline interpolation (SPL), kernel density estimation (KDE), an order statistics based model (ORD), and an independent quantile model (IND). The first two of these methods are nonparametric methods and neither of which include modeling the uncertainty of the quantiles.

The SPL method was used by Gerding et al. (2023) and Shandross et al. (2024) in order to estimate a CDF function given quantile forecasts from disease outbreak forecast hubs. SPL is an interpolation where monotonic cubic splines are fit to pass through each given quantile and predict a function for all values between given quantiles and beyond the extreme values. An `R` package used for fitting SPL and which we use in section 4 is `distfromq` (Ray and Gerding 2024).

The KDE method treats given quantiles as if they constitute a random sample from a distribution and applies kernel smoothing to estimate a density function. KDE requires selecting a kernel function, often a Gaussian kernel is chosen, and applying that function to each draw of a sample. Gyamerah et al. (2020) applied KDE smoothing with an Epanechnikov kernel to quantiles estimated via three different machine learning methods used for predicting crop yield. They then combined the estimated densities into an ensemble prediction. He et al. (2016) also use KDE to estimate density forecasts from quantiles to produce energy forecasts. We use the `evmix` package for performing QM using KDE (Hu and Scarrott 2018).

One of the more common methods for QM is to select a CDF or QF function and then select model parameters which give a least squares fit to the estimated quantiles as done by Li et al. (2019). The `R` package `rriskDistributions` performs this least squares fit for some standard distributions (Belgorodski et al. 2017). Independent random error may be included to the best fit QF allowing for estimation of the variability of the quantiles (Nirwan and Bertschinger 2020). In section 4, we analyze model (16) as a proxy for the IND model to compare with other methods.

$$F_\theta(\hat{Q}(p_k)) \overset{ind}{\sim} N(p_k, \sigma_\rho^2) \tag{16}$$

Here $\hat{Q}_k$ is the estimated sample quantile at probability p_k. The major difference between model (16) and model (10) is that the error for each quantile in a set of quantiles is considered independent thus ignoring the effects of sample size and correlation suggested by the CLT in (8). Note that the parameter σ_ρ is not a part of θ, the parameter of the "true" distribution estimated by QM, but instead is meant to capture independent error among the sample quantiles.

The final method we include is the ORD model introduced by Nirwan and Bertschinger (2020). This model relies on the definition of sample quantiles being order statistics. In the ORD model, the joint distribution of a set of order statistics is the model likelihood. The likelihood of a set of order statistics from a continuous distribution is a function of both

the CDF and PDF of a continuous distribution. ORD is the method most similar to QGP and it provides exact inference of quantile uncertainty, whereas the QGP provides asymptotic inference of uncertainty. We also note that the FM and TDPM constructions in (12) and (13) for estimating the true QF may immediately be incoporated into the IND and ORD models.

For assessing QM, there are two aspects we analyze. The first is a model's ability to estimate parameters and the second is how far away a fit QM distribution is from a true distribution. Of the methods previously listed, parameter inference is only possible for the QGP, IND, and ORD models. These are also the only methods which provide uncertainty quantification for quantile values. To analyze how far an estimated distribution is from a true distribution, a distance measure must be utilized. We outline three of these distances below.

3.4 Distance measures

A number of metrics for measuring the distance between two continuous univariate distributions exist. Those we consider in this manuscript are the Wasserstein distance (WD) and a slight modification of it, the total variation (TV) or the statistical difference, and the Kullback-Leibler divergence (KLD). A brief overview of these and other metrics, some of their statistical uses and properties, and relationships between metrics is found in Gibbs and Su (2002). These metrics were used in section 4 to assess how well different QM methods approximate a true distribution.

The WD is used to measure the distance between two univariate random variables by their CDFs or QFs and has many applications in mathematics, optimization, and statistics, see (Panaretos and Zemel 2019) for a review of the WD and its use in statistics. The p-WD is defined for two continuous random variables X and Y with respective CDFs F_X and F_Y. In terms of the two CDFs, the p-WD is defined in (17), and the p-WD in terms of the corresponding QFs is defined in (18).

$$WD_p(F_X, F_Y) = \left(\int_{\mathbb{R}} |F_X(t) - F_Y(t)|^p dt \right)^{1/p} \tag{17}$$

$$WD_p(F_X, F_Y) = \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|^p dt \right)^{1/p} \tag{18}$$

We define a modified version of the WD, the uniform 1-WD (UWD1) in (19). Here we take F_X to be the CDF of a continuous random variable X. For a different continuous random variable ξ with CDF F_ξ, $F_X(\xi)$ is a random variable with support on [0, 1]. We let $F_{X,\xi}$ be the CDF of the random variable $F_X(\xi)$. Then the measure in (19) measures how close $F_X(\xi)$ is to the CDF of a standard uniform distribution with CDF $\mathcal{U}$, noting that by the PIT, if $\xi \overset{D}{=} X$, $F_X(\xi)$ is the random variable of a standard uniform distribution. The integral in (19) takes values between 0 and 1/2 (Zhanxiong 2024), and multiplying by two ensures the measure takes values between 0 and 1. This makes the UWD1 more interpretable than the WD where a value near 0 suggests ξ and X are "near" each other, and a value near 1 suggests they are far apart.

$$UWD_1(F_{X,\xi}, \mathcal{U}) = 2 \int_0^1 |F_{X,\xi}(u) - x| du \tag{19}$$

The TV, defined in (20), is a distance between distributions measured in terms of the PDFs f and g. The TV takes values between 0 and 1, and it is closely related to the distance used in Sgouropoulos et al. (2015) to measure the goodness of QM.

$$TV(f, g) = \frac{1}{2} \int_{-\infty}^{\infty} |f(x) - g(x)| dx \tag{20}$$

The final metric between distributions we consider is the KLD defined in (21). The KLD is not a true metric but has many useful properties and applications. It may be interpreted as the expected divergence if one is using f to approximate g.

$$KLD(g||f) = \int_{\mathcal{X}} g(x) \log \left(\frac{g(x)}{f(x)} \right) dx \tag{21}$$

4 Quantile matching methods comparison on simulated data

In this section we present simulation studies assessing the QM of the QGP model and compare results with those of the SPL, KDE, IND, and ORD QM methods. The QGP, IND, and ORD models were each fit via Hamiltonian Monte Carlo (HMC) sampling using the `Stan` software and the `cmdstanr` package developed and maintained by the Stan Development Team (Stan Development Team 2024; Gabry et al. 2024). Simulation studies are done to assess parameter estimation and inference for known distribution families as well as QM estimation of a distribution for both known and unknown distribution families.

4.1 Parameter estimation and quantile matching for known distribution families

Where the distribution family is known, one goal of QM may be to estimate and make inference on model parameters. To that end, we analyzed parameter estimation and inference for normal and exponential family distributions with known parameters. We analyzed the QGP model from (10), the IND model, and the ORD model. We note that model (10) allows

one to do inference for both the model parameter θ and for an unknown sample size n. The ORD model also allows for estimation of an unknown n. For the QGP and ORD models, we include modeling cases where n is known and where n is unknown and estimated. We denote the models where n is known as QGP-n and ORD-n. We also compared the distance between the true model and the predictive distributions estimated via QGP, IND, ORD, SPL, and KDE, and we include a comparison of the QGP and ORD models where the true distribution is a quantile defined distribution.

4.1.1 Normal parameter estimation

In this simulation study, we fit the ORD and IND models to quantiles estimated from independent samples from a normal distribution and compare parameter estimation with fits from the QGP model. Quantiles were simulated by first drawing a sample of size n from a known distribution then estimating K quantiles given probabilities $\{p_1, ..., p_K\}$. The ORD, IND, and QGP models were then fit to the quantiles where F_θ was set to be the CDF of a normal distribution. For each of $n \in \{50, 150, 500, 1{,}000, 5{,}000\}$ and $K \in \{3, 5, 7, 9, 11, 15, 19, 23\}$, there were 500 simulation replicates. The data were simulated from a normal distribution with mean $\mu = 4$ and standard deviation $\sigma = 3.5$. For each of QGP, ORD, and IND models, we assigned the same prior distributions to μ and σ. When n is unknown, we also assigned it a prior distribution. As one will generally have access to the quantiles being fit, it is reasonable that an informative prior, centered near where the data is centered, be assigned to μ. We took this approach herein, and for scale parameters, we assigned normal priors truncated at 0 as recommended by Gelman (2006). The prior distributions were $\mu \sim N(5, 7^2)$, $\sigma \sim N(0, 6^2)\mathbb{1}\{\sigma > 0\}$, and $n \sim N(0, 3000^2)\mathbb{1}\{n > 0\}$. For the IND model the prior on the independent parameter σ_ρ from model (16) is $1/\sigma_\rho \sim N(0, 3000^2)\mathbb{1}\{\sigma_\rho > 0\}$, which is similar to the prior on n for the QGP and ORD models. For each fit, the HMC chain was run for 60,000 draws with the first 10,000 discarded as a burn-in.

The top of figure 2 shows the 95% coverage of the posterior distribution credible intervals as n increases for μ, σ, n and the true quantiles $Q(\boldsymbol{p})$ when $K = 23$ for five models. QGP-n and ORD-n models where n is known are included along with QGP, ORD, and IND. The 95% credible intervals of the parameters were calculated by computing the 0.025^{th} and 0.975^{th} quantiles from the posterior distribution samples. The coverage percentage is calculated as the percentage of the 500 replicates for which the true parameter value was within the 95% credible interval. As expected, the nominal coverage for the QGP and ORD models is better than that of the IND model, and for the two models where n is known the coverage is better than where it is unknown, particularly for lower values of n and K. Also of note is that the coverage of 95% predictive intervals of $Q(p)$ is a bit conservative and near 100%. The coverage for the ORD models appears to be slightly better than for the QGP models, though not by much, and the difference decreases as n increases. The superior performance of the ORD model is possibly because it provides exact inference whereas the QGP model provides asymptotic inference.

To analyze how well QM methods produced predictive distributions which approximate the true distribution $N(4, 3.5^2)$, we measured the distance between the estimated predictive distributions of QGP, ORD, and IND fits to the true distribution. We also measured the distances of the SPL and KDE QM fits to the true distribution. The distances were measured via the UWD1, TV, and KLD metrics. The UWD1 for each of QGP, ORD, QGP-n, ORD-n, and IND were calculated as follows. For $\{\theta_m\}^M$ being M draws from the posterior distribution of θ where $\theta_m = (\mu_m, \sigma_m)$, the QM posterior predictive distribution was simulated by repeatedly sampling a θ_m^* from $\{\theta_m\}^M$, then sampling X_m from a distribution with CDF $F_{\theta_m^*}$. Repeating this for $M = 50,000$ times gives a QM posterior predictive sample. For the CDF F_θ where θ is the true parameter, $\xi_m = F_\theta(X_m)$ was calculated and the empirical CDF $\hat{F}_\xi$ was calculated from the sample $\{\xi_m\}^M$. F_ξ in (19) was then replaced by $\hat{F}_\xi$ to calculate UWD1. For SPL and KDE $F_\xi(x)$ was replaced by the respective estimated CDFs. The `integrate` function in `R` was used for calculating the UWD1 in (19). The TV and KLD were both estimated as the mean score over 400 posterior draws of the model parameters. The draws were selected by thinning 400 equally spaced draws from the 50,000 total. The thinning was done to save time on computation.

The bottom of figure 2 shows the UWD1, TV, and KLD metrics averaged over the 500 simulation replicates as n increases for the five QM methods. The QGP, ORD, and IND are very similar to eachother and outperform the SPL and KDE in almost every case. For UWD1, the SPL performs similarly to the parametric QM methods, but for the TV and KLD, the SPL performs much worse, especially for smaller sample sizes. Additional results in the supplementary materials show model performance under different values of quantiles K. The results show that for the QGP, ORD, and IND models, the QM fits improve drastically from $K = 3$ to $K = 5$ and typically improve as K increases. However, the results also show that when estimated quantiles in the extreme tails are included, the QM fits may suffer due to the large tail uncertainty. For example, when $K = 11$ the most extreme quantiles included are the 0.05^{th} and 0.95^{th}, but for $K = 15$, the additional quantiles include the 0.01^{th}, 0.025^{th}, 0.975^{th}, and 0.99^{th}. The effect is present for the QGP, but the negative effect decreases if the sample size n is large. These methods greatly outperform the SPL and KDE methods, which inter-

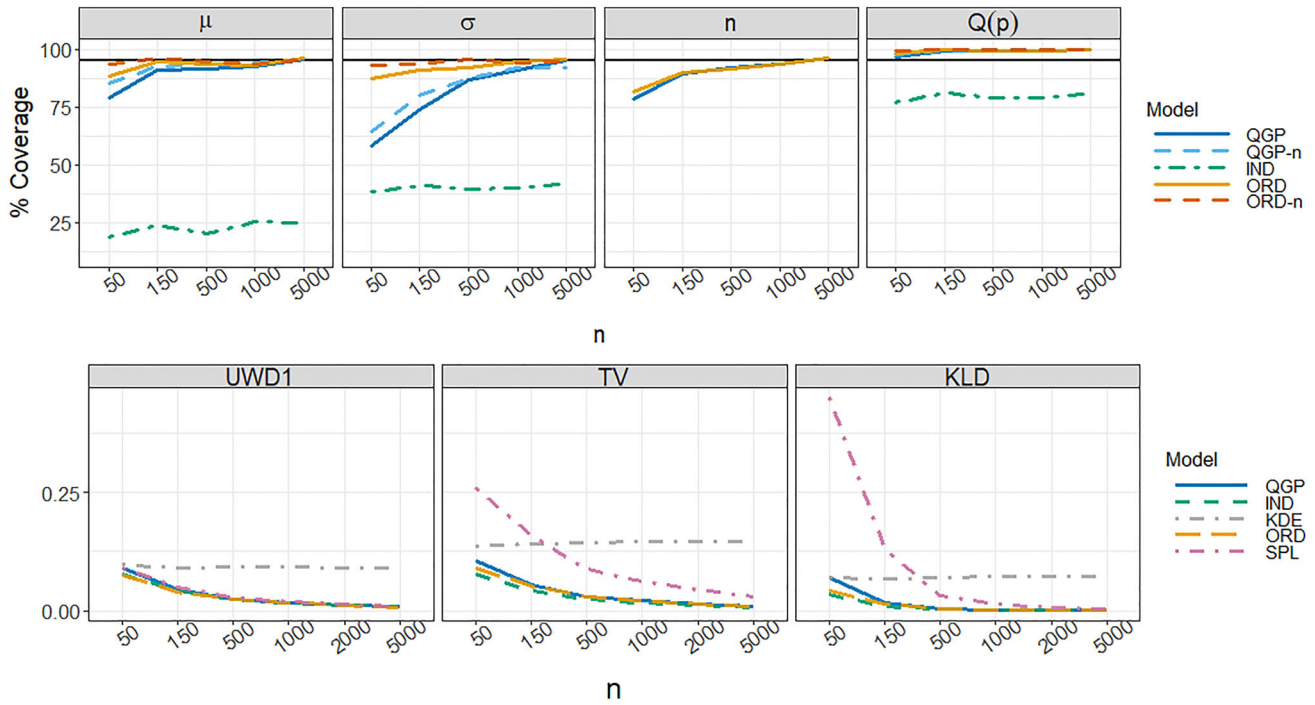

Fig. 2 Posterior coverage (top) for models of $K = 23$ quantiles calculated as the percentage of times the true parameter fell within the modeled 95% credible interval over the 500 replications. Coverage is faceted by the normal parameters μ, σ, and shown by increasing sample size (x-axis). The five models QGP, ORD, QGP-n, ORD-n, and IND are colored as shown the legend. The horizontal line (black) is at the nominal 95% level. Only QGP and ORD appear for inference for the parameter n as they are the only two models which estimate an unknown n. Distance between the true distribution and the estimated QM predictive distribution (bottom) averaged over the 500 replications. Distances include the UWD1, TV, and KLD for increasing sample size (x-axis)

estingly worsen as K increases. When fitting the Bayesian models both in the normal and in the exponential setting, the IND model tended to fit the fastest, followed by the ORD models, and finally the QGP models. The time to fit, however, was very fast, usually no more than seven seconds.

This study shows the ability of the QGP QM model to estimate and perform inference on model parameters where the true distribution family is known. It also shows the QGP's ability to produce a posterior predictive distribution which closely matches a true distribution, according to several metrics, relative to other QM methods, and it greatly outperforms SPL and KDE. The QGP also allows for making inference on the sample size n. Between parameter inference and predicting the true distribution, QGP is superior to all methods we compared it with except for the ORD model which has similar results to (or for small sample size slightly better than) the QGP. However, we feel the QGP's linear model form is easier to understand than the ORD model and is thus preferable for interpretation. Additional plots for this study and for a similar simulation study where the data were simulated from an exponential distribution are in the supplementary materials. The results of the exponential study are similar to those of the normal distribution study.

4.1.2 Quantile defined distributions

The normal and exponential families both have CDF and PDF functions which are either available in closed form or are easy to evaluate with software which is widely accessible. The study in this section was done to compare the QGP and ORD models where the distribution family was a quantile defined distribution which lacks a CDF that is easy to evaluate. The purpose of this study was to show an example where one may prefer to perform QM using the QGP rather than the ORD.

The GLD is reviewed in section 2.1.4, and a simulation carried out similarly to the previous studies was performed with GLD as the true distribution. However, QM was only done by the QGP and ORD models. Because of the relative difficulty of evaluating the CDF of the GLD and the ease of evaluating the QF, the CDF transformed QGP model in (9) instead of (10) was used. For the ORD model, evaluating the CDF was required for modeling. Evaluating the CDF of the GLD requires using an algebraic solver, and it took longer to code the ORD model in `Stan` due to a lack of readily available software which we were aware of. Once working, however, both QGP and ORD models fit the estimated quantiles well, however, the ORD usually required more time to fit than the QGP. Figure 3 shows boxplots for the 500 repli-

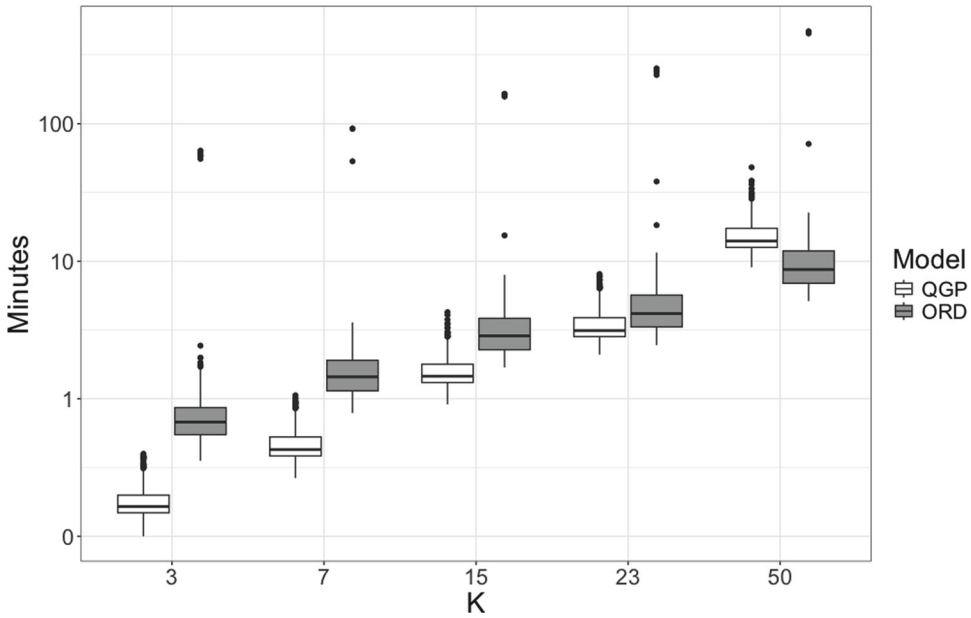

Fig. 3 Boxplots of time to fit a model in seconds (y-axis) for the 500 replicates in the simulation study for QM of the generalized lambda distribution (GLD) for $K \in \{3, 7, 15, 23, 50\}$ (x-axis). Boxplots are separated by QM methods QGP and ORD. The y-axis is on the $\log_{10}$ scale

cates for different values of K quantiles. In all cases except for $K = 50$ the QGP tends to fit much faster than ORD. Note that the y-axis is on the logarithmic scale. Even at $K = 50$ some outliers for the ORD model required hundreds of minutes to fit whereas the longest required time for fitting a QGP model was a few dozen minutes.

The MLD was also reviewed in section 2.1.4. We were able to fit a QGP model for the MLD which appeared to fit sample quantiles well, but we were unable to fit a working ORD model. The algebraic solver struggled to evaluate the CDF, and posterior sampling in `Stan` was extremely slow. When fitting finally finished, the result was a very poor fit. Perhaps with more time and effort we could have made a working model, but the QGP worked well enough that we did not feel it was worth the effort.

4.2 Matching unknown distributions

The simulation studies in this section are for the situation where the true distribution family is unknown, unimportant, or too complex to be practically evaluated but where one still wishes to recover or approximate the distribution. We selected three distributions, each with a different shape, to analyze how well the distributions can be approximated by QM using the QGP and competing methods.

From the three selected distributions, data were simulated, sample quantiles estimated, and QM methods were fit and assessed for accuracy. The distributions were the extreme value (EV) distribution with location 0 and scale 1 $EV(0, 1)$, Laplace (La) with location 0 and scale 1 $La(0, 1)$, and a two component normal mixture (MIX) $wN(-1, 0.9) + (1 - w)N(1.2, 0.6)$ where $w = 0.35$. From the three distributions, 200 replicates of $K \in \{3, 5, 11, 15, 19, 23\}$ quantiles were simulated for each of $n \in \{50, 150, 500, 1{,}000, 5{,}000\}$. QM for each replicate of simulated quantiles was done using QGP, ORD, IND, SPL, and KDE.

With the true distribution family being unknown, we set F_θ in the QGP model of (10) to be a mixture of normals distribution taking the form of either the FM (12) or the TDPM (13). For the QGP, ORD, and IND models, under the FM formulation, the parameters to be estimated are $\theta_c = \{\mu_c, \sigma_c^2, w_c\}$ for $c = 1, 2, ..., C$ and ν where $\nu = n$ for the QGP and ORD models and $\nu = 1/\sigma_\rho$ for the IND model. The prior distributions assigned were $\mu_c \overset{ind}{\sim} N(5, 7^2)$, $\sigma_c \overset{ind}{\sim} N(0, 6^2)\mathbb{1}\{\sigma_c > 0\}$, $w \sim \text{Dirichlet}(\mathbf{1}_C)$, and $\nu \sim N(0, 3000^2)\mathbb{1}\{\nu > 0\}$. Under the TDPM construction, $F_\theta = F_{\mathcal{G}}$ from (13) with the base prior distribution $G_0 = N(5, 7^2) \times N(0, 6^2)\mathbb{1}\{\sigma_c > 0\}$ and the total mass parameter $M = 1$.

Selecting the number of normal mixture components C for the FM model was based on simply analyzing the performance for several different values of C. Included in the supplementary materials are QM results of fitting a QGP model under the FM construction for several values of C. For the TDPM model, Ishwaran and James provide an $\mathcal{L}_1$ bound for the approximation of a TDPM to an infinite DPM which depends on the size of the data, the number of components C, and the total mass parameter M, see section 2.1 in Ishwaran and James (2002) for details. From their recommendation and given the specifications for K and M, $C = 20$ should provide the TDPM with a reasonably close approximation to the DPM. Figure 6 below, however, shows that QM results for $C \in \{8, 12, 20\}$ are similar when fitting the simulated data. The QM models were fit via HMC sampling using `Stan` software.

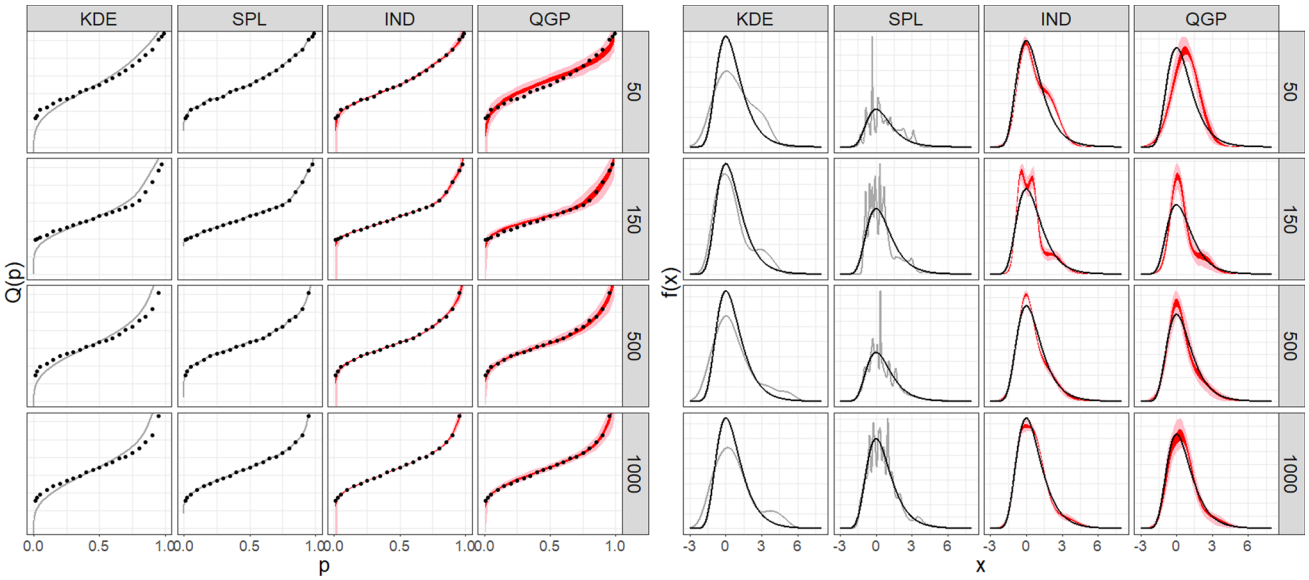

Fig. 4 QM fits of $K = 23$ quantiles by KDE, SPL, IND, and QGP for $n \in \{50, 150, 500, 1{,}000\}$. The quantiles were sampled from the extreme value distribution $EV(0, 1)$. The quantile fits (left) show the true quantiles (black) with either the QM fit line (grey) or the credible intervals of 50% (red) and 95% (pink). The estimated PDF plots (right) show the true PDF (black) with either the QM estimated PDF (grey) or the credible intervals of 50% and 95%

Often when implementing posterior sampling for a mixture distribution model, one deals with an issue called the label-switching problem. This is where for a mixture distribution parameter $\theta = \{\theta_1, ..., \theta_C\}$, the model likelihood is the same for different permutations of θ. This lack of identifiability for elements of θ makes parameter inference useless, but the predictive distribution may still be close to the true distribution (Stephens 2000). Because of this issue and because our interest is in QM, we did not consider parameter estimation as being critical and simply focus on how well the posterior predictive distribution approximates the true distribution. For estimating the posterior distribution, the HMC posterior sampling chain was run for 60,000 steps with the first 10,000 dropped as a burn-in.

Figure 4, shows examples of fits of $K = 23$ quantiles simulated from the EV distribution for different sample sizes n. Included QM methods are KDE, SPL, IND, and QGP under the TDPM construction where $C = 12$. ORD was excluded to make visualization easier, but the fits are very similar to the QGP fits. The KDE and SPL fits provide no uncertainty estimation of the quantiles, and return only a continuous function. The SPL fits show a lot of wiggle with the wiggle decreasing as n increases, whereas the KDE fits do not show as much wiggle. The KDE fits the distribution tails poorly even as n increases. The IND and QGP models provide uncertainty in the fits, but the QGP is much more conservative with wider intervals which provide superior coverage of the quantiles and the PDFs.

Figures 5 and 6 highlight differences between the FM and TDPM constructions for QGP, ORD, and IND models where sample quantiles are estimated from samples of the EV distribution. Figure 5 shows how the value C influences the goodness of QM fit in UWD1 for several values of K. The TDPM models are affected little by the size of C for each K, whereas for smaller values of K, the FM models are greatly affected. Specifically, the ORD suffers the most from setting C too high and the IND suffers the least. The results in this figure are for the EV distribution. Figure 6 shows how each of the six models performs in QM goodness of fit by UWD1 as the number of quantiles K increases from 3 to 23. For K less than 11, the goodness of fit of all models worsens rapidly but much more rapidly for FM than for TDPM. Under the FM construction the ORD suffers the most from small K. Overall, the regularization provided by the TDPM construction provides more robust fits than are provided by the FM. Results in the supplementary materials show how many components are actually used to fit the QM models when $C = 20$. In almost all cases, the TDPM selects no more than 6 components which contribute to the fit. Figures similar to 5 and 6 but for the La and MIX distributions are also included in the supplementary materials.

Figure 7 shows for $K = 23$ the goodness of QM fit for QGP, ORD, IND –under the TDPM construction– KDE, and SPL models in terms of KLD, TV, and UWD1. In nearly every case for n larger than 50, the KDE shows the worst fit by far. Otherwise the fits for QGP, ORD, IND, and SPL tend to be very similar. Figure 8 shows the empirical 95%

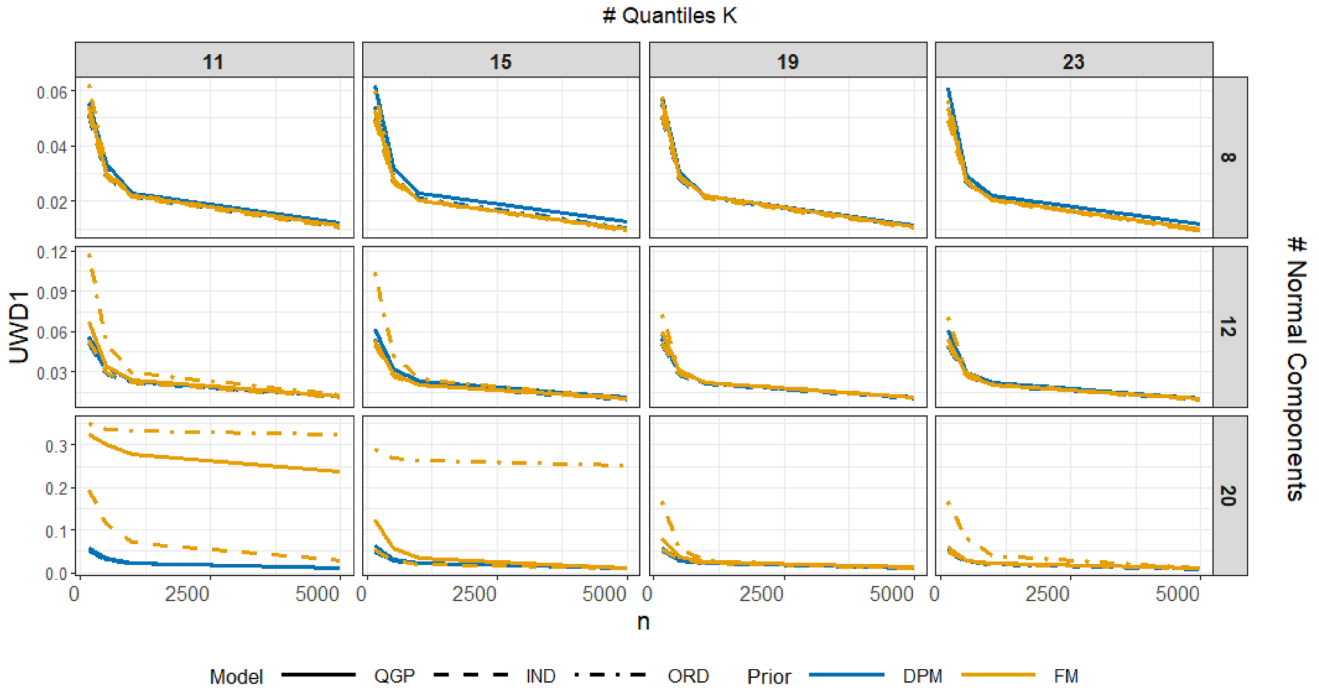

Fig. 5 UWD1 distance between extreme value distribution $EV(0, 1)$ and QM estimated distributions for increasing sample size n. Plots are faceted by number of quantiles K (columns) and number of mixture distribution components C (rows). Results are colored by mixture distribution model, and linetypes are by model type

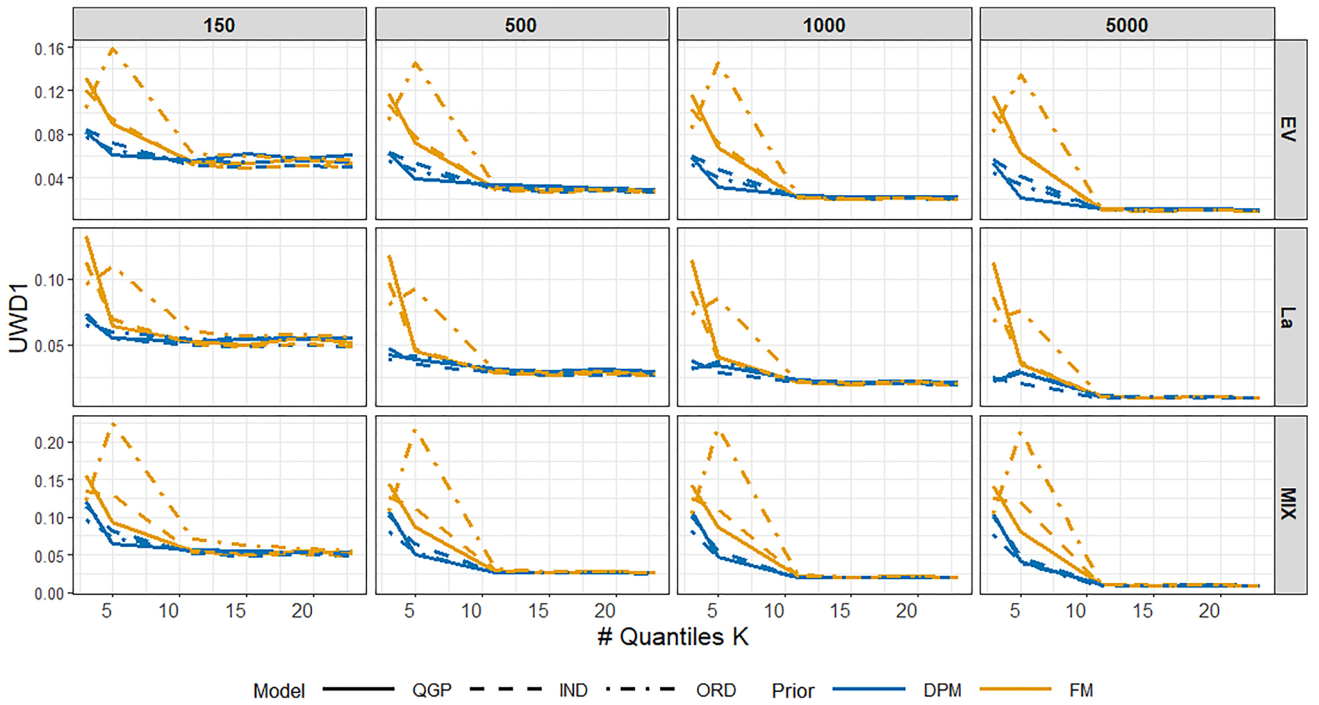

Fig. 6 UWD1 distance between extreme value distribution $EV(0, 1)$ and QM estimated distributions for increasing number of quantiles K. Plots are faceted by sample size n (columns) and by true distribution (rows). Results are colored by mixture distribution model, and linetypes are by model type

posterior interval coverage of the true quantiles for the QGP, ORD, and IND DPM models. The QGP and ORD models clearly outperform the IND model with coverage slightly too conservative approaching 100% whereas IND coverage is far below 95% especially for small K.

The simulation studies in this section show the ability of the QGP model to accurately perfom QM in various settings. Where the distribution family is known, the QGP and ORD models make more accurate parameter inference than the IND model, but in the quantile defined setting, the ORD models is much more difficult to fit making the QGP the preferrable choice. In the unknown distribution setting, the FM Bayesian QM models can suffer from model misspecification with the ORD model performing especially poorly. The regularization provided by the TDPM construction provides much more robust fits which provide better QM fits than KDE and fits generally on par with SPL. The QGP and ORD also provide reasonable quantile inference with conservative predictive intervals.

5 CDC flu forecasts analysis

In this section, the QGP and other models are used for QM of quantile forecasts targeting US flu hospitalizations. Beginning in 2013, the CDC began hosting a yearly forecast competition of the influenza outbreak in the US. This competition is known as FluSight. The flu epidemic typically begins in the fall and ends in late spring the following year, and the forecast competition lasts around 30 weeks starting in October and ending in May. FluSight involves a few dozen academic and industry research teams who each independently develop forecasts every week for predicting certain flu targets for future weeks (Biggerstaff et al. 2016). During the 2022-23 and 2023-24 seasons, the forecast targets were the 1, 2, 3, and 4-week ahead hospitalizations as reported by Health and Human Services (HHS) for the 50 US states, Puerto Rico, the District of Columbia, and the nation as a whole. The data for weekly hospitalizations of flu patients may be found at HealthData.gov (2024). The official guidelines of the 2023-24 FluSight and all submitted quantile forecasts are publicly available on Github (2024); Mathis et al. (2024).

Participating teams were free to create forecasts however they pleased, but forecasts for each target were required to consist of 23 quantiles for probability levels $\boldsymbol{p} = (0.01, 0.025, 0.5, 0.1, ..., 0.95, 0.975, 0.99)$, which were given by FluSight. Figure 9 shows examples of quantile forecasts of the same target from 12 participating teams. Each plot shows the log forecast for hospitalizations in the US for the week of January 13, 2024, and it is clear that there are major distributional differences between forecasts. We denote a quantile forecast for flu hospitalizations as $\hat{Q}^{(H)}(\boldsymbol{p}) = (\hat{Q}^{(H)}(0.01), ..., \hat{Q}^{(H)}(0.99))$. The general quantile forecast representation format allowed for forecasts to be compared by the same metric, the weighted interval score (WIS), and to be easily combined into a multi-model ensemble forecast (Mathis et al. 2024). However, under the quantile representation, the tools for scoring forecasts and for building ensemble models are limited as many of the existing tools for scoring forecasts and constructing ensembles require CDFs or PDFs (Wadsworth et al. 2023; Ranjan and Gneiting 2010). Because of these limitations, it may be desirable to approximate continuous distributions from the quantile forecasts to allow for more flexibility in scoring or ensemble building. In this section, we fit the QGP and other models for QM to forecasts submitted to FluSight during the 2023-24 season, and an analysis is made to compare the fits.

Before fitting a QGP model to a quantile forecast we transformed the quantiles $\hat{Q}^{(H)}(\boldsymbol{p})$ to $\log(\hat{Q}^{(H)}(\boldsymbol{p}) + 1)$ so that forecasts for all states were on a similar scale. The model fit was the same as the QGP fit in section 4.2, including the same prior distribution assignments. According to the official forecast competition rules, no forecast could include quantiles that were less than 0. As a result, there were forecasts with one or many quantiles equal to 0. When fitting the QGP model, the 0 values were removed so that some forecasts had $K < 23$ quantiles. Between 39 competing forecast teams, 53 locations, 29 forecast dates, and 4 horizons, QM fits were made by each of QGP, ORD, IND, SPL, and KDE on 151,041 quantile forecasts. For each forecast, the WIS was calculated for the quantile forecast and the continuous ranked probability score (CRPS) was calculated over the predictive distribution of the fit QM models.

The WIS and CRPS are both proper scoring rules, which is a class of scoring rule defined so as to keep a forecaster honest in their forecasts (Gneiting and Raftery 2007; Gneiting and Katzfuss 2014). The definition for the WIS is in (22) and is the same as that in Bracher et al. (2021). The WIS consists of the sum of multiple intervals scores (IS), the definition of which is in (23). Here α is the nominal level of an interval with l and r being the respective lower and upper bounds of the interval. R is the number of intervals in the quantile forecast, α_r is the nominal level for the r^{th} interval, and y^* is the observed value which one is attempting to forecast.

$$WIS_{0,R}(F, y^*) = \frac{1}{R + 1/2} \times (w_0 \times |y^* - median| + \sum_{r=1}^{R} \{w_r \times IS_{\alpha_r}(F, y^*)\}) \quad (22)$$

$$IS_{\alpha}(l, r; y^*) = (r - l) + \frac{2}{\alpha}(l - y^*)\mathbb{1}\{y^* < l\} + \frac{2}{\alpha}(y^* - r)\mathbb{1}\{y^* > r\} \quad (23)$$

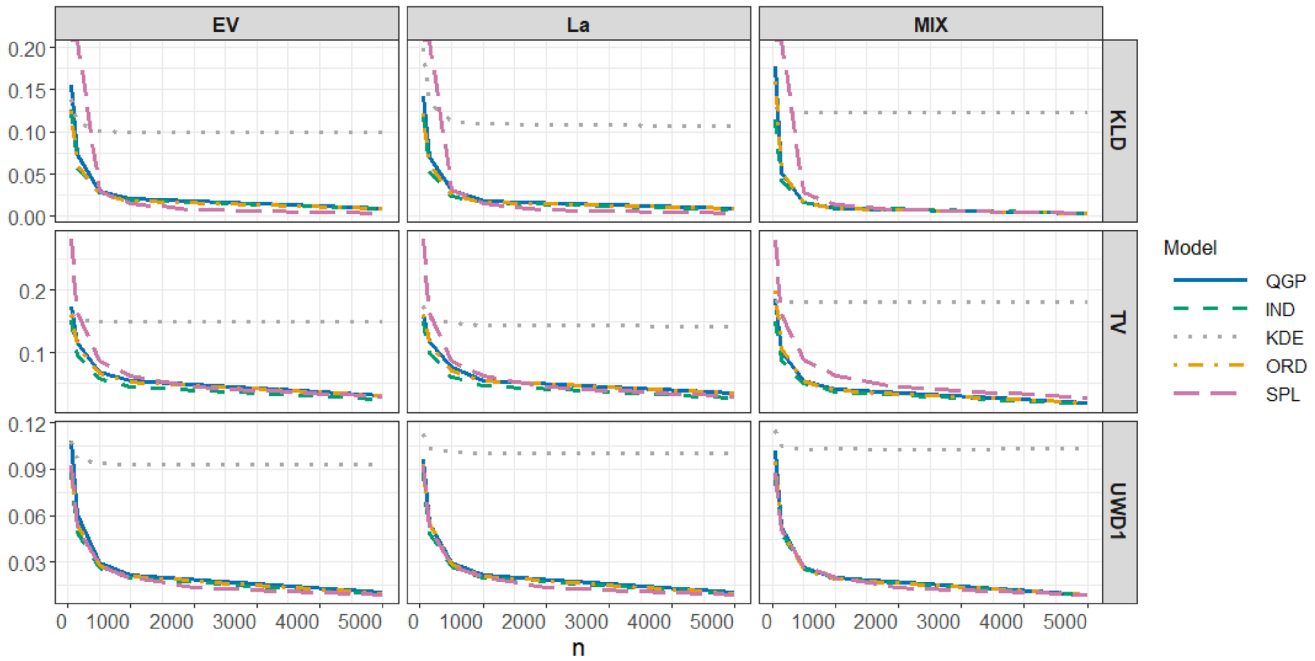

Fig. 7 Distance between the true distribution and the QM estimated distribution averaged over 200 simulation replicates and measured by KLD, TV and UWD1. Plots are faceted by distribution with the distributions being EV, La, and MIX. Lines are colored and shaped by QM method, and are drawn by increasing sample size n

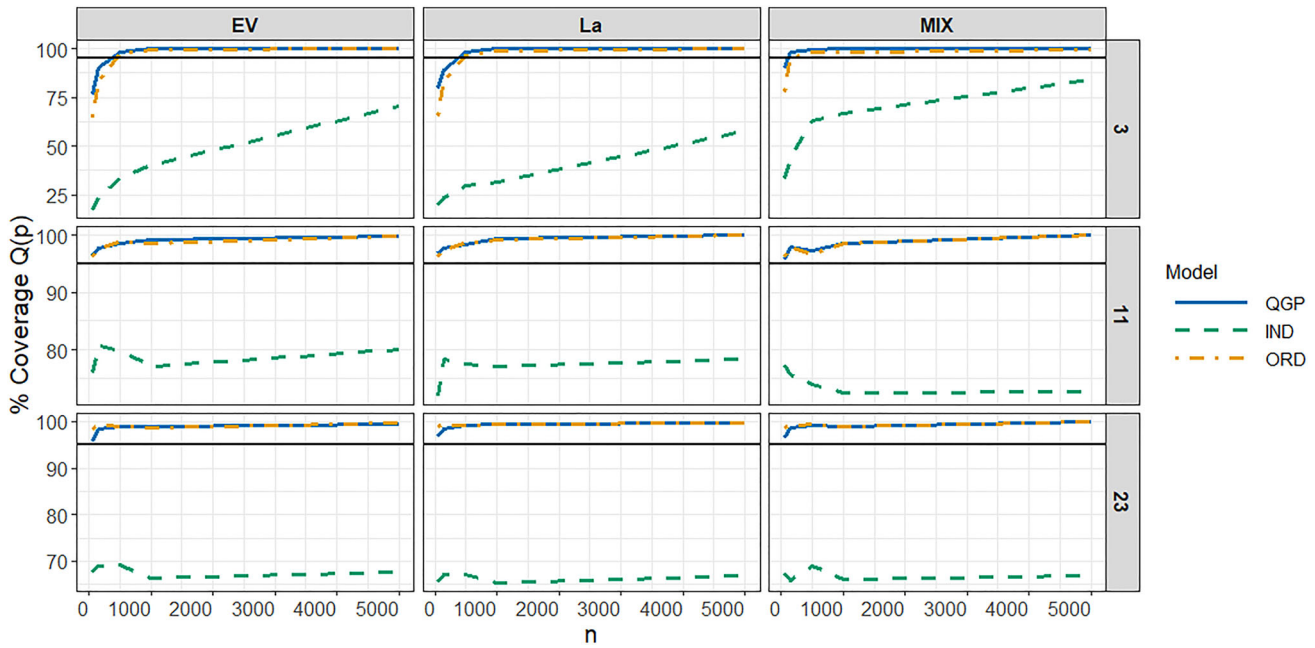

Fig. 8 Percent coverage of the true quantile values for select quantiles for 200 simulated replicates of sample quantiles averaged over all $K \in \{3, 11, 23\}$ quantiles faceted by the true EV, La, and normal mixture distributions. The plots shows percent coverage for the 95% credible intervals. The models included are QGP, ORD, and IND. The vertical bar (black) shows the nominal coverage level

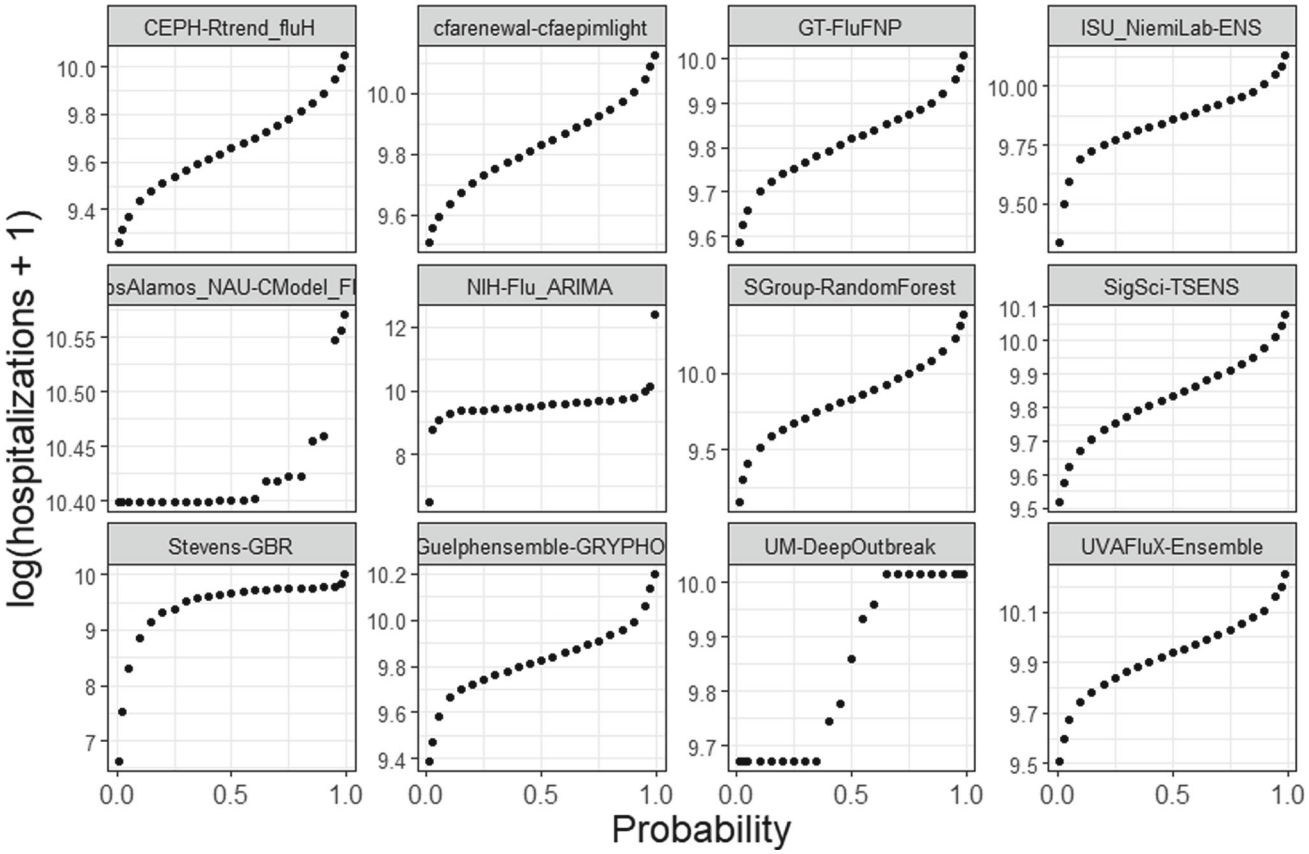

Fig. 9 1-week ahead log flu hospitalization quantile forecasts from 12 teams who participated in the 2023-24 CDC flu forecast competition. Forecasts are for the national level during the week of January 13, 2024

The CRPS is a widely used scoring rule for forecasting and is a function of the CDF of a continuous distribution, thus making it unavailable for scoring quantile functions. The CRPS is defined in (24) where F is the CDF of a forecast and y^* is the observed event one attempts to forecast. The definition is the same as in Gneiting and Katzfuss (2014).

$$\text{CRPS}(F, y^*) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}(y^* \leq x))^2 dx \tag{24}$$

Fitting continuous distributions to the quantile forecasts would allow for forecast comparison using the CRPS. The CRPS assesses a forecast across an entire distribution, including the tails, which the WIS is unable to do, giving more reason why one may want to perform QM on a quantile forecast. We fit the QGP model in (10) to the forecast competition forecasts from the 2023-24 season and compare the results of scoring the given quantile forecasts by the WIS and the CRPS calculated from the posterior predictive of the QGP model. Posterior predictive samples from the QM forecasts allow for approximate calculation of the CRPS. To calculate the WIS, we use the `evalcast R` package (Mcdonald et al. 2023), and to calculate the CRPS we use the `scoringutils` package (Jordan et al. 2019).

Figure 10 shows results of comparing the WIS from quantile forecasts and the CRPS from the fitted forecasts for QGP, ORD, IND, SPL, and KDE methods. The QGP, ORD, and IND fits were done using the TDPM construction with $C = 20$. The plots have in the x-axis the WIS and the CRPS is in the y-axis. Each point represents one forecast for any participating team, season week, and state. Unsurprisingly, the correlation between the QM CRPS and WIS is very high for each method. For all but the KDE method, there is a tendency for some CRPS values to go above the $x = y$ line suggesting the CRPS scores are enlarged due to the addition of distribution tails. The CRPS scores for the IND method also go below the $x = y$ line suggesting that the method's fit may simply be poorer than the other methods. The slightly lower correlation for the QGP compared to ORD and IND may also be a result of adding more uncertainty in the fitting of tail quantiles.

We do not know the true distributions from which each quantile forecast was made, but we can assess closeness of fit by measuring the difference between the QM quantiles and the original forecast quantiles given the FluSight probability values. Figure 11 shows the average over week mean absolute error (MAE) and mean sqaure error (MSE) between the

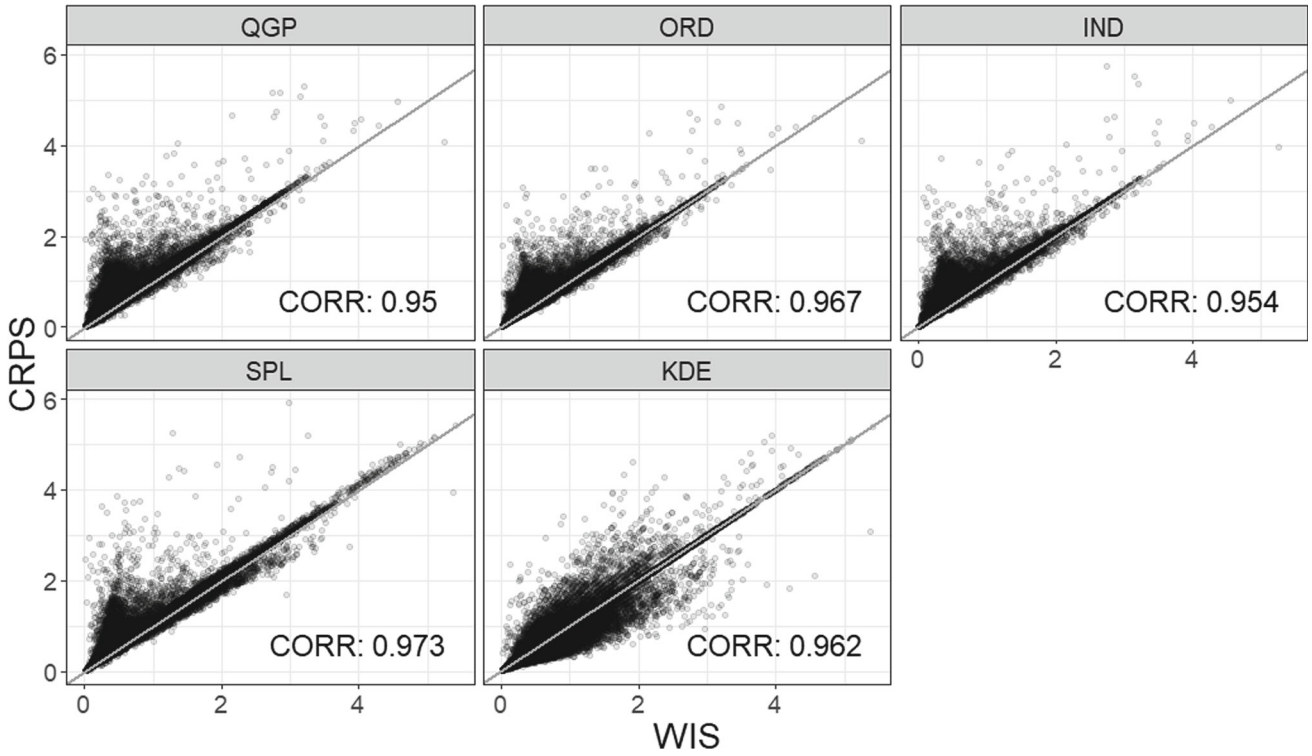

Fig. 10 Scatterplots of WIS and CRPS values for forecasts from 39 different teams for 2023-24 CDC flu forecast competition. Each plot is for a different QM method. Each point represents scores for a flu forecast for one week during the season, one state, and one competing team. Points are transparent to show where more scores tended to be. Overall linear correlation is also given in the corner of each plot

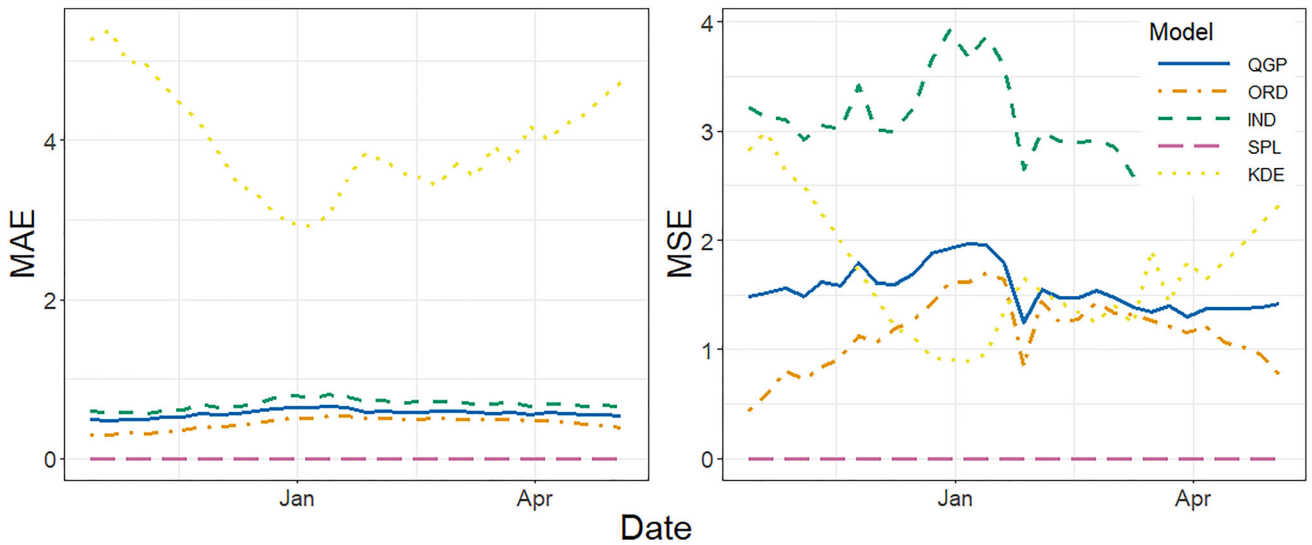

Fig. 11 Mean absolute error (MAE) and mean square error (MSE) over week for FluSight QM quantiles for each of QGP, ORD, IND, SPL, and KDE methods

the newly estimated QM quantiles and the original forecast quantiles. The MAE and MSE values for the SPL method are 0 for every week because the estimated quantile functions are interpolations of the original forecasts. For MAE, the QGP, ORD, and IND perform similiary well, and greatly outperform KDE. For MSE, the IND performs more poorly showing a lack of robustness. Figure 12 shows in minutes boxplots of time to fit the QGP, ORD, and IND models. The QGP fit time is, as expected, generally a bit slower than the

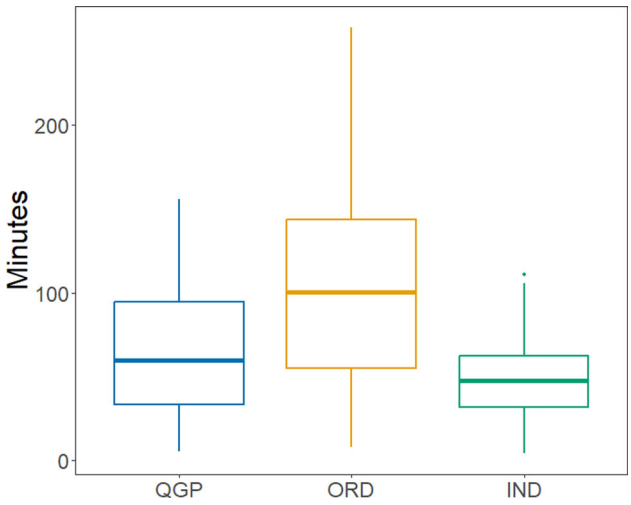

Fig. 12 Boxplots of time to fit in minutes of QM quantile forecasts for the QGP, ORD, and IND models

IND due to the Gaussian process uncertainty matrix, and the ORD has slower fittimes than the QGP.

6 Conclusion

In this manuscript we considered the situation where quantiles with known probability levels are given as data or the quantiles represent a distribution rather than raw data or a fully defined distribution. We reviewed basic properties of QFs, sample quantiles, and resulting CLTs which motivate a new model, the QGP model, for matching a set of quantiles to a continuous distribution. The QGP model is assessed for parameter estimation and inference, distribution approximation, and it is compared with other QM methods found in the literature. The QGP provides explicit and easily interpretable uncertainty quantification of QM distributions, and generally outperforms the methods to which it is compared, sometimes with the exception of the ORD. An added advantage, however, is that the QGP provides a way to perform QM for distributions defined by CDFs as well as quantile distributions where the CDF is difficult to evaluate. The time to fit the QGP model tends to be less than ORD, generally taking just over half as long to fit. For cases where the form of the CDF from which sample quantiles were estimated is unknown, we proposed modeling the unknown CDF as a mixture distribution and implemented a TDPM prior as a robust method of selecting contributing mixture distribution components.

An application of QM on quantile forecasts from the 2023-24 FluSight competition demonstrates that QM methods allow for the use of the CRPS for scoring forecasts. QGP and other QM forecasts may be scored using a variety of scoring rules unavailable to quantile forecasts, and independent forecasts may be combined using methods made only for combining distributions by CDF or PDF functions. QGP has been applied in Wadsworth and Niemi (Wadsworth and Niemi 2025) who used the QGP predictive distributions fit in section 5 to construct ensemble forecasts. In the analysis, it is shown that the QGP may offer a more robust fit to quantile distributions than other QM methods.

Approximating continuous distributions given only a set of estimated quantiles is not a new idea, but the QGP provides a method of doing so while also accurately accounting for asymptotic uncertainty in estimated or predicted quantiles. The CLTs in this manuscript are limited to the case where the quantiles are estimated given a distribution sample, however, similar results have been shown for the quantile regression case where quantiles are estimated in the presence of covariates (Kocherginsky et al. 2005; Koenker and Bassett 1978). These results could lead to additional QM modeling using the QGP model where covariates are given, and this may be more akin to the QM done by Sgouropoulos et al. (2015). Additional research for the QGP model may be on how the required distribution function is selected. Where a true model is unknown, we elected to model a distribution as a normal mixture distribution. Of course this has its limits, and using other nonparametric functions, similar to that used in Gasthaus et al. (2019) may provide needed flexibility. As long as the selected function is continuous and once differentiable, the theory herein applies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-026-10867-z.

Acknowledgements This work is partially supported by the National Science Foundation under Grant No. 2152117. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
We also acknowledge the anonymous reviewers whose comments and recommendations led to great improvements in the content and presentation of this work.

Author Contributions Spencer Wadsworth developed the methodology, implemented the computational methods, performed the analyses, and prepared the manuscript. Jarad Niemi contributed to the development and refinement of the methodology.

Data Availability All data used in this manuscript may be found at https://github.com/cdcepi/FluSight-forecast-hub

Code Availability All code used to obtain results in this manuscript, including code for accessing data, may be found at https://github.com/wadspen/quantile_fitting

Declarations

Competing interests The authors declare no competing interests.

References

Alvarez, L., Orestes, V.: Quantile mixture models: estimation and inference. Technical report, Working paper (2023)

Baran, S., Lerch, S.: Combining predictive distributions for the statistical post-processing of ensemble forecasts. Int. J. Forecast. **34**(3), 477–496 (2018)

Belgorodski, N., Greiner, M., Tolksdorf, K., Schueller, K.: rriskDistributions: fitting Distributions to Given Data or Known Quantiles. (2017). R package version 2.1.2. https://CRAN.R-project.org/package=rriskDistributions

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I.C.-H., Hickmann, K.S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., et al.: Results from the Centers for Disease Control and Prevention's predict the 2013–2014 Influenza Season Challenge. BMC Infect. Dis. **16**(1), 1–10 (2016)

Bogner, K., Liechti, K., Zappa, M.: Combining quantile forecasts and predictive distributions of streamflows. Hydrol. Earth Syst. Sci. **21**(11), 5493–5502 (2017)

Bracher, J., Ray, E.L., Gneiting, T., Reich, N.G.: Evaluating epidemic forecasts in an interval format. PLoS Comput. Biol. **17**(2), 1010592 (2021)

Casella, G., Berger, R.L.: Statistical Inference. Duxbury Advanced Series, Pacific Grove, CA, USA (2002)

CDC: CDC Extended BMI-for-age Growth Charts. https://www.cdc.gov/growthcharts/extended-bmi.htm. Accessed: 2024-10-22 (2022)

Chow, W.C.: Brownian bridge. Wiley Interdisciplinary Reviews: Computational Statistics **1**(3), 325–332 (2009)

Chung, Y., Neiswanger, W., Char, I., Schneider, J.: Beyond pinball loss: quantile methods for calibrated uncertainty quantification. Adv. Neural. Inf. Process. Syst. **34**, 10971–10984 (2021)

Cox, D.R., Snell, E.J.: A general definition of residuals. J. Roy. Stat. Soc.: Ser. B (Methodol.) **30**(2), 248–265 (1968)

Cramer, E.Y., Huang, Y., Wang, Y., Ray, E.L., Cornell, M., Bracher, J., Brennen, A., Castro Rivadeneira, A.J., Gerding, A., House, K., Jayawardena, D., Kanji, A.H., Khandelwal, A., Le, K., Niemi, J., Stark, A., Shah, A., Wattanachit, N., Zorn, M.W., Reich, N.G., Consortium, U.C.-.F.H: The United States COVID-19 forecast hub dataset. Scientific Data **9**(1), 462 (2022)

Cramér, H.: Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, USA (1951)

Cramer, E.Y., Ray, E.L., Lopez, V.K., Bracher, J., Brennen, A., Castro Rivadeneira, A.J., Gerding, A., Gneiting, T., House, K.H., Huang, Y., et al.: Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. Proc. Natl. Acad. Sci. **119**(15), 2113561119 (2022)

Dilger, M., Schneider, K., Drossard, C., Ott, H., Kaiser, E.: Distributions for time, interspecies and intraspecies extrapolation for deriving occupational exposure limits. J. Appl. Toxicol. **42**(5), 898–912 (2022)

Feldman, J., Kowal, D.R.: Nonparametric copula models for multivariate, mixed, and missing data. J. Mach. Learn. Res. **25**(164), 1–50 (2024)

Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 209–230 (1973)

Ferguson, T.S.: Bayesian density estimation by mixtures of normal distributions. In: Recent Advances in Statistics, pp. 287–302. Elsevier, New York, NY, USA (1983)

Gabry, J., Češnovar, R., Johnson, A.: cmdstanr: R Interface to 'CmdStan'. R package version 0.7.1, https://discourse.mc-stan.org (2024)

Gasthaus, J., Benidis, K., Wang, Y., Rangapuram, S.S., Salinas, D., Flunkert, V., Januschowski, T.: Probabilistic forecasting with spline quantile function RNNs. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 1901–1910 (2019). PMLR

Gelman, A.: Prior distributions for variance parameters in hierarchical models. Bayesian Anal. **1**(3), 515–533 (2006)

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis, 3rd edn. Chapman and Hall/CRC, Boca Raton, FL, USA (2013)

Gerding, A., Reich, N.G., Rogers, B., Ray, E.L.: Evaluating infectious disease forecasts with allocation scoring rules. arXiv preprint arXiv:2312.16201 (2023)

Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**(3), 419–435 (2002)

Gilchrist, W.: Statistical Modelling with Quantile Functions. Chapman and Hall/CRC, Boca Raton, FL, USA (2000)

Github: FluSight-forecast-hub. https://github.com/cdcepi/FluSight-forecast-hub. Accessed: 2024-10-22 (2024)

Gneiting, T., Katzfuss, M.: Probabilistic forecasting. Annual Rev. Stat. Appl. **1**, 125–151 (2014)

Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. **102**(477), 359–378 (2007)

Gneiting, T., Raftery, A.E., Westveld, A.H., Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Weather Rev. **133**(5), 1098–1118 (2005)

Gneiting, T., Balabdaoui, F., Raftery, A.E.: Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat Methodol. **69**(2), 243–268 (2007)

Gneiting, T., Wolffram, D., Resin, J., Kraus, K., Bracher, J., Dimitriadis, T., Hagenmeyer, V., Jordan, A.I., Lerch, S., Phipps, K., et al.: Model diagnostics and forecast evaluation for quantiles. Annual Rev. Stat. Appl. **10**, 597–621 (2023)

Gyamerah, S.A., Ngare, P., Ikpe, D.: Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov kernel function. Agricultural and Forest Meteorology **280** (2020)

He, Y., Xu, Q., Wan, J., Yang, S.: Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. Energy **114**, 498–512 (2016)

HealthData.gov: COVID-19 Reported Patient Impact and Hospital Capacity by State (RAW). https://healthdata.gov/dataset/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/6xf2-c3ie/about_data. Accessed: 2024-10-22 (2024)

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J.: Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Int. J. Forecast. **32**(3), 896–913 (2016)

Hu, Y., Scarrott, C.: Evmix: an R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation. J. Stat. Softw. **84**(5), 1–27 (2018). https://doi.org/10.18637/jss.v084.i05

Hyndman, R.J., Fan, Y.: Sample quantiles in statistical packages. Am. Stat. **50**(4), 361–365 (1996)

Ishwaran, H., James, L.F.: Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. J. Comput. Graph. Stat. **11**(3), 508–532 (2002)

Ishwaran, H., Zarepour, M.: Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. Biometrika **87**(2), 371–390 (2000)

Joiner, B.L., Rosenblatt, J.R.: Some properties of the range in samples from Tukey's symmetric lambda distributions. J. Am. Stat. Assoc. **66**(334), 394–399 (1971)

Jordan, A., Krüger, F., Lerch, S.: Evaluating probabilistic forecasts with scoringRules. J. Stat. Softw. **90**(12), 1–37 (2019). https://doi.org/10.18637/jss.v090.i12

Keelin, T.W.: The metalog distributions. Decis. Anal. **13**(4), 243–277 (2016)

Kocherginsky, M., He, X., Mu, Y.: Practical confidence intervals for regression quantiles. J. Comput. Graph. Stat. **14**(1), 41–55 (2005)
Koenker, R.: Quantile regression: 40 years on. Annual Rev Econom. **9**(1), 155–176 (2017)
Koenker, R., Bassett, G., Jr.: Regression quantiles. Econometrica **46**(1), 33–50 (1978)
Li, T., Wang, Y., Zhang, N.: Combining probability density forecasts for power electrical loads. IEEE Trans. Smart Grid **11**(2), 1679–1690 (2019)
Lo, A.Y.: On a class of Bayesian nonparametric estimates: I. Density estimates. The Annals of Statistics, 351–357 (1984)
Martin, R., Syring, N.: Direct Gibbs posterior inference on risk minimizers: construction, concentration, and calibration. In: Handbook of Statistics vol. 47, pp. 1–41. Elsevier, Cambridge, MA, USA (2022)
Mathis, S.M., Webber, A.E., León, T.M., Murray, E.L., Sun, M., White, L.A., Brooks, L.C., Green, A., Hu, A.J., Rosenfeld, R., et al.: Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations. Nat. Commun. **15**(1), 6289 (2024)
McDonald, D., Bien, J., O'Brien, M., Grabman, J., Colquhoun, S., Narasimhan, B., Tibshirani, R.: Evalcast: Tools For Evaluating COVID Forecasters. (2023). https://cmu-delphi.github.io/covidcast/evalcastR/, https://github.com/cmu-delphi/covidcast
Müller, P., Quintana, F.A., Jara, A., Hanson, T.: Bayesian Nonparametric Data Analysis, vol. 1. Springer, New York, NY, USA (2015)
Nguyen, H.D., McLachlan, G.: On approximations via convolution-defined mixture models. Commun. Stat.-Theory and Methods **48**(16), 3945–3955 (2019)
Nguyen, T.T., Nguyen, H.D., Chamroukhi, F., McLachlan, G.J.: Approximation by finite mixtures of continuous density functions that vanish at infinity. Cogent Math. Stat. **7**(1), 1750861 (2020)
Nirwan, R.-S., Bertschinger, N.: Bayesian quantile matching estimation. arXiv preprint arXiv:2008.06423 (2020)
Panaretos, V.M., Zemel, Y.: Statistical aspects of Wasserstein distances. Annual Rev. Stat. Appl. **6**(1), 405–431 (2019)
Parzen, E.: Quantile probability and statistical data modeling. Stat. Sci. **19**(4), 652–662 (2004)
Peel, D., MacLahlan, G.: Finite Mixture Models. John & Sons, New York, NY, USA (2000)
Perepolkin, D., Goodrich, B., Sahlin, U.: The tenets of quantile-based inference in Bayesian models. Comput. Stat. Data Anal. **187**, 107795 (2023)
Pohle, M.-O.: The Murphy decomposition and the calibration-resolution principle: a new perspective on forecast evaluation. arXiv preprint arXiv:2005.01835 (2020)
Ramberg, J.S., Schmeiser, B.W.: An approximate method for generating asymmetric random variables. Commun. ACM **17**(2), 78–82 (1974)
Ranjan, R., Gneiting, T.: Combining probability forecasts. J. R. Stat. Soc. Ser. B Stat Methodol. **72**(1), 71–91 (2010)
Ray, E., Gerding, A.: Distfromq: Reconstruct a Distribution from a Collection of Quantiles. (2024). R package version 1.0.4. http://reichlab.io/distfromq/
Sethuraman, J.: A constructive definition of Dirichlet priors. Statistica Sinica, 639–650 (1994)
Sgouropoulos, N., Yao, Q., Yastremiz, C.: Matching a distribution by matching quantiles estimation. J. Am. Stat. Assoc. **110**(510), 742–759 (2015)
Shandross, L., Howerton, E., Contamin, L., Hochheiser, H., Krystalli, A., Infectious Disease Modeling Hubs, C., Reich, N.G., Ray, E.L.: hubEnsembles: Ensembling methods in R. medRxiv, 2024–06 (2024)
Sherratt, K., Gruson, H., Johnson, H., Niehus, R., Prasse, B., Sandmann, F., Deuschel, J., Wolffram, D., Abbott, S., Ullrich, A., et al.: Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. Elife **12**, 81916 (2023)
Simpson, M., Holan, S.H., Wikle, C.K., Bradley, J.R.: Interpolating population distributions using public-use data: an application to income segregation using American Community Survey data. J. Am. Stat. Assoc. **118**(541), 84–96 (2023)
Stan Development Team: Stan Modeling Language Users Guide and Reference Manual, 2.34. https://mc-stan.org. Accessed: 2024-10-22 (2024)
Staudte, R.G.: The shapes of things to come: Probability density quantiles. Statistics **51**(4), 782–800 (2017)
Stephens, M.: Dealing with label switching in mixture models. J. R. Stat. Soc. Ser. B Stat Methodol. **62**(4), 795–809 (2000)
Tukey, J.W.: The practical relationship between the common transformations of percentages of counts and of amounts. Statistical techniques research group technical report **36** (1960)
Wadsworth, S., Niemi, J.: Bayesian stacking via proper scoring rule optimization using a Gibbs posterior. arXiv preprint arXiv:2509.04203 (2025)
Wadsworth, S., Niemi, J., Reich, N.: Mixture distributions for probabilistic forecasts of disease outbreaks. arXiv preprint arXiv:2310.11939 (2023)
Walker, A.: A note on the asymptotic distribution of sample quantiles. J. R. Stat. Soc. Ser. B Stat Methodol. **30**(3), 570–575 (1968)
Wang, X., Hyndman, R.J., Li, F., Kang, Y.: Forecast combinations: an over 50-year review. Int. J. Forecast. **39**(4), 1518–1547 (2023)
Wilkinson, D.J.: Stochastic Modelling for Systems Biology. Chapman and Hall/CRC, Boca Raton, FL, USA (2018)
Yang, L.: Double probability integral transform residuals for regression models with discrete outcomes. J. Comput. Graph. Stat. **33**, 1–17 (2024)
Zhanxiong: Upper bound for 1-Wasserstein distance between standard uniform and other distribution on [0, 1]. Cross Validated. (version: 2024-04-26) (2024). https://stats.stackexchange.com/q/645854