Submitted to the Annals of Applied Statistics

FAST INFERENCE FOR TIME-VARYING QUANTILES VIA FLEXIBLE DYNAMIC MODELS WITH APPLICATION TO THE CHARACTERIZATION OF ATMOSPHERIC RIVERS

BY RAQUEL BARATA[1], RAQUEL PRADO[1,*] AND BRUNO SANSÓ[1,†]

[1]*Department of Statistics, University of California Santa Cruz, rbarata@ucsc.udu;* [*]*raquel@soe.ucsc.edu;* [†]*bruno@soe.ucsc.edu*

Atmospheric rivers (ARs) are elongated regions of water vapor in the atmosphere that play a key role in global water cycles, particularly in western US precipitation. The primary component of many AR detection schemes is the thresholding of the integrated water vapor transport (IVT) magnitude at a single quantile over time. Utilizing a recently developed family of parametric distributions for quantile regression, this paper develops a flexible dynamic quantile linear model (exDQLM) which enables versatile, structured, and informative estimation of the IVT quantile threshold. A simulation study illustrates our exDQLM to be more robust than the standard Bayesian parametric quantile regression approach for non-standard distributions, performing better in both quantile estimation and predictive accuracy. In addition to a Markov chain Monte Carlo (MCMC) algorithm, we develop an efficient importance sampling variational Bayes (ISVB) algorithm for fast approximate Bayesian inference which is found to produce comparable results to the MCMC in a fraction of the computation time. Further, we develop a transfer function extension to our exDQLM as a method for quantifying non-linear relationships between a quantile of a climatological response and an input. The utility of our transfer function exDQLM is demonstrated in capturing both the immediate and lagged effects of El Niño Southern Oscillation Longitude Index on the estimation of the 0.85 quantile IVT.

1. Introduction. Motivated by the need to describe and quantify atmospheric rivers (ARs) in global climate and weather models, several techniques have been developed with the objective of detecting ARs (Rutz, Steenburgh and Ralph, 2014; Backes et al., 2015). To this end, an effective approach is to focus on the Integrated water vapor transport (IVT), a vector representing the total amount of water vapor being transported in an atmospheric column. This is increasingly used in the study of ARs because of its direct relationship with orographically induced precipitation (Neiman et al., 2009). One study in particular by Guan and Waliser (2015) presents a method for detection of ARs based on characteristics of the IVT magnitude. A key component of this and many other AR detection schemes is the thresholding of IVT magnitude at a specified quantile, specifically the 0.85 quantile in Guan and Waliser (2015). A sensitivity study found their AR detection scheme to be sensitive to the IVT threshold, thus accurate estimation of IVT quantile is crucial. However, the current approach for calculating the 0.85 quantile is unstructured, invariant from year to year, and incapable of including relevant climatological information. Motivated by the problem of modeling time-varying IVT thresholds in a way that provides richer quantitative information, we consider a class of models to describe the dynamics of a specific quantile of a time series. This prompts us to present several methodological and computational contributions for dynamic quantile modeling, and, more generally, non-Gaussian time-varying models.

Keywords and phrases: Dynamic quantile regression, asymmetric Laplace, variational Bayes, atmospheric river.

The first contribution of this paper is a novel model referred to as the extended dynamic quantile linear model (exDQLM). Our exDQLM utilizes a recently developed family of parametric error distributions for quantile regression, the extended asymmetric Laplace distribution (exAL; Yan and Kottas, 2017). In the Bayesian setting, parametric quantile regression models are almost exclusively based on the asymmetric Laplace (AL) distribution, a special case of the exAL. However the AL is known to have several drawbacks, which we discuss in detail in Section 2.1. For example, the skewness of the distribution as well as the location of the mode are fully dictated by the choice of the fixed quantile. More flexible error distributions for a single quantile have been considered extensively in the Bayesian non-parametric literature. The median regression case has been considered in the semi-parametric setting by Walker and Mallick (1999), Kottas and Gelfand (2001) and Hanson and Johnson (2002), with general quantile regression seen in Kottas and Krnjajić (2009) and Reich, Bondell and Wang (2009). Fully non-parametric nonlinear modeling of quantile regression functions is seen in Taddy and Kottas (2010) and Kottas and Krnjajić (2009), with simultaneous analysis of different quantiles seen in Reich and Smith (2013) and Tokdar et al. (2012). The literature on parametric approaches that lead to flexible quantile regression models is much less extensive. Wichitaksorn, Choy and Gerlach (2014) presents a new class of skew distributions with the AL as a special case, however the skewness remains fully determined by the fixed quantile. Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2011) present a four parameter family of asymmetric exponential power distributions for a fixed quantile, however, the mode of the distribution remains fixed at the quantile of interest. The exAL presented in Yan and Kottas (2017) overcomes these shortcomings in the current parametric methods. A detailed discussion of the properties of the exAL can be found Section 2.1. Our methods generalize the utility of the exAL to the time series setting and allow for time-varying quantile inference.

The second contribution of this paper is our importance sampling variational Bayes (ISVB) algorithm for fast, flexible inference of a time-varying quantile. Current methods for quantile regression with time-evolving parameters in both the parametric and semi-parametric approaches are almost exclusively based on the AL likelihood and check loss function, respectively (Gonçalves, Migon and Bastos, 2017; Bernardi et al., 2016; Paraschiv, Bunn and Westgaard, 2016; Koenker and Xiao, 2006). Nonparametric approaches are even more limited in the time series setting as defining likelihood functions for quantile-function-valued data is a non-trivial task (Chen et al., 2017). Further, a majority of these approaches, both parametric and non-parametric, are computationally expensive. This has prompted the development of efficient estimation algorithms. Although these alternative algorithms are faster computationally, many compromise the true underlying estimation problem in their original models. Our ISVB algorithm relieves the computational burden while preserving the underlying model structure, thus not compromising the interpretability of the resulting estimated quantile.

The final contribution of this paper is the development of a transfer function exDQLM as a method for quantifying associations that account for the cumulative effect of a time-varying input on a quantile of a response variable, e.g., a given climatological response. Most studies associating climate indices to specific atmospheric phenomena focus on simple linear associations, when in reality the relationships are much more complex. Numerous climate indices have been extensively studied as potential sources of predictability for precipitation and ARs. A few examples include the Arctic Oscillation (AO) index (Guan et al., 2013), the “Pineapple Express" (PE) index (Weller, Cooley and Sain, 2012), the Madden-Julian Oscillation (MJO) (Guan et al., 2012), the Niño3.4 index (Tziperman et al., 1998), as well as the recently developed El Niño Southern Oscillation (ENSO) Longitude Index (ELI; Williams and Patricola, 2018). In this paper, we demonstrate the practical utility of our transfer function exDQLM in capturing both immediate and lagged effects of ELI on the 0.85 quantile IVT magnitude.

The remainder of this paper is organized as follows. In Section 2, we begin with background on the exAL distribution and develop our exDQLM. We present a Markov chain Monte Carlo (MCMC) algorithm for posterior inference. We also discuss the computational challenges associated to posterior inference within this modeling framework, and provide an efficient ISVB algorithm that addresses such challenges. In Section 3, we present the results from a simulation study to compare the performance of the exDQLM with the dynamic quantile linear model (DQLM) introduced in Gonçalves, Migon and Bastos (2017), a special case of the exDQLM when the exAL is reduced to the AL. The results of the ISVB algorithm compared to the MCMC algorithm for the synthetic data are also included in Section 3. In Section 4, we develop our transfer function exDQLM with details on MCMC and ISVB algorithm augmentations for this new model. Section 5 demonstrates the utility of the transfer function exDQLM in capturing the non-linear effects of ELI on the 0.85 quantile of IVT magnitude in Santa Cruz, California. Lastly, Section 6 concludes with a final remarks and discussion of future work.

2. A flexible dynamic quantile linear model.

2.1. *Background.* As mentioned previously, Bayesian parametric quantile regression models are almost exclusively based around the asymmetric Laplace (AL) likelihood,

$$\text{AL}_p(y|\mu,\sigma) = \frac{p(1-p)}{\sigma}\exp\left\{-\frac{\rho_p(y-\mu)}{\sigma}\right\} \tag{1}$$

where $\rho_p(u) = u[p - I(u < 0)]$ is the check loss function and $I(\cdot)$ denotes the indicator function. $\sigma > 0$ is a scale parameter, $p \in (0, 1)$ is a skewness parameter typically fixed to be the quantile of interest, and the mode μ is the corresponding value of that p-th quantile. More explicitly, $\int_{-\infty}^{\mu} \text{AL}_p(y|\mu,\sigma)dy = p$. A model for quantile regression can be developed by allowing μ to be a function of covariates $\mathbf{x}$, such as $\mu = \mathbf{x}^T\boldsymbol{\beta}^p$ which yields a linear quantile regression structure. For a time-evolving y_t, a time-evolving mode $\mu_t = \mathbf{F}_t'\boldsymbol{\theta}_t^p$ yields a dynamic linear regression structure where $\mathbf{F_t}$ is the regression vector of the covariates corresponding to the parameter vector $\boldsymbol{\theta}_t^p$ at time t. In quantile regression, the parameter vectors are dependent on the fixed quantile p, however for notational simplicity we will omit the superscript p going forward.

The AL was first used for Bayesian quantile regression by Yu and Moyeed (2001) and Tsionas (2003). Kotz, Kozubowski and Podgorski (2001) presents several representations of the AL, one of which is a location-scale mixture which easily facilitates posterior simulation (Kozumi and Kobayashi, 2011). That is,

$$\text{AL}_p(y|\mu,\sigma) = \int_{\mathbb{R}^+} \text{N}(y|\mu + A(p)v, \sigma B(p)v)\text{Exp}(v|\sigma)dv \tag{2}$$

where $A(p) = \frac{1-2p}{p(1-p)}$, $B(p) = \frac{2}{p(1-p)}$ and $\text{Exp}(v|\sigma)$ denotes the exponential distribution with mean σ. Although the representation enables closed form posterior conditional distributions, the AL is known to have several limitations. Most notably, the skewness and quantile are fully dictated by choice of p, thus for a fixed quantile the skewness of the distribution is fully determined. In particular, when $p = 0.5$ the distribution is symmetric. Further, for any quantile, the mode of the distribution occurs at μ resulting in rigid tails for extreme percentiles.

To address the shortcomings of the AL parametrically, Yan and Kottas (2017) develop an extension of the AL which overcomes the restrictive aspects of the distribution. The new family of error distributions is constructed from an extension of the location-scale mixture representation of the AL in Equation (2). More specifically, replacing the Gaussian kernel in the mixture with a skew-normal distribution introduces an additional skewness parameter γ.

When $\gamma = 0$, the model reduces to the AL. The skew-normal density can also be written as a location normal mixture with mixing distribution given by the standard normal truncated to the positive real numbers, facilitating posterior simulation (Henze, 1986). Thus, the full mixture representation of the proposed family of error densities, $\text{exAL}(y|\mu, \sigma, \gamma)$, is

$$\int\int_{\mathbb{R}^+\times\mathbb{R}^+} \text{N}(y|\mu + C(p,\gamma)\sigma|\gamma|s + A(p)v, \sigma B(p)v)\text{Exp}(v|\sigma)\text{N}^+(s|0,1)dvds \tag{3}$$

where $\text{N}^+(s|0,1)$ denotes a normal distribution truncated to the positive reals with mean 0 and variance 1. Note that in this parameterization μ no longer corresponds to the p-th quantile of the distribution. To preserve the ability to fix the quantile of interest, which we will now denote to as p_0, Yan and Kottas (2017) defines the previously fixed parameter p such that $p = p(p_0, \gamma) = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$ where $g(\gamma) = 2\Phi(-|\gamma|)\exp(\gamma^2/2)$ and $\Phi(\cdot)$ denotes the standard normal CDF. The parameter γ has bounded support over the interval (L, U) where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$. Further, $A(p)$ and $B(p)$ are functions of p as defined in Equation (2) and $C(p,\gamma) = [I(\gamma > 0) - p]^{-1}$. Thus, by construction μ corresponds to the fixed quantile p_0 such that $\int_{-\infty}^{\mu} \text{exAL}(y|\mu, \sigma, \gamma)dy = p_0$.

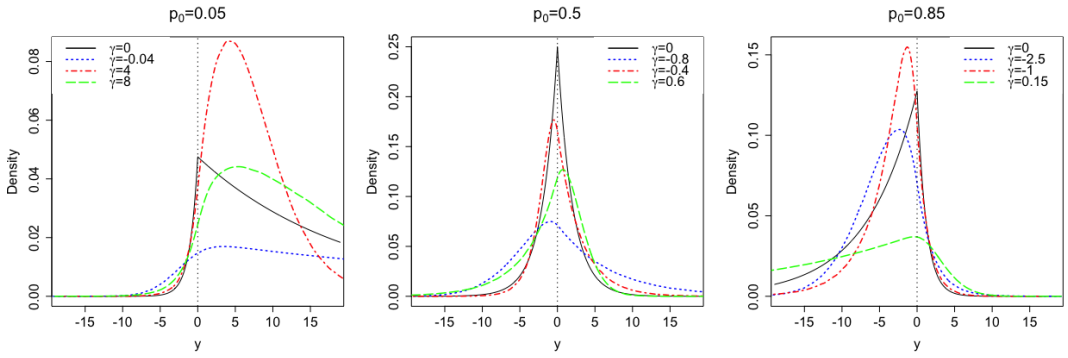

FIG 1. *Density functions of* $exAL_{p_0}(y|\gamma, \mu, \sigma)$ *defined in Equation* (3) *with* $\mu = 0$, $\sigma = 1$ *and different values of* γ *for fixed quantiles* $p_0 = 0.05$, 0.5, *and* 0.85. *The black solid line corresponds to the AL density, which is a special case when* $\gamma = 0$.

Figure 1 illustrates the flexibility induced by the additional skewness parameter γ for fixed quantiles $p_0 = 0.05$, 0.5, and 0.85. Recall γ has bounded support on the interval (L, U) which is dependent on γ, thus $\gamma = 0$ is the only skewness parameter which appears in all fixed quantiles of Figure 1. It can be seen that, when the median is fixed, γ enables both left and right skewness. The additional parameter controls the tail behavior allowing both heavier and lighter tails than the AL. Flexibility is also seen in the mode, which is no longer fixed at μ. Thus, the exAL is substantially more versatile than the AL while the hierarchical mixture representation preserves straight-forward posterior inference, making it a robust error distribution for our dynamic quantile model. For closed form representation as well as other properties of the new family of error distributions, see Yan and Kottas (2017). Note also that Yan and Kottas (2017) refers to the extension as the generalized asymmetric Laplace distribution, however we will refer to the distribution as the extended AL (exAL) to avoid confusion with the generalized asymmetric Laplace distribution defined in Kotz, Kozubowski and Podgorski (2001).

2.2. *The exDQLM.* Consider a set of time-evolving responses, y_t, for times $t = 1, \ldots, T$. For each t, a general dynamic model can be defined by

$$\text{Observation equation:} \qquad y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \epsilon_t \tag{4}$$

$$\text{System Equation:} \qquad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t. \tag{5}$$

Here $\mathbf{F}_t$ is the $q \times 1$ regression vector of the covariates corresponding to the $q \times 1$ regression parameter vector $\boldsymbol{\theta}_t$ at time t, and $\mathbf{G}_t$ is the q-dimensional evolution matrix defining the structure of the parameter vector evolution in time. We propose a extended dynamic quantile linear model (exDQLM) for inference of a single p_0-th quantile by specifying the observational errors of a dynamic linear model to be distributed from the exAL, which we denote exAL_{p_0}. That is, ϵ_t in Equation (4) are distributed independently from the exAL with quantile p_0 fixed such that $\int_{-\infty}^{0} \text{exAL}_{p_0}(\epsilon_t|0, \sigma, \gamma) = p_0$. Utilizing a Gaussian time-evolving structure on the system error vector, i.e. $\boldsymbol{\omega}_t \sim \text{N}(\mathbf{0}, \mathbf{W}_t)$ where $\mathbf{W}_t$ is the evolution variance matrix, our exDQLM model can be written

$$y_t|\boldsymbol{\theta}_t, \gamma, \sigma \sim \text{exAL}_{p_0}(\mathbf{F}_t'\boldsymbol{\theta}_t, \sigma, \gamma) \tag{6}$$

$$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t \sim \text{N}(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t) \tag{7}$$

where the normal distribution according to which $\boldsymbol{\theta}_t$ evolves is q-variate. The mixture representation of the exAL in Equation (3) can be exploited to rewrite the exDQLM as the following hierarchical model for $t = 1, \ldots, T$:

$$y_t|\boldsymbol{\theta}_t, \sigma, \gamma, v_t, s_t \sim \text{N}(y_t|\mathbf{F}_t'\boldsymbol{\theta}_t + C(p, \gamma)\sigma|\gamma|s_t + A(p)v_t, \sigma B(p)v_t) \tag{8}$$

$$v_t, s_t|\sigma \sim \text{Exp}(v_t|\sigma)N^+(s_t|0, 1) \tag{9}$$

$$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t \sim \text{N}(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t). \tag{10}$$

Here, $A(p)$, $B(p)$, $C(p, \gamma)$ are the functions of p and γ defined with Equation (3). A q-variate prior $\boldsymbol{\theta}_0 \sim \text{N}(\mathbf{m}_0, \mathbf{C}_0)$ is used at the initial stage. It is possible to place an inverse Wishart prior on the evolution covariance matrix $\mathbf{W}_t$, however for our analyses we utilize discount factors, which we discuss in Section 2.6. Yan and Kottas (2017) suggest a inverse-gamma prior for σ denoted $\text{IG}(a_\sigma, b_\sigma)$ and uniform prior for γ over the interval (L, U) denoted $\text{Uni}(L, U)$. Further discussion of the prior selection and posterior inference of σ and γ can be found in Section 2.7.

2.3. *Markov chain Monte Carlo algorithm.* The construction of the exAL through a structured mixture of normal distributions facilitates Bayesian posterior simulation using Markov Chain Monte Carlo (MCMC) with a Metropolis-Hastings (MH) step for the skewness parameter γ. Conditional on the latent variables $\mathbf{v} = \{v_1, \ldots, v_T\}$ and $\mathbf{s} = \{s_1, \ldots, s_T\}$, scale parameter σ and skewness parameter γ, the dynamic regression coefficients can be sampled using a forward filtering backwards sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). Full details of our FFBS can be found in the Appendix. MCMC posterior simulation is summarized in Algorithm 1.

Note that if a point mass prior at zero is used for skewness parameter γ, the model simplifies to the DQLM with fixed quantile $p = p_0$. The DQLM models the p-th quantile alternatively by specifying the observational errors ϵ_t in Equation (4) to be distributed independently from an AL (Gonçalves, Migon and Bastos, 2017). Similar to the exDQLM, a mixture representation can be exploited to rewrite the DQLM a hierarchical model to facilitate a MCMC algorithm for posterior inference. Such algorithm will follow very closely Algorithm 1 with a few changes: the posterior of σ reduces to an inverse gamma and all terms with skewness γ will simplify to 0. We compare the exDQLM with this special case in Section 3.

Algorithm 1: exDQLM MCMC

Initialize $\sigma^{(0)}, \gamma^{(0)}, \mathbf{v}^{(0)}, \mathbf{s}^{(0)}, \boldsymbol{\theta}_{1:T}^{(0)}$;

for *i = 0, . . . , I-1* **do**

1. Sample $\sigma^{(i+1)}|\boldsymbol{\theta}_{1:T}^{(i)}, \mathbf{v}^{(i)}, \mathbf{s}^{(i)}, \gamma^{(i)}$ from a generalized inverse Gaussian, denoted GIG$(\lambda_\sigma, \chi_\sigma, \psi_\sigma)$ where

$$\lambda_\sigma = -(a_\sigma + 1.5T), \quad \chi_\sigma = 2b_\sigma + 2\sum_{t=1}^{T} v_t^{(i)} + \sum_{t=1}^{T} \frac{(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t^{(i)} - A(p)^{(i)} v_t^{(i)})^2}{B(p)^{(i)} v_t^{(i)}},$$

$$\psi_\sigma = \sum_{t=1}^{T} \frac{(C(p)^{(i)}|\gamma^{(i)}|s_t^{(i)})^2}{B(p)^{(i)} v_t^{(i)}}.$$

2. Sample $\gamma^{(i+1)}|\boldsymbol{\theta}_t^{(i)}, \mathbf{v}^{(i)}, \mathbf{s}^{(i)}, \sigma^{(i)}$ using a Metropolis-Hastings step with a Gaussian random walk proposal on the logit scale.
3. **for** *t=1,...,T* **do**

   Sample $v_t^{(i+1)}|\boldsymbol{\theta}_t^{(i)}, s_t^{(i)}, \sigma^{(i)}, \gamma^{(i)} \sim$ GIG$(\lambda_{v_t}, \chi_{v_t}, \psi_{v_t})$ where $\lambda_{v_t} = 1/2$,

$$\chi_{v_t} = \frac{(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t^{(i)} - \sigma C(p)^{(i)}|\gamma^{(i)}|s_t^{(i)}))^2}{\sigma^{(i)} B(p)^{(i)}}, \quad \psi_{v_t} = \frac{2}{\sigma^{(i)}} + \frac{A(p)^{(i)2}}{\sigma^{(i)} B(p)^{(i)}}.$$

   end
4. **for** *t=1,...,T* **do**

   Sample $s_t^{(i)}|\boldsymbol{\theta}_t^{(i)}, v_t^{(i)}, \sigma^{(i)}, \gamma^{(i)} \sim \mathrm{N}^+(\mu_{s_t}, \sigma_{s_t}^2)$, where

$$\sigma_{s_t}^2 = \left[\frac{C(p)^{(i)2}\gamma^{(i)2}}{\sigma^{(i)} B(p)^{(i)} v_t^{(i)}} + 1\right]^{-1},$$

$$\mu_{s_t} = \sigma_{s_t}^2 \left[\frac{C(p)^{(i)}|\gamma^{(i)}|(y_t - \mathbf{F}_t'\boldsymbol{\theta}_t^{(i)} - A(p)^{(i)} v_t^{(i)})}{B(p)^{(i)} v_t^{(i)}}\right].$$

   end
5. **for** *t=1,...,T* **do**

   Sample $\boldsymbol{\theta}_t|\mathbf{v}^{(i)}, \mathbf{s}^{(i)}, \gamma^{(i)}, \sigma^{(i)}$ via FFBS. The forward part of the FFBS algorithm uses the forecast distribution $p(y_t|D_{t-1}) = \mathrm{N}(f_t, Q_t)$ where $D_{t-1} = \{y_1, \ldots, y_{t-1}\}$,

$$f_t = \mathbf{F}_t'\mathbf{a}_t + C(p)^{(i)}\sigma^{(i)}|\gamma^{(i)}|s_t^{(i)} + A(p)^{(i)} v_t^{(i)}, \quad Q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + \sigma^{(i)} B(p)^{(i)} v_t^{(i)}.$$

   end

end

2.4. *Importance sampling variational Bayes algorithm.* The addition of two latent parameters per observation in the hierarchical representation of the exDQLM makes model selection intractable even for shorter time series and the run-time of MCMC algorithm less than ideal for longer time series data. For example, a daily IVT magnitude time series at a single location from 1979 through 2015 consists of 13505 time points. With this many time points, running the proposed MCMC implies a significant computational burden and makes any model selection infeasible. We found many of the standard methods for fast inference of a non-Gaussian state space model (i.e., an expectation maximization algorithm, or state-space augmentation scheme) were unable to provide accurate inference or compromised the ability to fix the quantile of interest due to the complex structure of our exDQLM. In an effort to relieve the computational burden induced by the MCMC algorithm while preserving

the underlying parameter estimation problem, we present an efficient importance sampling variational Bayes (ISVB) algorithm.

Variational Bayes (VB) is an optimization method for fast, approximate posterior inference (Ostwald et al., 2014). Let $\boldsymbol{\xi} = \{\boldsymbol{\theta}_{1:T}, \sigma, \gamma, \mathbf{v}, \mathbf{s}\}$ denote the set of all parameters in the exDQLM. Within the VB framework, we approximate the posterior distribution $f(\boldsymbol{\xi}|\mathbf{y}_{1:T})$ with an arbitrary variational distribution $r(\boldsymbol{\xi})$ which minimizes the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) and equivalently maximizes the evidence lower bound (ELBO). For a full review of the VB approach, see Ostwald et al. (2014).

A common choice for the family of variational distributions over which we optimize the ELBO is a factorization over different sets of variables known as a mean-field approximation (Beal, 2003). In our particular model, we factorize as follows

$$r(\boldsymbol{\xi}) = r(\boldsymbol{\theta}_{1:T})r(\sigma, \gamma)r(\mathbf{v})r(\mathbf{s}). \tag{11}$$

Note, this reflects an assumption of stochastic independence between sets of variables. It has been shown that for each component of the factorization, the ELBO is maximized by the following

$$r(\boldsymbol{\xi}_c) \propto \exp\left\{\int \log f(\mathbf{y}_{1:T}, \boldsymbol{\xi}_{-c}) r(\boldsymbol{\xi}_{-c}) \mathrm{d}\boldsymbol{\xi}_{-c}\right\} \tag{12}$$

where $\boldsymbol{\xi}_c$ denotes the set of variables in the component being maximized and $\boldsymbol{\xi}_{-c}$ the variables not in that component of the partition (Tuckerman, 2010).

To implement this VB approach, we initialize the partitioned variational distributions seen in Equation (11) and iteratively maximize the ELBO using Equation (12) until convergence. For the exDQLM, the variational distribution updates at each iteration are recognizable, closed-form distributions with the exception of $r(\sigma, \gamma)$. Therefore, we propose to approximate the update of $r(\sigma, \gamma)$ at each iteration using importance sampling (IS). ISVB posterior inference for the exDQLM is summarized in Algorithm 2. For simplicity, we will use the following short-hand notation where $\boldsymbol{\xi}_c$ and $\boldsymbol{\xi}_{-c}$ are as defined in Equation (12)

$$\langle g(\boldsymbol{\xi}_c) \rangle = \int \log g(\boldsymbol{\xi}_c) r(\boldsymbol{\xi}_{-c}) \mathrm{d}\boldsymbol{\xi}_{-c}.$$

The resulting closed form integrals as well as complete details of the Forward Filtering Backwards Smoothing (FFBSm) and IS algorithms used to update the variational distributions can be found in the Appendix.

2.5. *Comparison criteria.* To evaluate the quantile inference and predictive performance of the exDQLM, we define several measures for comparison. Consider first the setting in which we know the true p_0 quantile, μ_t^{true}, for all t. To measure the fit of the quantile estimates, we compute the 95% credible interval (CrI) for the mean check loss (MCL),

$$\sum_t \rho_{p_0}(\mu_t^{true} - \mathbf{F}_t' \tilde{\boldsymbol{\theta}}_t)/T, \tag{13}$$

where $\tilde{\boldsymbol{\theta}}_t$ is a sample from the posterior distribution.

To evaluate the predictive ability of the exDQLM, we consider the Gelfand and Ghosh (1998) posterior predictive loss criterion (pplc) with check loss function ρ_{p_0}. Given the posterior replicate distribution of y_t, $p(y_t^{rep}|D_T)$,

$$\text{pplc} = \sum_t \mathbb{E}[\rho_{p_0}(y_t^{obs} - y_t^{rep})|D_T] \tag{14}$$

where $D_T = \{y_1, \ldots, y_T\}$.

Algorithm 2: exDQLM ISVB

Set $k=0$ and initialize $r^0(s_t)$, $r^0(v_t)$, $r^0(\boldsymbol{\theta}_t)$ and $r^0(\sigma,\gamma)$;
while *convergence has not been achieved* **do**

1. **for** *t=1,...,T* **do**

Update $r^{(k+1)}(v_t) = \text{GIG}(\lambda_{v_t}^{(k+1)}, \chi_{v_t}^{(k+1)}, \psi_{v_t}^{(k+1)})$ where $\lambda_{v_t} = 1/2$

$$\chi_{v_t} = \left\langle \frac{1}{\sigma B(p)} \right\rangle^{(k)} (y_t^2 - 2y_t \langle \mathbf{F}_t' \boldsymbol{\theta}_t \rangle^{(k)} + \langle (\mathbf{F}_t' \boldsymbol{\theta}_t)^2 \rangle^{(k)}) - 2\langle s_t \rangle^{(k+1)} \left\langle \frac{C(p)|\gamma|}{B(p)} \right\rangle^{(k)} (y_t - \langle \mathbf{F}_t' \boldsymbol{\theta}_t \rangle^{(k)}) + \langle s_t^2 \rangle^{(k+1)} \left\langle \frac{C(p)^2 \sigma |\gamma|^2}{B(p)} \right\rangle^{(k)}$$

$$\psi_{v_t} = 2\left\langle \frac{1}{\sigma} \right\rangle^{(k)} + \left\langle \frac{A(p)^2}{\sigma B(p)} \right\rangle^{(k)}.$$

end

2. **for** *t=1,...,T* **do**

Update $r^{(k+1)}(s_t) = \text{N}^+(\mu_{s_t}^{(k+1)}, \sigma_{s_t}^{2\,(k+1)})$, where

$$\sigma_{s_t}^{2(k+1)} = \left[\left\langle \frac{C(p)^2 \sigma \gamma^2}{B(p)} \right\rangle^{(k)} \left\langle \frac{1}{v_t} \right\rangle^{(k)} + 1 \right]^{-1}$$

$$\mu_{s_t}^{(k+1)} = \sigma_{s_t}^2 \left[(y_t - \langle \mathbf{F}_t' \boldsymbol{\theta}_t \rangle^{(k)}) \left\langle \frac{1}{v_t} \right\rangle^{(k)} \left\langle \frac{C(p)|\gamma|}{B(p)} \right\rangle^{(k)} - \left\langle \frac{C(p)|\gamma| A(p))}{B(p)} \right\rangle^{(k)} \right].$$

end

3. **for** *t=1,...,T* **do**

Update the smoothed distribution $r^{(k+1)}(\boldsymbol{\theta}_t) = \text{N}(\mathbf{m}_t^s, \mathbf{C}_t^s)$ using a FFBSm algorithm with forecast distribution $r^{(k+1)}(y_t|D_{t-1}) = \text{N}(f_t, Q_t)$ where $D_{t-1} = \{y_1, \ldots, y_{t-1}\}$,

$$f_t = \mathbf{F}_t' \mathbf{a}_t + \left[\left\langle \frac{C(p)|\gamma|}{B(p)} \right\rangle^{(k)} \langle s_t \rangle^{(k+1)} + \left\langle \frac{A(p)}{\sigma B(p)} \right\rangle^{(k)} \Big/ \left\langle \frac{1}{v_t} \right\rangle^{(k+1)} \right] \Big/ \left\langle \frac{1}{\sigma B(p)} \right\rangle^{(k)}$$

$$Q_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \left[\left\langle \frac{1}{v_t} \right\rangle^{(k+1)} \left\langle \frac{1}{\sigma B(p)} \right\rangle^{(k)} \right]^{-1}.$$

end

4. Update $r^{(k+1)}(\sigma,\gamma)$ using IS with proposal distributions $t_{(L,U)}(0,1)$ and $t_{(0,\infty)}(m_\sigma, v_\sigma)$ for γ and σ, respectively, where m_σ and v_σ denote the mean and variance of the prior distribution on σ. Further details of this IS step can be found in the Appendix.
5. Set $k = k+1$.

end

Lastly, as in Huerta, Jiang and Tanner (2003) and Prado, Molina and Huerta (2006) we use the one-step-ahead predictive distribution function introduced by Rosenblatt (1952) as a model diagnostic tool. If we define $\boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}} = \{\mathbf{v}, \mathbf{s}, \sigma, \gamma\}$, this distribution is given by

$$u_t = \Phi(y_t|D_{t-1}, \boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}}) = \Pr(Y_t \leq y_t | D_{t-1}, \boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}}). \tag{15}$$

Here u_t defines an independent sequence which is uniformly distributed on the interval $(0,1)$ (Rosenblatt, 1952). Conditional on $\boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}}$, the predictive distribution of y_t is normally distributed with mean f_t and variance Q_t seen in Algorithms 1 and 2, thus $u_t = \Phi(y_t|f_t, Q_t)$ where Φ denotes the normal CDF. We can obtain a point estimate for u_t conditionally on a posterior summary of $\boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}}$ from the MCMC or ISVB posterior samples. A diagnosis of the model performance can be done through the correlation of the estimated sequence $\{\hat{u}_t\}$ and their distribution shape. More specifically, transforming the values with a standard normal inverse CDF allows for examination of the distribution shape with a normal QQ-plot. To quantify the divergence from the standard normal distribution, we consider the KL divergence $\text{KL}(h,\phi) = \int_{-\infty}^{\infty} h(x) \log \frac{h(x)}{\phi(x)} \text{d}x$. We estimate the integrals using the numerically approximated densities of our transformed sample, which we denote h, and the standard normal density, ϕ.

2.6. *Discount factor selection.* A standard approach which allows us to specify the time-evolving covariance matrices $\mathbf{W}_t$ is the use of discount factors. (West and Harrison, 2006). Selection of discount factors is typically done by optimizing some model checking criterion. This criterion-based selection approach requires posterior inference for each set of discount factors which can become computationally expensive very quickly especially for large T. The ISVB algorithm makes this criterion-based selection approach computationally feasible. We propose selecting the discount factor, or combination of discount factors (see West and Harrison, 2006 for details on component discounting), that minimize the KL divergence calculated from the one-step-ahead predictive distribution functions u_t estimated using the MAP estimates of $\boldsymbol{\xi}_{-\boldsymbol{\theta}_{1:T}}$ from the ISVB algorithm, as discussed in Section 2.5. Fixing the discount factors within each quantile ensures consistent signal-to-noise ratios between differing models and algorithms.

2.7. *Comments on prior selection and inference of σ and γ.* We find that using a proper prior distribution on the skewness parameter γ facilitates reliable posterior inference by alleviating some of the inferential problems known to arise when utilizing the skew-normal family (Liseo and Loperfido, 2006). To this end, we implement a weakly informative Student-t distribution truncated to the interval (L,U) as the prior for γ, i.e. $\gamma \sim t_{(L,U)}(0,1)$ with 1 degree of freedom, in contrast to the flat prior suggest by Yan and Kottas (2017). Further, interaction between the parameters σ and γ can complicate posterior inference, particularly for extreme quantiles. Joint sampling of σ and γ with a random-walk MH step facilitates mixing and convergence within the MCMC algorithm.

The interaction between σ and γ is also prevalent within the ISVB algorithm, which commonly results in the variational distributions getting stuck in local optima. To facilitate fast posterior estimation with the ISVB algorithm, we place a point-mass prior on σ at the posterior mode of σ estimated from the DQLM. That is, for any fixed quantile of interest p_0, we set the prior on σ to be $\delta_{\hat{\sigma}^{p0}_{\gamma=0}}(\sigma)$ where δ denotes the Dirac delta function and $\hat{\sigma}^{p0}_{\gamma=0}$ is the posterior mode of σ under the DQLM for the p_0 quantile. Although this results in different posterior summaries for the skewness parameter γ from the ISVB algorithm than the MCMC algorithm, the posterior error distributions and modes (and therefore quantile estimates) are comparable.

3. Simulation study. We present results from a simulation study to compare the added flexibility of the exDQLM to the special case of the DQLM and the standard DLM for three different quantiles; 0.05, 0.50 and 0.85. For the underlying data-generating distributions, we consider three scenarios with different types of tail behavior and skewness.

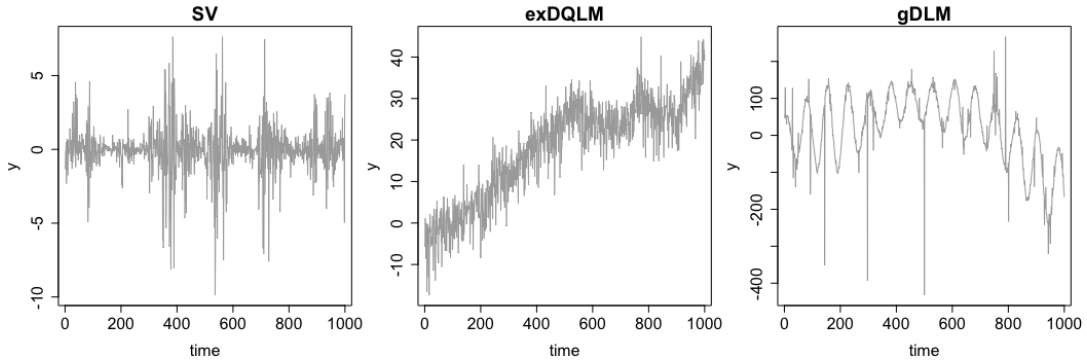

FIG 2. *Simulation study datasets. From left to right are the data simulated from the Stochastic Volatility (SV) model, exDQLM, and generalized DLM (gDLM) described in Section 3.*

Dataset 1: Stochastic Volatility. Stochastic Volatility (SV) models are commonly used to analyze returns (Kastner, 2016). These models are stochastic processes in which the log-variance is randomly distributed and follow an autoregressive structure. The SV model for $t = 1, \ldots, 1000$, where y_t denotes the return at time t, can be written as follows,

$$y_t | x_t \sim \text{N}(0, x_t) \tag{16}$$

$$\log x_t | \log x_{t-1}, \mu, \phi, \sigma \sim \text{N}(\mu + \phi\{\log x_{t-1} - \mu\}, \sigma^2) \tag{17}$$

$$\log x_0 | \mu, \phi, \sigma \sim \text{N}(\mu, \sigma^2/(1-\phi)) \tag{18}$$

Under SV models, posterior inference of the return distributions requires simulation of a latent time-varying process which can sometimes be difficult. To explore the posterior performance of our exDQLM with respect to this more complexly structured data, we generate time series of length 1000 from a SV model using the `stochvol` package in `R` with the level (μ), persistence (ϕ), and volatility (σ) of the log-variance to be 0, 0.95 and 0.5, respectively. We will utilize the exDQLM to model the $p_0 = 0.05, 0.5$, and 0.85 posterior quantiles with a first-order polynomial evolution structure,

$$y_t \sim \text{exAL}_{p_0}(\theta_t, \sigma, \gamma)$$
$$\theta_t \sim \text{N}(\theta_{t-1}, W_t).$$

Dataset 2: exDQLM. Next, we consider synthetic data from an exDQLM, for $t = 1, \ldots, 1000$,

$$y_t \sim \text{exAL}_{0.85}(\mathbf{F}'\boldsymbol{\theta}_t, \sigma, \gamma)$$
$$\boldsymbol{\theta}_t \sim \text{N}_2(\mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}).$$

With a slight abuse of notation, here $\mathbf{F}'\boldsymbol{\theta}_t$ denotes the $p_0 = 0.85$ quantile of the synthetic dataset at time t. The components $\mathbf{F}$ and $\mathbf{G}$ are specified with a second-order polynomial trend (West and Harrison, 2006),

$$\mathbf{F} = (1, 0)', \quad \mathbf{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

with

$$\mathbf{W} = \begin{bmatrix} 0.01 & 0.001 \\ 0.001 & 0.001 \end{bmatrix},$$

$\sigma = 1$, and skewness parameter $\gamma = -2.5$ causing the mode to be below the $p_0 = 0.85$ quantile for all t. We model each quantile with the same second-order polynomial evolutionary structure.

Dataset 3: Generalized DLM. For a dataset with extreme observations, we generate data from a non-Gaussian DLM (West and Harrison, 2006), for $t = 1, \ldots, 1000$,

$$y_t \sim \text{Cauchy}(\mathbf{F}'\boldsymbol{\theta}_t, \tau^2)$$
$$\boldsymbol{\theta}_t \sim \text{N}_4(\mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}).$$

Again, with a slight abuse of notation, here $\mathbf{F}'\boldsymbol{\theta}_t$ denotes the mean of the synthetic dataset at time t. The components $\mathbf{F}$ and $\mathbf{G}$ are specified with a second-order polynomial and Fourier form represented seasonality at frequency $\omega = 2\pi/75$ (West and Harrison, 2006),

$$\mathbf{F} = (1, 0, 1, 0)', \quad \mathbf{G} = \text{block-diag}\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \right\}$$

with $\tau^2 = 4$ and evolution covariance

$$\mathbf{W} = \text{block-diag}\left\{ \begin{bmatrix} 0.05 & 0.01 \\ 0.01 & 0.001 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right\}.$$

Again, we model the quantiles with the same trend and seasonal evolution structure.

3.1. *Results.* For all models, we set conjugate prior $\boldsymbol{\theta}_0 \sim \text{N}(\mathbf{m}_0, \mathbf{C}_0)$ and priors for σ and γ as discussed in Section 2.7. Table 1 reports the posterior results, with bold text indicating the model supported by the comparison criteria detailed in Section 2.5.

Overall, the exDQLM out-performs the standard DLM and DQLM. The exDQLM is favored with a lower MCL for all cases in which the true quantile is known with two exceptions where the MCL of the exDQLM is comparable to the MCL of the DQLM; the medians of the Stochastic Volatility and generalized DLM, both symmetric datasets. The one-step-ahead predictions assessed using the KL divergence also overwhelmingly favor the exDQLM for all quantiles except 0.5 of the symmetric datasets, in which the exDQLM is again comparable to the DQLM. Similarly, the Gelfand and Ghosh pplc favors the exDQLM for all extreme quantiles (0.05 and 0.85), and again is comparable to the DQLM for the median in the two cases for which the data is symmetric.This parallel between the DQLM and exDQLM for the median of the symmetric datasets is unsurprising, as the exAL reduces to the AL at the 0.5 quantile in the case when the distribution is symmetric. However, we find the pplc for the 0.5 quantile of the exDQLM dataset is also comparable between all three models and even slightly favors the DLM. The exAL distribution used to generate the dataset (seen in Figure 1) is only slightly left-skewed around the median with thin tails, therefore it is not unreasonable a normal observational distribution is able to produce equitable predictive results. With this exception, the exDQLM outperforms the other models in all cases for which the underlying distribution is skewed or the quantile of interest does not align with the mode. Figure 3 supports these findings where it can be seen, due to the rigidity of their observational error distributions, the estimated dynamic quantiles of the DLM and DQLM are more affected by extreme observations than the exDQLM particularly for extreme quantiles (i.e. 0.05). These results highlight the two main advantages of our exDQLM for parametric quantile inference on non-Gaussian dynamic models; robust estimation of any dynamic quantile and superior predictive accuracy for non-standard distributions.

The assessment criteria also illustrate the comparable accuracy of the ISVB exDQLM algorithm to the MCMC exDQLM, but with a fraction of the computational time (see Table 1). Both algorithms were implemented in the R programming language on a computer with a 2.5 GHz Intel Core i5 processor. The point-mass prior on parameter σ results in different

Model	γ	σ	MCL	KL	pplc	time
Stochastic Volatility ($\delta = 0.92, 0.99, 0.87$ for $p_0 = 0.05, 0.50, 0.85$, respectively)						
$p_0 = 0.05$						
DLM	—	—	—	3.464	839.7	0.1
DQLM	—	0.178 (0.17,0.19)	—	3.943	2012.6	19.6
exDQLM/MCMC	4.358 (3.96,4.77)	0.303 (0.29,0.32)	—	**3.641**	**697.5**	46.2
exDQLM/ISVB	6.980 (6.93,7.03)	0.178 (fixed)	—	3.911	774.3	4.1
$p_0 = 0.50$						
DLM	—	—	0.035	3.586	852.6	0.1
DQLM	—	0.486 (0.46,0.52)	**0.011**	**3.484**	**749.7**	21.3
exDQLM/MCMC	0.037 (-0.03,0.09)	0.487 (0.46,0.52)	0.012	3.498	757.4	48.7
exDQLM/ISVB	0.017 (-0.02,0.05)	0.486 (fixed)	0.012	3.524	769.6	0.8
$p_0 = 0.85$						
DLM	—	—	—	3.679	829.5	0.1
DQLM	—	0.299 (0.28,0.32)	—	3.979	1133.3	19.5
exDQLM/MCMC	-1.384 (-1.53,-1.18)	0.401 (0.376,0.423)	—	**3.667**	**740.6**	43.9
exDQLM/ISVB	-2.610 (-2.65,-2.57)	0.299 (fixed)	—	3.652	785.5	2.1
exDQLM (trend $\delta = 0.93$ for all p_0)						
$p_0 = 0.05$						
DLM	—	—	—	4.496	2282.7	0.1
DQLM	—	0.456 (0.43,0.48)	—	3.968	4276.3	18.6
exDQLM/MCMC	5.139 (4.75,5.52)	0.854 (0.81,0.94)	—	**3.910**	**1692.6**	46.2
exDQLM/ISVB	8.058 (8.02,8.09)	0.456 (fixed)	—	4.076	1799.9	4.6
$p_0 = 0.50$						
DLM	—	—	—	4.169	**2287.6**	0.1
DQLM	—	1.584 (1.49,1.68)	—	3.578	2343.8	19.0
exDQLM/MCMC	0.362 (0.27,0.48)	1.377 (1.20,1.53)	—	**3.548**	2328.8	45.9
exDQLM/ISVB	0.25 (0.21,0.28)	1.584 (fixed)	—	3.674	2345.0	0.7
$p_0 = 0.85$						
DLM	—	—	0.214	4.384	2283.3	0.1
DQLM	—	0.871 (0.82,0.93)	0.252	3.939	2963.1	18.9
exDQLM/MCMC	-2.514 (-2.72,-2.39)	0.967 (0.87,1.03)	**0.195**	**3.831**	**1489.7**	46.2
exDQLM/ISVB	-2.68 (-2.71,-2.65)	0.871 (fixed)	0.211	3.835	1510.2	0.7
Generalized DLM (trend $\delta = 0.98$, seasonality $\delta = 0.95$ for all p_0)						
$p_0 = 0.05$						
DLM	—	—	1.830	5.323	84242.3	0.1
DQLM	—	3.164 (2.98,3.37)	1.410	3.729	45607.3	19.0
exDQLM/MCMC	3.469 (3.23,3.76)	3.492 (3.35,3.71)	**0.472**	**3.603**	**10747.9**	45.0
exDQLM/ISVB	1.534 (1.45,1.62)	3.164 (fixed)	1.046	3.609	19189.5	2.4
$p_0 = 0.50$						
DLM	—	—	2.723	5.401	84316.5	0.1
DQLM	—	4.536 (4.26,4.84)	1.204	3.635	7400.3	18.9
exDQLM/MCMC	0.112 (0.06,0.18)	4.465 (4.21,4.71)	**1.203**	**3.581**	**7278.6**	45.5
exDQLM/ISVB	0.095 (0.06,0.13)	4.536 (fixed)	1.240	3.514	7616.1	0.7
$p_0 = 0.85$						
DLM	—	—	4.518	5.339	84347.7	0.1
DQLM	—	3.149 (2.95,3.36)	0.979	3.531	13454.3	18.6
exDQLM/MCMC	-1.139 (-1.24,-1.02)	3.703 (3.478,3.926)	**0.823**	**3.516**	**8873.2**	45.1
exDQLM/ISVB	-0.497 (-0.57,-0.42)	3.149 (fixed)	0.828	3.491	9836.2	1.1

TABLE 1

Posterior summaries for γ and σ (where applicable): mean (95% CrI). Mean check loss of the MAP dynamic quantile. KL divergences of the one-step-ahead distributions. Posterior predictive loss criterion (pplc) under the check loss function. Computation run-time (min).

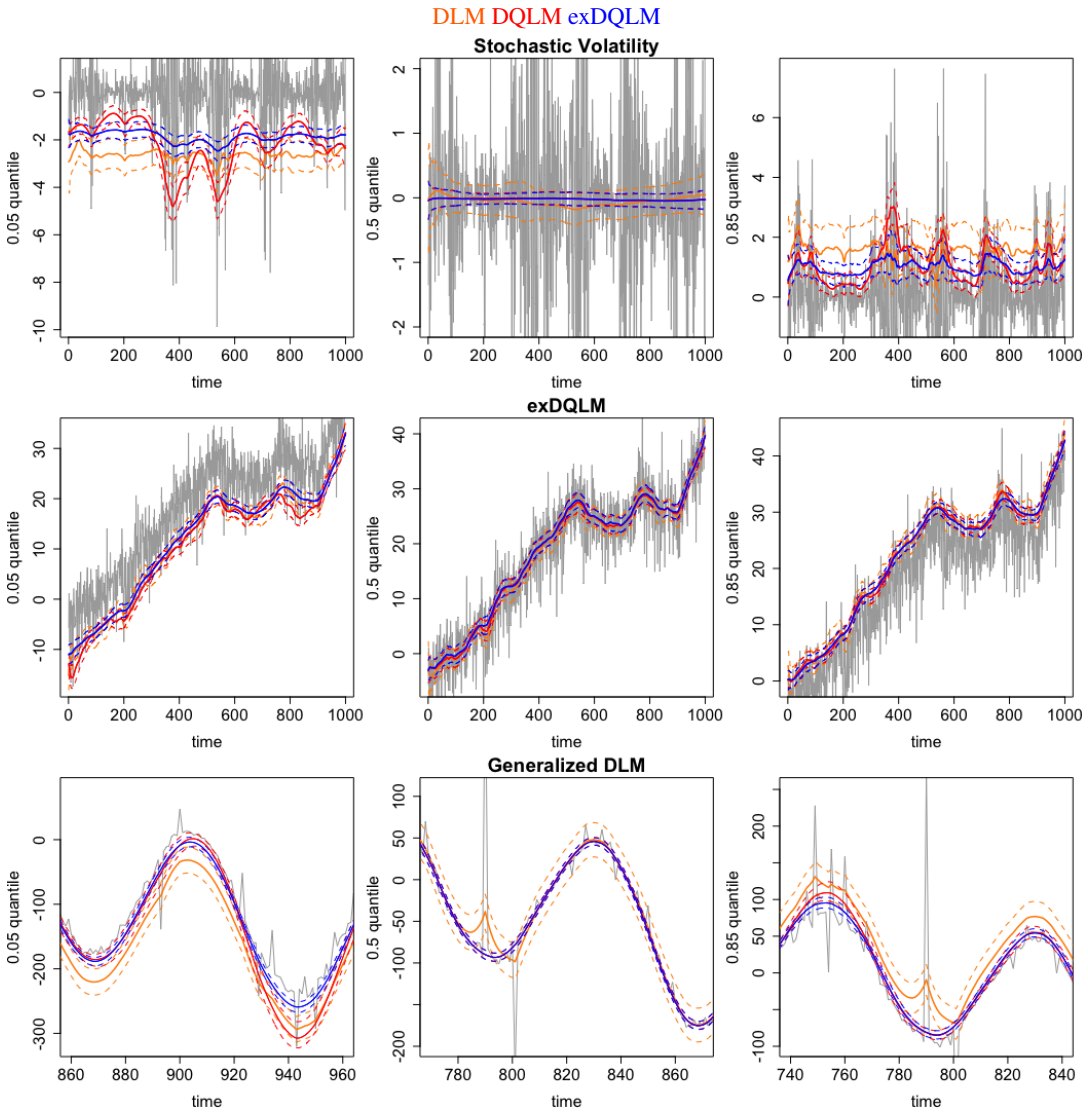

FIG 3. *MCMC results. Posterior distributions of dynamic quantiles* $p_0 = 0.05, 0.5, 0.85$. *exDQLM estimates in blue, DQLM in red, and DLM in orange. Dotted lines indicate the 95% CrI from the smoothed posterior distributions and solid lines indicate posterior mean estimates. Due to the scale of the gDLM data, we focus on time periods which highlight the similarities and/or differences of the models at the three different quantiles.*

posterior summaries for γ from the MCMC and ISVB algorithms, as discussed in Section 2.7 and seen in Table 1. However, the approximated posterior quantiles from the ISVB algorithm are almost entirely contained with the MCMC posterior 95% credible intervals (CrIs), as seen in Figure 4.

4. Transfer Function exDQLM. Quantifying the relationship between a climatological response and input at various quantiles is a non-trival task. In the mean-centric setting, transfer functions are a simple way to incorporate variables which measure the combined effect of current and past inputs or regression effects (West and Harrison, 2006). To capture both the immediate and lagged effects of a climatological variable, we expand the use of transfer functions to the dynamic quantile regression setting with the development of a transfer function extension to our exDQLM.

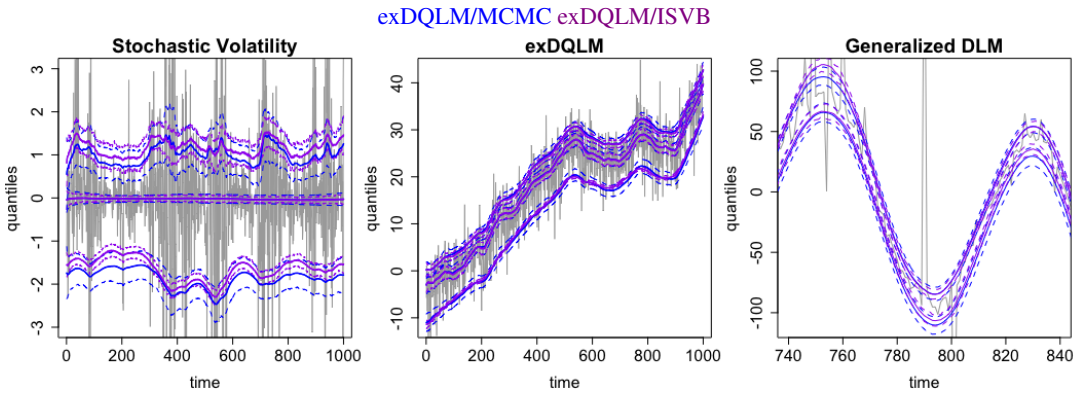

FIG 4. *ISVB and MCMC comparison results. Posterior distributions of dynamic quantiles* $p_0 = 0.05, 0.5, 0.85$. *MCMC exDQLM estimates in blue, ISVB exDQLM estimates in purple. Dotted lines indicate the 95% CrI from the smoothed posterior distributions and solid lines indicate posterior mean estimates. Due to the scale of the gDLM data, we focus on period of only 100 time points and omit the median estimates for visual clarity.*

For time $t = 1, \ldots, T$ and a single regression effect, X_t, a transfer function exDQLM with exponential decay is as follows:

$$y_t|\boldsymbol{\theta}_t, \gamma, \sigma \sim \text{exAL}_{p_0}(\mathbf{F}_t'\boldsymbol{\theta}_t + \zeta_t, \sigma, \gamma) \tag{19}$$

$$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{W}_t \sim \text{N}(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t) \tag{20}$$

$$\zeta_t|\zeta_{t-1}, \psi_{t-1}, \omega_t \sim \text{N}(\lambda\zeta_{t-1} + X_t\psi_{t-1}, \omega_t) \tag{21}$$

$$\psi_t|\psi_{t-1}, \nu_t \sim \text{N}(\psi_{t-1}, \nu_t). \tag{22}$$

Here ζ_t captures the effect of the current and past regression effects, as seen in Equation (21). The parameter ψ_t determines the immediate effect X_t has on the quantile. Alternatively the parameter λ is a quantity in the unit interval which represents the memory of the regression effect up to time t. This effect decays at an exponential rate, reducing by a factor of λ at every time step. To see this more explicitly we can derive the transfer function effect k steps ahead, that is:

$$\zeta_{t+k} = \lambda^k\zeta_t + \sum_{r=0}^{k} \lambda^{k-r}\psi_{t+r-1}X_{t+r} + \partial\zeta_{t+k}. \tag{23}$$

Thus, the effect of X_t on the quantile at time $t + k$ is $\lambda^k\psi_{t-1}X_t$. This effect is negligible when $\lambda^k|\psi_{t-1}X_t| \leq \epsilon$ for small ϵ. Using this we can derive a series, k_t, representing a lower bound for the number of time steps until the effect of X_t is less than or equal to a fixed ϵ. That is, for $t = 1, \ldots, T$

$$k_t \geq \frac{\log(\epsilon) - \log(|\psi_{t-1}X_t|)}{\log(\lambda)}. \tag{24}$$

To complete the model, conjugate priors are available for the additional transfer function parameters; normal conjugate priors for $\zeta_0 \sim N(m_{\zeta_0}, C_{\zeta_0})$ and $\psi_0 \sim N(m_{\psi_0}, C_{\psi_0})$, and a conjugate normal truncated to the unit interval prior for $\lambda \sim N_{(0,1)}(m_\lambda, v_\lambda)$.

4.1. *MCMC and ISVB Algorithm Augmentations.* This transfer function exDQLM can equivalently be rewritten in the form of a standard exDQLM

$$y_t|\gamma, \boldsymbol{\theta}_t, \sigma \sim \text{exAL}_{p_0}(\tilde{\mathbf{F}}_t'\tilde{\boldsymbol{\theta}}_t, \sigma, \gamma) \tag{25}$$

$$\tilde{\boldsymbol{\theta}}_t|\tilde{\boldsymbol{\theta}}_{t-1}, \tilde{\mathbf{W}}_t \sim \text{N}(\tilde{\mathbf{G}}_t\tilde{\boldsymbol{\theta}}_{t-1}, \tilde{\mathbf{W}}_t) \tag{26}$$

where $\tilde{\mathbf{F}}_t' = (\mathbf{F}_t', 1, 0)$, $\tilde{\boldsymbol{\theta}}_t' = (\boldsymbol{\theta}_t', \zeta_t, \psi_t)$, $\tilde{\mathbf{G}}_t = \text{blockdiag}\left\{\mathbf{G}_t, \left(\begin{smallmatrix}\lambda & X_t \\ 0 & 1\end{smallmatrix}\right)\right\}$, and $\tilde{\mathbf{W}}_t = \text{blockdiag}\left\{\mathbf{W}_t, \left(\begin{smallmatrix}\omega_t & 0 \\ 0 & \nu_t\end{smallmatrix}\right)\right\}$.

Using this representation, the exDQLM MCMC Algorithm 1 can easily be augmented to incorporate the transfer function structure as follows: (1) Replace all $\{\mathbf{F}_t, \boldsymbol{\theta}_t, \mathbf{G}_t, \mathbf{W}_t\}$ with $\{\tilde{\mathbf{F}}_t, \tilde{\boldsymbol{\theta}}_t, \tilde{\mathbf{G}}_t, \tilde{\mathbf{W}}_t\}$, where $\tilde{\mathbf{G}}_t$ will be conditional on $\lambda^{(i)}$; (2) For each iteration i, add an additional step to sample $\lambda^{(i+1)}|\boldsymbol{\zeta}^{(i+1)}, \boldsymbol{\psi}^{(i+1)} \sim N_{(0,1)}(\mu_\lambda, \sigma_\lambda^2)$ with

$$\sigma_\lambda^2 = [\sum_{t=1}^{T} \frac{\zeta_{t-1}^{(i+1)2}}{\omega_t} + \frac{1}{v_\lambda}]^{-1},$$

$$\mu_\lambda = \sigma_\lambda^2[\sum_{t=1}^{T} \frac{\zeta_{t-1}^{(i+1)}(\zeta_t^{(i+1)} - \psi_t^{(i+1)} X_t)}{\omega_t} + \frac{1}{v_\lambda}].$$

Augmenting the ISVB algorithm is not as straight-forward. The random parameter λ within the evolution matrix $\tilde{\mathbf{G}}_t$ compromises our ability to update the state parameter variational distributions within the FFBS while using discount factors to specify $\tilde{\mathbf{W}}_t$. To preserve the ability to utilize discount factors, we propose optimizing the parameter λ with respect to the KL divergence of the one-step-ahead predictive distribution functions as discussed in Sections 2.5 and 2.6. For optimal λ, say $\tilde{\lambda}$, the ISVB algorithm can be augmented to incorporate the transfer function structure by simply replacing all $\{\mathbf{F}_t, \boldsymbol{\theta}_t, \mathbf{G}_t, \mathbf{W}_t\}$ with $\{\tilde{\mathbf{F}}_t, \tilde{\boldsymbol{\theta}}_t, \tilde{\mathbf{G}}_t, \tilde{\mathbf{W}}_t\}$ where $\tilde{\mathbf{G}}_t = \text{blockdiag}\left\{\mathbf{G}_t, \left(\begin{smallmatrix}\tilde{\lambda} & X_t \\ 0 & 1\end{smallmatrix}\right)\right\}$.

5. Estimating the 0.85 quantile IVT threshold. The primary dataset used to calculate IVT in the study by Guan and Waliser (2015) is the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis (ERA-Interim). ERA-Interim produces 6-hourly atmospheric fields at a $1.5^\circ \times 1.5^\circ$ spatial resolution beginning in 1979 and is continuously updated (Berrisford et al., 2011; Dee et al., 2011). At each time and location, IVT is derived from observational products of humidity and windspeed, resulting in a large (over both time and space) dataset. The method presented in Guan and Waliser (2015) for detection of ARs from the calculated IVT is as follows. For each of the 12 months, the 0.85 quantile IVT is calculated over all time steps during the 5 month windows centered on that month over the period from 1997 to 2014 at a specific location. Comparison to the estimated 0.85 quantile in combination with a minimum threshold is used to isolate regions of enhanced IVT as possible ARs. Criteria are then applied to the length and width of these regions, resulting in a defined set of ARs. Finally, the coastal location intersecting with an AR at which the IVT magnitude is highest is defined as the cell in which the AR makes landfall. For more details on the full AR detection algorithm, see Guan and Waliser (2015). The top panel of Figure 5 illustrates the average daily IVT magnitude in Santa Cruz, CA, of which we examine the 0.85 quantile in this analysis. For illustration, the times at which ARs detected to make landfall at that location and in the neighboring coastal locations are illustrated in the middle panels of Figure 5 for two time periods; years 1982 to 1985 in which CA saw an exceptional amount of rain, and years 2012 to 2015 which were exceptionally dry for CA.

Although many climate indices other than ELI have been studied as potential sources of predictability for ARs, initial examination of several indices with our transfer function exDQLM did not demonstrate significant associations. Therefore in this analysis, we focus solely on the association between IVT and ELI. ELI is a single metric which captures the spatial diversity of ENSO, created utilizing the monthly ECMWF twentieth century reanalysis (ERA-20C). In particular, ELI is the average longitude at which tropical Pacific deep

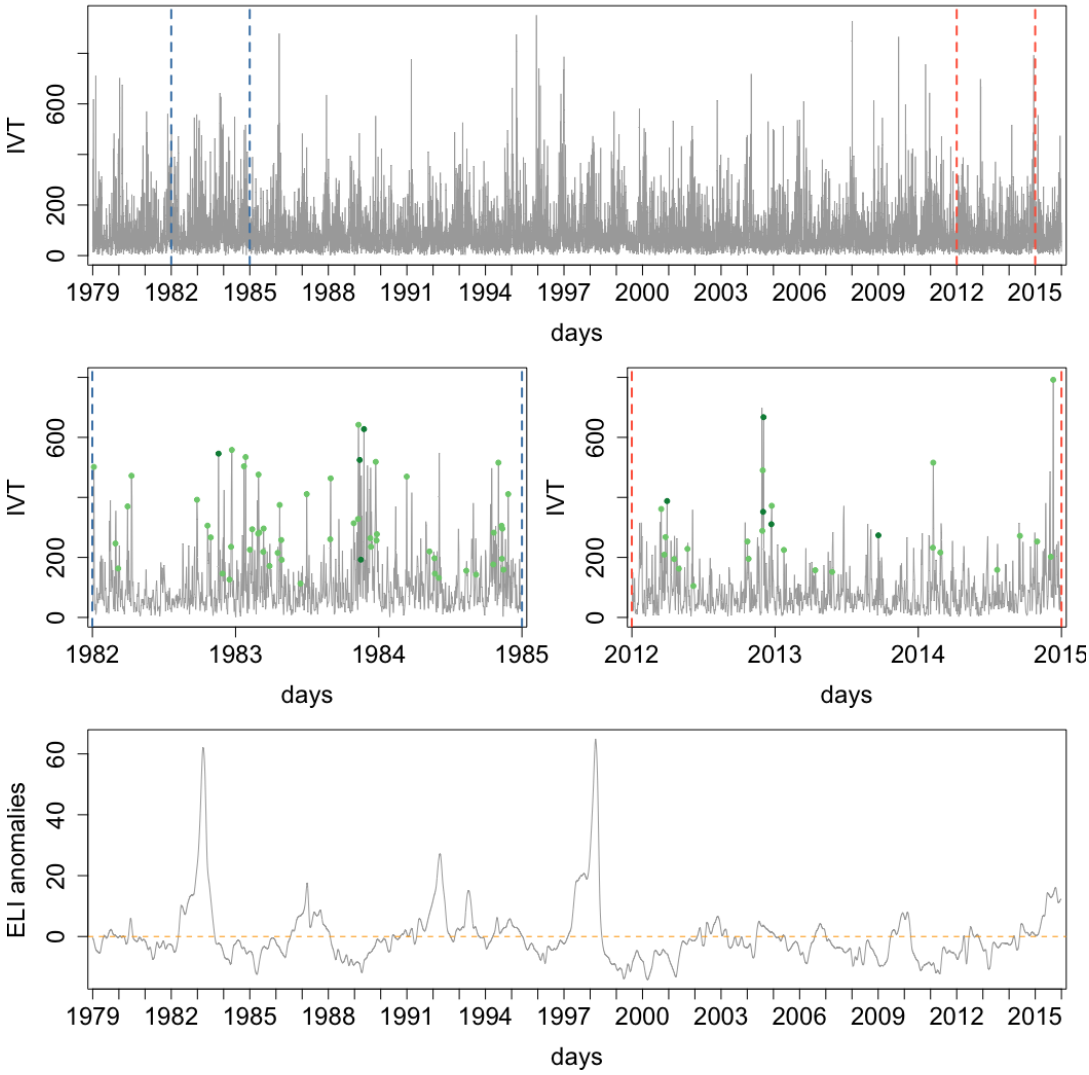

FIG 5. *Top panel: Average daily IVT magnitude in the grid cell containing Santa Cruz, CA. The blue vertical lines enclose a time period in which CA saw an exceptional amount of rain. The red vertical lines encloses as time period in which CA was exceptionally dry. Middle panel: A closer look at the two time periods highlighted in the top panel. Years 1982 to 1985 illustrate a time period which saw an exceptional amount of rain. Years 2012 to 2015 illustrate a period which was exceptionally dry. ARs detected by the scheme proposed in Guan and Waliser (2015) in the Santa Cruz grid cell are indicated with dark green points. ARs detected in neighboring coastal grid cells are indicated with light green points. Bottom panel: ELI anomalies resulting from the de-seasonalization of the interpolated ELI. The dashed, orange horizontal line is at zero, for reference.*

convection occurs at a given month. For further details on the development ELI and the relationship of ELI with precipitation see Williams and Patricola (2018) and Patricola et al. (2020), respectively. The monthly ELI dataset is available online beginning in 1854 and is frequently updated (Williams and Patricola, 2018). For our analysis, we interpolate the ELI to the daily time scale and de-seasonalize the time series by removing the smoothed posterior mean estimates from a standard DLM with constant trend, annual and semi-annual components. De-seasonalizing in this way ensures the variability in the 0.85 quantile described by the ELI component of our model is not an artifact of the seasonality in the original ELI time

series. This de-seasonalization results in a time series of ELI anomalies, also seen in Figure 5, which we use as the input in our analysis of the 0.85 quantile of the IVT magnitude.

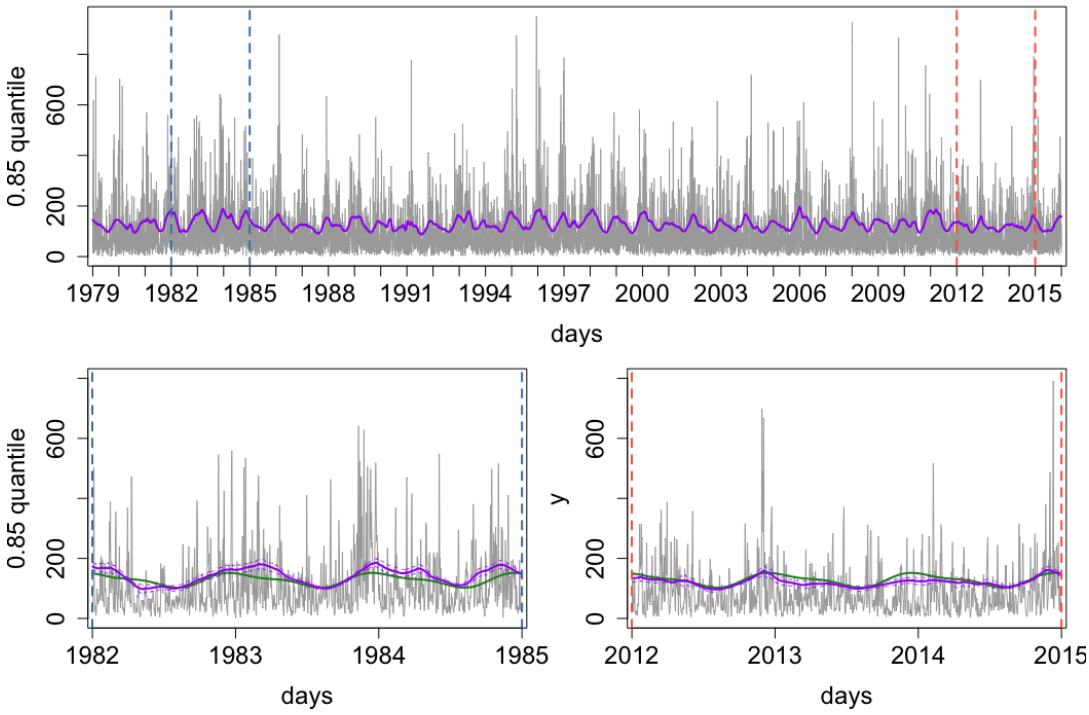

FIG 6. *Top panel: In purple, M_1 MAP 0.85 quantile threshold of the average daily IVT magnitude in Santa Cruz, CA. The blue vertical lines enclose a period in which CA saw an exceptional amount of rain. The red vertical lines encloses a period in which CA was exceptionally dry. Bottom panels: Years 1982 to 1985 (left, exceptional amount of rain) and years 2012 to 2015 (right, exceptionally dry). The MAP 0.85 quantile from M_1 and M_0 are seen in purple and green, respectively.*

5.1. *IVT analysis.* We fit two separate models to estimate the 0.85 quantile of the IVT magnitude. The first, M_0, is a simplified version of the second, M_1. The simplified M_0 includes only a baseline level and seasonal effects without any input from the ELI time series. Alternatively, M_1 includes the ELI input utilizing our transfer function exDQLM. Our analysis will focus on the results of M_1, only highlighting features of M_0 for comparison purposes.

The state parameters for the baseline component and the seasonal components in both models are denoted by η_t, (baseline) and $\alpha_{1,t}^{(1)}$, $\alpha_{2,t}^{(1)}$, $\alpha_{1,t}^{(2)}$ and $\alpha_{2,t}^{(2)}$ (seasonal). We describe the baseline component in the model, η_t, with a first-order polynomial structure

$$\eta_t = \eta_{t-1} + \boldsymbol{\omega}_t^{\eta}, \quad \boldsymbol{\omega}_t^{\eta} \sim N(\mathbf{0}, \mathbf{W}_t^{\eta}) \,. \tag{27}$$

A second-order polynomial component was also considered but found to be not significant. Here the system evolution error vectors $\boldsymbol{\omega}_t^{\eta}$, are assumed to be independent over time. We include seasonal components $\alpha_{1,t}^{(l)}$ for harmonics $l = 1, 2$ for a period of 365 days. We found only the annual ($l = 1$) and semi-annual ($l = 2$) harmonics to be significant, and model them using Fourier form seasonal components (West and Harrison, 2006) as follows for $l = 1, 2$,

$$\begin{pmatrix} \alpha_{1,t}^{(l)} \\ \alpha_{2,t}^{(l)} \end{pmatrix} = \begin{pmatrix} \cos(\frac{2\pi}{365}l) & \sin(\frac{2\pi}{365}l) \\ -\sin(\frac{2\pi}{365}l) & \cos(\frac{2\pi}{365}l) \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1}^{(l)} \\ \alpha_{2,t-1}^{(l)} \end{pmatrix} + \boldsymbol{\omega}_t^{\alpha,l}, \quad \boldsymbol{\omega}_t^{\alpha,l} \sim N_2(\mathbf{0}, \mathbf{W}_t^{\alpha,l}) \,. \tag{28}$$

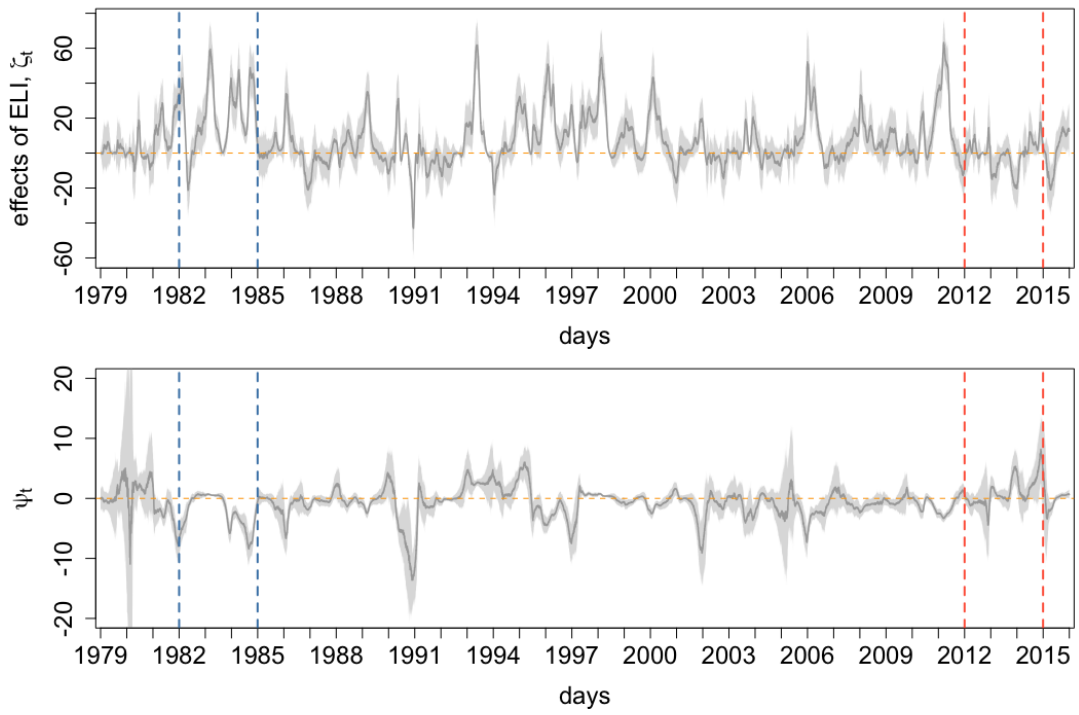

FIG 7. *Top panel: Effects of ELI captured by the transfer function component, ζ_t. Bottom panel: Instantaneous effects of ELI, ψ_t. In both panels, dark grey lines indicate the MAP estimates. 95% CrI are indicated by the grey shaded regions. Dashed horizontal orange lines are at zero, for reference. Blue vertical lines enclose years 1982 to 1985 in which CA saw an exceptional amount of rain. Red vertical lines enclose 2012 to 2015 in which CA was exceptionally dry.*

We denote the lth seasonal evolution matrix $\mathbf{G}^{\alpha,l} = \begin{pmatrix} \cos(\frac{2\pi}{365}l) & \sin(\frac{2\pi}{365}l) \\ -\sin(\frac{2\pi}{365}l) & \cos(\frac{2\pi}{365}l) \end{pmatrix}$. Again, it is assumed that $\boldsymbol{\omega}_t^{\alpha,l}$ are independent over time, as well as independent of $\boldsymbol{\omega}_t^{\eta}$ for $t = 1, \ldots, T$. Using superposition the resulting state vector $\boldsymbol{\theta}_t$, $\mathbf{F}'$, $\mathbf{G}$ and $\mathbf{W}_t$ in Equations (19)-(22) are defined, respectively, as $\boldsymbol{\theta}_t = (\eta_t, \alpha_{1,t}^{(1)}, \alpha_{2,t}^{(1)}, \alpha_{1,t}^{(2)}, \alpha_{2,t}^{(2)})'$, $\mathbf{F}' = (1,1,0,1,0)$, $\mathbf{G} = \text{blockdiag}(1, \mathbf{G}^{\alpha,1}, \mathbf{G}^{\alpha,2})$, and $\mathbf{W}_t = \text{blockdiag}(\mathbf{W}_t^{\eta}, \mathbf{W}_t^{\alpha,1}, \mathbf{W}_t^{\alpha,2})$.

We choose to model the baseline and seasonal components in both models as non-time-varying, thus any variation in the 0.85 quantile from year to year will solely be attributed to the effects of the ELI as input to our transfer function model. This is easily done utilizing component discounting to specify $\mathbf{W}_t$ with discount factor values of 1 (West and Harrison, 2006), which also preserves our ability to update the state parameter using FFBSm. Note, under this modeling choice the baseline and seasonal parameters in the state vector are non-time-varying, thus we omit their subscripts t going forward.

In addition to the baseline and seasonal components, in M_1 we utilize the exponentially decaying transfer function exDQLM as specified in Equations (19)-(22) to capture both the immediate and lagged effects of ELI on the 0.85 quantile. We complete the models with conjugate priors, where applicable; $\boldsymbol{\theta}_0 \sim \text{N}(\mathbf{m}_0, \mathbf{C}_0)$ with $\mathbf{m}_0 = \mathbf{0}$ and $\mathbf{C}_0 = 100 I_q$, $\zeta_0 \sim N(m_{\zeta_0}, C_{\zeta_0})$ and $\psi_0 \sim N(m_{\psi_0}, C_{\psi_0})$ with $m_{\zeta_0} = m_{\psi_0} = 0$ and $C_{\zeta_0} = C_{\psi_0} = 10$, and $\gamma \sim t_{(-5.137, 0.213)}(0, 1)$ with 1 degree of freedom. The parameter σ is fixed at $\hat{\sigma}_{\gamma=0}^{0.85}$, as discussed in Section 2.7. Lastly, in M_1, the parameter λ as well as the discount factors for the evolution of ζ_t and ψ_t, δ_ζ and δ_ψ respectively, are optimized using the KL divergence of the one-step-

	M_0	M_1
$(\delta_\zeta, \delta_\psi)$	—	(0.96,0.97)
$\tilde{\lambda}$	—	0.37
$\hat{\sigma}^{0.85}_{\gamma=0}$	22.02	21.27
γ	-0.75 (-0.77,-0.73)	-0.68 (-0.70,-0.65)
η	128.33 (127.63,129.04)	121.65 (120.13,122.79)
A_1	22.27 (21.52,23.26)	20.46 (19.03,22.02)
P_1	-0.23 (-0.28,-0.19)	-0.12 (-0.20,-0.05)
A_2	7.89 (6.93,8.75)	8.02 (6.74,9.19)
P_2	1.30 (1.19,1.44)	1.20 (1.04,1.32)
pplc	784445.6	767292.5
KL	1.566	0.839
run-time	2.43	3.31

TABLE 2

IVT analysis results for M_0 and M_1. Optimal δ_ζ, δ_ψ, and $\tilde{\lambda}$, as discussed in Section 2.7. Values of $\hat{\sigma}^{0.85}_{\gamma=0}$ used in the priors on scale parameter σ, also discussed in Section 2.7. Posterior summaries (format: mean (95% CrI)) for skewness parameter γ, baseline η, annual amplitude A_1, annual phase P_1, semi-annual amplitude A_2, and semi-annual phase P_2. pplc: Posterior predictive loss criterion under the check loss function. KL: Kullback-Liebler divergences of the one-step-ahead distributions. Run-time: ISVB run-times in minutes.

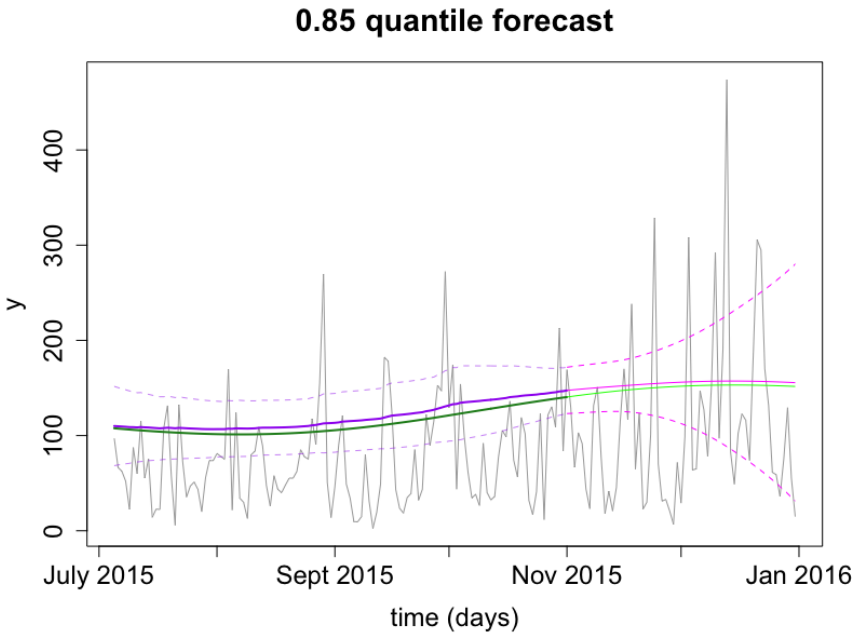

FIG 8. *60-step-ahead quantile forecast beginning November 2, 2015 through December 31, 2015. From M_1: The solid magenta line indicates the forecast means and the dashed magenta indicate the 95% CrI of the forecasted 0.85 quantile. Also included are the filtered means and 95% CrI up to November 2, 2015 in purple solid and dashed lines, respectively. From M_0: The solid light green line indicates the forecast means and the solid dark green line indicates the filtered means. The IVT magnitude data is seen in grey.*

ahead predictive distribution functions as discussed in Section 4.1. Optimal λ, δ_ζ, δ_ψ can be found in Table 2, as well as $\hat{\sigma}^{0.85}_{\gamma=0}$.

We apply our ISVB algorithm to estimate the 0.85 quantile of the daily IVT magnitude in Santa Cruz, CA from 1979 to 2016. For this time series of length 13505 the ISVB computation times are under four minutes for both models; exact times can be found in Table 2. Figure 6 illustrates the MAP 0.85 quantile of M_1 for the entire time period, as well as the MAP estimates from both models for the two time periods, 1982 through 1985 and 2012 through 2015, in which CA saw drastically different amounts of precipitation. It can be seen that the M_1 quantile is generally higher than the M_0 quantile when CA experienced an exceptional amount of rain, with the opposite true when CA was exceptionally dry. Further, we note the ARs detected to make landfall near Santa Cruz, CA by Guan and Waliser (2015) are entirely above the estimated thresholds in both time periods.

From the posterior estimates of the annual and semi-annual harmonic components of the models we compute the amplitude and phase, $A_l = \sqrt{(\alpha_1^l)^2 + (\alpha_2^l)^2}$ and $P_l = \arctan(-\alpha_2^l/\alpha_1^l)$ respectively. Posterior summaries of these as well as the baseline parameter, η, can be found in Table 2. It can be seen the baseline and annual amplitude of M_0 are significantly larger than those of M_1. Further, the annual phase of M_0 is significantly smaller than that of M_1. Alternatively, the semi-annual harmonic components are indistinct between the two models. This suggests there is a substantial amount of variability in the 0.85 quantile that can be associated with the ELI time series specifically on the annual time scale, with the distinction less clear at the semi-annual scale.

The amount of variability in the 0.85 quantile attributed to the effects of ELI captured with transfer function component in M_1, ζ_t, are seen in Figure 7. The effects of ELI are overall significant and are dramatically more pronounced between 1982 and 1985 than between 2012 and 2015. In particular, a majority of the effects between 1982 and 1985 (in which CA received heavy precipitation) are distinctly positive whereas the effects between 2012 to 2015 (when drought was severe) are negative or not significant. The instantaneous effects of ELI at time t, ψ_t, also exhibit very different behavior in the two time periods, seen in Figure 7. Upon computing the series k_t from Equation (24) for $\epsilon = 1\text{e}{-3}$ (not pictured), we find the lagged effects of ELI persist for around 8.5 days, on average.

To assess the predictive value added by the transfer function component capturing the effects of ELI in M_1, we compare the pplc and KL divergence of the one-step-ahead forecast distributions for M_1 to those of M_0, seen in Table 2. M_1 is favored with smaller values of both model comparison criterion, suggesting the inclusion of ELI improves both the predictive and forecasting power of M_1. To further examine the information added by ELI, we can examine the k-step-ahead quantile forecast distributions. That is, for each time t the k-step-ahead future marginal distribution of the quantile is

$$\tilde{\mathbf{F}}_{t+k}'\tilde{\boldsymbol{\theta}}_{t+k}|y_1,\ldots,y_t \sim \text{N}(\tilde{\mathbf{F}}_{t+k}'a_t(k), \tilde{\mathbf{F}}_{t+k}'\mathbf{R}_t(k)\tilde{\mathbf{F}}_{t+k}) \tag{29}$$

where $\mathbf{a}_t(k) = \tilde{\mathbf{G}}_{t+k}\mathbf{a}_t(k-1)$, $\mathbf{R}_t(k) = \tilde{\mathbf{G}}_{t+k}\mathbf{R}_t(k-1)\tilde{\mathbf{G}}_{t+k}' + \tilde{\mathbf{W}}_{t+k}$, $\mathbf{a}_t(0) = \mathbf{m}_t$, and $\mathbf{R}_t(0) = \mathbf{C}_t$, with $\mathbf{m}_t$ and $\mathbf{C}_t$ denoting the filtered mean and covariance of $\tilde{\boldsymbol{\theta}}_t$, respectively. The posterior means and 95% CrIs of these distributions for 60-steps-ahead can be seen in Figure 8. The MAP quantile forecast of M_1, which takes the ELI into consideration, suggests the 0.85 quantile will be slightly higher than the seasonal average projected by M_0.

6. Conclusion. Motivated by the need for versatile estimation of a single quantile over time, we have presented several methodological and computational contributions for dynamic quantile modeling. Our exDQLM has two main advantages; the model facilitates more flexibility in the estimation of the quantile than standard Bayesian parametric quantile regression approaches, and relevant features such as seasonality or structured long-term variability are easily included in the evolution structure of the quantile. Further, the development of our efficient ISVB algorithm facilitates fast posterior inference, making our methodology accessible

even in applications with very long time series data. Finally in contrast to current schemes, our transfer function exDQLM develops a straight-forward method for quantifying non-linear relationships between a response and input at a specified quantile. Our methodology is immediately beneficial not only in climatological applications such as AR detection detailed in this work, but more generally in any application with non-Gaussian time-varying models.

We illustrated the utility of our methods in the analysis of the IVT magnitude 0.85 quantile threshold in Santa Cruz, CA; an analysis made possible by our ISVB algorithm. In contrast to many current thresholding approaches, estimating the 0.85 quantile with our model provides rich inference about the structure of the time series and thus enhances the tools for characterization of ARs. The results demonstrated not only a significant annual and semi-annual seasonal structure in the quantile, but also a significant non-linear relationship with the climate index ELI captured by the transfer function component of our model. Through several model checking criteria, we were able to show the inclusion of ELI in the model was advantageous both in forecast and predictive accuracy. We saw the effects of ELI on the 0.85 quantile varied substantially between two time periods which experienced drastically different amounts of precipitation; results that are relevant to understanding the roll of IVT magnitude and ARs in the global water cycle and regional weather.

We consider only univariate dynamic quantile modeling in this work. However, multivariate and spatial interaction between the IVT magnitudes at various locations motivate extensions of our exDQLM to these settings. Non-time-varying multivariate and spatial quantile regression has been considered, non-parametrically (Reich, Fuentes and Dunson, 2011) and parametrically (Lum et al., 2012). Some work has been done in the spatio-temporal setting, both parametric (Neelon et al., 2015) and semi-paramteric (Reich, 2012), however again, the parametric approaches are exclusively based on the AL. Our more flexible methodology naturally scales to the multivariate and spatial time-varying settings, making this the clear next step in our work.

APPENDIX

A.1. MCMC Forward Filtering Backwards Sampling. Let $D_t = \{y_1, \ldots, y_t\}$. To simplify the notation, we leave out conditional parameters $\mathbf{v}, \mathbf{s}, \gamma, \sigma$. For $t = 1, \ldots, T$, sample $\boldsymbol{\theta}_t | D_T = \boldsymbol{\theta}_t | D_T, \mathbf{v}, \mathbf{s}, \gamma, \sigma$ using the following FFBS updates:

- Forward filtering, for $t = 1, \ldots, T$ compute:
  - Prior $p(\boldsymbol{\theta}_t | D_{t-1})$: Given $\boldsymbol{\theta}_{t-1} | D_{t-1} \sim \mathrm{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$,

$$p(\boldsymbol{\theta}_t | D_{t-1}) = \int \mathrm{N}(\boldsymbol{\theta}_t | \mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) \mathrm{N}(\boldsymbol{\theta}_{t-1} | \mathbf{m}_{t-1}, \mathbf{C}_{t-1}) d\boldsymbol{\theta}_{t-1}.$$

    Thus, $p(\boldsymbol{\theta}_t | D_{t-1}) = \mathrm{N}(\mathbf{a}_t, \mathbf{R}_t)$, with $\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t^T \mathbf{C}_{t-1} \mathbf{G}_t + \mathbf{W}_t$.
  - Forecast $p(y_t | D_{t-1})$: Given $\boldsymbol{\theta}_t | D_{t-1} \sim \mathrm{N}(\mathbf{a}_t, \mathbf{R}_t)$,

$$p(y_t | D_{t-1}) = \int \mathrm{N}(y_t | \mathbf{F}_t' \boldsymbol{\theta}_t + C(p)\sigma|\gamma|s_t + A(p)v_t, \sigma B(p) v_t) \mathrm{N}(\boldsymbol{\theta}_t | \mathbf{a}_t, \mathbf{R}_t) d\boldsymbol{\theta}_t.$$

    Thus, $p(y_t | D_{t-1}) = \mathrm{N}(f_t, q_t)$, with $f_t = \mathbf{F}_t' \mathbf{a}_t + C(p)\sigma|\gamma|s_t + A(p)v_t$ and $q_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \sigma B(p) v_t$.
  - Posterior $p(\boldsymbol{\theta}_t | D_t)$: Given the prior and forecast distributions, the joint distribution can be written

$$\begin{bmatrix} \boldsymbol{\theta}_t \\ y_t \end{bmatrix} \Big| D_{t-1} \Big] \sim \mathrm{N}_{p+1}\Big(\begin{bmatrix} \mathbf{a}_t \\ f_t \end{bmatrix}, \begin{bmatrix} \mathbf{R}_t & \mathbf{R}_t \mathbf{F}_t \\ \mathbf{F}_t' \mathbf{R}_t & q_t \end{bmatrix} \Big).$$

    The conditional distribution of the multivariate normal $p(\boldsymbol{\theta}_t | D_t) = \mathrm{N}(\mathbf{m}_t, \mathbf{C}_t)$ with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (y_t - f_t)/q_t$ and $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t / q_t$

- Backwards sampling, $p(\boldsymbol{\theta}_t|D_T)$:
  - For T, sample $\boldsymbol{\theta}_T|D_T \sim \mathrm{N}(\mathbf{m}_T, \mathbf{C}_T)$.
  - For $t = T-1, \ldots, 1$, sample $\boldsymbol{\theta}_t|D_T \sim \mathrm{N}(\mathbf{m}_t^s, \mathbf{C}_t^s)$ with $\mathbf{m}_t^s = \mathbf{m}_t + \mathbf{C}_t\mathbf{G}_t'\mathbf{R}_{t+1}^{-1}(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1})$ and $\mathbf{C}_t^s = \mathbf{C}_t - \mathbf{C}_t\mathbf{G}_t'\mathbf{R}_{t+1}^{-1}\mathbf{G}_t\mathbf{C}_t$.

A.2. ISVB Forward Filtering Backwards Smoothing. Similarly, we can update $r^{(k+1)}(\boldsymbol{\theta}_t) = \mathrm{N}(\mathbf{m}_t^s, \mathbf{C}_t^s)$ using FFBSm as follows.

- Forward filter, for $t = 1, \ldots, T$:
  - Prior: $r^{(k+1)}(\boldsymbol{\theta}_t|D_{t-1}) = \mathrm{N}(\mathbf{a}_t, \mathbf{R}_t)$, with $\mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t^T\mathbf{C}_{t-1}\mathbf{G}_t + \mathbf{W}_t$.
  - Forecast: $r^{(k+1)}(y_t|D_{t-1}) = \mathrm{N}(f_t, Q_t)$, with

$$f_t = \mathbf{F}_t'\mathbf{a}_t + \left[\left\langle \frac{C(p)|\gamma|}{B(p)} \right\rangle^{(k)} \langle s_t \rangle^{(k+1)} + \left\langle \frac{A(p)}{\sigma B(p)} \right\rangle^{(k)} \Big/ \left\langle \frac{1}{v_t} \right\rangle^{(k+1)}\right] \Big/ \left\langle \frac{1}{\sigma B(p)} \right\rangle^{(k)}$$

$$Q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + \left[\left\langle \frac{1}{v_t} \right\rangle^{(k+1)} \left\langle \frac{1}{\sigma B(p)} \right\rangle^{(k)}\right]^{-1}$$

  .
  - Posterior: $r^{(k+1)}(\boldsymbol{\theta}_t|D_t) = \mathrm{N}(\mathbf{m}_t, \mathbf{C}_t)$, with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t(y_t - f_t)/Q_t$ and $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t\mathbf{F}_t'\mathbf{R}_t/Q_t$
- Backward smoother:
  - For T, $r^{(k+1)}(\boldsymbol{\theta}_T|D_T) = \mathrm{N}(\mathbf{m}_T^s = \mathbf{m}_T, \mathbf{C}_T^s = \mathbf{C}_T)$
  - For $t = T-1, \ldots, 1$, $r^{(k+1)}(\boldsymbol{\theta}_t|D_T) = \mathrm{N}(\mathbf{m}_t^s, \mathbf{C}_t^s)$ with $\mathbf{m}_t^s = \mathbf{m}_t + \mathbf{B}_t(\mathbf{m}_{t+1} - \mathbf{a}_{t+1})$ and $\mathbf{C}_t^s = \mathbf{C}_t + \mathbf{B}_t(\mathbf{C}_{t+1}^s - \mathbf{R}_{t+1})\mathbf{B}_t'$, where $\mathbf{B}_t = \mathbf{C}_t\mathbf{G}_t'\mathbf{R}_{t+1}^{-1}$

A.3. ISVB Importance Sampling. The variational distribution $r^{(k+1)}(\sigma, \gamma)$ can be computed up to a proportionality constant,

$$\begin{aligned} r^{(k+1)}(\sigma,\gamma) \propto & f_0(\gamma)f_0(\sigma)\sigma^{1.5T}\exp\Big\{ -\sum \langle u_t \rangle^{(k+1)}/\sigma \\ & - \frac{1}{2}\sum \Big[\frac{1}{\sigma B(p)} \left\langle \frac{1}{v_t} \right\rangle^{(k+1)} (y_t^2 - 2y_t\langle \mathbf{F}_t'\boldsymbol{\theta}_t \rangle^{(k+1)} + \langle (\mathbf{F}_t'\boldsymbol{\theta}_t)^2 \rangle^{(k+1)}) \\ & - 2\Big(\frac{C(p)|\gamma|}{B(p)} \left\langle \frac{1}{v_t} \right\rangle^{(k+1)} \langle s_t \rangle^{(k+1)} + \frac{A(p)}{\sigma B(p)}\Big)(y_t - \langle \mathbf{F}_t'\boldsymbol{\theta}_t \rangle^{(k+1)}) \\ & + 2\frac{C(p)|\gamma|A(p)}{B(p)}\langle s_t \rangle^{(k+1)} + \frac{C(p)^2\sigma|\gamma|^2}{B(p)}\langle s_t^2 \rangle^{(k+1)} \left\langle \frac{1}{v_t} \right\rangle^{(k+1)} + \frac{A(p)^2}{\sigma B(p)}\langle v_t \rangle^{(k+1)} \Big] \Big\} \end{aligned}$$

where $f_0(\sigma)$ and $f_0(\gamma)$ denote the prior distributions of σ and γ, respectively. Therefore, we can update $r^{(k+1)}(\sigma, \gamma)$ with importance sampling as follows:

- For n in $1, ..., N$, sample $(\sigma^n, \gamma^n) \sim l(\sigma, \gamma)$ where $l(\sigma, \gamma)$ denotes the chosen proposal distribution.
- Compute the weights

$$w(\sigma^n, \gamma^n) = \frac{r^{(k+1)}(\sigma^n, \gamma^n)}{l(\sigma^n, \gamma^n)} \tag{30}$$

The variational distribution $r^{(k+1)}(\sigma, \gamma)$ can be approximated by

$$r^{(k+1)}(\sigma,\gamma) \approx \frac{\sum_{n=1}^N w(\sigma^n, \gamma^n)\delta_{(\sigma^n,\gamma^n)}(\sigma,\gamma)}{\sum_{n=1}^N w(\sigma^n, \gamma^n)}. \tag{31}$$

Similarly, for any function $h(\sigma, \gamma)$,

$$\mathbb{E}[h(\sigma,\gamma)] \approx \frac{\sum_{n=1}^{N} h(\sigma^n,\gamma^n) w(\sigma^n,\gamma^n)}{\sum_{n=1}^{N} w(\sigma^n,\gamma^n)}. \tag{32}$$

A.4. ISVB Closed Form Integrals. For notational simplicity, we omit the superscript indicating the VB iteration.

If $r(s_t) = \mathrm{N}^+(\mu_{s_t}, \sigma^2_{s_t})$, $\phi(\cdot)$ is the probability density function of the standard normal distribution, and $\Phi(\cdot)$ is its cumulative distribution function, then

$$\langle s_t \rangle = \mu_{s_t} + \sigma_{s_t} \frac{\phi(\mu_{s_t}/\sigma_{s_t})}{\Phi(\mu_{s_t}/\sigma_{s_t})} \tag{33}$$

$$\langle s_t^2 \rangle = \mu_{s_t}^2 + \sigma_{s_t}^2 + \mu_{s_t}\sigma_{s_t} \frac{\phi(\mu_{s_t}/\sigma_{s_t})}{\Phi(\mu_{s_t}/\sigma_{s_t})} \tag{34}$$

If $r(v_t) = \mathrm{GIG}(\lambda_{v_t}, \chi_{v_t}, \psi_{v_t})$ and $K_\lambda(\cdot)$ is a modified Bessel Function of the second kind with order λ, then

$$\langle v_t \rangle = \frac{\sqrt{\chi_{v_t}} K_{\lambda_{v_t}+1}(\sqrt{\chi_{v_t}\psi_{v_t}})}{\sqrt{\psi_{v_t}} K_{\lambda_{v_t}}(\sqrt{\chi_{v_t}\psi_{v_t}})} \tag{35}$$

$$\left\langle \frac{1}{v_t} \right\rangle = \frac{\sqrt{\psi_{v_t}} K_{\lambda_{v_t}+1}(\sqrt{\chi_{v_t}\psi_{v_t}})}{\sqrt{\chi_{v_t}} K_{\lambda_{v_t}}(\sqrt{\chi_{v_t}\psi_{v_t}})} - \frac{2\lambda_{v_t}}{\chi_{v_t}} \tag{36}$$

If $r(\boldsymbol{\theta}_t) = \mathrm{N}(\mathbf{m}_t^s, \mathbf{C}_t^s)$, then

$$\langle \mathbf{F}_t' \boldsymbol{\theta}_t \rangle = \mathbf{F}_t' \mathbf{m}_t^s \tag{37}$$

$$\langle (\mathbf{F}_t' \boldsymbol{\theta}_t)^2 \rangle = \mathbf{F}_t' \mathbf{C}_t^s \mathbf{F}_t + (\mathbf{F}_t' \mathbf{m}_t^s)^2 \tag{38}$$

Lastly, if $r(\sigma, \gamma)$ is approximated with IS according to Equation (31), the following expectations can be approximated using Equation (32);

$$\left\langle \frac{C(p)^2\sigma|\gamma|^2}{B(p)} \right\rangle, \left\langle \frac{1}{\sigma} \right\rangle, \left\langle \frac{C(p)|\gamma|}{B(p)} \right\rangle, \left\langle \frac{C(p)|\gamma|A(p)}{B(p)} \right\rangle, \left\langle \frac{1}{\sigma B(p)} \right\rangle, \left\langle \frac{A(p)}{\sigma B(p)} \right\rangle, \text{ and } \left\langle \frac{A(p)^2}{\sigma B(p)} \right\rangle.$$

Supplemental Material. The R code of Algorithms 1 and 2, as well as the Santa Cruz IVT and ELI time series data used to estimate the 0.85 quantile IVT thresholds in Section 5 are available in the Supplemental Material.

Acknowledgements. The authors wish to thank Bin Guan and Duane Waliser at NASA Jet Propulsion Laboratory for sharing their IVT and AR datasets. The AR database is available at https://ucla.app.box.com/v/ARcatalog. We also thank Christina Patricola at Lawrence Berkeley National Laboratory for helpful conversation about the ELI. The ELI is available at https://portal.nersc.gov/archive/home/projects/cascade/www/ELI.

REFERENCES

BACKES, T. M., KAPLAN, M. L., SCHUMER, R. and MEJIA, J. F. (2015). A climatology of the vertical structure of water vapor transport to the Sierra Nevada in cool season atmospheric river precipitation events. *Journal of Hydrometeorology* **16** 1029–1047.

BEAL, M. J. (2003). Variational algorithms for approximate Bayesian inference, PhD thesis, UCL (University College London).

BERNARDI, M., CASARIN, R., MAILLET, B. and PETRELLA, L. (2016). Dynamic Model Averaging for Bayesian Quantile Regression. *arXiv preprint arXiv:1602.00856*.

BERRISFORD, P., KÅLLBERG, P., KOBAYASHI, S., DEE, D., UPPALA, S., SIMMONS, A., POLI, P. and SATO, H. (2011). Atmospheric conservation properties in ERA-Interim. *Quarterly Journal of the Royal Meteorological Society* **137** 1381–1399.

CARTER, C. K. and KOHN, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81** 541–553.

CHEN, W. Y., PETERS, G. W., GERLACH, R. H. and SISSON, S. A. (2017). Dynamic quantile function models. *arXiv preprint arXiv:1707.02587*.

DEE, D. P., UPPALA, S., SIMMONS, A., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M., BALSAMO, G., BAUER, P. et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* **137** 553–597.

FRÜHWIRTH-SCHNATTER, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis* **15** 183–202.

GELFAND, A. E. and GHOSH, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* **85** 1–11.

GONÇALVES, K., MIGON, H. S. and BASTOS, L. S. (2017). Dynamic quantile linear model: a Bayesian approach. *arXiv preprint arXiv:1711.00162*.

GUAN, B. and WALISER, D. E. (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres* **120** 12514–12535.

GUAN, B., WALISER, D. E., MOLOTCH, N. P., FETZER, E. J. and NEIMAN, P. J. (2012). Does the Madden–Julian oscillation influence wintertime atmospheric rivers and snowpack in the Sierra Nevada? *Monthly Weather Review* **140** 325–342.

GUAN, B., MOLOTCH, N. P., WALISER, D. E., FETZER, E. J. and NEIMAN, P. J. (2013). The 2010/2011 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resources Research* **49** 6731–6743.

HANSON, T. and JOHNSON, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97** 1020–1033.

HENZE, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scandinavian journal of statistics* 271–275.

HUERTA, G., JIANG, W. and TANNER, M. A. (2003). Time series modeling via hierarchical mixtures. *Statistica Sinica* 1097–1118.

KASTNER, G. (2016). Dealing with Stochastic Volatility in Time Series Using the R Package stochvol. *Journal of Statistical Software, Articles* **69** 1–30.

KOENKER, R. and XIAO, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association* **101** 980–990.

KOTTAS, A. and GELFAND, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* **96** 1458–1468.

KOTTAS, A. and KRNJAJIĆ, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics* **36** 297–319.

KOTZ, S., KOZUBOWSKI, T. and PODGORSKI, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.

KOZUMI, H. and KOBAYASHI, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation* **81** 1565–1578.

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics* **22** 79–86.

LISEO, B. and LOPERFIDO, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* **136** 373–389.

LUM, K., GELFAND, A. E. et al. (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis* **7** 235–258.

NEELON, B., LI, F., BURGETTE, L. F. and BENJAMIN NEELON, S. E. (2015). A spatiotemporal quantile regression model for emergency department expenditures. *Statistics in medicine* **34** 2559–2575.

NEIMAN, P. J., WHITE, A. B., RALPH, F. M., GOTTAS, D. J. and GUTMAN, S. I. (2009). A water vapour flux tool for precipitation forecasting. In *Proceedings of the Institution of Civil Engineers-Water Management* **162** 83–94. Thomas Telford Ltd.

OSTWALD, D., KIRILINA, E., STARKE, L. and BLANKENBURG, F. (2014). A tutorial on variational Bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology* **60** 1–19.

PARASCHIV, F., BUNN, D. and WESTGAARD, S. (2016). Estimation and application of fully parametric multi-factor quantile regression with dynamic coefficients.

PATRICOLA, C. M., OÂĂŹBRIEN, J. P., RISSER, M. D., RHOADES, A. M., OÂĂŹBRIEN, T. A., ULLRICH, P. A., STONE, D. A. and COLLINS, W. D. (2020). Maximizing ENSO as a source of western US hydroclimate predictability. *Climate Dynamics* **54** 351–372.

PRADO, R., MOLINA, F. and HUERTA, G. (2006). Multivariate time series modeling and classification via hierarchical VAR mixtures. *Computational Statistics & Data Analysis* **51** 1445–1462.

REICH, B. J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61** 535–553.

REICH, B. J., BONDELL, H. D. and WANG, H. J. (2009). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* **11** 337–352.

REICH, B. J., FUENTES, M. and DUNSON, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association* **106** 6–20.

REICH, B. J. and SMITH, L. B. (2013). Bayesian quantile regression for censored data. *Biometrics* **69** 651–660.

ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics* **23** 470–472.

RUTZ, J. J., STEENBURGH, W. J. and RALPH, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Monthly Weather Review* **142** 905–921.

TADDY, M. A. and KOTTAS, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business & Economic Statistics* **28** 357–369.

TOKDAR, S. T., KADANE, J. B. et al. (2012). Simultaneous linear quantile regression: A semiparametric Bayesian approach. *Bayesian Analysis* **7** 51–72.

TSIONAS, E. G. (2003). Bayesian quantile inference. *Journal of statistical computation and simulation* **73** 659–674.

TUCKERMAN, M. (2010). *Statistical mechanics: theory and molecular simulation*. Oxford university press.

TZIPERMAN, E., CANE, M. A., ZEBIAK, S. E., XUE, Y. and BLUMENTHAL, B. (1998). Locking of El Nino's peak time to the end of the calendar year in the delayed oscillator picture of ENSO. *Journal of climate* **11** 2191–2199.

WALKER, S. and MALLICK, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55** 477–483.

WELLER, G. B., COOLEY, D. S. and SAIN, S. R. (2012). An investigation of the pineapple express phenomenon via bivariate extreme value theory. *Environmetrics* **23** 420–439.

WEST, M. and HARRISON, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

WICHITAKSORN, N., CHOY, S. B. and GERLACH, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics* **42** 579–596.

WILLIAMS, I. N. and PATRICOLA, C. M. (2018). Diversity of ENSO events unified by convective threshold sea surface temperature: a nonlinear ENSO index. *Geophysical Research Letters* **45** 9236–9244.

YAN, Y. and KOTTAS, A. (2017). A new family of error distributions for Bayesian quantile regression. *arXiv preprint arXiv:1701.05666*.

YU, K. and MOYEED, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* **54** 437–447.

ZHU, D. and GALBRAITH, J. W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions. *Journal of Empirical Finance* **18** 765–778.

ZHU, D. and ZINDE-WALSH, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of econometrics* **148** 86–99.