Computational Statistics and Data Analysis 54 (2010) 1138–1150

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# Bayesian nonparametric quantile regression using splines

Paul Thompson [a,*], Yuzhi Cai [a], Rana Moyeed [a], Dominic Reeve [b], Julian Stander [a]

[a] School of Computing & Mathematics, 2-5 Kirby Place, University of Plymouth, Devon, PL4 8AA, UK

[b] C-CoDE, School of Engineering, Reynolds Building, University of Plymouth, Devon, PL4 8AA, UK

ARTICLE INFO

*Article history:*
Received 2 September 2008
Received in revised form 7 September 2009
Accepted 7 September 2009
Available online 15 September 2009

ABSTRACT

A new technique based on Bayesian quantile regression that models the dependence of a quantile of one variable on the values of another using a natural cubic spline is presented. Inference is based on the posterior density of the spline and an associated smoothing parameter and is performed by means of a Markov chain Monte Carlo algorithm. Examples of the application of the new technique to two real environmental data sets and to simulated data for which polynomial modelling is inappropriate are given. An aid for making a good choice of proposal density in the Metropolis–Hastings algorithm is discussed. The new nonparametric methodology provides more flexible modelling than the currently used Bayesian parametric quantile regression approach.



## 1. Introduction

Quantile regression can be described as a method that provides a more complete inferential picture than ordinary least-squares regression. The latter technique estimates the conditional mean of some response variable $Y$ given the value $t$ of a covariate, while quantile regression takes a different approach by estimating the conditional quantiles of $Y$ given $t$. More precisely, in quantile regression we are interested in estimating quantile functions $q_p(t)$, $0 < p < 1$, such that $\Pr(Y \leq q_p(t)$ given that the covariate takes the value $t) = p$. This allows the full range of the data to be modelled and so can be beneficial when large values are of particular interest. It also means that quantile regression can be viewed as a data exploration technique. Koenker (2005) presents a wide range of discussion about the use of quantile regression. Quantile regression can be implemented in a range of different forms, and Yu et al. (2003) provide an overview of some commonly used quantile regression techniques.

Bayesian inference for quantiles has been considered by several authors, including Yu and Moyeed (2001) and Dunson and Taylor (2005). Dunson and Taylor (2005) discuss Bayesian inference for quantiles when the likelihood function is not fully specified. They present an example based on a linear quantile regression function. Kottas and Gelfand (2001) consider Bayesian semiparametric median regression modelling under a Dirichlet process mixing framework. Kottas and Krnjajić (2009) extend this approach to general quantiles. Laurini and Pauli (2009) discuss smoothing sample extremes using a mixed model approach. Other relevant articles from this journal include Lejeune and Sarda (1988), Ng (1996), Doksum and Koo (2000), and Wang et al. (2009).

The Bayesian quantile regression (BQR) methodology developed in Yu and Moyeed (2001) adopts a parametric approach based on a polynomial quantile regression function. Inference about the posterior distribution of the parameters of this regression function is made by means of a Markov chain Monte Carlo (MCMC) algorithm. Although Yu and Moyeed (2001) present excellent results, there are certain drawbacks associated with using polynomials. These include the influence of outliers and the need to choose the degree of the polynomial, possibly for each quantile considered. Also, the data may

* Corresponding author. Tel.: +44 1752 232640; fax: +44 1752 586900.
*E-mail addresses:* p1thompson@plymouth.ac.uk, Paul.Thompson1@plymouth.ac.uk (P. Thompson).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.09.004

have a limited local effect on the shape of a polynomial regression curve especially when modelling high quantiles. Yu and Moyeed (2001) work with low order polynomials; problems associated with using very high order polynomials may include over-fitting and poor MCMC convergence.

In this article, we present a nonparametric alternative to the parametric approach of Yu and Moyeed (2001) based on using natural cubic splines (NCS)rather than polynomials. Our approach provides a more versatile and flexible method of fitting a quantile regression curve. Section 2 of the article gives an introduction to NCS. It then presents our Bayesian nonparametric quantile regression methodology by describing the posterior density of the spline and an associated smoothing parameter and outlining a MCMC algorithm for making inferences from this posterior. Section 3 presents the applications of our methodology to two real environmental data sets and to simulated data. The first data set comprises of coastal wave conditions from near the Selsey Bill area and were generated using a hindcasting technique (see Reeve et al. (2004)) using wind records. The data comprise hourly hindcast measurements of the variables significant wave height, wave period and wave direction over an approximate time span of 27 years. A better understanding of this type of data is important for the coastal design process, as illustrated by Thompson et al. (2008). Here, our variable $Y$ of interest will be wave height, while the covariate $t$ will be the cosine of wave direction. In this example, we take a random sample of 10,000 observations for computational and presentational reasons. The resulting data set is shown later in Fig. 1 and will be denoted by $(t_1, y_1), \ldots, (t_n, y_n)$, where sample size $n = 10{,}000$. Our second application is similar, but consists of hindcast offshore wave data from Poole Bay, UK. The simulated data for our third application are based on a well-known published data set for which polynomial modelling is inappropriate. Our applications allow us to illustrate learning about model parameters from the data. We also discuss the advantages offered by our nonparametric methodology. In Section 4 we, discuss the performance of our MCMC procedure and present an aid for making a good choice of proposal density in the Metropolis–Hastings algorithm, thereby improving its efficiency. Finally, Section 5 is a short conclusion on the study.

## 2. Bayesian nonparametric quantile regression methodology

When fitting a curve through a bivariate data set, one important consideration is the roughness of the curve, i.e. how "wiggly" it is. More specifically, we tend to prefer smooth curves that have a reduced amount of rapid fluctuation. We are able to quantify the roughness of a curve $g$ with continuous second derivative on the interval $[a, b]$ by means of a roughness penalty, which is defined in this article as the integrated squared second derivative $\int_a^b g''(t)^2 \mathrm{d}t$; see Green and Silverman (1994). A standard approach to curve fitting is based on a trade-off between the lack-of-fit of a curve to the data and its roughness, or, equivalently, between goodness-of-fit and smoothness, as discussed in Green and Silverman (1994). These authors have also shown how this approach can be formalized within the Bayesian framework (see Gamerman (1997), for example) by having a prior distribution which quantifies probabilistically the roughness of the fitted curve; we describe this in detail in Section 2.2.

In Section 2.1, we define NCS by following the standard approach given by Green and Silverman (1994). The aim of this article is to include the natural cubic spline in the BQR methodology of Yu and Moyeed (2001), thereby extending and making their parametric technique more flexible. Full details of our proposed methodology are given in Section 2.2.

### 2.1. The natural cubic spline and associated results

We say that a curve $g$ is a cubic spline with $N \geq 2$ knots $\tau_1 < \cdots < \tau_N$ if $g$ is a cubic polynomial between knots $\tau_{i-1}$ and $\tau_i$, $i = 2, \ldots, N$, and if $g$ has continuous first and second derivatives at $\tau_i$, $i = 2, \ldots, N-1$. Let $a < \tau_1$ and $b > \tau_N$. The cubic spline $g$ is said to be a NCS on $[a, b]$ if it is linear on the intervals $[a, \tau_1]$ and $[\tau_N, b]$ and if it has continuous first and second derivatives at $\tau_1$ and $\tau_N$. This definition of a NCS is equivalent to the one given by Green and Silverman (1994, pages 11–12); see also Hastie et al. (2001, Section 5.2).

We now introduce notation and present results that we will use later. Let $\boldsymbol{g} = (g_1, \ldots, g_N)^{\mathrm{T}}$ be a column vector of values $g_i = g(\tau_i)$, $i = 1, \ldots, N$, of a NCS $g$ at its knots $\tau_1, \ldots, \tau_N$. Further, let $h_i = \tau_{i+1} - \tau_i$, $i = 1, \ldots, N-1$, let $Q$ be the $N \times (N-2)$ banded matrix with entries $q_{ij}$, $i = 1, \ldots, N$ and $j = 2, \ldots, N-1$, given by $q_{j-1,j} = 1/h_{j-1}$, $q_{jj} = -1/h_{j-1} - 1/h_j$, $q_{j+1,j} = 1/h_j$ and $q_{ij} = 0$ for $|i-j| \geq 2$, and let $R$ be the $(N-2) \times (N-2)$ banded positive definite symmetric matrix with entries $r_{ii} = (h_{i-1} + h_i)/3$, $i = 2, \ldots, N-1$, $r_{i,i+1} = r_{i+1,i} = h_i/6$, $i = 2, \ldots, N-2$, and $r_{ij} = 0$ for $|i-j| \geq 2$. We can now define the $N \times N$ symmetric matrix $K$ with rank $N-2$ as $K = QR^{-1}Q^{\mathrm{T}}$. We will make use of Theorem 2.1 of Green and Silverman (1994) that tells us that the roughness penalty satisfies

$$\int_a^b g''(t)^2 \mathrm{d}t = \boldsymbol{g}^{\mathrm{T}} K \boldsymbol{g}. \tag{1}$$

We will also use Theorem 2.2 of Green and Silverman (1994) that establishes that, given any values $g_1, \ldots, g_N$, there is a unique NCS $g$ with knots at $\tau_1, \ldots, \tau_N$ satisfying $g(\tau_i) = g_i$, $i = 1, \ldots, N$.

### 2.2. Bayesian nonparametric quantile modelling and inference

In this section, we present a framework for Bayesian nonparametric quantile regression using splines rather than polynomials as in Yu and Moyeed (2001). In our approach, we model a quantile function of a covariate $t$ using a NCS with

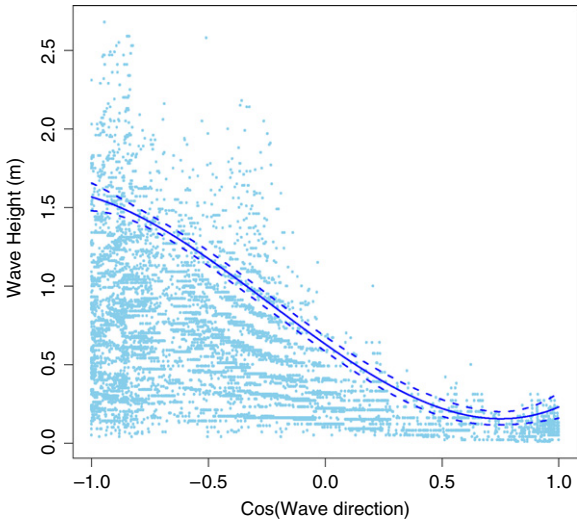


**Fig. 1.** Scatter plot of the coastal wave data showing the $p = 0.9$ BQR curve using a cubic polynomial. A 95% credible envelope is also presented.

$N$ fixed knots at points $\tau_1, \ldots, \tau_N$ that cover the range of $t$. The NCS is uniquely determined by its values $\boldsymbol{g} = (g_1, \ldots, g_N)^{\mathrm{T}}$ at these knots, since, as explained in Section 2.1, there is a unique NCS that can be drawn through the points $(\tau_i, g_i)$, $i = 1, \ldots, N$. As our approach is Bayesian, we begin by defining the prior density for $\boldsymbol{g}$ as a multivariate normal; see Green and Silverman (1994, page 51) for a discussion about the use of the multivariate normal density as a prior in this context.

Our prior for $\boldsymbol{g}$ is defined by means of the multivariate normal density

$$\pi(\boldsymbol{g}|\lambda) = \frac{\lambda^{(N-2)/2}}{(2\pi)^{(N-2)/2}(\mu_1 \cdots \mu_{N-2})^{1/2}} \exp\left(-\frac{1}{2}\lambda\, \boldsymbol{g}^{\mathrm{T}} K \boldsymbol{g}\right), \tag{2}$$

in which $\mu_1, \ldots, \mu_{N-2}$ are the inverses of the $N-2$ non-zero eigenvalues of $K$ and $\lambda > 0$ is an unknown parameter. More detail about this multivariate normal distribution can be found in Rao (2002, page 528). Note that (2) depends through (1) on the roughness $\int_a^b g''(t)^2 \mathrm{d}t = \boldsymbol{g}^{\mathrm{T}} K \boldsymbol{g}$ of the NCS $g$ uniquely defined by $\boldsymbol{g}$. As larger values of $\lambda$ result in more probability density being given to less rough curves $g$, we will refer to $\lambda$ as a smoothing parameter.

We next require a priority on the smoothing parameter $\lambda$, which is constrained by a lower limit of zero. Hence, we follow standard practice by using the gamma density as our prior for $\lambda$ which takes the form

$$\pi(\lambda) = \frac{\lambda^{\alpha-1}\exp(-\lambda/\beta)}{\Gamma(\alpha)\beta^{\alpha}}, \quad \lambda > 0, \tag{3}$$

in which $\Gamma$ is the usual gamma function. The user is able to specify the hyperparameters $\alpha$ and $\beta$. Under this prior $\mathrm{E}[\lambda] = \alpha\beta$ and $\mathrm{Var}[\lambda] = \alpha\beta^2$, results can be used to guide hyperparameter choice.

The final step in our Bayesian approach is to define the likelihood of the data $(t_i, y_i)$, $i = 1, \ldots, n$, given $\boldsymbol{g}$. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$. We proceed in accordance with the BQR approach of Yu and Moyeed (2001) by substituting our NCS $g$ for their polynomial. The resulting likelihood takes the form:

$$L(\boldsymbol{y}|\boldsymbol{g}) = p^n(1-p)^n \exp\left\{-\sum_{i=1}^{n} \rho_p(y_i - g(t_i))\right\} \tag{4}$$

where $p$ is the probability corresponding to the quantile of interest, $0 < p < 1$, and $\rho_p$ is the standard quantile regression loss function

$$\rho_p(u) = u(p - I(u < 0)) \tag{5}$$

in which $I$ is the usual indicator function. The values of $g(t_i)$, $i = 1, \ldots, n$, in (4) are uniquely determined by $\boldsymbol{g}$. We note that the likelihood is not dependent on $\lambda$. Combining $\pi(\lambda)$, $\pi(\boldsymbol{g}|\lambda)$ and $L(\boldsymbol{y}|\boldsymbol{g})$, we can write the posterior density function of $\boldsymbol{g}$ and $\lambda$ as

$$\pi(\boldsymbol{g}, \lambda|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{g})\pi(\boldsymbol{g}|\lambda)\pi(\lambda). \tag{6}$$

We now simulate realizations of $\boldsymbol{g}$ and $\lambda$ from this posterior density using an MCMC approach implemented through the Metropolis–Hastings algorithm; see Gamerman (1997), for example. Our inferences will be based on these posterior realizations. In particular, we shall use the posterior mean $(\overline{g_1}, \ldots, \overline{g_N})$ of $\boldsymbol{g}$ to produce our estimated quantile. Our algorithm can be summarized as follows:

(i) Assign initial values $\boldsymbol{g}^{(0)}$ and $\lambda^{(0)}$ to $\boldsymbol{g}$ and $\lambda$. We set $\boldsymbol{g}^{(0)}$ to be the values at $\tau_1, \ldots, \tau_N$ of the posterior mean cubic quantile regression curve obtained using the methodology of Yu and Moyeed (2001). The cubic quantile regression curve was

chosen as this is also an example of a cubic spline, although a very constrained one. We obtain the value of $\lambda^{(0)}$ by applying generalized cross validation (GCV) to the usual mean smoothing spline; see Green and Silverman (1994). We chose this value, which we shall refer to as GCV(mean spline), because it can be found easily and quickly using R's (R Development Core Team, 2009) `smooth.spline` function (see Venables and Ripley (2002), for example); Section 4.2 provides brief further discussion. We set iteration number $j = 1$.

(ii) We generate a candidate vector $\boldsymbol{g}^\star$ from the multivariate normal distribution

$$\boldsymbol{g}^\star|\boldsymbol{g}^{(j-1)} \sim MVN(\boldsymbol{g}^{(j-1)}, \Sigma) \tag{7}$$

with mean $\boldsymbol{g}^{(j-1)}$ and variance–covariance matrix $\Sigma = \sigma^2 K^-/\lambda$, where $K^-$ is the generalized inverse of $K$. The constant $\sigma^2$ is specified by the user; see Section 4.2.

(iii) We set $\boldsymbol{g}^{(j)}$ to $\boldsymbol{g}^\star$ with probability:

$$\alpha(\boldsymbol{g}^{(j-1)}, \boldsymbol{g}^\star) = \min\left\{1, \frac{L(\boldsymbol{y}|\boldsymbol{g}^\star)\pi(\boldsymbol{g}^\star|\lambda^{(j-1)})q(\boldsymbol{g}^{(j-1)}|\boldsymbol{g}^\star)}{L(\boldsymbol{y}|\boldsymbol{g}^{(j-1)})\pi(\boldsymbol{g}^{(j-1)}|\lambda^{(j-1)})q(\boldsymbol{g}^\star|\boldsymbol{g}^{(j-1)})}\right\} \tag{8}$$

where the proposal density $q(\boldsymbol{g}^\star|\boldsymbol{g}^{(j-1)})$ is the probability density function of the multivariate normal specified in (7). Sincee $q$ is symmetric in its arguements, it cancels out of (8). Otherwise, $\boldsymbol{g}^{(j)} = \boldsymbol{g}^{(j-1)}$.

(iv) We now generate a candidate $\lambda^\star$ from the log-normal distribution as follows:

$$\mu^\star \sim N(\log(\lambda^{j-1}), \sigma_\lambda^2), \quad \lambda^\star = \exp(\mu^\star) \tag{9}$$

where the normal distribution has mean $\log(\lambda^{(j-1)})$ and variance $\sigma_\lambda^2$, which can be specified by the user; see Section 4.2.

(v) We set $\lambda^{(j)}$ to $\lambda^\star$ with probability:

$$\alpha(\lambda^{(j-1)}, \lambda^\star) = \min\left\{1, \frac{\pi(\boldsymbol{g}^{(j)}|\lambda^\star)\pi(\lambda^\star)q(\lambda^{(j-1)}|\lambda^\star)}{\pi(\boldsymbol{g}^{(j)}|\lambda^{(j-1)})\pi(\lambda^{(j-1)})q(\lambda^\star|\lambda^{(j-1)})}\right\} \tag{10}$$

where $q$ is the log-normal probability density function specified through (9). In this case, cancellation of the $q$ terms in (10) is not possible as $q$ is not symmetric in its arguements. Otherwise, $\lambda^{(j)} = \lambda^{(j-1)}$.

(vi) We now increment $j$ by 1, and repeat steps (ii)–(vi) for a total of $d$ iterations.

Whilst the methodology of Yu and Moyeed (2001) updates the parameters of a fixed degree regression polynomial at each iteration of the Metropolis–Hastings algorithm, our methodology updates both the entire vector of values $\boldsymbol{g}$ at the fixed knots of the NCS and the smoothing parameter $\lambda$. We set the number of iterations $d$ to 500,000. We allow a burn-in of 50,000 iterations. Inference is based on thinned values of $\boldsymbol{g}$ and $\lambda$ produced by the Metropolis–Hastings algorithm after burn-in; we used every tenth value. Convergence issues are discussed in detail in Section 4.2. All codes were written in R (R Development Core Team, 2009), using R's random number generating functions.

A considerable advantage of the Bayesian approach is that we can calculate associated credible intervals to provide an idea of the associated posterior uncertainty. These credible intervals are obtained by ordering the thinned $g^{(j)}(\tau_i)$ sequence over $j > 50,000$ and extracting the values which correspond to, for example, the 2.5% and 97.5% quantiles. A 95% posterior credible interval for $\lambda$ can be obtained in a similar way. In the next section, examples of this methodology applied to the data sets discussed in Section 1 are presented.

Although we have adopted the Metropolis–Hastings algorithm to simulate realizations from our posterior, we note that due to its multivariate nature, potentially more efficient samplers may be available. Neal (2003) presents an alternative sampling technique called slice sampling, based on the principle that we can simulate from a distribution by sampling uniformly from the area below its plotted density function. The algorithm proceeds by alternating two steps: uniform sampling in the 'vertical' direction at the current 'horizontal' point, and uniform sampling from the 'horizontal' slice defined by the the current 'vertical' position. This latter step can be computationally very demanding with the consequence that the computational expense of slice sampling may outweigh any potential advantages over our more simple Metropolis–Hastings algorithm.

We finish this section by remarking that another approach to quantile regression is based on the minimization over curves $g$ of

$$\sum_{i=1}^{n} \rho_p(y_i - g(t_i)). \tag{11}$$

Often $g$ is taken to be a B-spline (Hastie et al., 2001) or a NCS with pre-specified knots and hence smoothness. The minimizing $g$ can be found using the `quantreg` package (Koenker, 2008) running under R (R Development Core Team, 2009); see Koenker (2005) for an example. Some other authors have considered the problem of minimizing over curves $g$ belonging to a suitable space a version of (11) penalized for roughness such as

$$\sum_{i=1}^{n} \rho_p(y_i - g(t_i)) + \lambda \int_a^b g''(t)^2 \mathrm{d}t; \tag{12}$$

see Bosch et al. (1995) and reference therein, and Koenker et al. (1994) for further discussion. Koenker et al. (1994) also describe a similar minimization approach based on a total variation roughness penalty; software for this is again available in Koenker (2008). As far as we know, none of these approaches routinely yield confidence envelopes for the estimated curve.

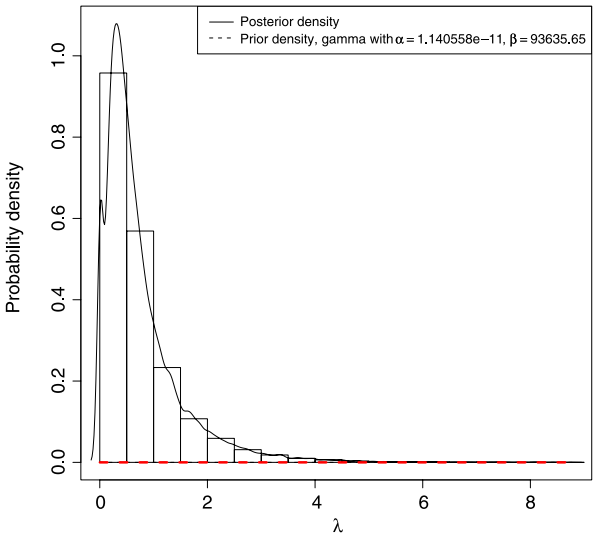


**Fig. 2.** Plot comparing the prior and posterior densities of $\lambda$ given the coastal wave data. The prior is a gamma density with parameters $\alpha \approx 10^{-11}$ and $\beta \approx 10^{5}$ and is effectively flat over a large range of $\lambda$ values. The posterior density is very different from the prior, clearly showing that learning about $\lambda$ has taken place.

## 3. Applied examples

### 3.1. Application to coastal wave data

To illustrate the practical effectiveness of the approach described in Section 2, we present the results obtained by applying our methodology to the hindcast coastal wave data discussed in Section 1 and plotted in Fig. 1. This plot also shows the parametric Bayesian quantile cubic regression curve of Yu and Moyeed (2001) for $p = 0.9$ together with a 95% credible envelope. In our spline based approach, the number $N$ of knots and their position $\tau_1, \ldots, \tau_N$ can be chosen by the user. We have found that taking $N = 30$ knots at $\tau_1 < \cdots < \tau_{30}$ equally spaced over the range of covariate values $t_1, \ldots, t_n$ allows flexible modelling without imposing a very high computational burden. We remark that in the context of mean regression some authors such as Denison et al. (1998) and Dias and Gamerman (2002) have also made inference about the number and position of knots. The resulting algorithm is based on the reversible jump MCMC method of Green (1995) and can be computationally highly demanding.

We set the gamma prior hyperparameters $\beta = 0.1/\text{GCV(mean spline)} \approx 10^{5}$ and $\alpha = \text{GCV(mean spline)}/\beta \approx 10^{-11}$ in which $\text{GCV(mean spline)} \approx 10^{-6}$. With these hyperparameters the prior mean and variance of $\lambda$ are $\text{E}[\lambda] \approx 10^{-6}$ and $\text{Var}[\lambda] \approx 0.1$, representing a large amount of prior uncertainty about $\lambda$. We set the hyperparameters to yield an sensible expected $\lambda$, which is comparable with the GCV value for $\lambda$ from the usual mean smoothing spline. Fig. 2 displays the prior and posterior densities of $\lambda$ for this example. The difference between the prior and the posterior of $\lambda$ clearly shows that learning about $\lambda$ has been achieved. This is to be expected for such a diffuse prior. Learning about $\lambda$ can still be achieved with an even more informative prior as shown in Fig. 3, with hyperparameters $\alpha = 160$ and $\beta = 0.025$. Here, the marginal posterior of $\lambda$ lies between the prior and the posterior shown in Fig. 2 based on much larger uncertainty about $\lambda$.

Fig. 4 presents the resulting Bayesian nonparametric quantile regression curve and 95% credible envelope for our first choice of gamma prior hyperparameters. To obtain the regression curve shown in Fig. 4, we drew the unique NCS through the points $(\tau_i, \overline{g_i})$, $i = 1, \ldots, N$, using R's `spline` function (R Development Core Team, 2009). Similarly, we produce our 95% credible envelope by drawing NCSs through the 2.5% and 97.5% posterior quantiles found in Section 2.2. The more local nature of the fitting procedure is easily seen from Fig. 4. In order to judge the goodness-of-fit of both approaches, we found empirical and fitted quantiles on a grid of 100 sections along the covariate and calculated 'residuals' as:

$$\text{residual} = \text{empirical quantile} - \text{fitted quantile}, \tag{13}$$

in which for each grid section the empirical quantile is the $p$th quantile of the $y$ data values in the section and the fitted quantile is the value produced by our model at the centre of the section. As usual, smaller residuals in absolute value are associated with better fits. Fig. 5 shows the absolute value of the residuals from both the cubic polynomial quantile regression curve shown in Fig. 1 and the spline based curve shown in Fig. 4 against the cosine of wave direction. A robust locally linear smoother provided by R's (R Development Core Team, 2009) `loess` function (see Venables and Ripley (2002), for example) was added through each set of (covariate, |residual|) points. These smoothers indicate that the spline based quantile curve gives a better quality of fit through almost the full covariate range than the cubic polynomial quantile. This is due to the more local nature of the spline based fitting procedure. We also calculated the mean square error based on the residuals for each

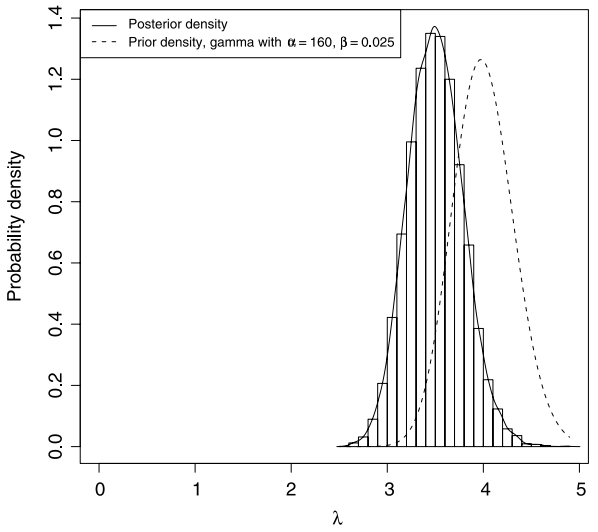


**Fig. 3.** Plot comparing the prior and posterior densities of $\lambda$ given the coastal wave data for a stronger prior with $\alpha = 160$ and $\beta = 0.025$. The posterior density is different from the prior, showing that learning about $\lambda$ has taken place.

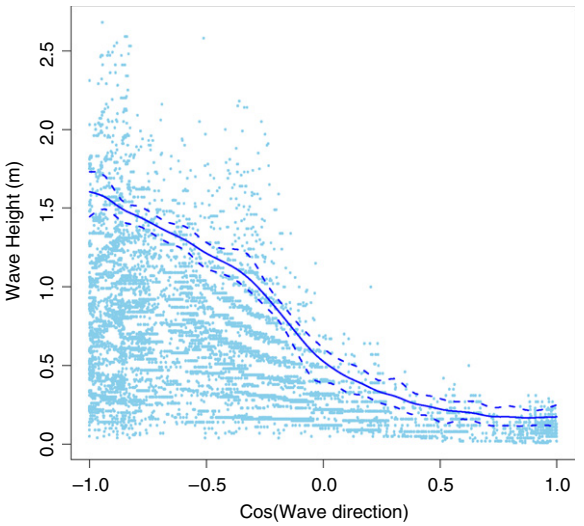


**Fig. 4.** Scatter plot of the coastal wave data showing the $p = 0.9$ Bayesian nonparametric quantile regression curve using splines. A 95% credible envelope is also presented.

model as a further method of assessing goodness-of-fit. We obtained mean square error values of 0.010 and 0.016 for the spline and polynomial based approach, respectively. This is a further indication of the improvement that the nonparametric approach provides over its parametric counterpart. We should, however, bear in mind that in general goodness-of-fit and smoothing are competing aims in curve fitting.

Finally, we remark that the $p = 0.9$ Bayesian nonparametric quantile regression curve using splines obtained with the gamma hyperparameters $\alpha = 160$ and $\beta = 0.025$ was very similar to that shown in Fig. 4. The credible envelope was, however, somewhat smoother. Our experience is that an estimated quantile curve is relatively insensitive to the hyperparameter choice.

### *3.2. Application to offshore wave data*

In our second example, we use hindcast offshore wave data from Poole Bay, UK to further illustrate and validate our approach. There are three variables: wave height, wave period and wave direction, each having 86,384 observations at 3-h intervals, which amounts to just over 29 years of data. The data that we use are shown in Fig. 6. We can see that this data set has a different underlying structure from the coastal wave data as there is less variation in the magnitude of values (including high values) over the direction covariate. We also show in Fig. 6, the same $p = 0.9$ Bayesian quantile regression curves and

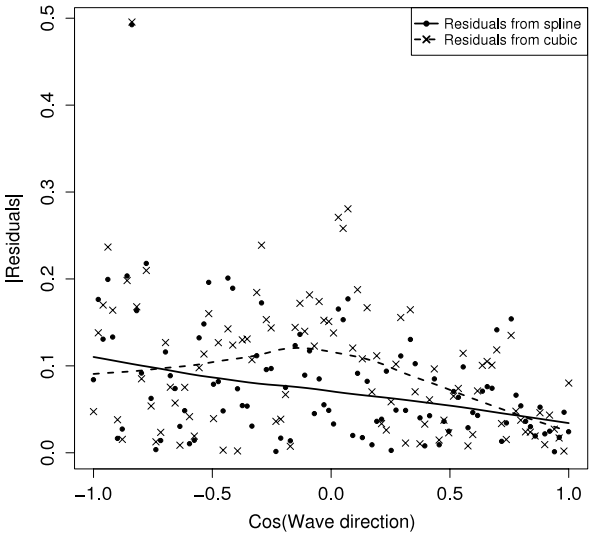


**Fig. 5.** The absolute values of the residuals against the cosine of wave direction with associated `loess` smoother from both the spline (dots, unbroken line) and the cubic (crosses, dashed line) quantile regressions. A grid of 100 sections along the covariate was used in the calculation of the residuals.

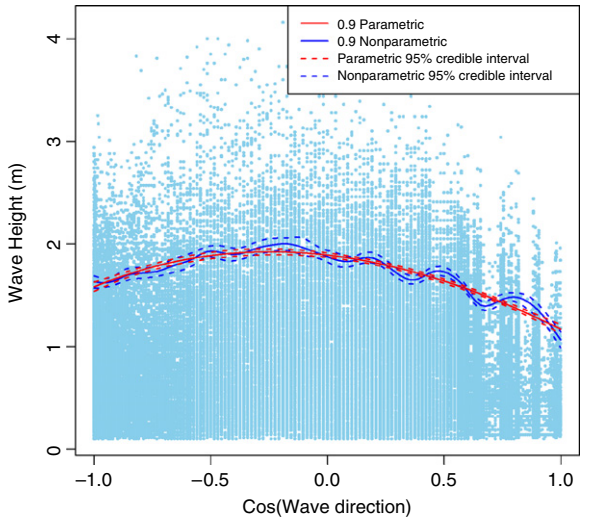


**Fig. 6.** Scatter plot of the offshore wave data showing the $p = 0.9$ Bayesian nonparametric quantile regression curve using splines and $p = 0.9$ parametric BQR curve. 95% credible envelopes are also presented.

associated credible intervals as before. Our nonparametric quantile regression curve using splines provides us with a better understanding of the fine features of the $p = 0.9$ quantile than the cubic quantile regression curve. This advantage can be particularly helpful with the data sets of this size and visual complexity.

Fig. 7 shows the absolute value of the residuals from both the cubic polynomial quantile regression curve and spline based curve against the cosine of wave direction. We can clearly see that our spline based approach again provides a better quality of fit through the full covariate range than the cubic polynomial quantile curve. Again this is as a result of the more local nature of the spline based fitting procedure. This is apparent in this example as we have a large amount of data to work with a meaning that the local variation can be better identified than in smaller data sets.

### *3.3. Application to a simulated data set bases on the motorcycle accident data*

In our third example, we apply our Bayesian nonparametric quantile regression spline based methodology to a simulated data set generated from the famous motorcycle accident data, discussed by Silverman (1985) and presented in Fig. 8. The data set comprises the head acceleration in multiples of the acceleration due to gravity $g$ at 133 times in milliseconds after a simulated motorcycle accident used to test crash helmets. This well-known data set has been used frequently to motivate and demonstrate spline based methodology, since the nature of the underlying process makes polynomial modelling inappropriate. It provides a suitable test for our methodology.

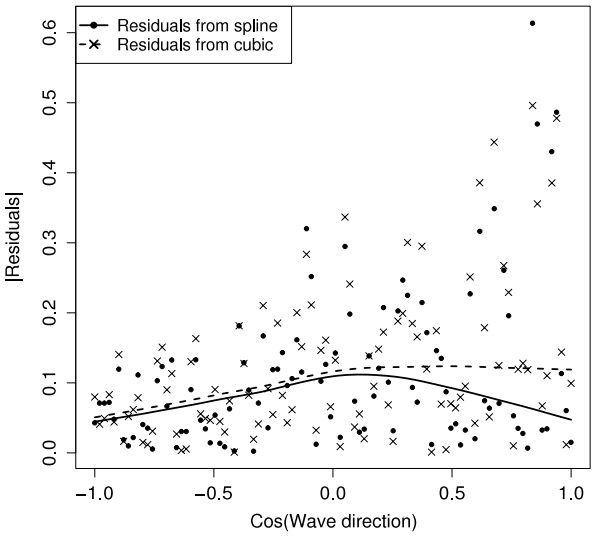


**Fig. 7.** The absolute values of the residuals against the cosine of wave direction with associated loess smoother from both the spline (dots, unbroken line) and the cubic (crosses, dashed line) quantile regressions. A grid of 100 sections along the covariate was used in the calculation of the residuals.

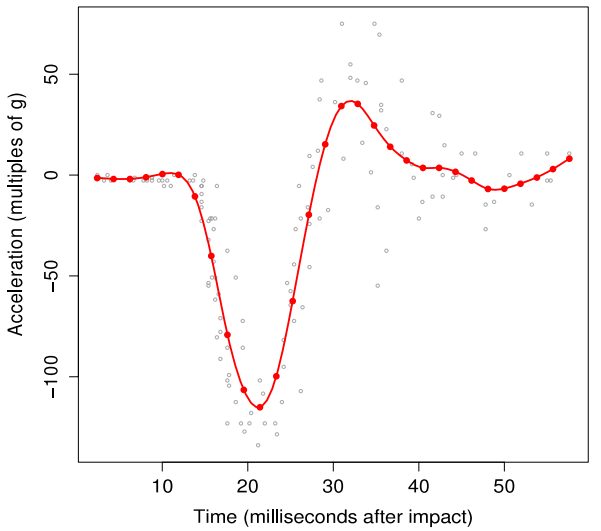


**Fig. 8.** Scatter plot of the motorcycle accident data from Silverman (1985). A smoothing spline has been added using the R package `splines` (R Development Core Team, 2009). The values of the spline at 30 equally spaced times are shown using filled dots.

Fig. 8 also shows a smoothing spline found using the R's `spline` function (R Development Core Team, 2009); see Green and Silverman (1994) for a detailed discussion about the definition and calculation of smoothing splines. We simulated 100 values of acceleration at each of 30 equally spaced time points from a normal distribution with mean equal to the value of the smoothing spline at the time point, as shown in Fig. 8 using filled dots, and standard deviation set to 20. We present the simulated data together with the smoothing spline in Fig. 9. It is straightforward to calculate the true $p = 0.95$ quantile using mean + 1.96 standard deviation; this is also shown in Fig. 9, together with the empirical 0.95 quantile at each time point. We now apply our Bayesian nonparametric quantile regression spline based methodology to this simulated data set. We again used $N = 30$ knots, one at each time point at which the data are generated. The resulting curve is shown in Fig. 10. It recovers the true quantile function well, so confirming the effectiveness of our methodology.

## 4. MCMC performance

### 4.1. Choosing the proposal density and acceptance rate

In step (ii) of the Metropolis–Hastings algorithm presented in Section 2.2 the candidate vector $\boldsymbol{g}^\star$ was drawn from a multivariate normal distribution with variance–covariance matrix $\Sigma = \sigma^2 K^- / \lambda$. In this way, a candidate $\boldsymbol{g}^\star$ has similar

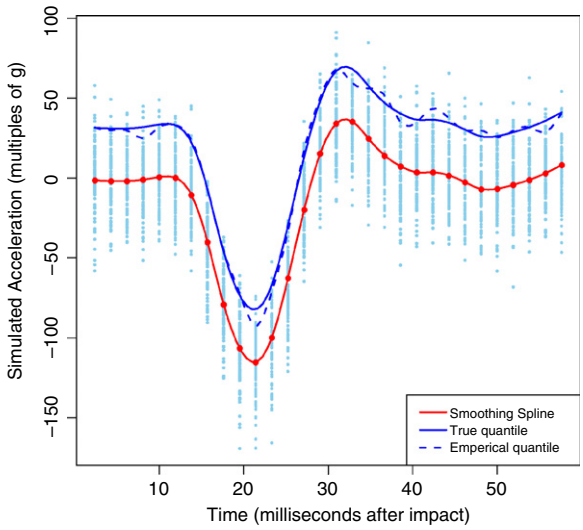


**Fig. 9.** Scatter plot of simulated data based on the smoothing spline shown in Fig. 8. The true $p = 0.95$ quantile function is shown together with the empirical $p = 0.95$ quantile at each time point.

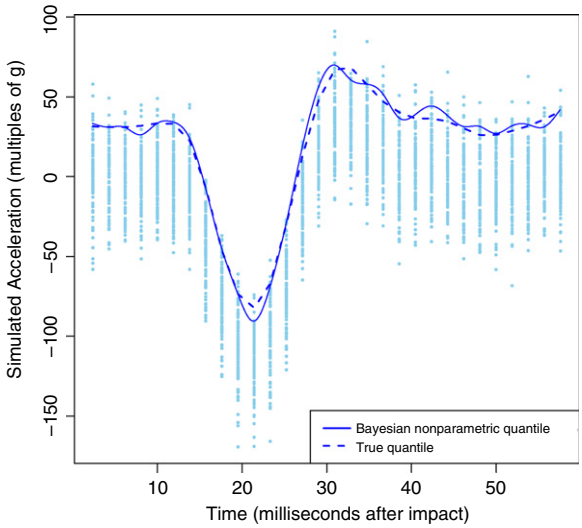


**Fig. 10.** Scatter plot of simulated data based on the smoothing spline shown in Fig. 8. The true $p = 0.95$ quantile function is shown together with the $p = 0.95$ Bayesian nonparametric quantile regression curve using splines.

structure to a $\boldsymbol{g}$ from the prior term $\pi(\boldsymbol{g}|\lambda)$. We also considered generating $\boldsymbol{g}^\star$ from a multivariate normal distribution with $\Sigma = \sigma^2 I_N$ where $I_N$ is the $N \times N$ identity matrix. As a third possibility, we updated a random subset of $g_1, \ldots, g_N$ again using independent normal distributions with variance $\sigma^2$. All three possibilities of generating $\boldsymbol{g}^\star$ performed similarly, with the choice of $\sigma^2$ having the greatest effect on the convergence of the Metropolis–Hastings algorithm.

Bédard (2008) introduced a technique that can be applied here to optimally choose the parameter $\sigma^2$ that controls the variance $\Sigma = \sigma^2 K^-/\lambda$ of the proposal density $q$ for $\boldsymbol{g}$ in the Metropolis–Hastings algorithms. The technique plots an efficiency criterion against acceptance rates from the Metropolis–Hastings algorithm or against $\sigma^2$. The acceptance rate or value of $\sigma^2$ that corresponds to the maximum efficiency can then be chosen.

The key to this procedure is the use of the first order efficiency criterion, which measures the average squared jumping distance for each parameter from one iteration to the next. In the case of the polynomial model of Yu and Moyeed (2001) in which the parameters $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are updated individually, Bédard (2008) would define the first order efficiency criterion (FOE) for the $i$th parameter as

$$\mathrm{FOE}_i = E\left[\left(\beta_i^{(j+1)} - \beta_i^{(j)}\right)^2\right], \tag{14}$$

where the expectation is over iterations $j$.

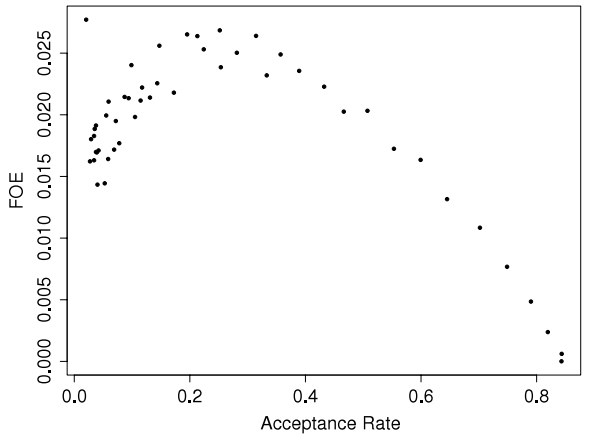


**Fig. 11.** FOE against acceptance rate when updating $\boldsymbol{g}$ in the Metropolis–Hastings algorithm.

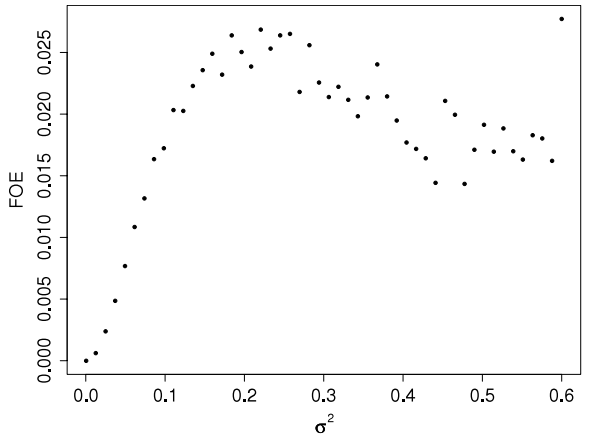


**Fig. 12.** FOE against $\sigma^2$ for updating $\boldsymbol{g}$ in the Metropolis–Hastings algorithm.

The definition can be easily extended to the case of the spline, in which all the parameters $\boldsymbol{g} = (g_1, \ldots, g_N)^{\mathrm{T}}$ are updated simultaneously, by using squared Euclidean distance as follows:

$$\mathrm{FOE} = E\left[\sum_{i=1}^{N}\left(g_i^{(j+1)} - g_i^{(j)}\right)^2\right], \tag{15}$$

where again the expectation is over iterations $j$.

Figs. 11 and 12 show plots of FOE against acceptance rate and against $\sigma^2$ for updating $\boldsymbol{g}$. These plots allow the user to choose the acceptance rate or $\sigma^2$ corresponding to the highest value of FOE. From Fig. 11 it can be seen that an acceptance rate of about 0.24 is most appropriate. This may seem rather low, but is due to the fact that we are updating a whole vector of parameters $\boldsymbol{g}$ and not just an individual parameter. It is also in agreement with some of the literature about optimal acceptance rates; see Bédard (2008), Bédard (submitted for publication) and references therein for example. A relatively low acceptance rate corresponds to a relatively high proposal variance, which itself allows larger possible jumps for the vector of parameters $\boldsymbol{g}$. A similar approach can be used to choose the value of $\sigma_\lambda^2$ for updating the smoothing parameter $\lambda$ in step (v) of the Metropolis–Hastings algorithm. In our application, we fixed a value for $\sigma_\lambda^2$ and tuned $\sigma^2$. We then fixed our chosen $\sigma^2$ and tuned $\sigma_\lambda^2$. Finally, we fixed our chosen $\sigma_\lambda^2$ and re-tuned $\sigma^2$. We found that we were able to achieve good convergence for both $\boldsymbol{g}$ and $\lambda$ with these tuned values of $\sigma^2$ and $\sigma_\lambda^2$, as we will discuss in Section 4.2. We also found that this approach yielded a value of $\sigma_\lambda^2$ that was relatively insensitive to the value of $\sigma^2$.

### *4.2. Assessing MCMC convergence*

Visual assessment of the convergence of the Metropolis–Hastings algorithm was found to be difficult as the simulated elements included $N = 30$ points along the spline rather than just a few model parameters. We found that the combination

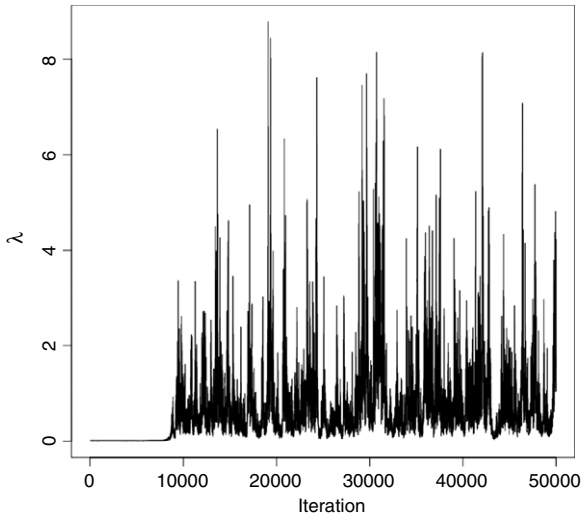


**Fig. 13.** Plot of $\lambda^{(j)}$ against thinned iteration $j$.

of a large number of sub-chains and an acceptance step based on a vector of points rather than a individual parameter could cause some convergence issues, although these could be overcome with good choices of $\sigma^2$ and $\sigma_\lambda^2$ as discussed in Section 4.1. Convergence for our nonparametric approach is generally slower than for parametric models. However, this computational cost is balanced by the improved localized fitting of the model that we have seen. The visual assessment of convergence of $\lambda$ was also difficult as the parameter took a wide range of values as shown in Fig. 13, where we can see that the time series converges around a lower value, with a tendency to jump to higher values (corresponding to smoother curves). We see that the time series has moved away from the low initial value of $\lambda^{(0)} = 10^{-6}$. In fact, values of $\lambda$ as low as $10^{-6}$ produce curves (not shown) that are visually far too rough.

After initially examining the time series plots of $\lambda^{(j)}$ and of the individual $g_i^{(j)}$ sub-chains as shown in Fig. 14, we used the more formal Gelman–Rubin statistic, discussed in Gelman and Rubin (1992), Gelman (1996) and Brooks and Gelman (1998), to assess convergence of $\boldsymbol{g}$ and of $\lambda$. The Gelman–Rubin procedure compares the variances between and within chains to monitor convergence and is based on the 'estimated potential scale reduction factor' $\hat{R}^{1/2}$, which represents the estimated factor by which a credible interval for a parameter of interest may shrink if further simulation is carried out. Good performance is indicated by values of $\hat{R}^{1/2}$ close to 1. The value of $\hat{R}^{1/2}$ should certainly not exceed 1.2 as suggested in Kass et al. (1998). We calculated $\hat{R}^{1/2}$ for each sub-chain $g_i, i = 1, \ldots, N$, and for $\lambda$ and found that $\hat{R}^{1/2}$ took values between 1.0006 and 1.0152. Thinning was applied by taking every tenth value as particular sub-chains showed strong autocorrelations. Thinning also reduced storage requirements.

Our examination of time series plots together with satisfactory values of the Gelman–Rubin statistic gave us confidence that the Metropolis–Hastings algorithm was producing realizations from the posterior distribution $\pi(\boldsymbol{g}, \lambda|\boldsymbol{y})$.

## 5. Conclusions

In this article, we have developed a methodology within the Bayesian framework to extend fixed degree polynomial based quantile regression to nonparametric quantile regression by using a spline based approach. We sampled from the posterior density of a NCS and an associated smoothing parameter by means of a specially tuned Metropolis–Hastings algorithm and used our sample to make inferences that include the quantification of uncertainty.

We have presented applications of our Bayesian nonparametric quantile regression methodology to two real environmental data sets, providing favourable comparisons with an existing parametric method and illustrating that learning about model parameters from data has taken place. We have confirmed the effectiveness of our methodology using simulated data based on a well-known published data set for which polynomial modelling is inappropriate.

## Acknowledgement

The authors acknowledge the support of a doctoral scholarship from the University of Plymouth and funding from the EPSRC projects RF-PeBLE (grant No. EP/C005392/1), LEACOAST 2 (grant No: EP/C013085/1) and BVANG (grant No: EP/C002172/1). The authors would also like to thank Dr. Peter Hawkes from HR Wallingford for the coastal wave data set, and Dr. Anna Zacharoudaki from the University of Plymouth for the offshore wave data set, and supplementary information provided. We are very grateful to the editor and two reviewers for helpful suggestions that have substantially improved this article.

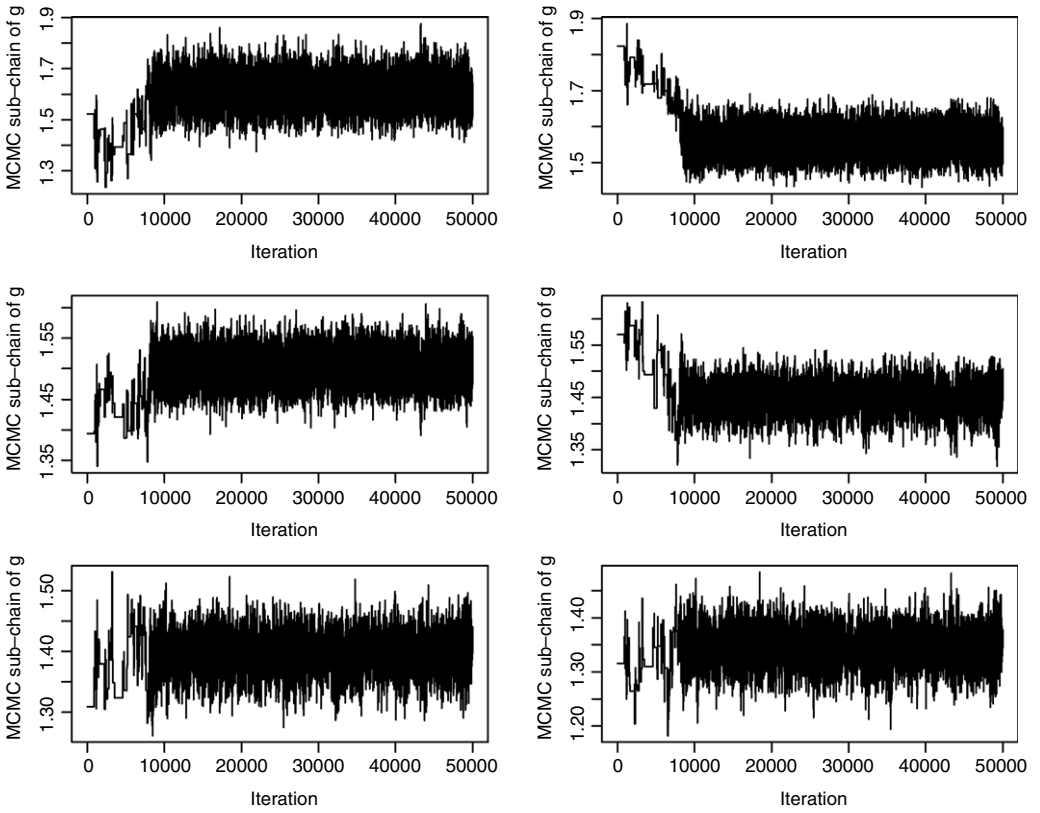


**Fig. 14.** Plots of $g_i^{(j)}$ against thinned iteration $j$ for six values of $i$.

## References


Bédard, M., 2006. Optimal acceptance rates for Metropolis–Hastings algorithms: Moving beyond 0.234. Annals of Statistics (submitted for publication).

Bédard, M., 2008. Efficient sampling using algorithms. Journal of Computational and Graphical Statistics 17 (2), 312–332.

Bosch, R., Ye, Y., Woodworth, G.G., 1995. A convenient algorithm for quantile regression with smoothing splines. Computational Statistics and Data Analysis 19, 613–630.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7 (4), 434–455.

Denison, D.G.T., Mallick, B.K., Smith, A.F.M., 1998. Automatic Bayesian curve fitting. Journal of the Royal Statistical Society: Series B 60 (2), 333–350.

Dias, R., Gamerman, D., 2002. A Bayesian approach to hybrid splines non-parametric regression. Journal of Statistical Computing and Simulation 72 (4), 285–297.

Doksum, K., Koo, J., 2000. On spline estimators and prediction intervals in nonparametric regression. Computational Statistics & Data Analysis 35 (1), 67–82.

Dunson, D.B., Taylor, J.A., 2005. Approximate Bayesian inference for quantiles. Nonparametric Statistics 17 (3), 385–400.

Gamerman, D., 1997. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall, London.

Gelman, A., 1996. Inference and monitoring convergence. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 131–143.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7, 457–511.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Green, P.J, Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models. Chapman and Hall, London.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.

Kass, R.E., Calin, B.P., Gelman, A., Neal, R.M., 1998. MCMC in practice: A roundtable discussion. The American Statistician: Statistical Practice 52 (2), 93–100.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge.

Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81 (4), 673–680.

Koenker, R., 2008. quantreg: Quantile Regression (version 4.17). Vienna, Austria. http://www.r-project.org.

Kottas, A., Gelfand, A.E., 2001. Bayesian semiparametric median regression modeling. Journal of the American Statistical Association 96 (456), 1458–1468.

Kottas, A., Krnjajić, M., 2009. Bayesian semiparametric modelling in quantile regression. Scandinavian Journal of Statistics 36, 297–319.

Laurini, F., Pauli, F., 2009. Smoothing sample extremes: The mixed model approach. Computational Statistics & Data Analysis 53 (11), 3842–3854.

Lejeune, M.G., Sarda, P., 1988. Quantile regression: A nonparametric approach. Computational Statistics & Data Analysis 6 (3), 229–239.

Neal, R.M., 2003. Slice sampling. Annals of Statistics 31 (3), 705–767.

Ng, P., 1996. An algorithm for quantile smoothing splines. Computational Statistics & Data Analysis 22 (2), 99–118.

R Development Core Team,, 2009. R: A language and environment for statistical computing. Vienna, Austria. ISBN: 3-900051-07-0. http://www.R-project.org.

Rao, R.C., 2002. Linear Statistical Inference and its Applications, 2nd ed. John Wiley & Sons, New York.

Reeve, D., Chadwick, A., Fleming, C., 2004. Coastal Engineering: Processes, Theory and Design Practice. SPON, London.

Silverman, B.W., 1985. Some aspects of the spline smoothing approach to nonparametric curve fitting. Journal of the Royal Statistical Society: Series B 47, 1–52.

Thompson, P., Reeve, D., Cai, Y., Stander, J., Moyeed, R., 2008. Bayesian nonparametric quantile regression using splines for modelling wave heights. In: FloodRisk 2008 Conference. Keble College, University of Oxford, UK.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4th ed. Springer-Verlag, New York.
Wang, Y., Shao, Q., Zhu, M., 2009. Quantile regression without the curse of unsmoothness. Computational Statistics & Data Analysis 53 (10), 3696–3705.
Yu, K., Moyeed, R., 2001. Bayesian quantile regression. Statistics and Probability Letters 54, 437–447.
Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: Application and current research areas. Journal of the Royal Statistical Society: The Statistician 52 (3), 331–350.