ESTIMATION OF MONOTONE FUNCTION FOR DATA ON RECORDS

Jean-Francois Angers[1] and **Atanu Biswas**[2]

[1] *Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succ. "Centre-ville", Montréal, Québec H3C 3J7. Email: angers@dms.umontreal.ca*

[2] *Applied Statistical Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700108, India, Email: atanu@isical.ac.in*

Summary
In several regression problems monotonicity is a key feature of the underlying regression function, although in some cases the observations are not strictly monotonic due to random error. There are some other cases in which the observations are monotonic by nature. In all such cases the fitted curve should possess this monotonicity in order to explain the dataset. In the present paper, the dataset on the development of the world record on men's 100 m sports is considered for analysis. Using Bayesian methodology the fittings of the data is described by two methods, namely using monotone spline and the local regression technique of O'Hagan (1978). A Bayesian prediction for the future world record is also considered.

Key Words: Importance sampling; local regression; monotone decreasing function; monotone spline; Monte Carlo numerical integration.

1 Introduction

Statistical regression and prediction has long been a fundamental problem in statistics. Quite a large variety of regression techniques have been proposed and explored in statistical literature, both in the frequentist and Bayesian view point to explain and model different types of datasets which arise in different kinds of real life problems. Monotonicity, as we know, is an important feature of several mathematical functions. Several examples are available in literature where monotonicity is a key feature of the underlying regression curves. For example, Box and Cox (1964) considered breaking

strength of yarns and Ramsay (1988) considered a city gasoline consumption data to describe monotone spline regression. Recently, He and Shi (1998) considered two examples of which one is the degradation curves for the roof condition index of EPDM roofs, and the other being a classical example of theoretical power curve of any standard test where computations may be through simulations. In many such examples, although monotonicity is a desirable part of the fitted function, the observations may not be always monotonic due to random error. But there are situations where the observations, despite this randomness, are also monotonic. In the present paper we consider one such example and our focus is to analyze the data through the Bayesian approach.

The world champion in men's 100 m sport is called the fastest man on the earth, and his speed is the greatest speed on this planet. For 100 m run the record time is the minimum time in which a human being has ever run 100 m till date. A new record will be a time smaller than the current world record time. Thus if we consider the world record times over years corresponding to the time points they were recorded we get a decreasing sequence of observations.

An interesting job is of course to employ a suitable and appropriate regression technique for such a kind of data. The problem will be more difficult if, in addition to the time on which the regression has to be made, some other covariates come into play. For example, in track and field the wind speed might be such an important covariate. Running in favour of wind is an advantage to a runner, and quite obviously his time will be reduced. But any speed of wind is not permissible to account a performance to be a new world record. For example, a wind speed of $+2$ m/s is the upper permissible limit for such a run, and a performance with the help of a wind speed greater than that is not considered as a record value.

The plan of the present paper is as follows. In Section 2, we provide a brief review of the existing literature dealing with the record values. In the present paper we analyze the world record data on men's 100 m by using Bayesian regression techniques for monotone decreasing functions. Two such cases are considered. In Section 3, we discuss the monotone regression spline in a Bayesian setting, while in Section 4 the problem is addressed by local regression. Finally in Section 5 some numerical computations and comparisons are provided for the 100 m data.

2 A review of literature on records and the present data

Statistical analysis on record values have been an interesting problem over years. Most of the works available in the literature are in the frequentist view point, and they deal with distribution of the extreme values and relevant stochastic processes. The theory of record times started with

the works of Chandler (1952) and Foster and Stuart (1954). A detailed description of the relevant results are provided in the book by Galambos (1978). Thorough investigation of the records themselves started with Tata (1969) and Pickands (1971). Subsequently several results were obtained including results on extremal processes. In a recent discussion Arnold and Villasenor (1996) considered the problem of finding the tallest man of the world and provided some theoretical developments. They considered a branching process having an associated measured attribute, and studied the maximal value of the attribute in a given generation and in the entire process.

In several real life examples on records the observations may not be homogeneous, there may be so many covariates responsible for the observations. Extremal processes obtained as limits of independent but not identically distributed values are introduced and studied by Weisman (1975a, 1975b, 1975c). Regression model with extreme value distribution for the error terms is considered in Paula and Rojas (1997).

Table 1
World record data for men's 100m sport.

Name	Record time	Date	Wind speed
Robert Hayes	10.06	1964-10-15	+1.03
Jim Hines	10.03	1968-08-11	+0.88
James Sanford	10.02	1980-05-11	+1.0
Carl Lewis	10.00	1981-05-16	±0.0
Carl Lewis	9.97	1983-05-14	+1.48
Mel Lattany	9.96	1984-05-05	+0.06
Carl Lewis	9.93	1987-08-30	+0.95
Carl Lewis	9.92	1988-09-24	+1.1
Leroy Burrell	9.90	1991-06-14	+1.9
Leroy Burrell	9.88	1991-08-25	+1.2
Carl Lewis	9.86	1991-08-25	+1.2
Leroy Burrell	9.85	1994-07-06	+1.2
Donovan Bailey	9.84	1996-07-27	+0.7
Maurice Greene	9.79	1999-06-16	+0.1

We consider the dataset describing the development of 100 m world record. In the dataset, we have 14 observations from 1964 to 1999 which are the last 14 developments in men's 100m sport. Exact dates of the records are also recorded. In October 15, 1964 Robert Hayes ran in 10.06s with an wind speed of +1.03m/s, while in June 16, 1999 Maurice Greene set the new world record of 9.79s and he had +0.1m/s wind speed in his

favour. 12 other world records within this period has been recorded and the data comprises the wind speed as a covariate. The data is presented in Table 1. Monotone regression splines and local regression technique are used to analyze this data. These are illustrated in the subsequent sections.

3 Monotone spline

It is supposed that there exists a monotone decreasing function $g(x)$ that summarizes how the response variable y depends on x, that is

$$y_i = g(x_i) + \varepsilon_i, \tag{3.1}$$

for $i = 1, 2, \ldots, n$, where ε_i represents a Gaussian random noise. Suppose we observe $(x_i, y_i) \in \mathbf{R^2}$ for $i = 1, 2, \ldots, n$. Being records, our data are obviously not independent. However, since we propose a Bayesian model on $g(x)$, this will create marginal dependency between the observations. Hence, the Bayesian model can be seen as a way to model dependency into our setting.

There is several possible choices of function $g(x)$. In this paper, only two cases are considered, that is: monotone regression splines (*cf.* Ramsey, 1988) discussed in this section and local regression (*cf.* O'Hagan, 1978) provided in Section 4.

3.1 Classical setting

Note that usually the monotone splines are taken to be monotone increasing. However, since the monotone regression splines are bounded, transforming them into monotone decreasing function is easy.

To ease the presentation, let assume that the time domain is $[L, U]$ and let $L = t_1 \leq t_2 \leq \ldots \leq t_{m+k} = U$ be a sequence of knots where m is the number of free parameters and k is the order of the spline (*cf.* Ramsey, 1988). Furthermore, it is also assumed that

$$g(L) = 1 \text{ and } g(U) = 0. \tag{3.2}$$

Let $\underset{\sim}{t} = (t_1, t_2, \ldots, t_{m+k})$ and $\{J_j(x, k, \underset{\sim}{t})\}_{j=1}^{m+k}$ be the sequence of monotone splines. Then

$$g(x_i) = \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j. \tag{3.3}$$

Since the $J_j(x, k, \underset{\sim}{t})$'s are monotone decreasing functions between 0 and 1, if we impose the constraints $\theta_j \geq 0$ for $j = 1, 2, \ldots, m + k$, the $g(x)$ function will also be a monotone decreasing function. Furthermore if we impose that $\sum_{j=1}^{m+k} \theta_j = 1$, then equation (3.2) will also be satisfied. It

should be noted that if we assume additivity of the covariates effects, this model can be easily adapted to more than one covariate. For example, if two covariates are available, equation (3.3) is replaced by

$$g(x_{1,i}, x_{2,i}) = \sum_{j=1}^{m_1+k_1} J_j(x_{1,i}, k_1, \underset{\sim}{t}_1)\theta_j + \sum_{j=1}^{m_2+k_2} J_j(x_{2,i}, k_2, \underset{\sim}{t}_2)\theta_{m_1+k_1+j},$$

where $\underset{\sim}{t}_l$ are the knots sequence for the l^{th} covariate ($l = 1, 2$). The constraints on the θ_j's are the same as before, *i.e.* $\theta_j \geq 0$ for $j = 1, 2, \ldots, m_1 + m_2 + k_1 + k_2$ and $\sum_{j=1}^{m_1+m_2+k_1+k_2} \theta_j = 1$.

The selection of knots is usually crucial for fitting unconstrained splines. However, according to He and Shi (1998), the monotonicity constraints make the selection of knots easier because it eliminates sharp changes in the curve. He and Shi (1998) proposed the following method to select the knots $\underset{\sim}{t}$.

1. Start with $t_1 = \ldots = t_k = L$, $t_{k+j} = x_{[j/m]}$ for $j = 1, 2, \ldots, m$, and $t_{m+k+1} = \ldots = t_{m+2k} = U$ where $x_{[p]}$ denotes the $p \times 100\%$ percentile of $x_1, x_2, \ldots, x_n$.

2. Let $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_{m+k})$. Solve

$$\min_{\underset{\sim}{\theta}} \sum_{i=1}^{n} \left(y_i - \left[\sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j \right] \right)^2, \tag{3.4}$$

   under the constraints $\theta_j \geq 0$ for $j = 1, 2, \ldots, (m+k)$ and $\sum_{j=1}^{m+k} \theta_j = 1$. Let $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_{m+k}$ be the solution of equation (3.4).

3. Compute

$$\text{IC}(m) = \log \left(\sum_{i=1}^{n} \left(y_i - \left[\sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\widehat{\theta}_j \right] \right)^2 \right) + \frac{2(m + 2k + 2)}{n}.$$

4. Choose m to be the smallest minimizer of $\text{IC}(m)$.

Note that, in practice, m should not be bigger than n. In fact, asymptotic theory shows that m should be in the order of $n^{1/5}$ (*cf.* He and Shi, 1998). The $\text{IC}(m)$ function proposed in He and Shi (1998) is a variation of the information criterion used by Kooperberg and Stone (1992) and it acts as penalty for models with large number of coefficients.

Application of this method is discussed in Section 5. In this application, we have used the monotone regression spline of order 2. Using Ramsey (1988), they are given by

$$J_j(x,2,\underset{\sim}{t}) = \begin{cases} 1 & \text{if } x < t_j, \\ 1 - \frac{(x-t_j)^2}{\prod_{l=1}^{2}(t_{j+l}-t_j)} & \text{if } t_j \le x < t_{j+1}, \\ \frac{(x-t_{j+2})^2}{\prod_{l=1}^{2}(t_{j+2}-t_{j+l-1})} & \text{if } t_{j+1} \le x < t_{j+2}, \\ 0 & \text{if } x \ge t_{j+2}. \end{cases}$$

Note that due to the nature of the functions J_j's, some rescaling of the observations has to be done. A "theoretical" minimum for the y_i's needs also to be elicit (see Section 5). According to Ramsey (1988), the J_j splines can be expressed in terms of B-splines. Hence the equation (3.3) can also be written using integrated B-splines. The method given in this paper can also be used with any monotone spline.

3.2 Bayesian setting

Let us again consider the model given in equations (3.1) and (3.3) with the same constraints on the θ_j's, that is

$$\underset{\sim}{\theta} \in \mathcal{S}^{m+k} = \left\{ \underset{\sim}{\theta} \mid \theta_j \ge 0 \, \forall j = 1, 2, \ldots, m+k \text{ and } \sum_{j=1}^{m+k} \theta_j = 1 \right\}.$$

A natural prior for $\underset{\sim}{\theta}$ which will satisfy the above constraints is the Dirichlet density with parameters $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{m+k})$, denoted by $\underset{\sim}{\theta} \sim \mathrm{D}(\underset{\sim}{\alpha})$ and given by

$$\pi(\underset{\sim}{\theta}) = \begin{cases} \frac{\Gamma(\sum_{j=1}^{m+k} \alpha_j)}{\prod_{j=1}^{m+k} \Gamma(\alpha_j)} \prod_{j=1}^{m+k} \theta_j^{\alpha_j - 1} & \text{if } \underset{\sim}{\theta} \in \mathcal{S}^{m+k}, \\ 0 & \text{otherwise.} \end{cases}$$

In equation (3.1), the random error ε_i are supposed to be Gaussian, that is $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, independent for $i = 1, 2, \ldots, n$. Hence, the likelihood function is given by

$$\begin{aligned} f(\underset{\sim}{\varepsilon} \mid \sigma^2) &= f(\underset{\sim}{y} \mid \underset{\sim}{\theta}, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left(y_i - \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t}) \theta_j \right)^2 \right\} \end{aligned}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j \right)^2 \right\}.$$

Since $\sum_{j=1}^{m+k} \theta_j = 1$, the exponent of the previous equation can be written as:

$$\begin{aligned}
&\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j \right)^2 \\
&\quad = \sum_{i=1}^{n} \left(\sum_{j=1}^{m+k} \theta_j y_i - \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j \right)^2 \\
&\quad = \sum_{i=1}^{n} \left(\sum_{j=1}^{m+k} \theta_j z_{ij} \right)^2,
\end{aligned}$$

where $z_{ij} = y_i - J_j(x_i, k, \underset{\sim}{t})$. Hence

$$\begin{aligned}
&\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m+k} J_j(x_i, k, \underset{\sim}{t})\theta_j \right)^2 \\
&\quad = \sum_{i=1}^{n} \left[\sum_{j=1}^{m+k} \theta_j^2 z_{ij}^2 + 2 \sum_{l>j} \theta_j \theta_l z_{ij} z_{il} \right] \\
&\quad = \sum_{i=1}^{n} \underset{\sim}{\theta}^t \mathbf{A}_i \underset{\sim}{\theta},
\end{aligned}$$

where $\mathbf{A}_i$ is a $(m+k) \times (m+k)$ matrix whose (j,l) component is given by

$$\begin{aligned}
(\mathbf{A}_i)_{jl} &= z_{ij} z_{il} \\
&= \left(y_i - J_j(x_i, k, \underset{\sim}{t})\right) \left(y_i - J_l(x_i, k, \underset{\sim}{t})\right).
\end{aligned}$$

Consequently, the likelihood function is given by

$$\begin{aligned}
f(\underset{\sim}{y} \mid \underset{\sim}{\theta}, \sigma^2) &= f(\underset{\sim}{z} \mid \underset{\sim}{\theta}, \sigma^2) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{n}{2\sigma^2} \underset{\sim}{\theta}^t \overline{\mathbf{A}} \underset{\sim}{\theta} \right\},
\end{aligned}$$

where $\overline{\mathbf{A}} = n^{-1}\sum_{i=1}^{n}\mathbf{A}_i$. In this context, σ^2 is considered to be a nuisance parameter and it should be integrated out. As prior on σ^2, an inverse gamma with parameter γ and β, denoted by $\sigma^2 \sim \Pi\Gamma(\gamma,\beta)$ is chosen. If there is no prior information available on σ^2, we choose $\beta = 0$ and $\gamma \in \{0, 1/2, 1\}$. (These choices correspond respectively to the constant prior (*c.f.* Berger, 1985, Section 4.9), the Jeffreys' prior and the left invariant Haar's prior.) The marginal posterior density of $\underset{\sim}{\theta}$ is given in the next theorem.

Theorem 1. *Let*

$$y_i = \sum_{j=1}^{m+k} J_j(x_i \mid k, \underset{\sim}{t})\theta_j + \epsilon_i,$$

where $\epsilon_i \sim N(0,\sigma^2)$, *independent for* $i = 1, 2, \ldots, n$. *If* $\underset{\sim}{\theta} \sim \mathrm{D}(\underset{\sim}{\alpha})$ *and* $\sigma^2 \sim \Pi\Gamma(\gamma,\beta)$, *then the posterior density of* $\underset{\sim}{\theta}$ *is*

$$\pi(\underset{\sim}{\theta} \mid \underset{\sim}{y}) \propto \frac{\prod_{j=1}^{m+k}\theta_j^{\alpha_j-1}}{\left(\frac{2}{n}\beta + \underset{\sim}{\theta}^t\overline{\mathbf{A}}\underset{\sim}{\theta}\right)^{\frac{n}{2}+\gamma}}.$$

The proof of this theorem can be obtained using straightforward calculation and so it is omitted.

Consequently if one uses the squared error loss, the posterior means and covariances are given by

$$\mathbb{E}[\theta_j \mid \underset{\sim}{y}] = \frac{I(\underset{\sim}{e_j})}{I(\underset{\sim}{0})},$$

$$\mathrm{Cov}(\theta_j,\theta_l \mid \underset{\sim}{y}) = \frac{I(\underset{\sim}{e_j}+\underset{\sim}{e_l})}{I(\underset{\sim}{0})} - \left(\frac{I(\underset{\sim}{e_j})}{I(\underset{\sim}{0})}\right) \times \left(\frac{I(\underset{\sim}{e_l})}{I(\underset{\sim}{0})}\right),$$

for $j, l = 1, 2, \ldots, m+k$ where $\underset{\sim}{e_j}$ denotes a vector of 0 with a 1 only at the j coordinate and

$$I(\underset{\sim}{a}) = \int\cdots\int_{\mathcal{S}^{m+k}} \frac{\prod_{j=1}^{m+k}\theta_j^{a_j+\alpha_j-1}}{\left(\frac{2}{n}\beta + \underset{\sim}{\theta}^t\overline{A}\underset{\sim}{\theta}\right)^{\frac{n}{2}+\gamma}} d\theta_1\ldots d\theta_{m+k}. \tag{3.5}$$

This integral cannot be evaluated analytically but it can be computed easily using Monte Carlo with importance sampling integration technique or Gibbs sampling. In Section 5, we use the Monte Carlo technique with the prior $\mathrm{D}(\underset{\sim}{\alpha})$ as importance sampling function. Note that to generate a random vector from a Dirichlet density, we can do the following :

i) $U_j \sim \Gamma(\alpha_j, 1)$ independent for $j = 1, 2, \ldots, m+k$,

ii) $\theta_j = \frac{U_j}{\sum_{l=1}^{m+k} U_l}$.

(*cf.* Lange, 1998).

4 Local regression

In this section, we propose an alternative to the spline introduced in O'Hagan (1978). Let us consider the standard linear model

$$y_i = \underset{\sim}{f}^t(\underset{\sim}{x}_i)\underset{\sim}{\beta}(\underset{\sim}{x}_i) + \varepsilon_i, \tag{4.1}$$

where $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ independent for $i = 1, 2, \ldots, n$, $\underset{\sim}{f}(\underset{\sim}{x})$ is a $q \times 1$ vector of known functions of independent variables $\underset{\sim}{x}$ and $\underset{\sim}{\beta}$ is a $q \times 1$ unknown parameters. Since it is supposed that the y_i are observations from a decreasing function, the functions f will be taken to be decreasing. In the first subsection no constraint is put on the β vector, and hence the approach of O'Hagan (1978) is applied directly. In the second subsection we will put nonnegativity constraints on $\underset{\sim}{\beta}$ forcing the local regression function to be decreasing.

4.1 Unconstrained local regression

In order to provide an adequate approximation to the regression function locally, it is assumed that $\underset{\sim}{\beta}(\underset{\sim}{x})$ and $\underset{\sim}{\beta}(\underset{\sim}{x}^*)$ are highly correlated when $\|\underset{\sim}{x} - \underset{\sim}{x}^*\|$ is "small", where $\|\cdot\|$ represents a norm of a vector. As in O'Hagan (1978), we assume that the prior information about $\underset{\sim}{\beta}(\underset{\sim}{x})$ is the same for all values of $\underset{\sim}{x}$. In particular, we assume that

$$\mathbb{E}[\underset{\sim}{\beta}(\underset{\sim}{x})] = \underset{\sim}{b}_0 \ \forall \underset{\sim}{x},$$

and

$$\mathrm{Cov}(\underset{\sim}{\beta}(\underset{\sim}{x}), \underset{\sim}{\beta}(\underset{\sim}{x}^*) \mid \underset{\sim}{b}_0) = \rho(\|\underset{\sim}{x} - \underset{\sim}{x}^*\|)\mathbf{B}_0,$$

where $\rho(d)$ is a monotonic decreasing function of $d \in [0, \infty)$ and $\rho(0) = 1$ and $\mathbf{B}_0$ is a known $q \times q$ symmetric positive definite matrix. If $\underset{\sim}{b}_0$ is not known, O'Hagan (1978) recommend to estimate it by

$$\widehat{\underset{\sim}{b}_0} = \left(\mathbf{G}^t\mathbf{A}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^t\mathbf{A}^{-1}\underset{\sim}{y}, \tag{4.2}$$

where

$$\begin{aligned}
\mathbf{G} &= (\underset{\sim}{f}(\underset{\sim}{x}_1), \underset{\sim}{f}(\underset{\sim}{x}_2), \ldots, \underset{\sim}{f}(\underset{\sim}{x}_n))^t, \\
\mathbf{A} &= \sigma^2\mathbf{I}_n + \mathbf{C}, \\
\mathbf{C} &= (c_{ij})_{i,j\in\{1,2,\ldots,n\}}, \\
c_{ij} &= \rho(\|\underset{\sim}{x}_i - \underset{\sim}{x}_j\|)\underset{\sim}{f}^t(\underset{\sim}{x}_i)\mathbf{B}_0\underset{\sim}{f}(\underset{\sim}{x}_j),
\end{aligned}$$

and $\mathbf{I}_n$ denotes the identity matrix of order n. Assuming the normality for $\underset{\sim}{\beta}(\underset{\sim}{x})$ and using equation (4.2), the posterior mean is

$$\widehat{\underset{\sim}{\beta}}(x) = \mathbf{S}^t(\underset{\sim}{x})\mathbf{A}^{-1}\underset{\sim}{y} + \mathbf{Q}^t(\underset{\sim}{x})\widehat{\underset{\sim}{b}_0}, \tag{4.3}$$

where

$$\mathbf{S}(\underset{\sim}{x}) = \begin{pmatrix} \rho(\|\underset{\sim}{x} - \underset{\sim}{x}_1\|)\underset{\sim}{f}^t(\underset{\sim}{x}_1)\mathbf{B}_0 \\ \rho(\|\underset{\sim}{x} - \underset{\sim}{x}_2\|)\underset{\sim}{f}^t(\underset{\sim}{x}_2)\mathbf{B}_0 \\ \vdots \\ \rho(\|\underset{\sim}{x} - \underset{\sim}{x}_n\|)\underset{\sim}{f}^t(\underset{\sim}{x}_n)\mathbf{B}_0 \end{pmatrix},$$

$$\mathbf{Q}(\underset{\sim}{x}) = \mathbf{I}_q - \mathbf{G}^t\mathbf{A}^{-1}\mathbf{S}(\underset{\sim}{x}),$$

and the posterior covariance matrix is given by

$$\mathrm{Cov}(\underset{\sim}{\beta}(\underset{\sim}{x}), \underset{\sim}{\beta}(\underset{\sim}{x}^*) \mid \underset{\sim}{y}) = \mathbf{B}_1(\underset{\sim}{x}, \underset{\sim}{x}^*) + \mathbf{Q}^t(\underset{\sim}{x}) \left(\mathbf{G}^t\mathbf{A}^{-1}\mathbf{G}\right)^{-1} \mathbf{Q}(\underset{\sim}{x}^*), \qquad (4.4)$$

where

$$\mathbf{B}_1(\underset{\sim}{x}, \underset{\sim}{x}^*) = \rho(\|\underset{\sim}{x} - \underset{\sim}{x}^*\|)\mathbf{B}_0 - \mathbf{S}^t(\underset{\sim}{x})\mathbf{A}^{-1}\mathbf{S}(\underset{\sim}{x}^*).$$

4.2 Constrained local regression

If we assume that the y_i's are observations from a decreasing but unknown function. The main difference between O'Hagan (1978) and this paper is that it is known that the y_i's are from a monotone function. In order to use the methodology discussed in the previous subsection, one has to put constraints on the $\beta(x)$'s coefficients. Hence, if the model given by equation (4.1) is valid, we can suppose that the $f(\underset{\sim}{x})$ function is a decreasing function and that the $\beta(\underset{\sim}{x})$'s coefficients are nonnegative. Consequently, the $\beta(x)$'s coefficients cannot be estimated directly using equation (4.3) since there is no guarantee that this equation will be nonnegative. The Bayes rule for $\underset{\sim}{\beta}(\underset{\sim}{x})$ under the squared error loss is therefore given by

$$\widehat{\underset{\sim}{\beta}}(\underset{\sim}{x}) = \int_0^\infty \cdots \int_0^\infty \underset{\sim}{\beta}\, \pi(\underset{\sim}{\beta} \mid \underset{\sim}{x}, \underset{\sim}{y}) d\beta_1 \ldots d\beta_q, \qquad (4.5)$$

where $\pi(\underset{\sim}{\beta} \mid \underset{\sim}{x}, \underset{\sim}{y})$ is a Gaussian process on $\mathbb{R}_+^q = [0, \infty)^q$ with mean function given by equation (4.3) and covariance kernel given by equation (4.4). Note that the integral given in equation (4.5) cannot be evaluated analytically but it can be computed easily using Monte Carlo numerical integral. Application of this procedure is discussed in Section 5.

5 Example and comparison

We first standardize the time (x_1) in a new scale which corresponds to 0.0792 and 3.5605 for the two time points when the first and last world record of our list were recorded. First we fit just an exponential model to

our data using ordinary least squares (OLS) omitting the first two observations. The fitted model is thus

$$\widehat{y}_{OLS} = \exp(2.32222 - 0.0114204x_1).$$

Figure 1 provides a plot of the fitted OLS curve and the observed points. In the figure yy stands for the y values. Note that this regression is not intended to model the data as much as to give a lower bound needed to standardise the observations. The first two observations were omitted in order to obtain a better fit by the least squared method. Using $x_1 = 14$ (after standardization) which corresponds somewhere in 2100 A.D., the limiting value for this model is 8.791 s. If we want to predict the world record around 2020 A.D., we have $x_1 = 5.6236$ in the standardized scale, and we obtain the predicted world record by the exponential model is 9.563903. Later we will compare these values with the corresponding values by O'Hagan method.

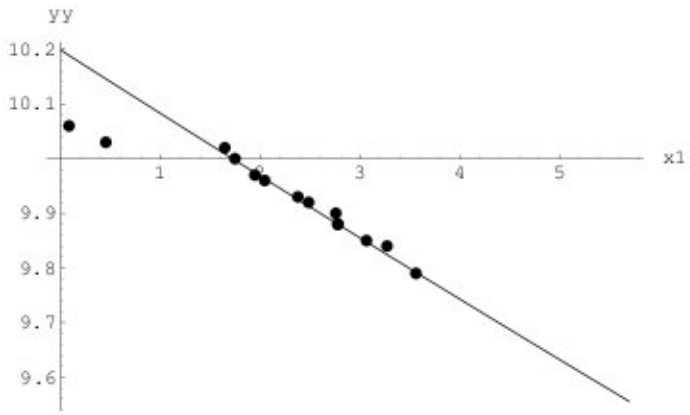

Figure 1
Fitted OLS curve and the observed points.

We use the second one (*i.e.* $x_1 = 5.6236$ and $y = 9.563903$) in the spline and then compare the spline and the O'Hagan methods. As discussed in Section 3, the values corresponding to x_1 (time), x_2 (wind speed) and y

are transformed within the range $[0,1]$. An S-Plus program was used to fit the monotone spline using 2500 iterations.

To choose the number of knots we apply the technique of He and She (1998). For the given data we have $IC(1) = -1.322203$, $IC(2) = 1.662231$, $IC(3) = -0.337568$, $IC(4) = -0.633286$, $IC(5) = -1.501385$, $IC(6) = -1.049112$ and $IC(7) = -1.250209$. Thus we choose $m = 5$, as it minimizes the IC-values.

Fitting the data in a spline with $m = 5$ we observe that the sum of squares of the classical and Bayesian residuals are respectively 0.046291 and 0.033341. With $m = 5$, the limiting value for the world record in year 2020 is equal to 9.575415 s. Figure 2 provides a graphical representation of the spline technique.

To compute $\widehat{\beta}$ (*cf.* equation (4.5)) Monte Carlo technique is used using truncated normal random variable density on $\mathbf{R}_q^+$ as importance sampling function. Let $SSE = \sum_{i=1}^n (y_i - f(\underline{x}_i)\widehat{\beta})^2$ and $SSE_0 = \sum_{i=1}^n (y_i - f(\underline{x}_i)\widehat{b}_0)^2$ where $\widehat{b}_0$ is given by equation (4.2). Using 1000 iterations, we obtain $SSE = 0.01727$ and $SSE_0 = 0.0090533$. Figure 3 provides a graphical representation for the O'Hagan technique. Note that, the constrained local regression provide a better fit than the constrained spline as shown by their SSE.

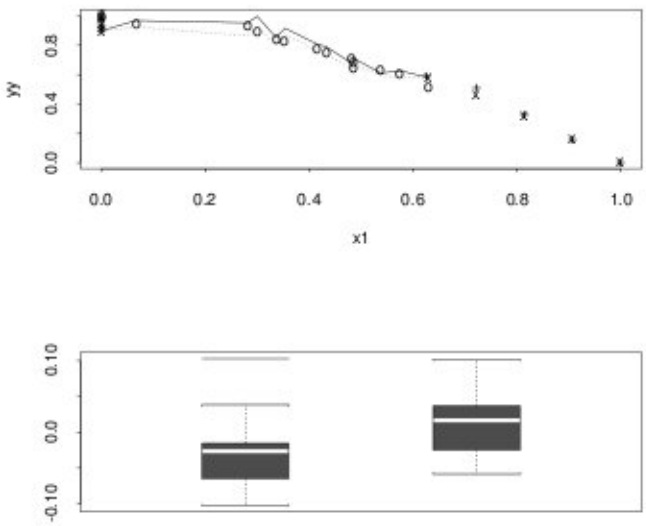

Figure 2: Spline. o=observed, +=classical spline, x=Bayesian splin

By the O'Hagan model, the predicted record at $x_1 = 14$ (somewhere in 2100 A.D.) is 8.788 s, and the predicted record at $x_1 = 5.6236$ (in 2020

A.D.) is 9.575415 s. Note that these limiting values are pretty close to the corresponding values obtained by OLS method and the spline technique. Hence both techniques predict approximately the same value for the world record in 2020. However, before to use the spline technique, we need to find a "theoretical" limiting value.

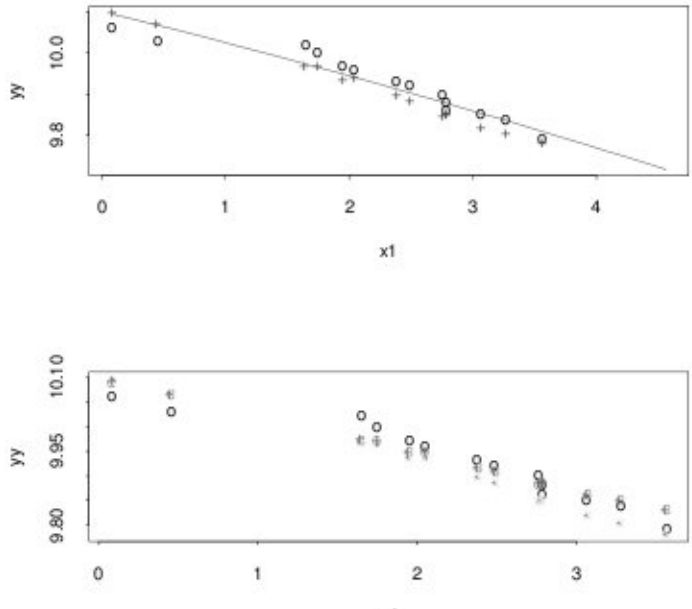

Figure 3: O'Hagan's method. o=obs; c=OLS, *=O'Hagan, x=Pos, +=B0

Acknowledgements

The authors wish to thank the referee for his valuable suggestions which led to an improvement over an earlier version of the manuscript.

(Received September, 2000. Revised August, 2001.)

References

Arnold, B. C. and Villasenor, J. A. (1996). The tallest man in the world. In *Statistical theory and applications.* H. N. Nagaraja, P. K. Sen and D. F. Morrison (eds.). New York: Springer-Verlag.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society* B, **26**, 211–252.

Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society* B, **14**, 220–228.

Foster, F. G. and Stuart, A. (1954). Distribution free tests in time-series based on the breaking of records. *Journal of the Royal Statistical Society* B, **16**, 1–13.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* New York: John Wiley and Sons.

He, X. and Shi, P. (1998). Monotone B-spline smoothing. *Journal of the American Statistical Association*, **93**, 643–650.

Kooperberg, C. and Stone, C.J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, 301–328.

Lange, K. (1998). *Numerical Analysis for Statisticians.* New York: Springer-Verlag.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society* B, **40**, 1–42.

Paula, G. A. and Rojas, O. V. (1997). On the restricted hypotheses and extreme value regression models. *Computational Statistics and Data Analysis*, **25**, 143–157.

Pickands, J. III (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, **8**, 745–756.

Ramsey, J. (1988). Monotone regression spline in action. *Statistical Science*, **3**, 425–461.

Tata, M. N. (1969). On outstanding values in a sequence of random variables. *Zeitschrift fur Wahrschein. verw. Geb.*, **12**, 9–20.

Weisman, I. (1975a). Extremal processes generated by independent nonidentically distributed random variables. *Ann. Probability*, **3**, 172–177.

Weisman, I. (1975b). On location and scale functions of a class of limiting processes with application to extreme value theory. *Ann. Probability*, **3**, 178–181.

Weisman, I. (1975c). Multivariate extreme processes generated by independent nonidentically distributed random variables. *J. Appl. Probability*, **12**, 477-487.