ELSEVIER

Journal of Hydrology 256 (2002) 106–119

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Local models for exploratory analysis of hydrological extremes

N.I. Ramesh[a,*], A.C. Davison[b]

[a]*School of Computing and Mathematical Sciences, The University of Greenwich, Maritime Greenwich Campus, Park Row, Greenwich, London SE10 9LS, UK*

[b]*Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland*

Received 20 September 2000; revised 10 July 2001; accepted 7 September 2001

Abstract

Trend analysis is widely used for detecting changes in hydrological data. Parametric methods for this employ pre-specified models and associated tests to assess significance, whereas non-parametric methods generally apply rank tests to the data. Neither approach is suitable for exploratory analysis, because parametric models impose a particular, perhaps unsuitable, form of trend, while testing may confirm that trend is present but does not describe its form. This paper describes semi-parametric approaches to trend analysis using local likelihood fitting of annual maximum and partial duration series and illustrates their application to the exploratory analysis of changes in extremes in sea level and river flow data. Bootstrap methods are used to quantify the variability of estimates.

Keywords: Annual maximum method; Bootstrap; Generalized extreme-value distribution; Generalized Pareto distribution; Local likelihood; Partial duration series

1. Introduction

Parametric models are widely used to describe the dependence of a hydrological variable on time or other covariates. However parametric modelling can constrain the range of possible fits in ways that are too inflexible for exploratory modelling. Consider Fig. 1, for example. The left panel shows the 10 largest sea levels in Venice for each year of the period 1887–1981, apart from years 1922 and 1935 in which just one and six observations were available, respectively. There is evidently an upward trend, particularly from about 1920, but it is not obvious that this trend is linear throughout the range of the data, nor that the variability of extremes is constant. The right panel shows the Nidd river flow exceedances of a threshold of 65 m^3/s from 1934 to 1969 at Hunsingore Weir. The data suggest a slight increase in later years, but this may be more apparent than real. In both cases it would be possible to fit models with polynomial trends, but it is widely recognized that the resulting fits may be poor, for several reasons. First, they are non-local, in the sense that the fit in one part of the observation space may depend heavily on data that are distant. This is clearly undesirable: why should an observation gathered 30 years ago be more influential for the fitted value today than one only five years old? Second, they tend to be sensitive to outliers. Third, there is no guarantee that a trend will have polynomial form, so polynomials of high degree may need to be fitted. Difficulties such as these have led to the widespread use of time- and space-localised methods such as wavelets in signal processing and other areas of the physical sciences (Percival and Walden, 2000). In this

* Corresponding author.

E-mail addresses: n.i.ramesh@greenwich.ac.uk (N.I. Ramesh), anthony.davison@epf1.ch (A.C. Davison).

0022-1694/02/$ - see front matter
PII: S0022-1694(01)00522-4

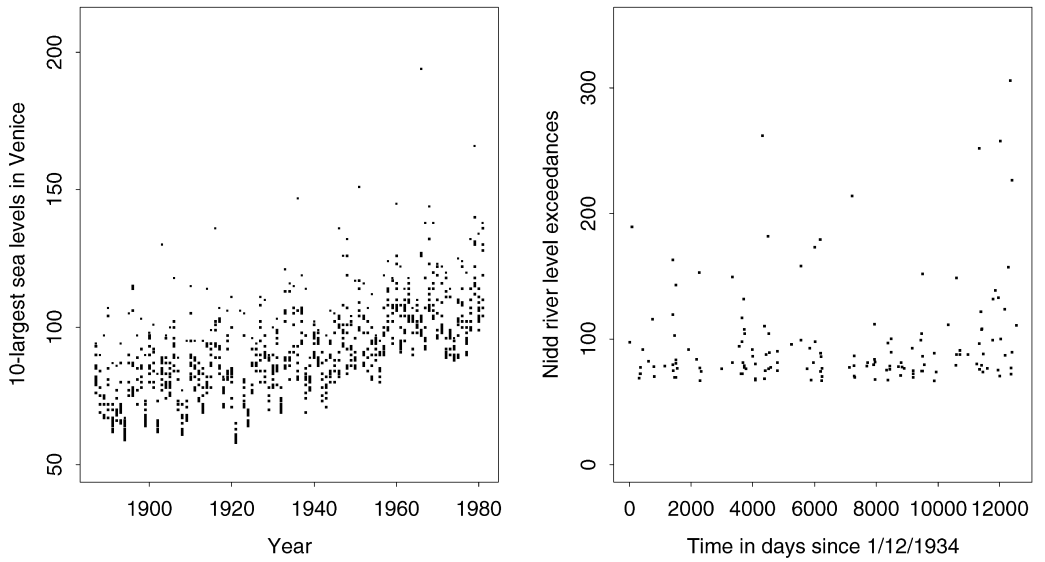

Fig. 1. Data sets used in this paper. Left: 10 largest annual sea levels (cm) in Venice from 1887 to 1981 (Tawn, 1988; Pirazzoli, 1982). Right: Nidd river flow exceedances (m^3/s) of a threshold 65 m^3/s (NERC, 1975, vol. 4, pp. 235–236.

paper we describe a different alternative to polynomial fitting for exploratory analysis, namely use of local likelihood, and illustrate its use by application to hydrological extremes.

Local likelihood is related to kernel density estimation, as applied to flood frequency data by Adamowski (1985) and extended to incorporate historical information by Adamowski and Feluch (1990), who showed that its inclusion can improve extrapolation beyond the record length; see also Bardsley (1989). As recently shown by Adamowski et al. (2000) and Adamowski (2000), nonparametric estimation of this sort may improve over less flexible parametric modelling using annual maximum or partial duration approaches when data are bimodal. However it does not allow extrapolation outside the data, unlike suitable parametric models.

Many authors have used parametric techniques to assess trend in extremes. Typical examples in the statistical literature involve global (i.e. non-local) likelihood estimation of overall trends, with distributional assumptions taken from the classical theory of statistics of extremes; see for example Coles and Tawn (1990) and Robinson and Tawn (1995). Bardsley et al. (1990) describe a general method for estimating extremal quantiles in this context, which can be used with a wide variety of fitting procedures, though simulation from the fitted model is a competing approach that is not restricted to independent extremes. Applications of parametric trend modelling in hydrology and related fields include Suppiah and Hennessy (1998), who supplement linear regression of extremes with non-parametric tests for trend in a study of extreme rainfall in Australia. Whether classical or robust regression is used for trend detection (Changon and Kunkel, 1995; Lettenmaier et al., 1994), a major difficulty with regional studies is the presence of spatial correlation across data series as well as temporal correlation within them. As pointed out by Douglas et al. (2000), this can dramatically reduce the effective sample size and if unaccounted for can lead to grossly overstated claims for significance of regional trends. Similar phenomena are well-known when dealing with single time series (Bloomfield, 1992; Smith, 1993), where ignorance of or misspecification of the form of dependence can be catastrophic for inference. In such cases it may be useful to adopt a form of block bootstrap resampling for time series, though this will not account for long-range dependence if it is present. For details and further references see Chapter 8 of Davison and Hinkley (1997). A quite different approach to expression of uncertainty in parametric trend modelling is through Bayes' theorem; see for example the method

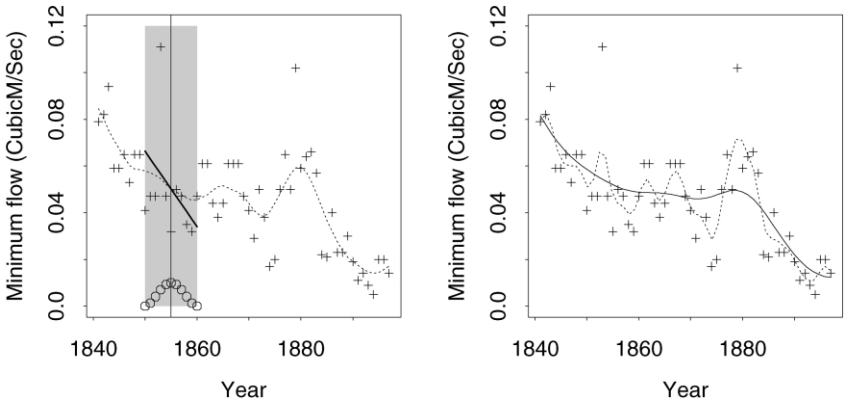

Fig. 2. Local likelihood fitting using Eq. (2) to annual minima of monthly mean flow data (+) from Wendover Springs. The left panel shows the local likelihood estimate (dotted line) and the fitted local straight line at $t_0 = 1855$ (dark line) using $h = 0.1$. The kernel at $t_0 = 1855$ is shown by circles at the foot. In the right panel, the bandwidth $h = 0.2$ for the solid line is chosen by eye, while the dotted line shows the result of using likelihood cross-validation, which gives $h = 0.04$.

for change-point analysis in hydrometrological time series proposed by Perreault et al. (2000).

Our semi parametric approach has elements of both parametric and nonparametric modelling. We fit standard models for trend in partial duration and annual maximum series and their extensions, but do so using local fitting, whereby the parameters of these models are estimated separately at each time by weighting the data appropriately. This yields parameter estimates that depend upon time, and gives local estimates of extreme quantiles. The purpose of the procedure is to provide an exploratory tool for assessing changes in parameters and which can be used both to suggest suitable parametric models and to test their fit. Two resampling methods are used to assess the variability of fits. Although we present the methodology with particular emphasis on extremes, potential extensions to other situations will be obvious to the reader; indeed, these and related approaches to modelling are now widespread in statistical analysis, with numerous books devoted to them (Bowman and Azzalini, 1997; Fan and Gijbels, 1996; Green and Silverman, 1994; Hastie and Tibshirani, 1990; Simonoff, 1996).

Sections 2 and 3 outline notions of local likelihood and of statistics of extremes, which are then merged in Section 4. Section 5 describes a point process approach to extremal modelling. As with kernel density estimation, the localness of the fit is controlled by a bandwidth, whose choice is discussed briefly in Section 2.2. Section 6 describes the application of the bootstrap in this context. Subsequent sections describe the application of these methods, and the paper is rounded off with a brief discussion.

2. Local likelihood

2.1. Basic idea

Suppose that we have n independent data pairs $(t_1, y_1), \ldots, (t_n, y_n)$, where we suppose that the t_j are fixed, and that y_j is a realization of a random variable Y_j, with density $f(y; \theta)$, where θ is a function that depends smoothly on t. That is, we suppose that $\theta = \theta(t)$, so that the density of Y in turn depends smoothly on t. In the most common setting for a hydrological application, t represents time. We wish to use the data pairs to make inferences about $\theta(t)$ at a particular value $t = t_0$ that lies within the range $(\min t_j, \max t_j)$ of the observed t_j. The local likelihood method assumes that $\theta(t_0)$ can be estimated by using the pairs (t_j, y_j) to fit a polynomial in t, but with pairs weighted so that those closer to t_0 assume more importance in the fitting. In most applications it is enough to take a linear polynomial for $\theta(t)$. That is, we suppose that $\theta(t_j) = \beta_0 + \beta_1(t_j - t_0)$, and use the estimated β_0 as the estimate of $\theta(t_0)$. Although we confine

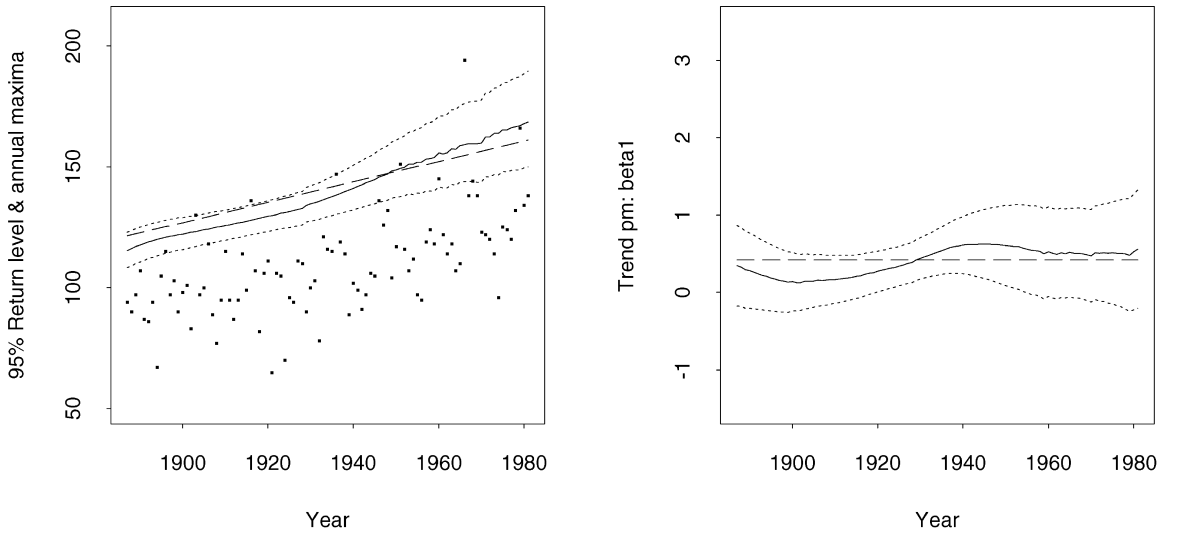

Fig. 3. Local linear GEV fit ($h = 0.4$) for annual maxima. Left: annual maximum sea levels and 20-year return levels for models M2 (solid) and M1 (dashes). Right: estimated trends for M2 (solid) and M1 (dashes). The 95% pointwise confidence bands (dots) from standard errors of M2 are shown in both panels.

ourselves to linear polynomials in this paper, the method extends easily to polynomials of higher degree; we discuss this below.

To fit this model locally, we choose a symmetric function $w(u)$ known as a kernel, with the properties that $w(0) = 1$, that $w(u)$ falls monotonically to zero as $|u| \to 1$ and that $w(u) = 0$ when u is outside the interval $[-1, 1]$. Examples are the biweight or Epanechnikov kernels

$$w(u) = (1 - u^2)^2, \quad w(u) = (1 - u^2)_+, \quad |u| \le 1,$$

where subscript + denotes the positive part function, but there are many other possibilities; see for example Section 3.3.2 of Silverman (1986). Let $d = \max t_j - \min t_j$ be the range of the t-values. We also choose a bandwidth $h > 0$, which will control the smoothness of the estimated curve. Given a value of h, we assign weight $w_j(t_0) = w\{(t_j - t_0)/(\mathrm{d}h)\}$ to the likelihood contribution from y_j when estimating the parameters at t_0. Hence only the observations at values t_j that lie inside the interval $[t_0 - \mathrm{d}h,\ t_0 + \mathrm{d}h]$ will contribute to estimation at t_0, and owing to the shape of $w(\cdot)$, values of y in the middle of the interval will contribute more than those at its endpoints. We discuss the choice of h below.

The maximum likelihood estimate for $\theta(t_0)$ is obtained by maximising the local log likelihood

$$\ell_h(\beta_0, \beta_1; t_0) = \sum_{t_j \in [t_0 - \mathrm{d}h, t_0 + \mathrm{d}h]} w_j(t_0) \log f\{y_j; \beta_0 + \beta_1(t_j - t_0)\}, \tag{1}$$

say, with respect to β_0 and β_1; the estimate is then taken to be $\hat{\beta}_0$. Maximization is performed for every value of t_0 of interest, giving a series of values of $\hat{\beta}_0$ which are then interpolated to give a curve that estimates how $\theta(t)$ depends on t. In particular, we may estimate $\theta(t)$ at $t_1, \ldots, t_n$, resulting in $\hat{\theta}(t) = \hat{\beta}_0(t)$ for the given values of t_j. Examples of such curves are shown in Figs. 2–6.

A special case of Eq. (1) occurs when $f(y; \theta)$ is normally distributed with constant variance σ^2. Then Eq. (1) becomes

$$\ell_h(\beta_0, \beta_1; t_0) = -\sum_{j=1}^{n} w_j(t_0) \left\{ \frac{1}{2\sigma^2} [y_j - \beta_0 - \beta_1(t_j - t_0)]^2 + \frac{1}{2} \log \sigma^2 \right\} \tag{2}$$

and differentiation shows that the local likelihood

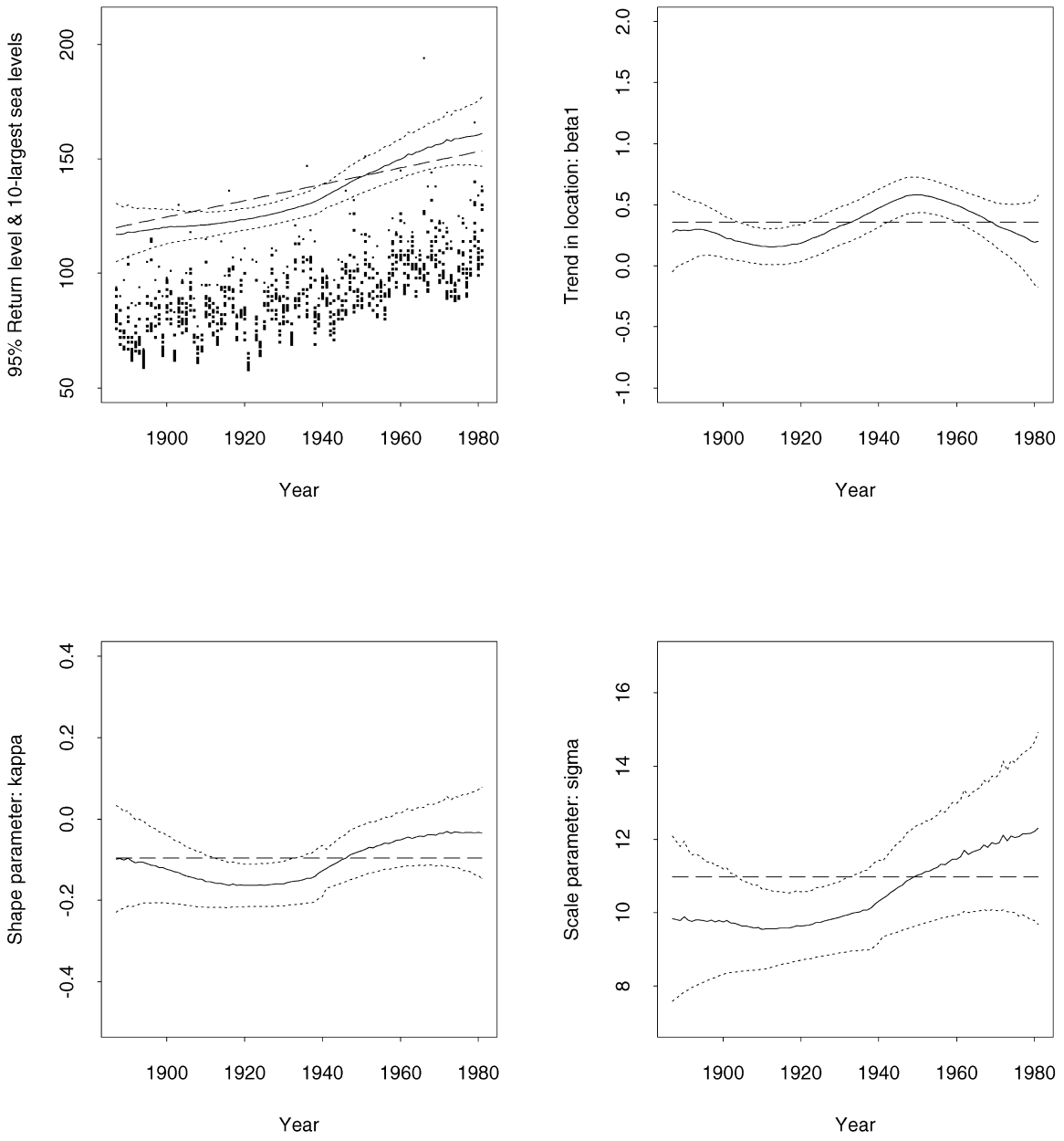

Fig. 4. Local linear 10-largest fit to Venice data ($h = 0.3$). The estimated curves from M4 (solid lines) and 95% confidence bands (dotted lines) are shown in all panels, together with estimates from the global model M3 (dashes). The upper left panel shows the 10 largest annual sea levels and the estimated 20-year return level. The upper right panel shows a smooth estimate of the trend $\hat{\beta}_1$, with smooth estimates of shape and scale parameters κ and σ shown in the lower panels.

estimate of $\theta(t_0)$ is the weighted least squares estimate of β_0, with weights that depend on the distance of t_j from t_0. Hence $\theta(t_0)$ is estimated unbiasedly if $\theta(t)$ is actually linear over the interval $[t_0 - \mathrm{d}h, t_0 + \mathrm{d}h]$; if not it has a bias that depends on the size of the first non-zero non-linear coefficient of a polynomial expansion of $\theta(t)$ at t_0, be this the quadratic, cubic or higher term. The simplest approach would be to

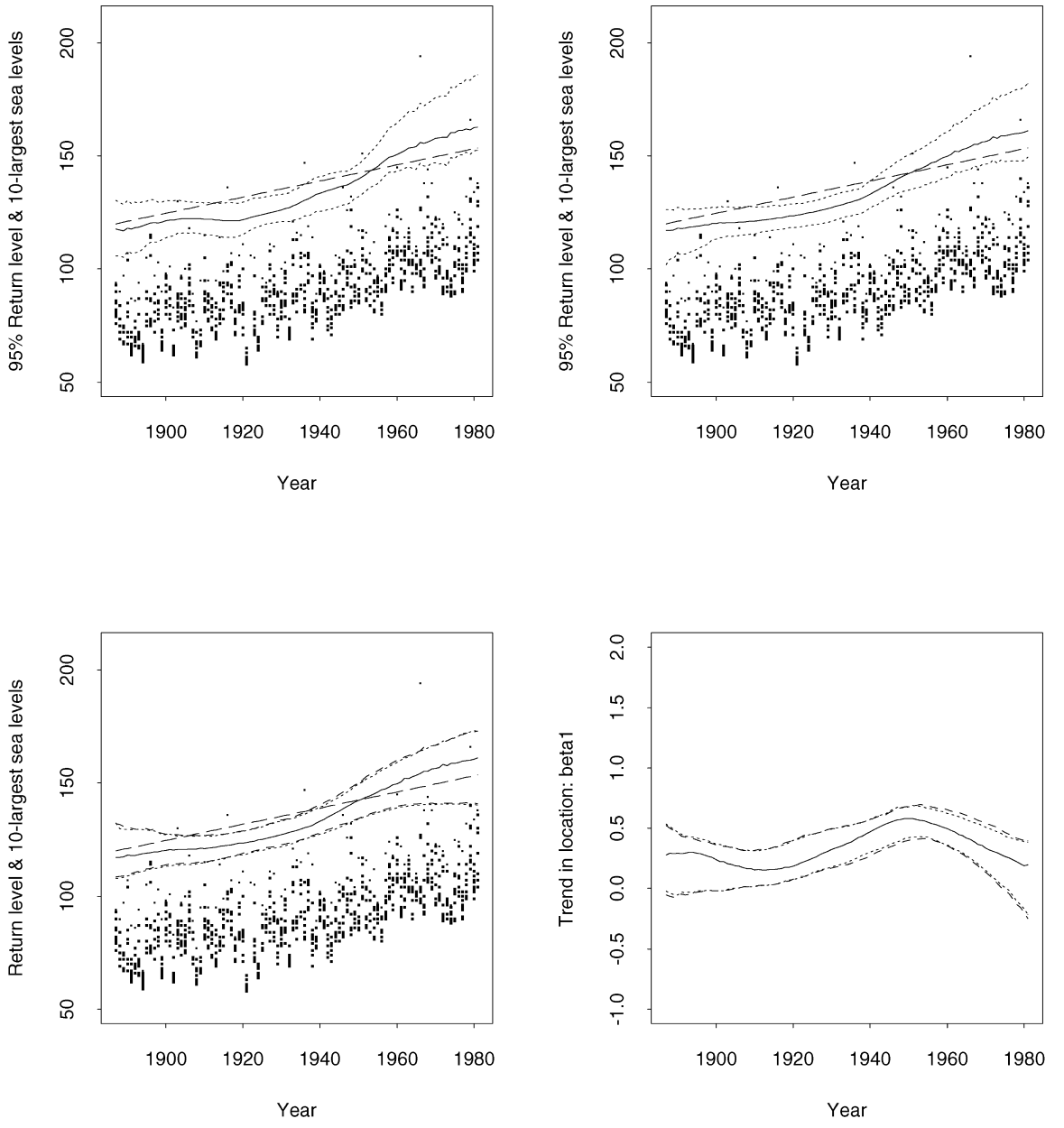

Fig. 5. Local linear 10-largest fit (M4: $h = 0.2$ upper left panel, $h = 0.3$ other panels): The top panels show the 10-largest sea levels and estimated 20-year return level curves for M3 (large dashes) and M4 (solid), together with 90% basic bootstrap bands for the latter using model-based resampling (dots). The bottom panels show both model-based (dots) and model-robust (small dashes) 90% bootstrap percentile bands for smooth estimates of ψ and β_1. In all panels $R = 499$ simulated series were used to construct confidence bands.

fit a local constant, but it turns out that taking a local linear fit reduces the bias when t_0 is close to the ends of the interval $(\max_j t_j, \min_j t_j)$. Similar conclusions apply more generally (Fan and Gijbels, 1996). In the context of extremes, one expects the function $\theta(t)$ to change rather slowly, and hence a local linear fit generally works well in practice.

The left panel of Fig. 2 shows the local likelihood

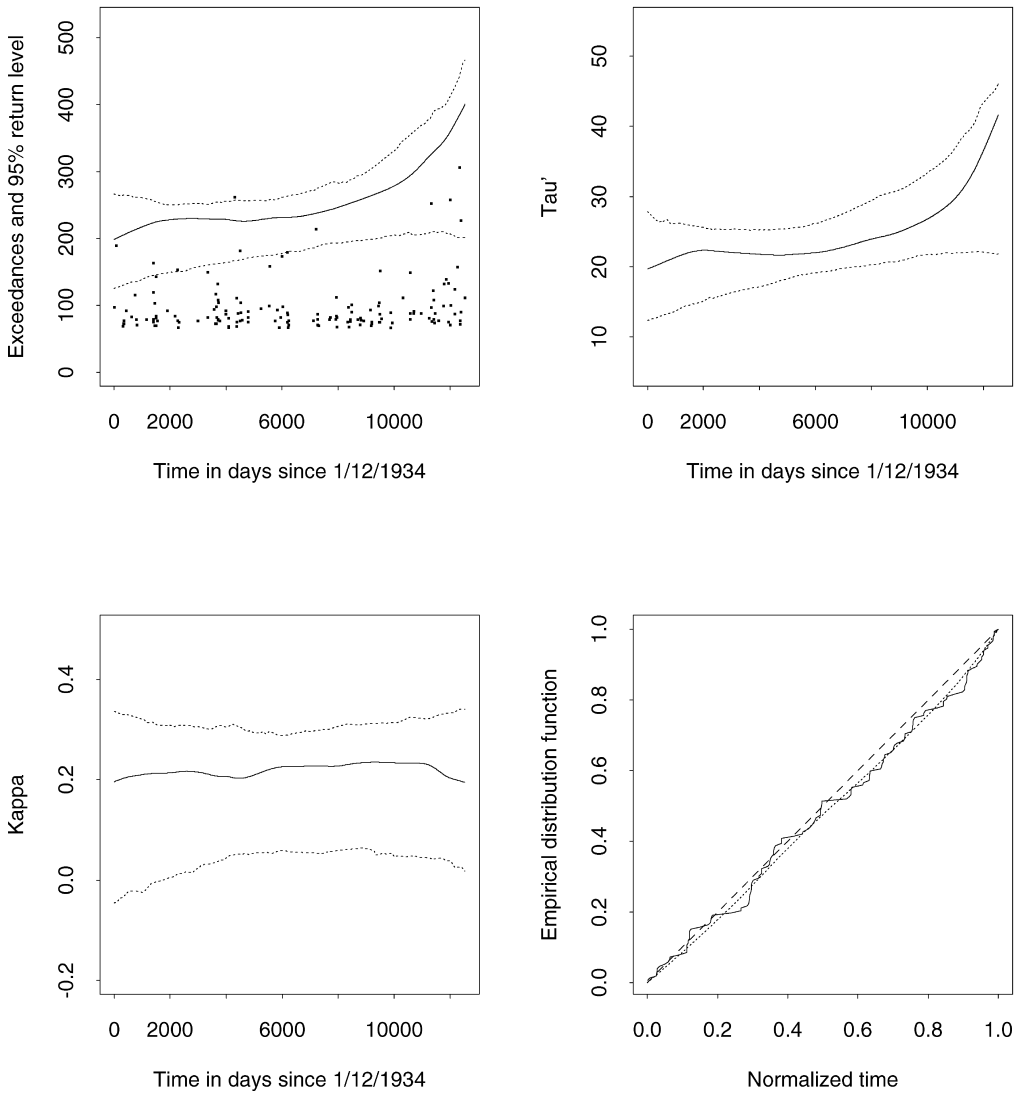

Fig. 6. Analysis of Nidd data. Top left: River level exceedances with estimated 20-year return level (solid) and 90% pointwise bootstrap confidence bands (dots), with $h = 0.6$. Values of $\hat{\tau}'(t)$ and $\hat{\kappa}'(t)$, with 90% pointwise confidence bands, are shown in the top right and bottom left panels. Lower right: Empirical distribution function for times of occurrence, with fitted generalized additive model (dots) and line $y = x$ for comparison (dashes).

estimate for annual minima of monthly mean flows at Wendover Springs, at $t_0 = 1855$, using $h = 0.1$ and the Epanechnikov kernel. Only the observations falling inside the shaded region contribute to the local fit, and the weight given to them is shown at the bottom scaled by 100 for clarity. Note how the very large value in 1853 is down weighted by the kernel, though it still strongly affects the locally fitted slope.

2.2. Choice of bandwidth

Previous studies on local fitting (Cleveland and Devlin, 1988; Tibshirani and Hastie, 1987; Hastie and Loader, 1993) have shown that the fitted parameters depend more on the bandwidth h than on the kernel. Smaller values of h result in overly variable curves, but larger h gives excessively smooth curves

and a biased fit, so there is the usual bias-variance trade-off. An example of this is given in the right panel of Fig. 2, where the dotted line shows a fit with $h = 0.04$ chosen by likelihood cross-validation as outlined below, while the solid line with $h = 0.2$ clearly gives a more plausible overall picture, being less sensitive to local fluctuations in the data.

A mathematical discussion of the issue is given in Fan and Gijbels (1996), though they do not discuss multiparameter problems of the type considered below, which are treated by Aerts and Claeskens (1997). In such cases a natural approach to selection of an overall bandwidth is to choose the $h > 0$ that maximizes the cross-validated log likelihood

$$\ell_{\mathrm{CV}}(h) = \sum_{j=1}^{n} \log f(y_j; \hat{\beta}_{-j}),$$

where $\hat{\beta}_{-j}$ is the local estimate of θ at t_j obtained when the jth observation is dropped from the dataset and bandwidth h is used. For brevity here we suppress the polynomials appearing in Eq. (1) and simply write the corresponding density as $f(y; \beta)$; note that this is not $f(y; \theta)$ with $\theta = \beta$. At first sight computation of $\ell_{\mathrm{CV}}(h)$ looks very computer-intensive, because $\hat{\beta}_{-j}$ must be computed for each j and over a grid of values of h. However Taylor series expansion shows that $\ell_{\mathrm{CV}}(h)$ may be approximated by

$$\sum_{j=1}^{n} \left\{ \log f(y_j; \hat{\beta}_j) + (\hat{\beta}_{-j} - \hat{\beta}_j)^{\mathrm{T}} \frac{\partial \log f(y_j; \hat{\beta}_j)}{\partial \beta} + \frac{1}{2} (\hat{\beta}_{-j} - \hat{\beta}_j)^{\mathrm{T}} \frac{\partial^2 \log f(y_j; \hat{\beta}_j)}{\partial \beta \, \partial \beta^{\mathrm{T}}} (\hat{\beta}_{-j} - \hat{\beta}_j) \right\},$$

where $\hat{\beta}_j$ is shorthand for $\hat{\beta}(t_j)$, the estimate of θ at $t = t_j$ based on the entire dataset, while $\hat{\beta}_{-j}$ is the solution to the weighted likelihood equation

$$\sum_{i \neq j} w_i(t_j) \frac{\partial \log f(y_j; \hat{\beta}_{-j})}{\partial \beta} = 0,$$

and Taylor series arguments and the fact that $w(0) = 1$ give

$$\hat{\beta}_{-j} - \hat{\beta}_j \doteq \left[\sum_{i \neq j} w_i(t_j) \frac{\partial^2 \log f(y_i; \hat{\beta}_j)}{\partial \beta \, \partial \beta^{\mathrm{T}}} \right]^{-1} \frac{\partial \log f(y_j; \hat{\beta}_j)}{\partial \beta}$$

Thus $\ell_{\mathrm{CV}}(h)$ can be approximated using elements of the original fit, though not without additional programming. In the single-parameter case simplications are possible, but they do not seem to be extend to multiparameter models. General experience with cross-validation suggests that the value $\hat{h}_{\mathrm{CV}}$ that maximizes $\ell_{\mathrm{CV}}(h)$ will be close to unbiased for the optimal bandwidth but will be rather variable. It is tempting to speculate that these properties will also hold here, but little seems to have been published for multiparameter problems. Asymptotic approaches to the local choice of optimal bandwidth are described by Aerts and Claeskens (1997), but in practice they choose to maximize $\ell_{\mathrm{CV}}(h)$.

The variance and hence the standard errors of the local maximum likelihood estimates are obtained either from the observed information matrix or using the sandwich variance matrix $V(t)$ evaluated at $\theta = \hat{\beta}(t)$ (Davison and Hinkley, 1997, Section 2.7), given by

$$V(t) = J^{-1}(t) \times \left\{ \sum_{j=1}^{n} w_j(t)^2 \frac{\partial \log f(y_j; \beta)}{\partial \beta} \frac{\partial \log f(y_j; \beta)}{\partial \beta^{\mathrm{T}}} \right\} J^{-1}(t), \tag{3}$$

where

$$J(t) = -\left\{ \sum_{j=1}^{n} w_j(t) \frac{\partial^2 \log f(y_j; \beta)}{\partial \beta \, \partial \beta^{\mathrm{T}}} \right\} \tag{4}$$

is the weighted observed information matrix; both $J(t)$ and $V(t)$ are evaluated at $\theta = \hat{\beta}(t)$. Sandwiches such as these, though with different fillings, give variance estimates for use in potentially misspecified situations or those involving estimating functions.

3. Extremal models

The local likelihood approach of Section 2 has been applied to many special classes of model, examples of which are given by Bowman and Azzalini (1997), Fan and Gijbels (1996) and Loader (1999). Here we describe its application to the class of extremal models, important in hydrology and in areas such as

insurance and risk assessment. Further statistical details are given by Davison and Ramesh (2000).

In statistical analysis of extreme value data it is common to model maxima using the generalized extreme value (GEV) distribution. Its probability density function is

$$f(y;\mu,\sigma,\kappa) = \frac{1}{\sigma}\left[1+\kappa\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1-1/\kappa} \times \exp\left\{-\left[1+\kappa\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1/\kappa}\right\} \quad (5)$$

where $\sigma > 0$, $-\infty < \mu$, $\kappa < \infty$ and the range of y is such that $1 + \kappa(y-\mu)/\sigma > 0$. The parameters μ, σ and κ determine the location, scale and shape of the distribution. If we have a random sample of annual maxima $y_1, \ldots, y_n$, the log likelihood is

$$\ell(\mu,\sigma,\kappa) = -n\log\sigma - \left(\frac{1}{\kappa}+1\right)\sum_{i=1}^{n}\log \times\left[1+\kappa\left(\frac{y_i-\mu}{\sigma}\right)\right]_+ - \sum_{i=1}^{n}\left[1+\kappa\left(\frac{y_i-\mu}{\sigma}\right)\right]_+^{-1/\kappa}, \quad (6)$$

from which maximum likelihood estimates and their standard errors may be obtained using standard routines.

In most applications we are interested in the quantiles of the fitted distribution, in order to make predictions about the level exceeded once every $1/(1-p)$ years on average, the return level

$$y_p = \mu + \frac{\sigma}{\kappa}[\{-\log(1-p)\}^{-\kappa} - 1]; \quad (7)$$

this is estimated by replacing parameters with their estimates. Its standard error is obtained in the usual way. In a non-stationary context the usual interpretation of y_p is no longer valid and it is better to think of it as a quantile of the distribution of $y$in the current year. For consistency we continue to use the term return level, however.

The analysis described above uses only one observation per year, but information is also contained in other large values, and more precise inference is possible when the r largest members of a independent sample are used (Tawn, 1988). This involves selecting independent flood peaks or other events and this may involve difficult judgements in practice. If such events can be identified then their distribution is obtained from the asymptotic joint density of the r largest order statistics, with r an integer normally in the range 1–10; the value $r = 5$ is often used. If $y^1 \geq ,\ldots y^r$ are the r-largest annual values, their joint probability density is

$$f(y^1,\ldots,y^r;\mu,\sigma,\kappa) = \frac{1}{\sigma^r}\prod_{j=1}^{r}\left[1+\kappa\left(\frac{y^j-\mu}{\sigma}\right)\right]_+^{-1-1/\kappa} \times\exp\left\{-\left[1+\kappa\left(\frac{y^r-\mu}{\sigma}\right)\right]_+^{-1/\kappa}\right\}, \quad (8)$$

where previous restrictions on the ranges of the parameters apply, but now $1 + \kappa(y^j - \mu)/\sigma$ must be positive for each $j = 1, \ldots, r$. The log likelihood for n independent years of annual r-largest values $y_i^1, \ldots, y_i^r$, $i = 1, \ldots, n$, is

$$\ell(\mu,\sigma,\kappa) = -nr\log\sigma - \left(\frac{1}{\kappa}+1\right)\sum_{i=1}^{n}\sum_{j=1}^{r}\log \times\left[1+\kappa\left(\frac{y_i^j-\mu}{\sigma}\right)\right]_+ - \sum_{i=1}^{n}\left[1+\kappa\left(\frac{y_i^r-\mu}{\sigma}\right)\right]_+^{-1/\kappa}. \quad (9)$$

The likelihood function (6) is the special case of Eq. (9) obtained when $r = 1$. Once estimates of the parameters have been obtained, y_p may be estimated as before.

4. Local likelihood for extremal models

We now illustrate the use of local likelihood in assessing trend in extremes. Consider a series of maxima $y_1, \ldots, y_n$ at times $t_j = j/n$, with $j = 1, \ldots, n$. Suppose that the location, scale and shape parameters of the density Eq. (3) are not constant, but are smooth functions $\mu(t)$, $\sigma(t)$ and $\kappa(t)$, so the probability density of maxima $Y(t)$ at time t, for $0 \leq t \leq 1$, becomes $f\{y;\mu(t),\sigma(t),\kappa(t)\}$; the $Y(t)$ are assumed independent but not identically distributed. In our application we take the functions $\mu(t)$, $\log\sigma(t)$ and $\kappa(t)$ to be

polynomials linear in t. Hence Eq. (3) has $\mu(t) = \beta_0 + \beta_1(j - nt)$, $\sigma(t) = \mathrm{e}^{\gamma_0+\gamma_1(j-nt)}$, and $\kappa(t) = \delta_0 + \delta_1(j - nt)$ in place of μ, σ and κ, respectively, for $j = 1, \ldots, n$ and t in [0,1]. We denote the new parameters by $\eta = (\beta_0, \gamma_0, \delta_0, \beta_1, \gamma_1, \delta_1)$.

The local log likelihood for η at time t is

$$\ell_h(\eta, t) = \sum_{j=1}^{n} w_j(t)\ell(\eta; y_j), \tag{10}$$

where $\ell(\eta; y_j)$ is the log likelihood contribution from the maximum of the jth year, y_j. It is enough to sum over values $j = [n(t - h)] + 1, \ldots, [n(t + h)]$ for which $w_j(t)$ is positive. If n is small, local parameter estimates can easily be obtained at $t = t_1, \ldots, t_n$ by maximizing Eq. (10) successively, while for larger series it may be sufficient to estimate parameters every few time steps and then interpolate the $\hat{\eta}(t)$. A detailed account of the methodology and its implementation is presented in Davison and Ramesh, (2000).

5. Point process model

The basic parametric forms of the models described in Section 3 can be derived from a point process model for extremes that results from a limiting characterization of exceedances of stationary data over a high threshold (Smith, 1989; Chavez-Demoulin and Davison, 2001). Suppose that we have an underlying stationary series $X_1, \ldots, X_m$, and consider the pattern in the plane with points at (x,y) coordinates $(j/(m + 1), a_m(X_j - b_m))$, for $j = 1, \ldots, m$. Then if a_m and b_m are chosen in such a way that $a_m(X_j - b_m)$ has a non-degenerate limiting distribution as $m \to \infty$, the pattern of events above a threshold at $y = y_0$ converges to an inhomogeneous Poisson process with intensity

$$\Lambda\{(t_1, t_2) \times (y_0, \infty)\} = (t_2 - t_1)\left(1 + \kappa\frac{y - \eta}{\tau}\right)_+^{-1/\kappa}, \tag{11}$$

where the interval (t_1, t_2) on the x-axis is a subset of [0,1], and $y > y_0$. Thus events in non-overlapping subsets of $[0, 1] \times (y_0, \infty)$ are independent, and if $y > y_0$, the probability of no events in the region $(t_1, t_2) \times (y, \infty)$ is

$$\exp\left\{-(t_2 - t_1)\left(1 + \kappa\frac{y - \eta}{\tau}\right)_+^{-1/\kappa}\right\}.$$

This enables a likelihood for η, τ and κ based on such data to be written down, following Smith (1989), who used parametric forms for η, τ and κ in analysis of extreme ozone data. If events are observed at $(t_1, y_1), \ldots, (t_n, y_n)$ in the region $[0, 1] \times (y_0, \infty)$, then the likelihood is

$$\exp[-\Lambda\{[0, 1] \times (y_0, \infty)\}] \prod_{j=1}^{n} \mathrm{d}\Lambda\{(t_j, y_j)\}$$

$$= \exp(-\Lambda_0)\Lambda_0^n \times \prod_{j=1}^{n} \frac{\mathrm{d}\Lambda\{(t_j, y_j)\}}{\Lambda_0},$$

where we have temporarily written Λ_0 for $\Lambda\{[0, 1] \times (y_0, \infty)\}$. Expressed thus, we see that the likelihood has two parts. First, there is a contribution for the number of events N observed, whose distribution is Poisson with mean Λ_0. Second, conditional on $N = n$, there are independent contributions for each of the (t_j, y_j). These are generalized Pareto with survivor function

$$\Pr(Y \geq y | Y \geq y_0) = \left\{1 + \frac{\kappa}{\tau'}(y - y_0\right\}_+^{-1/\kappa}, \tag{12}$$

where $\tau' = \tau + \kappa(y_0 - \eta)$, and the range of y is that for which this probability is valid. Davison and Smith (1990) give a detailed account of inference based on Eq. (12).

To extend our ideas on local estimation to this setting, we suppose that exceedances of y_0 occur as a univariate inhomogeneous Poisson process with rate $\lambda(t)$. Given that an observation y has occurred at time t, its distribution is taken to be generalized Pareto, with survivor function $\{1 + \kappa(t)(y - y_0)/\tau'(t)\}^{-1/\kappa(t)}$; note that the threshold exceedance is $y - y_0$. We now use smoothers to estimate the functions $\lambda(t)$, $\tau'(t)$, and $\kappa(t)$.

A natural way to estimate $\lambda(t)$ is to divide the time axis into equal-sized boxes, and to average the counts in each, weighted according to their distance from t, using the kernel $w_j(t)$. This is precisely the local likelihood estimate of $\lambda(t)$, using a constant polynomial, but it will be better to use a local linear fit, in effect assuming that close to t, $\log \lambda(u)$ is of

form $\gamma_0 + \gamma(u - t)$. Such an estimate is readily obtained by repeated fitting of a Poisson regression model, with weights $w_j(t)$. Special programming can be reduced or avoided altogether by using a generalized linear or additive model routine to estimate $\lambda(t)$ (McCullagh and Nelder, 1989; Hastie and Tibshirani, 1990).

As before, we estimate $\tau'(t)$ and $\kappa(t)$ using local polynomial smoothing. Suppose that exceedances of sizes $y_1 - y_0, \ldots, y_n - y_0$ occur at times $t_1, \ldots, t_n$; note that these t_j are not the same as those in Section 4. Then the local log likelihood at t is

$$\mathcal{L}(\eta; t) = -\sum_{j=1}^{n} h^{-1} w\left(\frac{t_j - t}{h}\right)\left[\left\{\frac{1}{\kappa(t_j)} + 1\right\} \times \log\{1 + \kappa(t_j)(y_j - y_0)/\tau'(t_j)\}_+ + \log \tau'(t_j)\right] \quad (13)$$

The parameters $\tau'(t)$ and $\kappa(t)$ are estimated from Eq. (13), with the same considerations applying as previously.

6. Bootstrap assessment of uncertainty

Confidence intervals based on large sample properties of maximum likelihood estimates are widely used, but with the advent of bootstrap methods (Efron, 1979), other techniques have been developed that use computer simulation to assess uncertainty. A comprehensive and up-to-date account of bootstrap methods may be found in Davison and Hinkley (1997). Brief descriptions of the basic bootstrap and bootstrap percentile bands used in this paper are given below.

Suppose that we use an estimator T to estimate a parameter θ. If the quantiles of $T - \theta$ are denoted by a_p, then, when T is continuous, the $100(1 - 2\alpha)\%$ equi-tailed interval has limits

$$\hat{\theta}_\alpha = t - a_{1-\alpha}, \qquad \hat{\theta}_{1-\alpha} = t - a_\alpha.$$

As the distribution of $T - \theta$ is usually unknown, approximation of the quantiles of $T - \theta$ are considered. We mimic the distribution of $T - \theta$ by repeatedly simulating data from the fitted model and refitting the model to the fake data. For a given set of data $y_1, \ldots, y_n$, let t be the observed value of T. Let T^* denote the estimate of the parameter θ for a simulated data set $y_1^*, \ldots, y_n^*$. Then the distribution of $T - \theta$ can be mimicked by that of $T^* - t$, whose quantiles can be estimated from the empirical quantiles of R copies of this computed from the simulated data. Then $100(1 - 2\alpha)\%$ basic bootstrap confidence limits for the true parameter θ are

$$\hat{\theta}_\alpha = 2t - t^*_{((R+1)(1-\alpha))}, \qquad \hat{\theta}_{1-\alpha} = 2t - t^*_{((R+1)\alpha)},$$

where $t^*_{(1)} \le \ldots \le t^*_{(R)}$ are ordered empirical quantiles of T^*. The bootstrap percentile interval for θ is

$$\left[t^*_{((R+1)\alpha)}, t^*_{((R+1)(1-\alpha))}\right].$$

Other more complicated approaches are described in Chapter 5 of Davison and Hinkley (1997).

Successful application of the above demands choosing appropriate resampling schemes for generation of artificial data. The two schemes that we use are model-based and model-robust resampling. In model-based resampling we use the fitted model to generate new datasets. Suppose that the local model Eq. (8) is fitted to a set of annual maxima $y_1, \ldots, y_n$. If $\hat{\mu}_j$, $\hat{\sigma}_j$ and $\hat{\kappa}_j$ are the fitted parameter values at $j = 1, \ldots, n$, the corresponding residuals are $e_j = \{1 + \hat{\kappa}_j(y_j - \hat{\mu}_j)/\hat{\sigma}_j\}^{-1/\hat{\kappa}_j}$. We now draw a bootstrap sample $\epsilon_1^*, \ldots, \epsilon_n^*$ by sampling with replacement from the set of rescaled residuals, creating new data

$$Y_j^* = \hat{\mu}_j + \hat{\sigma}_j\left\{\left(\epsilon_j^*\right)^{-\hat{\kappa}_j} - 1\right\}/\hat{\kappa}_j, \quad j = 1, \ldots, n.$$

In the model-robust scheme we split the n years of annual maxima $y_1, \ldots, y_n$ into blocks of about the same length, k years, say, and perform stratified sampling with replacement of the observations within blocks. Finally we allocate the k resampled observations in each block to a random year within that block. This scheme allows for the possibility of gradual changes in the variability of extremes, even if this is not explicitly modelled.

When resampling the r-largest values under these schemes the set of r residuals or observations of a year are treated as a single sampling unit, so their correlation structure within that year is unchanged. All schemes presuppose that data in different years are independent; for further discussion see Davison and Ramesh (2000).

7. Venice data analysis

7.1. Models for maxima

Visual inspection and preliminary analysis suggest that the scale and shape parameters of the maxima of the Venice data exhibit little variation. Therefore we first model the annual maxima by fitting the GEV distribution with a linear trend in location only, taking $\mu_j = \beta_0 + \beta_1 j$, $\sigma_j = e^{\gamma_0}$ and $\kappa_j = \kappa$, for $j = 1, \ldots, 95$. The parameter estimates when this model (M1) is fitted to all 95 maxima are $\hat{\beta}_0 = 82.38(3.47)$, $\hat{\beta}_1 = 0.42(0.06)$, $\hat{\sigma} = 15.33(1.14)$ and $\hat{\kappa} = -0.11(0.04)$, giving estimated 20-year return level $\psi_j = 120.99 + 0.42j$ in year $1886 + j$. The standard errors from observed information, obtained by numerical differentiation of the log likelihood at its maximum, are given in brackets.

We now look at the GEV model for the maxima with local linear trend in location but locally constant scale and shape parameters (M2). For $0 \le t \le 1$ and $1 \le j \le n$, therefore, we take $\mu(t) = \beta_0 + \beta_1(j - nt)$, $\sigma(t) = e^{\gamma_0}$ and $\kappa(t) = \kappa$. The cross-validation technique outlined in Section 2.2 gave an optimal value for h in the range (0.4,0.45), and we take $h = 0.4$. The left panel of Fig. 3 shows the 20-year return levels and 95% confidence interval for our locally-linear fit (M2) using $h = 0.4$. The confidence interval is obtained using a sandwich variance estimate (Davison and Hinkley, 1997, Section 2.7). Also shown are the return level curve of model M1. The right panel shows the estimates of trend parameter β_1 for M1 and M2, together with its pointwise 95% confidence band. Compared with M2, model M1 seems to overestimate the return level slightly for the first half of the data and underestimate it for the second half, but the difference lies within the confidence band. Experimentation with different values of h showed that the return levels for M2 approach the return level line of M1 when $h \to \infty$, as one would expect, as in that case all the data are weighted equally and M2 approaches M1.

7.2. Models for r-largest observations

We now fit local r-largest models for the Venice data. We begin with $r = 10$ and a linear trend in location only, i.e. we fit $\mu_j = \beta_0 + \beta_1 j$, $\sigma_j = e^{\gamma_0}$ and $\kappa_j = \kappa$. This global model (M3) is fitted to all the available data, giving maximum likelihood estimates and standard errors $\hat{\beta}_0 = 91.29(1.19)$, $\hat{\beta}_1 = 0.36(0.02)$, $\hat{\sigma} = 10.98(0.36)$ and $\hat{\kappa} = -0.10(0.02)$.

The estimated 20-year return level line for the data under this model is $\hat{\psi}_j = 119.64 + 0.36j$, reducing the value of $\hat{\psi}$ in 1981 by 7.05 cm when compared with the corresponding model (M1) for annual maxima. This estimated return level is based on more data, and so should indeed has smaller variance.

We now turn to the local models, first considering the model with a locally linear trend in location but constant shape and scale parameters; we drop γ_1 and δ_1. The fitted parameter curves of this 10-largest model (M4) with $h = 0.3$ are displayed in Fig. 4, with 95% confidence intervals from observed information. The upper left panel shows the fitted 20-year return levels. The confidence interval is narrower than for the fit based on maxima alone. Also shown is the 20-year return level for the global model M3.

In the upper right panel, the trend in location $\hat{\beta}_1$ increases slowly from 1910 until 1950 and then declines gradually, with $\beta_1 = 0$ outside the 95% confidence interval for most of the period. The curves for $\hat{\kappa}$ and $\hat{\sigma}$ show little variation in shape or scale, except a slight increase in σ from about 1940. It might be worth exploring a model with linear trend in σ. All the confidence intervals are a factor 0.5 shorter than those of M2 as we are using more data values.

As mentioned in Section 6, we use two resampling schemes to construct confidence bands for the return levels and parameter estimates, using $R = 499$ resampled series. In this application we found stratified sampling useful in the model-based scheme, rather than simple random sampling. The upper panels of Fig. 5 display estimates of the 20-year return level curve for M4 for $h = 0.2$ and 0.3 and the corresponding 90% model-based basic bootstrap bands, together with the estimated return levels from M3. The bottom panels display the 90% model-based and model-robust bootstrap percentile bands of the estimates for model M4 with $h = 0.3$. The 20-year return levels for M4 and M3 are shown in the left panel, and the right panel shows the estimated local trend in location.

In general the return level curve of M4 lies below that of M3 from 1887 to 1950 but subsequently exceeds it. The confidence bands for the trend curves do not include zero, except at the edges. Furthermore the trend does not appear to be constant throughout

the period, as modelled by M3, but varies slowly with time. The local changes in the trend have been captured by the curve of β_1. This suggests ways in which M3 might be regarded as inadequate and hence improved, if it was thought necessary.

When using model-robust resampling, we split the 95 years of data into blocks of lengths 5,10,…,10 years. The confidence bands are similar to those of the model-based resampling scheme. The bootstrap bands of the scale and shape parameters, under both resampling schemes, show much the same level of variation as in the normal confidence intervals in Fig. 5. Judging from the above analysis, M3 tends to underestimate the return level for 1950–80.

In conclusion, the analysis based on *r*-largest models and the assessment of uncertainty using bootstrap methods suggest that the linear model M3 does not entirely capture the pattern of trend in the Venice sea levels. Although it models the overall trend in sea levels reasonably well, model M4 suggests local variation in trend.

8. Nidd data analysis

We illustrate the ideas in Section 5 using the data in the right panel of Fig. 1. There appears to be a slight increase in the size and number of exceedances towards the end of the time period, but it is difficult to assess its significance.

For a simple analysis of the times of exceedances, we fit to their monthly counts a generalized additive model with 4° of freedom, giving an estimate of the daily rate of exceedances, $\hat{\lambda}(t)$. The corresponding cumulative rate, $\hat{\Lambda}(t) = \int_0^t \hat{\lambda}(v)\mathrm{d}v$, agrees well with the empirical distribution of exceedance times, normalized to lie in the unit interval and shown in the lower right panel of Fig. 6. These times would be uniform if the process of exceedances was stationary, but there is a possible slight decrease in their rate in the second half of the data, though it appears to rise again toward the end of the period. The Kolmogorov–Smirnov test (Cox and Lewis, 1966, Chapter 6), however, shows that this is not significant.

The lower left and upper right panels show the estimated shape and scale parameters for the sizes of the exceedances, with smoothing parameter $h = 0.6$. The estimate $\hat{\kappa}(t)$ lies at the the same level close to 0.2 whereas $\hat{\tau}'(t)$ shows a sharp increase towards the end. The exceedances and their 95% return level are shown in the upper left panel of the figure. The return level increases fairly sharply in the last quarter of the data period. Despite the increase in the estimates of $\lambda(t)$, $\psi(t)$ and $\tau'(t)$ towards the end of the period, the 90% equitailed pointwise bootstrap confidence bands, based on $R = 249$ simulated copies of the process, suggest that these changes may be spurious.

The bootstrap replicates for this analysis were obtained by applying the probability integral transform with $\hat{\lambda}(t)$, $\hat{\kappa}(t)$ and $\hat{\tau}(t)$ to both sizes and times of exceedances, and then bootstrapping the resulting point process in the unit square by taking blocks of side 0.05×1, using a block bootstrap. The resulting resampled points were back-transformed to the original scale using $\hat{\lambda}(t)$, $\hat{\kappa}(t)$ and $\hat{\tau}(t)$.

To sum up, it appears that a stationary model is statistically adequate for the Nidd data despite the large values late in the series.

9. Discussion

Smoothing methods are now widely used in statistics, but their potential for hydrological applications is not yet well-explored. Typical uses are in the exploratory phase of analysis, when they are used to suggest possible models, to assess stationarity of data and so forth, and in the confirmatory stage, when they can be used as part of the phase of model criticism. Though more flexible than simple parametric models, they are more expensive to fit owing to their local nature.

There is a huge and somewhat inconclusive statistical literature on the choice of the bandwidth h, which is typically performed by minimizing some criterion such as integrated mean squared error, prediction error or some information criterion. Automatic choice of this sort is important if the fitted curve is to play a role in further analysis, or if a very large number of fits are to be performed on similar datasets, but in exploratory analysis it seems more useful to view different values of h as giving different potential insights into the data.

Local likelihood is relatively easily-motivated and straightforward to apply, but it is certainly worthwhile to investigate which among the many other

approaches to smoothing are best-suited to hydrological data.

Acknowledgement

This work was supported by the Swiss National Science Foundation.

References

Adamowski, K., 1985. Nonparametric kernel estimation of flood frequencies. Water Resources Research 21, 1585–1590.

Adamowski, K., 2000. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. Journal of Hydrology 229, 219–231.

Adamowski, K., Feluch, W., 1990. Nonparametric flood-frequency analysis with historical information. Journal of Hydraulic Engineering 116, 1035–1047.

Adamowski, K., Liang, G., Patry, G., 2000. Annual maxima and partial duration flood series analysis by parametric and nonparametric methods. Hydrological Processes 12, 1685–1699.

Aerts, M., Claeskens, G., 1997. Local polynomial estimation in multiparameter likelihood models. Journal of the American Statistical Association 92, 1536–1545.

Bardsley, W.E., 1989. Using historical data in nonparametric flood estimation. Journal of Hydrology 108, 249–255.

Bardsley, W.E., Mitchell, W.M., Lennon, G.W., 1990. Estimating future sea level conditions under conditions of sea level rise. Coastal Engineering 14, 295–303.

Bloomfield, P., 1992. Trends in global temperature. Climatic Change 21, 1–16.

Bowman, A.W., Azzaini, A., 1997. Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustration. Clarendon Press, Oxford.

Changon, S.A., Kunkel, K.E., 1995. Climate-related fluctuations in midwestern floods during 1921–1985. Journal of Water Resources Planning and Management 121, 326–334.

Chavez-Demoulin, V., Davison, A.C., 2001. Generalized additive models for sample extremes. Submitted for publication.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fiting. Journal of the American Statistical Association 83, 596–610.

Coles, S.G., Tawn, J.A., 1990. Statistics of coastal flood prevention. Philosophical Transactions of the Royal Society, Series A 332, 457–476.

Cox, D.R., Lewis, P.A.W., 1966. The Statistical Analysis of Series of Events. Chapman and Hall, London.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press, Cambridge.

Davison, A.C., Ramesh, N.I., 2000. Local likelihood smoothing of sample extremes. Journal of the Royal Statistical Society, Series B 62, 191–208.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with discussion). Journal of the Royal Statistical Society, Series B 52, 393–442.

Douglas, E., Vogel, R.M., Kroll, C.N., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. Journal of Hydrology 240, 90–105.

Efron, B., 1979. Bootstrap methods: another look at te jackknife. Annals of Statistics 7, 1–26.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.

Green, P.J., Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall, London.

Hastie, T.J., Loader, C., 1993. Local regression: automatic kernal carpentry (with discussion). Statistical Science 8, 120–143.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London.

Lettenmaier, D.P., Wood, E.F., Wallis, J.R., 1994. Hydro-climatological trends in the continental United States, 1948–88. Journal of Climate 7, 586–607.

Loader, C., 1999. Local Regression and Likelihood. Springer, New York.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. . 2nd edChapman and Hall, London.

NERC, 1975. The Flood Studies Report. The Natural Environment Research Council, London.

Pervical, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge.

Perreault, L., Vernier, J., Bobee, B., Parent, E., 2000. Bayesian change-point analysisi in hydrometeorological time series. Part 1. The normal model revisited. Journal of Hydrology 235, 22–241.

Pirazzoli, P.A., 1982. Maree estreme a Venezia (periodo 1972–1981). Acqua Aria 10, 1023–1039.

Robinson, M.E., Tawn, J.A., 1995. Statistics for exceptional athletics records. Applied Statistics 44, 499–511.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer, New York.

Smith, R.L., 1989. Extreme value analysis of environmental time series: an example based on ozone data (with discussion). Statistical Science 4, 367–393.

Smith, R.L., 1993. Long-range dependence and global warming. In: Barnett, V.D., Turkman, K.F. (Eds.). Statistics for the Environment. Wiley, Chichester, pp. 141–161.

Suppiah, R., Hennessy, K.J., 1998. Trends in total rainfall, heavy rain events and number of dry days in Australia, 1910–1990. International Journal of Climatology 10, 1141–1164.

Tawn, J.A., 1988. An extreme value theory model for dependent observations. Journal of Hydrology 101, 227–250.

Tibshirani, R.J., Hastie, T.J., 1987. Local likelihood estimation. Journal of the American Statistical Association 82, 559–567.