ORIGINAL PAPER

Ath. Kehagias · Ev. Nidelkou · V. Petridis

A dynamic programming segmentation procedure for hydrological and environmental time series

Abstract We present a procedure for the segmentation of hydrological and environmental time series. The procedure is based on the minimization of Hubert's segmentation cost or various generalizations of this cost. This is achieved through a dynamic programming algorithm, which is guaranteed to find the globally optimal segmentations with $K = 1$, 2, ..., K_{max} segments. Various enhancements can be used to speed up the basic dynamic programming algorithm, for example recursive computation of segment errors and "block segmentation". The "true" value of K is selected through the use of the Bayesian information criterion. We evaluate the segmentation procedure with experiments which involve artificial as well as temperature and river discharge time series.

Keywords Time series · Segmentation · Change point · Dynamic programming · River discharge

Introduction

In this note we present a *dynamic programming* (DP) solution to the following *time series segmentation* problem: divide a given time series into several *segments* (i.e. blocks of contiguous data) so that each segment is homogeneous, while contiguous segments are heterogeneous (homogeneity is defined in terms of appropriate segment statistics).

The problem of time series segmentation appears often in hydrology and environmetrics (some times it is called *change point detection and estimation*). In a previous paper (Kehagias 2004) we have addressed the problem using a *hidden Markov model*(HMM) formulation. The reader can consult (Kehagias 2004) for a

Ath. Kehagias (✉) · Ev. Nidelkou · V. Petridis
Faculty of Engineering, Aristotle University of Thessaloniki,
Box 464, 54124 Thessaloniki, Greece
E-mail: kehagiat@gen.auth.gr
Tel.: + 30-2310-995944

review of the literature on hydrological time series segmentation; here we simply mention that both the current work and (Kehagias 2004) have been motivated by *Hubert's segmentation procedure* (Hubert 1997, 2000); we also note the recent appearance of (Fortin et al. 2004, 2004) (where time series segmentation is treated using the *shifting-means* model and Markov Chain Monte Carlo methods) and (Chen 2002; Sveinsson et al. 2003).

A review of the statistical and engineering literature on change point detection and estimation is beyond the scope of this paper. We simply refer the reader to the excellent book (Basseville and Nikiforov 1993) which contains a great number of useful references.

We consider hydrological time series segmentation as a *computational* problem and our main concern is with fast and efficient implementation[1]. We follow (and generalize) Hubert's formulation of segmentation as an optimization problem. The HMM approach we have proposed in (Kehagias 2004) is quite fast but is not guaranteed to produce the *globally optimal* segmentation. We have noted in (Kehagias 2004) that the DP solution yields the globally optimal segmentation at the cost of longer execution time. Here we attempt to streamline the DP approach so that it is computationally viable for time series of several thousand observations. Hence the DP approach can be seen as an improvement of Hubert's procedure for the segmentation of time series with multiple change points.

Dynamic programming has often been used in the past for the segmentation of curves (Bellman 1961) and waveforms (Jackson 2005; Kay 1998; Kay and Han, unpublished manuscript; Pavlidis and Horowitz 1974). We are not aware of any work on DP segmentation of hydrological time series. DP has been used in computational biology for segmentation of DNA sequences (see for example (Akaike 1974; Braun and Mueller 1998; Braun et al. 2000) which apply to *discrete* valued sequences some algorithms similar to the one presented

---

[1]On the other hand, we make no claims about the validity of our approach as a realistic hydrological *model!*

here). Interest in time series segmentation problems has recently increased in the literature of *data mining* (see Himberg et al. 2001; Keogh 2001 and, especially, Keogh et al. 2003 which presents a good survey of segmentation methods appearing in the data mining literature). In some of these works it is claimed that DP is *not* a computationally viable approach. The current paper may serve as a partial rebuttal of such claims.

This paper is organized as follows. In Sect. 2 we present a formulation (which closely follows Hubert) of time series segmentation as an optimization problem. In Sect. 3 we present the segmentation procedure. Some segmentation experiments are presented in Sect. 4. In Sect. 5 our results are summarized. Some mathematical details are presented in Appendices 1, 2 and 3

Segmentation as optimization

The formulation and notation introduced in this Section is heavily based on (Hubert 1997, 2000). A *segmentation* of the integers $\{1, 2, ..., T\}$ is a sequence $\mathbf{t} = (t_0, t_1, ..., t_K)$ which satisfy $0 = t_0 < t_1 < \cdots < t_{K-1} < t_K = T$. The intervals of integers $[t_0 + 1, t_1], [t_1 + 1, ..., t_2], ..., [t_{K-1} + 1, t_K]$ are called *segments*, the times $t_0, t_1, ..., t_K$ are called *segment boundaries* or *breaks* and K, the number of segments, is called the *order* of the segmentation. We will denote the set of all segmentations of $\{1, 2, ..., T\}$ by T and the set of all segmentations of order K by T_K. Clearly $T = \cup_{K=1}^{T} T_K$. Denoting by card $(\cdot)$ the number of elements of a set, we have card $(T_K) = \binom{T-1}{K-1}$ and card $(T) = \sum_{K=1}^{T} \mathrm{card}(T_K) = 2^{T-1}$.

In many applications a time series $x_1, x_2, ..., x_T$ is given and we seek a segmentation of $\{1, 2, ..., T\}$ which corresponds to changes of the behavior of $x_1, x_2, ..., x_T$ (in this case segment boundaries are also called *change points*). This can be formulated as an optimization problem. We define the *segmentation cost* $J(\mathbf{t})$ as follows:

$$J(\mathbf{t}) = \sum_{k=1}^{K} d_{t_{k-1}+1, t_k}, \tag{1}$$

where $d_{s,t}$ (for $0 \leq s < t \leq T$) is the *segment error* corresponding to segment $[s,t]$. The *optimal segmentation*, denoted as $\widehat{\mathbf{t}} = (\widehat{t}_0, \widehat{t}_1, \ldots, \widehat{t}_K)$ is defined as

$$\widehat{\mathbf{t}} = \arg\min_{\mathbf{t} \in T} J(\mathbf{t})$$

and the *optimal segmentation of order K*, denoted by $\widehat{\mathbf{t}}^{(K)} = \left(\widehat{t}_0^{(K)}, \widehat{t}_1^{(K)}, \ldots, \widehat{t}_K^{(K)}\right)$, is defined as

$$\widehat{\mathbf{t}}^{(K)} = \arg\min_{\mathbf{t} \in T_K} J(\mathbf{t}).$$

Obviously, the segment error $d_{s,t}$ depends on the data $x_1, x_2, ..., x_T$. A variety of $d_{s,t}$ functions can be used in (1). In Sect. 3 we will introduce several such functions, all of which will have the form (for $0 \leq s < t \leq T$)

$$d_{s,t} = \sum_{\tau=s}^{t} (x_\tau - \widehat{x}_\tau)^2, \tag{2}$$

where $\widehat{x}_\tau$ will be *some regression estimate* of x_τ, obtained from the *within-segment* data $x_s, x_{s+1}, ..., x_t$. Hence $d_{s,t}$ is the sample variance, over the segment $[s,t]$, multiplied by $(t - s)$. The special case

$$\hat{x}_\tau = \frac{\sum_{\tau=s}^{t} x_t}{t - s + 1}$$

(here $\hat{x}_t$ is the segment-mean and corresponds to regression by a constant) is used by Hubert (1997, 2000) and other authors.

Given a particular $d_{s,t}$ function, we can find the optimal segmentation by exhaustive enumeration of all possible segmentations. But this is not computationally practical because the total number of segmentations increases exponentially in T. Hubert uses a *branch-and-bound* approach to search efficiently the set of all possible segmentations. In [15, p.299] is stated that this approach "currently" (in 2000) can segment time series with several tens of terms but is not able "... to tackle series of much more than a hundred terms ..." because of the combinatorial increase of computational burden. As will be seen in Sect. 3, dynamic programming allows the computation of the optimal segmentation in time $O(T^2)$ and yields in a few seconds the optimal segmentation of time series with over a thousand terms.

Dynamic programming segmentation

The segment error function

A complete definition of segmentation cost requires the specification of segment error $d_{s,t}$ for every (s,t) pair such that $0 \leq s < t \leq T$. We use linear regression to introduce a quite general family of segment error functions. Suppose that the time series $x_1, x_2, ..., x_T$ is generated by a process of the form

$$x_\tau = a^1 u_\tau^1 + a^2 u_\tau^2 + \cdots + a^M u_\tau^M + e_\tau,$$

where $u_\tau^1, u_\tau^2, ..., u_\tau^M$ are appropriate input variables, e_τ is zero-mean white noise and the model parameters $a^1, a^2, ..., a^M$ remain constant within each segment but change between segments (examples will be given presently). In other words, we assume that within the k-th segment the time series satisfies the relation (for $\tau = t_{k-1} + 1, t_{k-1} + 2, ..., t_k$)

$$x_\tau = a_k^1 u_\tau^1 + a_k^2 u_\tau^2 + \cdots + a_k^M u_\tau^M + e_\tau,$$

and a_k^m (for $m = 1, 2, ..., M$) are appropriate coefficients.

We define $d_{s,t}$ to be the *error* of the regression model fitted to the samples $x_s, x_{s+1}, ..., x_t$. If we assume $[s,t]$ to be a segment, an estimate of x_τ for $\tau \in [s,t]$ is given by

$$\widehat{x}_\tau = \widehat{a}_{s,t}^1 u_\tau^1 + \widehat{a}_{s,t}^2 u_\tau^2 + \cdots + \widehat{a}_{s,t}^M u_\tau^M, \tag{3}$$

where $\widehat{a}_{s,t}^m$ are the least squares estimates obtained from $x_{s+1}, x_{s+2}, ..., x_t$ (hence $\widehat{x}_\tau$ depends on s and t—this dependence is suppressed in the interest of brevity). Then the segment error is

$$d_{s,t} = \sum_{\tau=s}^t (x_\tau - \widehat{x}_\tau)^2, \tag{4}$$

where $\widehat{x}_\tau$ is given by Eq. 3. Here are some specific examples (many other choices of $u_\tau^1, u_\tau^2, ..., u_\tau^M$ are possible).

1. *Regression by constants*: $M=1$, $u_t^1=1$ and the model is

$$x_t = a_k^1 \cdot 1 + e_t.$$

It is easy to check that in this case the optimal $\widehat{a}_{s,t}^1$ is given by $\widehat{a}_{s,t}^1 = \sum_{\tau=s}^t x_\tau / (t-s+1)$, i.e. the empirical mean; the corresponding segment cost is the within-segment variance (multiplied by $t-s$).

2. *Regression by linear functions*: $M=2$, $u_t^1=1$, $u_t^2=t$ and the model is

$$x_t = a_k^1 \cdot 1 + a_k^2 \cdot t + e_t. \tag{5}$$

3. *Auto-regression*: $u_t^1=x_{t-1}$, $u_t^2=x_{t-2}$, ..., $u_t^M=x_{t-M}$ and the model is

$$x_t = a_k^1 x_{t-1} + a_k^2 x_{t-2} + \ldots + a_k^M x_{t-M} + e_t.$$

When s and t are true segment boundaries, $d_{s,t}$ reflects only data noise (for instance, if e_τ were 0 for $\tau=s, s+1, ..., t$, then $d_{s,t}$ would also equal zero); when $[s,t]$ straddles two true segments, then $d_{s,t}$ also contains systematic error (since we attempt to fit two regression relations with one set of coefficients). Hence, smaller $d_{s,t}$ values are associated with true segments.

What if $[s,t]$ is a subset of a true segment? In general, segmentation cost can always be reduced by splitting true segments into sub-segments. In fact, we can make a stronger statement. Denote by $\widehat{J}(K)$ the *optimal* segmentation cost obtained by using segmentations with K segments. In other words, we set

$$\widehat{J}(K) = \min_{\mathbf{t} \in T_K} J(\mathbf{t}).$$

It is easily checked that $K > K'$ implies $\widehat{J}(K) \leq \widehat{J}(K')$ [for every $d_{s,t}$ of the form (Eq. 4)]. As an (extreme) example, note that $\widehat{J}(T) = 0$; this is obtained from the trivial segmentation $(0, 1, 2, ..., T)$, where every segment contains a single sample. But this segmentation does not convey *any* information about the structure of the time series. Hence we must guard against using "too many" segments. This is the problem of determining the *segmentation order*, and it is a special case of *model order selection*, a problem studied extensively in the statistical literature. We will address this problem in Sect. 3.3.

We now present explicit formulas to compute $d_{s,t}$ for all segments $[s,t]$. First, let us introduce some notation. We write $\mathbf{u}_t = \begin{bmatrix} u_t^1 & u_t^2 & \ldots & u_t^M \end{bmatrix}$ and

$$\mathbf{X}_{s,t} = \begin{bmatrix} x_s \\ x_{s+1} \\ \cdots_t \end{bmatrix}, \quad \widehat{\mathbf{X}}_{s,t} = \begin{bmatrix} \widehat{x}_s \\ \widehat{x}_{s+1} \\ \cdots \\ \widehat{x}_t \end{bmatrix}, \quad \mathbf{U}_{s,t} = \begin{bmatrix} \mathbf{u}_s \\ \mathbf{u}_{s+1} \\ \cdots \\ \mathbf{u}_t \end{bmatrix}. \tag{6}$$

We also write $\mathbf{A}_{s,t} = \begin{bmatrix} \widehat{a}_{s,t}^1 & \widehat{a}_{s,t}^2 & \ldots & \widehat{a}_{s,t}^M \end{bmatrix}$. Finally, when $\mathbf{f}$ is a vector, $|\mathbf{f}|$ is its Euclidean norm and we have $|\mathbf{f}| = \sqrt{\mathbf{f}'\mathbf{f}}$ (where $'$ denotes matrix transpose). Using the above notation we have $\widehat{x}_t = \mathbf{u}_t \mathbf{A}_{s,t}$, $\widehat{\mathbf{X}}_{s,t} = {}_{s,t}\mathbf{A}_{s,t}$ and

$$\begin{aligned} d_{s,t} &= \left| \mathbf{X}_{s,t} - \widehat{\mathbf{X}}_{s,t} \right|^2 \\ &= \left| \mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t} \right|^2 = \left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t} \right)' \left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t} \right). \end{aligned}$$

From the theory of least squares linear regression, we know that

$$\mathbf{A}_{s,t} = \left(\mathbf{U}_{s,t}'\mathbf{U}_{s,t} \right)^{-1} \mathbf{U}_{s,t}'\mathbf{X}_{s,t}. \tag{7}$$

The computation of $d_{s,t}$ for a particular segment $[s,t]$ requires the computation of the matrix products $\mathbf{U}_{s,t}'\mathbf{U}_{s,t}$ and $\mathbf{U}_{s,t}'\mathbf{X}_{s,t}$; these require $O(t-s)$ operations. Hence the straightforward computation of $d_{s,t}$ for all possible segments $[s,t]$ requires $O(T^3)$ computations.

However, we can reduce computation time. We now present a *recursive* scheme which computes $d_{s,t}$ for all possible $[s,t]$ in time $O(T^2)$. The correctness of this scheme is proved in Appendix 1. We define

$$\mathbf{Q}_{s,t} = \mathbf{U}_{s,t}'\mathbf{U}_{s,t}, \quad \mathbf{R}_{s,t} = \mathbf{U}_{s,t}'\mathbf{X}_{s,t}. \tag{8}$$

Then we can rewrite Eq. 7 as

$$\mathbf{A}_{s,t} = \mathbf{Q}_{s,t}^{-1}\mathbf{R}_{s,t}. \tag{9}$$

We have (see Appendix 1)

$$\mathbf{Q}_{s,t+1} = \mathbf{Q}_{s,t} + \mathbf{u}_{t+1}'\mathbf{u}_{t+1}, \tag{10}$$

$$\mathbf{R}_{s,t+1} = \mathbf{R}_{s,t} + \mathbf{u}_{t+1}'x_{t+1}. \tag{11}$$

Furthermore, with $\boldsymbol{\delta}\mathbf{A} = \mathbf{A}_{s,t+1} - \mathbf{A}_{s,t}$, we have (Appendix 1)

$$\begin{aligned} d_{s,t+1} &= d_{s,t} + \boldsymbol{\delta}\mathbf{A}'\mathbf{Q}_{s,t}\boldsymbol{\delta}\mathbf{A} \\ &+ \left(x_{t+1} - \mathbf{u}_{t+1}\mathbf{A}_{s,t+1} \right)' \left(x_{t+1} - \mathbf{u}_{t+1}\mathbf{A}_{s,t+1} \right). \end{aligned} \tag{12}$$

In short, we have the following algorithm, which involves two loops and hence runs in $O(T^2)$ time.

The dynamic programming algorithm

We now present the actual segmentation algorithm. The motivation for this algorithm is the following. Consider an optimal segmentation of $x_1, x_2, ..., x_t$ which contains K segments and suppose its last segment is $[s+1, t]$.

Segment error computation
Input
The time series $x_1, x_2, ..., x_T$;
the time series $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_T$.
Main
For $s = 1, 2, ..., T$
  $\mathbf{Q}_{s,s} = \mathbf{u}_s \mathbf{u}_s$
  $\mathbf{R}_{s,s} = \mathbf{u}_s x_s$
  $\mathbf{A}_{s,s} = \mathbf{Q}_{s,s}^{-1} \mathbf{R}_{s,s}$
  $d_{s,s} = 0$
  For $t = s, s+1, ..., T-1$
    $\mathbf{Q}_{s,t+1} = \mathbf{Q}_{s,t} + \mathbf{u}_{t+1} \mathbf{u}_{t+1}$
    $\mathbf{R}_{s,t+1} = \mathbf{R}_{s,t} + \mathbf{u}_{t+1} x_{t+1}$
    $\mathbf{A}_{s,t+1} = \mathbf{Q}_{s,t+1}^{-1} \mathbf{R}_{s,t+1}$
    $\boldsymbol{\delta A} = \mathbf{A}_{s,t+1} - \mathbf{A}_{s,t}$
    $d_{s,t+1} = d_{s,t} + \boldsymbol{\delta A}' \mathbf{Q}_{s,t} \boldsymbol{\delta A} + | x_{t+1} - \mathbf{u}_{t+1} \mathbf{A}_{s,t+1} |^2$.
  End
End

Then the remaining $K-1$ segments form an optimal segmentation of $x_1, x_2, ..., x_s$. More specifically, if $c_t^{(k)}$ is the minimum segmentation cost of $x_1, x_2, ..., x_t$ into k segments, we have

$$c_t^{(k)} = c_s^{(k-1)} + d_{s+1,t}.$$

This is a typical dynamic programming approach and can be implemented by the following algorithm.

On termination, the algorithm has computed the optimal segmentation cost $c_{T,K} = \widehat{J}(K) = \min_{\mathbf{t} \in T_K} J(\mathbf{t})$ and, by backtracking, the optimal segmentation $\widehat{\mathbf{t}}^{(K)} = (t_0^{(K)}, t_1^{(K)}, ..., t_K^{(K)})$; these quantities have been

Dynamic programming segmentation
Input
The time series $x_1, x_2, ..., x_T$;
the segment errors $d_{s,t}$ ($0 \leq s < t \leq T$);
the max number of segments $K_{\max}$.
Initialization
$c_0^{(1)} = 0$
For $t = 1, 2, ..., T$
  $c_t^{(1)} = d_{1,t}$
  $z_t^{(1)} = 0$
End
Minimization
For $K = 2, ..., K_{\max}$
  $c_0^{(K)} = 0$
  For $s = 1, 2, ..., T$
  For $t = 1, 2, ..., s-1$
    $e_{t,s} = c_t^{(K-1)} + d_{t+1,s}$
  End
  $c_s^{(K)} = \min_{1 \leq t \leq s-1}(e_{t,s})$
  $z_s^{(K)} = \arg\min_{1 \leq t \leq s-1}(e_{t,s})$
  End
End
Backtracking
For $K = 1, 2, ..., K_{\max}$
  $\widehat{t}_K^{(K)} = T$
  For $k = K, K-1, ..., 1$
    $\widehat{t}_{k-1}^{(K)} = z_{\widehat{t}_k^{(K)}}^{(k)}$
  End
End

computed for $K = 1, 2, ..., K_{\max}$, in other words a *sequence* of minimization problems has been solved recursively.

Selection of segmentation order

After having obtained the optimal segmentation for every value of K, it remains to select the "correct" vaue of K, i.e. the "true" number of segments. This is a difficult problem. It is a special case of the *model selection problem*, which has been researched extensively and, apparently, has not yet been solved satisfactorily (for relatively recent reviews see (Chen and Gupta 1998; Kearns et al. 1997)). Hubert's solution to this problem consists in testing, by the Scheffe criterion, the null hypothesis that successive segments do not have significantly different means; he takes the true value $\hat{K}$ to be the largest K for which the null hypothesis can be rejected for segments k and $k+1$ (with $k = 1, 2, ..., K-1$). There are several difficulties with this approach. Some of these are noted by Hubert (2000); in our opinion the main difficulty is that successive samples are not independent; also, it is not immediately obvious how to generalize Scheffe's criterion when the segment statistic is not the mean.

We use a different approach, based on Schwarz's *Bayesian information criterion* (BIC). We define

$$\widetilde{J}(K) = T \log\left(\frac{\widehat{J}(K)}{T-1}\right) + R_K \log(T) \tag{13}$$

(where R_K is the number of parameters in the K-th order model[2]) and take $\widehat{K}$ to be the value of K which minimizes $\widetilde{J}(K)$. The general motivation for the use of BIC can be found in (Schwarz 1978); a classical application to change point problems appears in (Yao 1988). While BIC does not always yield a good estimate of the true number of segments, in our experience it performs much better than alternative criteria such as Akaike's (Auger and Lawrence 1989) and Ninomiya's (Ninomiya 2003); we discuss the issue further in Sects. 4.5 and 5.

Segmentation experiments

In the following sections we present several experiments to evaluate our algorithm.

In some of these experiments we use artificially synthesized time series for which the correct segmentation is known; hence we can form a very accurate picture of the performance of our algorithm. More specifically, when the correct segmentation of a time series is known, the *error* of a hypothetical segmentation can be evaluated using (among other measures)

---

[2] For example, in the regression-by-constants model $R_K = 2K$ (the $t_1^{(K)}, ..., t_{K-1}^{(K)}$, the parameters $a_1^{(K)}, ..., a_1^{(K)}$ and the standard deviation $\widehat{\sigma}^{(K)}$ of e_t), in the regression-by-lines model $R_K = 3K$ and so on.

Fig. 1 A noiseless ($\sigma = 0$, *dotted line*) and a noisy ($\sigma = 0.2$, *solid line*) version of the synthetic time series used in Sect. 4.1
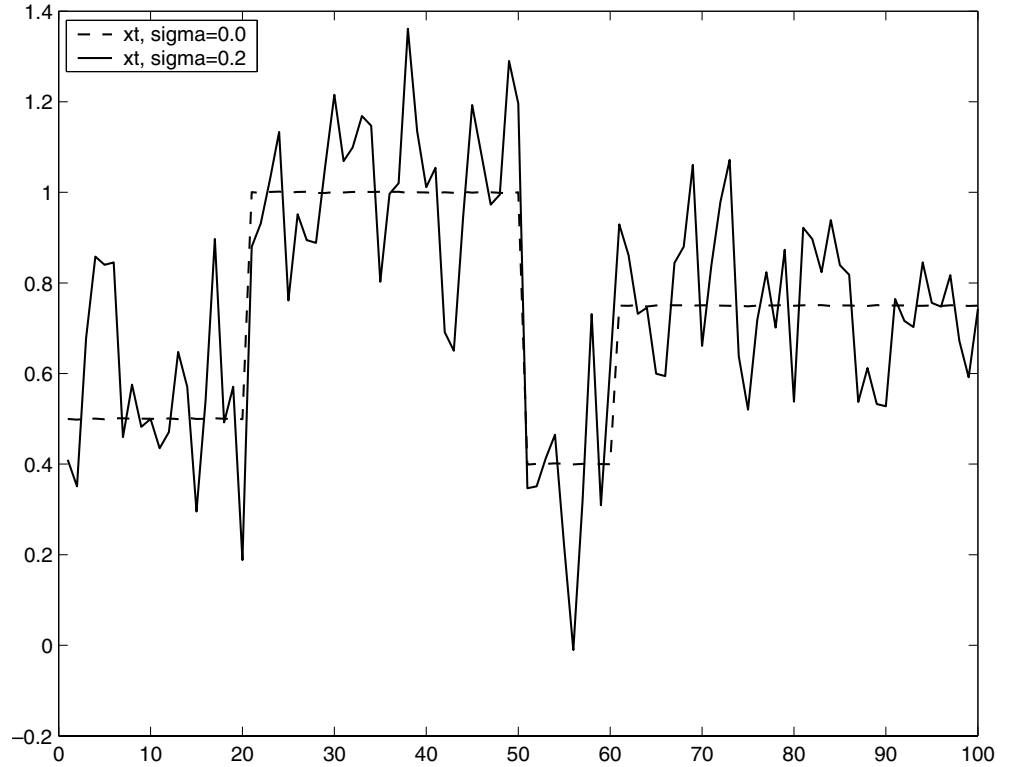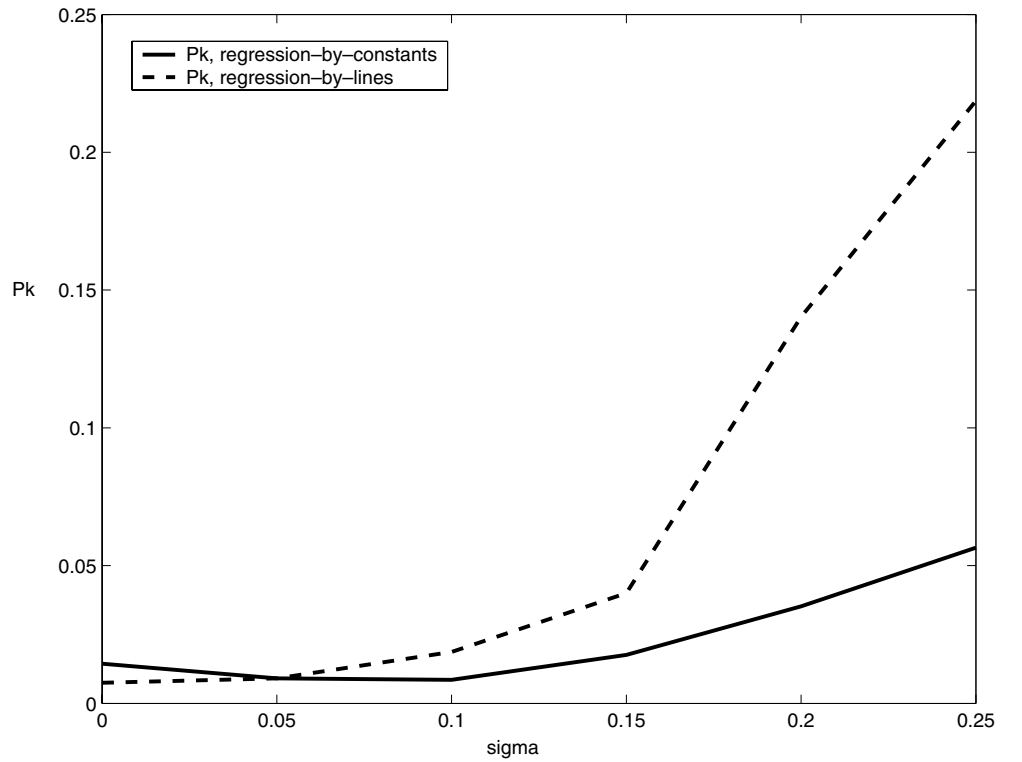

Fig. 2 Plots of P_k values averaged over the various σ values used in Sect. 4.1. The *solid line* corresponds to regression-by-constants and the *dotted line* corresponds to regression-by-lines

Beeferman's segmentation metric P_k (Beeferman et al. 1999). A small value of P_k indicates low segmentation error ($P_k = 0$ indicates a perfect segmentation). P_k is widely used for evaluation of *text* segmentation algorithms but we believe it is a good measure for time series segmentation as well. The definition of P_k is given in Appendix 3; the interested reader can find more details in (Beeferman et al. 1999).

In the remaining experiments we use some real-world time series; while the correct segmentations are not known for these cases, it is important to evaluate the algorithm performance on realistic problems.

Fig. 3 A synthetic time series x_t (*solid line*) and its approximation $\hat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-constants ($\sigma = 0$)

Fig. 4 A synthetic time series x_t (*solid line*) and its approximation $\hat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-constants ($\sigma = 0.15$)

Synthetic data: short time series

In the first group of experiments we start with a piece-wise constant time series of length $T = 100$, which contains four segments. To this basic time series we superimpose additive white noise with a Gaussian distribution of zero mean and standard deviation σ; we use $\sigma = 0.00, 0.05, 0.10, ..., 0.25$.

Fig. 5 A synthetic time series x_t (*solid line*) and its approximation $\widehat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-constants ($\sigma = 0.25$)
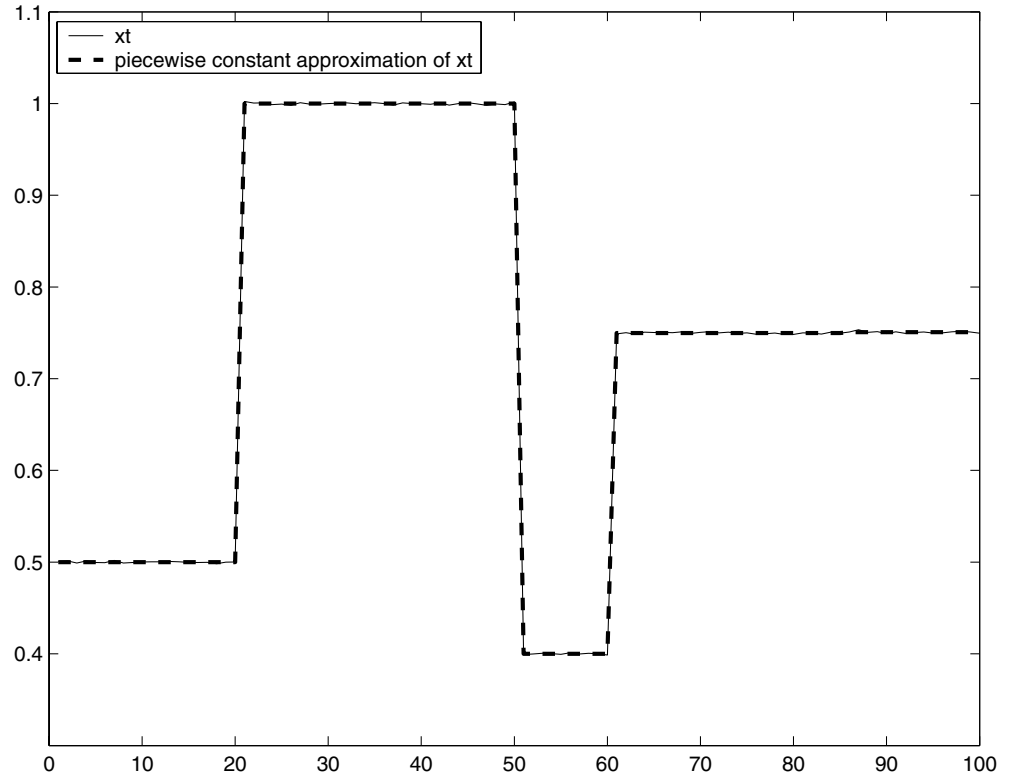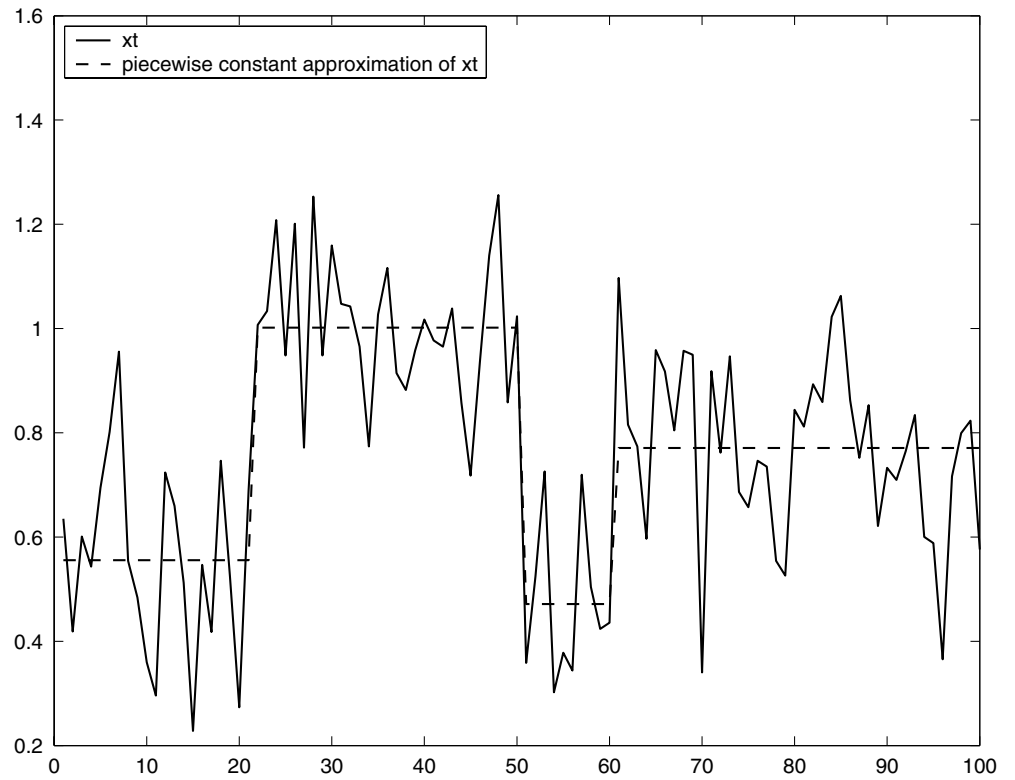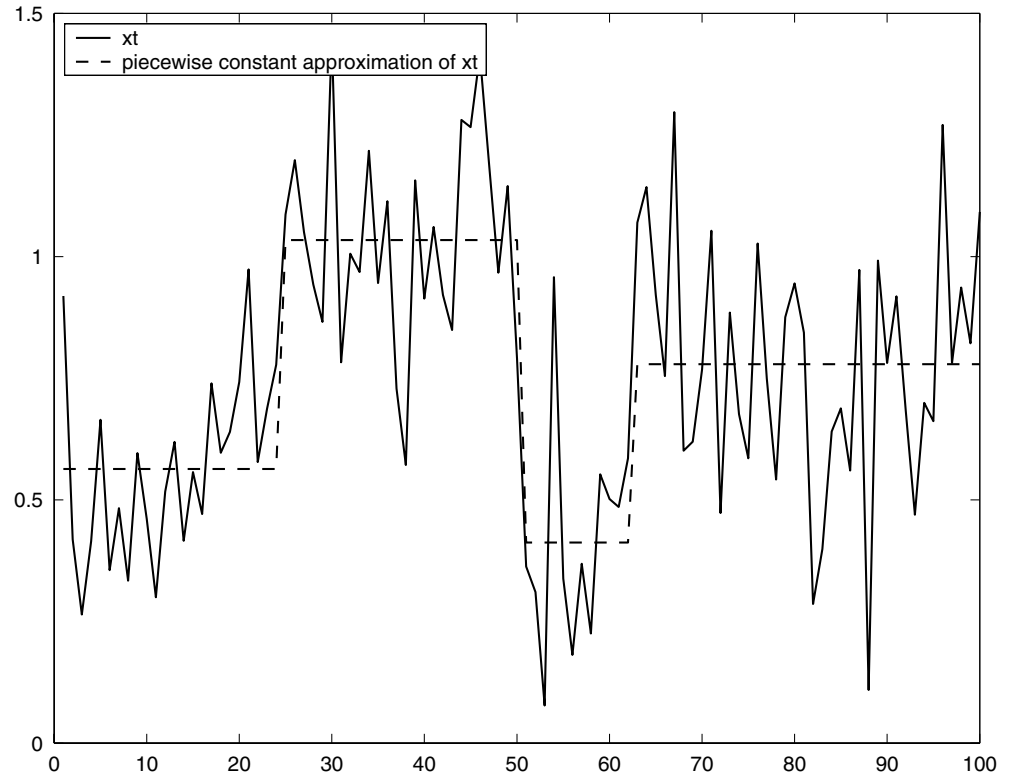

Fig. 6 A synthetic time series x_t (*solid line*) and its approximation $\widehat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-lines ($\sigma = 0.15$)
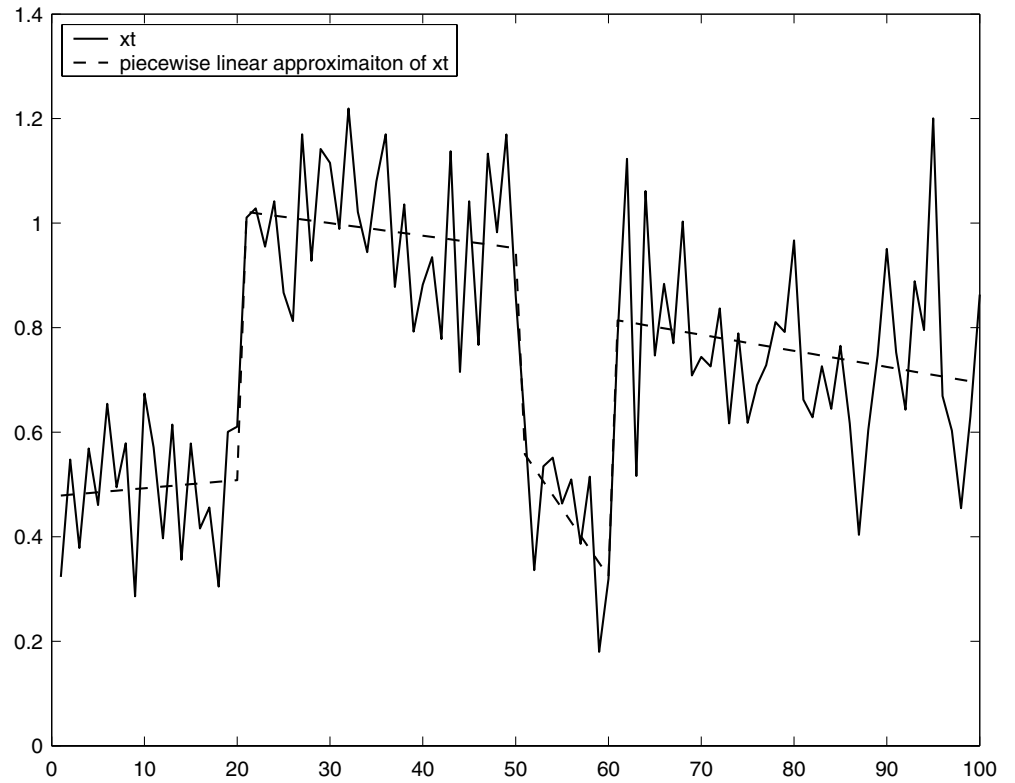
For each σ value we create 50 noisy versions of the original (noiseless) time series. The original and a noisy version are plotted in Fig. 1.

There is a total of $6 \times 50 = 300$ time series. We apply our algorithm to each time series using *regression-by-constants:*

Fig. 7 Annual discharge of the Senegal river (*solid line*) and its regression-by-constants approximation (*dotted line*) obtained from the optimal segmentation of order $K = 5$
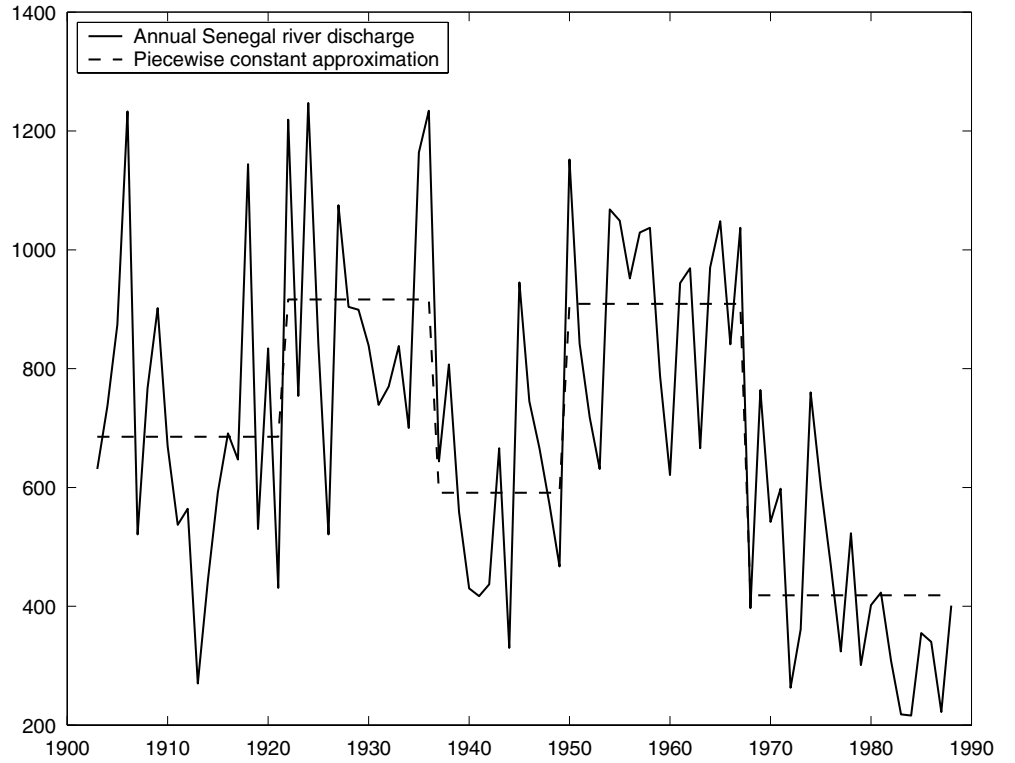
$$\widehat{x}_\tau = a_1 + e_t. \tag{14}$$

We compute segmentations of orders $K = 1, 2, ..., 10$ and determine the optimal order using the BIC. Hence for each of the 300 time series we obtain a "best" segmentation and compute the corresponding P_k value; we average P_k values obtained at the same σ value and plot the average P_k values against σ in Fig. 2. Typical segmentations at three noise levels appear in Figs. 3, 4 and 5. We create an additional 300 time series (50 for each σ value) and apply our algorithm again, using *regression-by-lines*:

$$\widehat{x}_\tau = a_1 t + a_2 + e_t.$$

The average P_k values are plotted in Fig. 2. A typical segmentation appears in Fig. 6.

It is clear that regression-by-constants performs better than regression-by-lines. This is not surprising, since a piecewise constant function conforms exactly to the regression-by-constants equation 14. However, note that regression-by-lines can also reproduce a piecewise constant function by using slope coefficients a_2 (nearly) equal to zero. This effect can be seen in Fig. 6. In fact, if we consider the resulting segmentation as a discontinuous, noiseless representation of the original noisy time series (see the dotted lines in Figs. 3, 4, 5, 6) then the representation of Fig. 6 has a lower sum squared error than the one of Fig. 4 (in both cases $\sigma = 0.15$). However, note that the *segmentation* of Fig. 4 is just as accurate as that of Fig. 6.

It can be seen from Fig. 2 that our algorithm yields highly accurate segmentations even at the highest noise levels. The average execution time for one segmentation is 0.9 s for regression-by-constants and 1.17 s for regression-by-lines (with a MATLAB implementation of the segmentation algorithm on a Pentium III 1 GHz personal computer).

Table 1 Optimal segmentations of orders $K = 1, 2, ..., 10$ for the Senegal river time series, as obtained from regression-by-constants

1	1902	1988								
2	1902	1967	1988							
3	1902	1949	1967	1988						
4	1902	1938	1949	1967	1988					
5	**1902**	**1921**	**1936**	**1949**	**1967**	**1988**				
6	1902	1921	1936	1949	1967	1978	1988			
7	1902	1909	1917	1934	1936	1949	1967	1988		
8	1902	1909	1917	1934	1936	1949	1967	1978	1988	
9	1902	1905	1906	1921	1934	1936	1949	1967	1978	1988
10	1902	1905	1906	1921	1924	1934	1936	1949	1967	1978	1988

The BIC selected segmentation is listed in bold. Note the re-appearance of "important" change points in segmentations of several orders

Table 2 Optimal segmentations of orders $K = 1, 2, ..., 10$ for the Senegal river time series, as obtained from regression-by-lines

1	1902	1988									
2	1902	1953	1988								
3	**1902**	**1938**	**1967**	**1988**							
4	1902	1906	1936	1967	1988						
5	1902	1906	1936	1944	1967	1988					
6	1902	1906	1936	1944	1949	1967	1988				
7	1902	1906	1921	1934	1944	1949	1967	1988			
8	1902	1906	1917	1921	1934	1944	1949	1967	1988		
9	1902	1906	1917	1921	1934	1936	1944	1949	1967	1988	
10	1902	1906	1917	1921	1934	1936	1944	1949	1953	1967	1988

The BIC selected segmentation is listed in bold. Note the re-appearance of "important" change points in segmentations of several orders. Also compare with Table 1

Annual discharge of the Senegal river

In the second group of experiments we use the time series of the Senegal river annual discharge data, measured at the Bakel station for the years 1903–1988. This time series has been previously used in (Fortin et al. 2004; Hubert 1997, 2000; Kehagias 2004); it consists of 86 points and its graph appears in Fig. 7.

Since in this case the "true" segmentation is not known, we cannot compute P_k values. Instead, for comparison purposes, we list the optimal segmentations of *all* orders $K = 1, 2, ..., 10$ as obtained from the regression-by-constants (Table 1) and as obtained from regression-by-lines (Table 2). In each table, the segmentation selected by BIC is indicated by bold letters. In Fig. 7 we plot the actual time series and the optimal segmentation obtained from regression-by-constants and the BIC.

It can be seen that there is a certain consistency between segmentations of different orders and also between tables; for instance the years 1917–1921, 1934–1936, 1949 and 1967 appear in most segmentations of both tables. Hence it appears likely that some hydrological change (which has been detected by both models) occurred around years 1936, 1949 and 1967. A change also seems likely to have occurred around the year 1921. The overall best segmentation in Table 1 is (1902, 1921, 1936, 1949, 1967, 1988). This is the same as the one obtained by Hubert's branch-and-bound procedure (Hubert 1997, 2000) and Kehagias' HMM (2004). In Table 2 the selected segmentation is (1902, 1938, 1967, 1988). Regarding the discrepancy of the two segmentations, the following remarks can be made. From a computational point of view, each of the above mentioned segmentations is optimal with respect to the corresponding $d_{s,t}$ function. As remarked in (Kehagias

Fig. 8 Annual mean global temperature (*solid line*) and its autoregression approximation (*dotted line*) obtained from the optimal segmentation of order $K = 4$
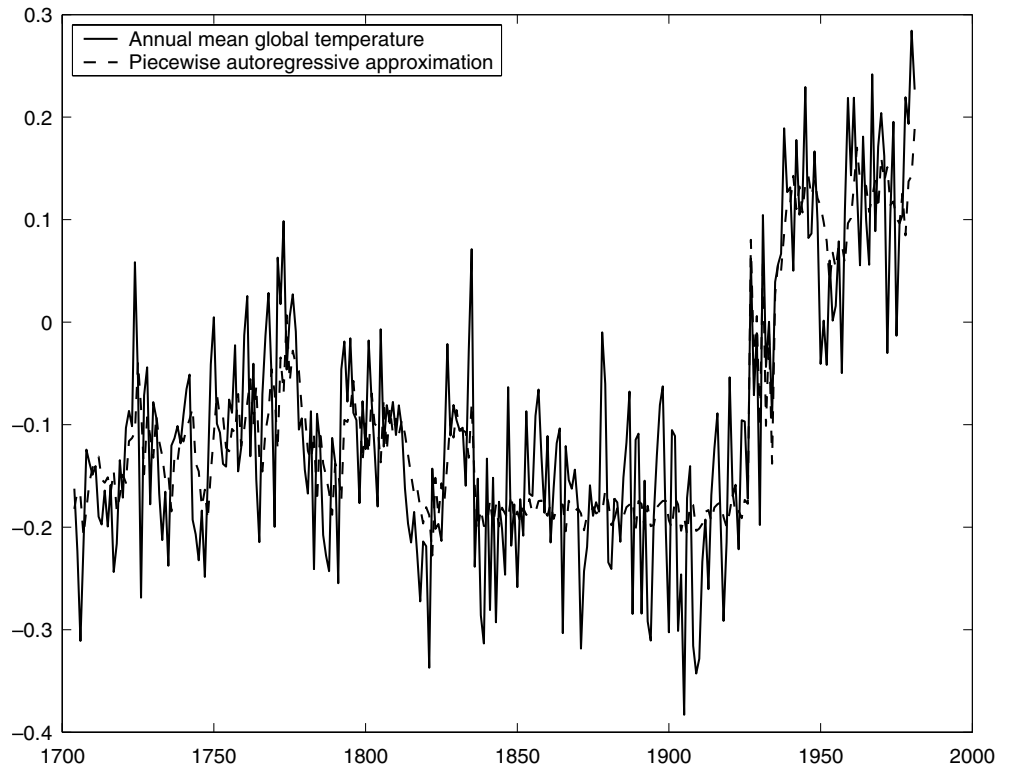
Table 3 Optimal segmentations of orders $K = 1, 2, ..., 10$ for the annual mean global temperature time series, as obtained from regression-by-constants

1	1699	1981									
2	1699	1930	1981								
3	1699	1812	1930	1981							
4	**1699**	**1720**	**1812**	**1930**	**1981**						
5	1699	1720	1812	1926	1935	1981					
6	1699	1720	1812	1926	1934	1977	1981				
7	1699	1720	1812	1825	1835	1926	1935	1981			
8	1699	1720	1812	1825	1835	1926	1934	1977	1981		
9	1699	1720	1812	1825	1835	1926	1935	1949	1957	1981	
10	1699	1718	1770	1777	1813	1825	1835	1926	1934	1977	1981

The BIC selected segmentation is listed in bold. Note the re-appearance of "important" change points in segmentations of several orders

2004), there is no a priori reason that the two segmentations should be the same; each one optimizes a different function (i.e. a function obtained from a different $d_{s,t}$ error).

Obtaining the sequence of ten segmentations with the MATLAB implementation of the segmentation algorithm required 0.76 s for Table 1 and 0.93 s for Table 2 (both on a Pentium III 1 GHz personal computer). For comparison purposes, Hubert reports execution time around 1 minute (but this is probably on a slower machine).

Annual mean global temperature

In this group of experiments we use the time series of annual mean global temperature for the years 1700–1981; only the temperatures for the period 1902–1981 come from actual measurements; the remaining temperatures were *reconstructed* according to a procedure described in (Mann et al. 1999) and also at the Internet address http://www.ngdc.noaa.gov/paleo/ei/ei_intro.html. This time series has been used in (Kehagias 2004). The time series consists of 282 points, its graph appears in Fig. 8.

We apply the segmentation algorithm in conjunction with regression-by-constants (Table 3) and a third-order autoregression (Table 4) given by

$$\widehat{x}_\tau = a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + e_t.$$

In each table, the correct segmentation (as obtained by the BIC) is indicated by bold.

In Table 3 BIC selects the segmentation (1699, 1720, 1812, 1930, 1981); this is the same as the segmentation obtained in (Kehagias 2004) using an HMM and regression-by-constants. In Table 4 BIC selects the segmentation (1699, 1832, 1923, 1931, 1981); this is similar to the one obtained in (Kehagias 2004) using an HMM and autoregression. In Fig. 8 we plot the actual time series and the optimal segmentation obtained from regression-by-constants and the BIC. Again, the two segmentations are not identical; perusing Table. 3 and 4 hydrological changes appear likely around 1812 and 1930 and perhaps also in some year between 1720 and 1770. Obtaining the sequence of ten segmentations with a MATLAB implementation of the segmentation algorithm required 8.35 s for regression-by-constants and 11.04 s for autoregression.

Synthetic data: long time series

In this group of experiments we use artificial time series with 1,200 samples each. When $T = 1,200$, the computational load becomes excessive and some modification of the original segmentation algorithm is necessary. Hence in this case we use *block* segmentation, i.e. we operate on blocks of samples, each block containing N samples. This results in a modification of of the $d_{s,t}$ function, but the DP algorithm remains the same. The details are presented in Appendix 2.

Table 4 Optimal segmentations of orders $K = 1, 2, ..., 10$ for the annual mean global temperature time series, as obtained from autoregression

1	1699	1981									
2	1699	1923	1981								
3	1699	1832	1920	1981							
4	**1699**	**1832**	**1923**	**1931**	**1981**						
5	1699	1832	1849	1923	1931	1981					
6	1699	1832	1899	1910	1923	1931	1981				
7	1699	1832	1849	1901	1910	1923	1931	1981			
8	1699	1832	1849	1901	1910	1923	1931	1973	1981		
9	1699	1745	1767	1774	1832	1899	1910	1923	1931	1981	
10	1699	1745	1767	1774	1832	1849	1901	1910	1923	1931	1981

The BIC selected segmentation is listed in bold. Note the re-appearance of "important" change points in segmentations of several orders. Also compare with Table 3

Fig. 9 A noiseless ($\sigma=0$, *dotted line*) and a noisy ($\sigma=0.15$, *solid line*) version of the long synthetic time series used in Sect. 4.4
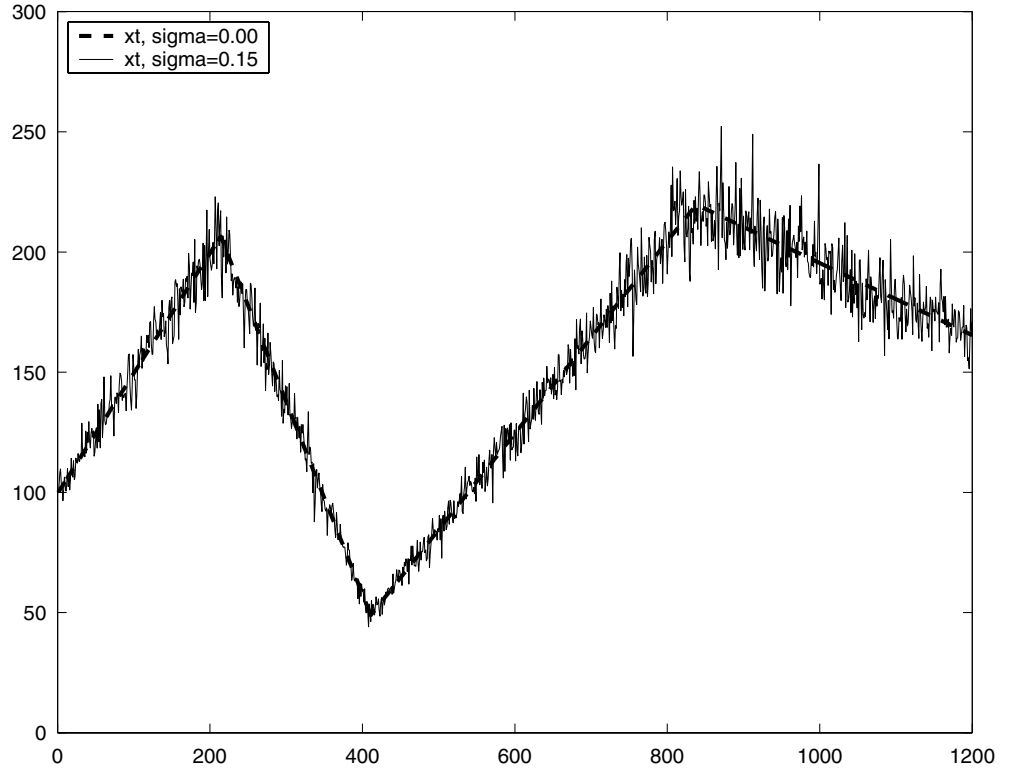
Our time series is generated by the equations

$$z_t = a_1^K z_{t-1} + a_2^K, \quad x_t = (1 + e_t) \cdot z_t, \qquad (15)$$

where e_t is white, zero mean Gaussian noise with standard deviation equal to σ (again we use $\sigma=0.00, 0.05, ..., 0.25$). The time series contains four segments. Two typical time series (for $\sigma=0.00$ and $\sigma=0.15$) are plotted in Fig. 9.

As can be seen, the noiseless series consists of continuously joined up- and down-sloping straight line segments—this was achieved by appropriate selection of the coefficients a_1^k, a_2^k (for $k=1, 2, ..., K$). Note that noise is added *multiplicatively* to the basic time series, hence Eq. 15 is *not* the same as Eq. 5. This was a deliberate choice, to check the performance of the algorithm when using "misspecified" models.

The experimental procedure is the same as in Sect. 4.1. We use regression-by-constants and regression-by-lines and we take $\sigma=0.00, 0.05, ..., 0.25$. For each combination of regression function and noise level we generate 50 time series to which we apply our algorithm.

Regression-by-constants generally yields segmentations which are not in good correspondence to the true segmentation. On the other hand, our algorithm in conjunction with regression-by-lines always discovers the correct fourth order segmentation or a very close approximation. However, the BIC criterion fails for this long time series: $\widetilde{J}(K)$ is decreasing with K in the range $K=1, 2,..., 25$ (we have not tried values of K greater than 25). If we followed the BIC criterion we would assign $K=25$ or higher and thus obtain incorrect segmenta-tions. We have obtained similar results with other criteria, such as Scheffe's, Akaike's and Ninomiya's.

It must be stressed that the abovementioned problem is not a shortcoming of the *computational* procedure. The fourth order segmentations obtained by our algorithm are very accurate, as can be seen in Fig. 11 (average P_k values *for the fourth order segmentation*, plotted against σ for regression-by-lines). On the other hand, in Fig. 10 we see that P_k values for regression-by-constants are not good. The reason for the difference between segmentation with constants and segmentation with lines can be understood by looking at two typical segmentations, in Figs. 12 and 13. Namely, the segmentation by constants does not conform to the piecewise linear structure of the time series. An additional fact which emerges from Fig. 11 is that the segmentation results are quite similar for different values of block size N; in other words the algorithm is quite robust with respect to N.

The average execution time for one segmentation is between 0.45 and 1.16 s for regression-by-constants and between 0.55 and 1.90 s for regression-by-lines.

Annual minimum flow of the Nile

The final experiment uses the time series of annual minimum level of the Nile river, for the years 622–1921.[3]

---

[3]This time series comes from the book (Hipel and McLeod 1994) and can be found in the Web at http://www-personal.buseco.mo-nash.edu.au/hyndman/TSDL/annual/minimum.dat.

Fig. 10 Plots of P_k values averaged over the various σ values used in Sect. 4.4. All curves were obtained with regression-by-constants
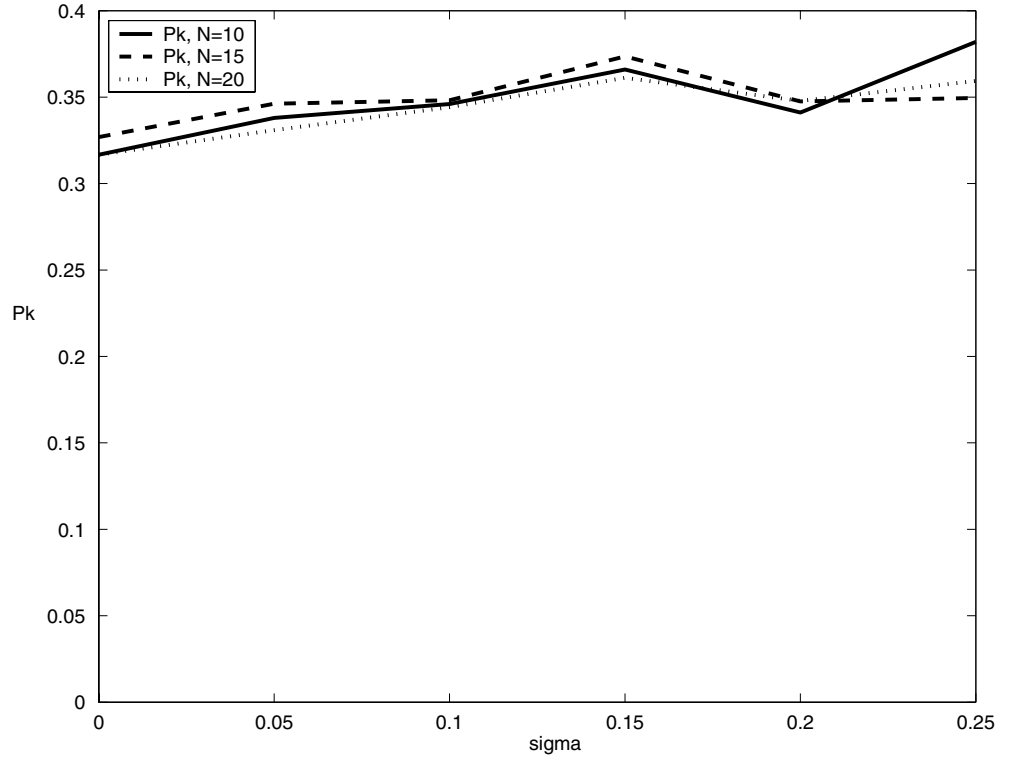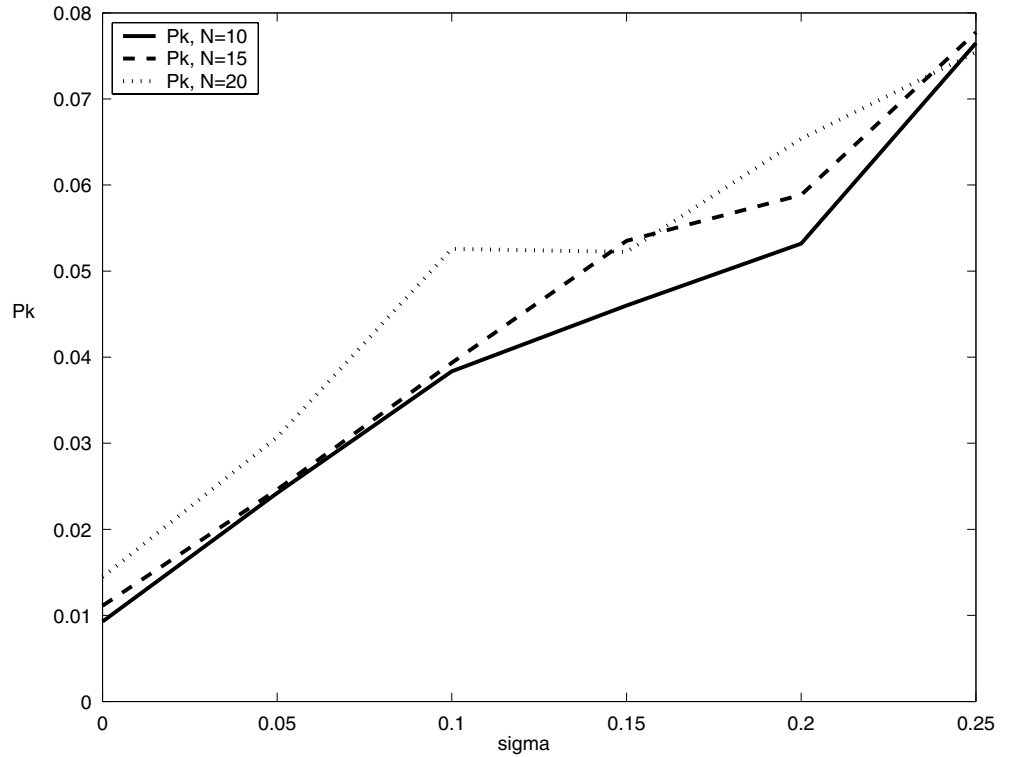

Fig. 11 Plots of P_k values averaged over the various σ values used in Sect. 4.4. All curves were obtained with regression-by-lines

This time series consists of 1,298 points; its graph appears in Fig. 14. The straight line segments in the period 1471–1702 are probably imperfect recordings.

For this *long* time series we use *block* segmentation in conjunction with regression-by-constants and regression-by-lines. To check robustness with respect to block size N, we repeat each of the experiments using $N = 10$, 15, 20 and compare the results.

Using the regression-by-constants, the optimal segmentations of orders $K = 1, 2, ..., 10$ are presented in

Fig. 12 A long synthetic time series x_t (*solid line*) and its approximation $\hat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-constants ($\sigma = 0.10$, $K = 4$)
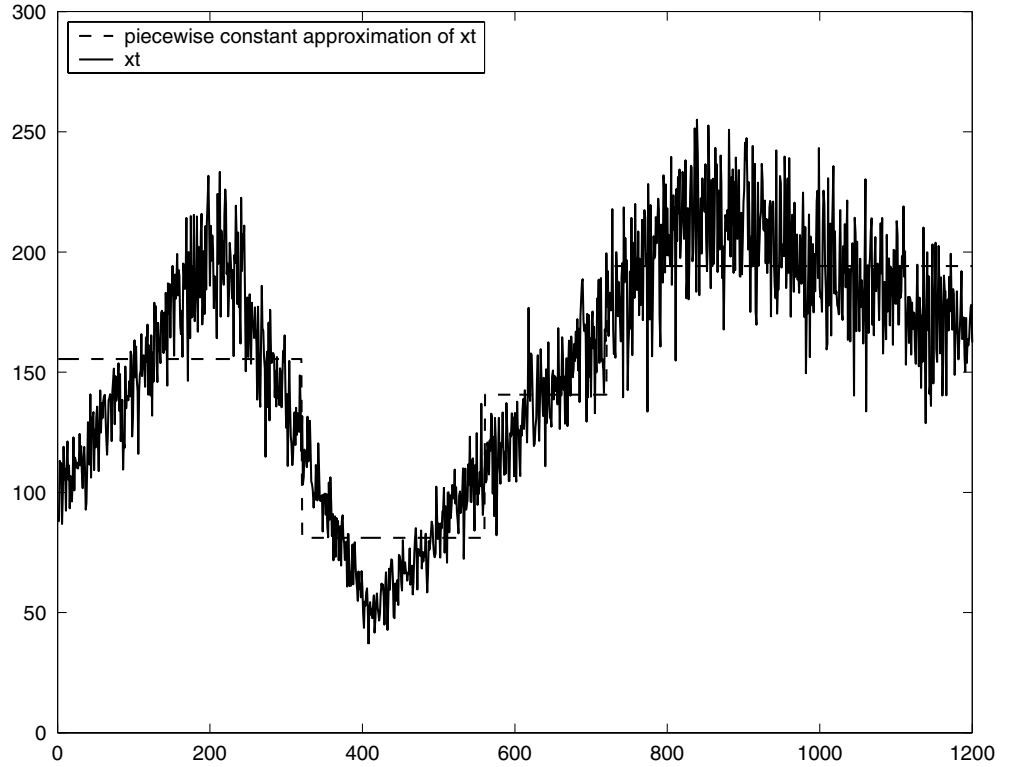
Fig. 13 A long synthetic time series x_t (*solid line*) and its approximation $\hat{x}_t$ (*dotted line*) obtained from segmentation with regression-by-lines ($\sigma = 0.15$, $K = 4$)
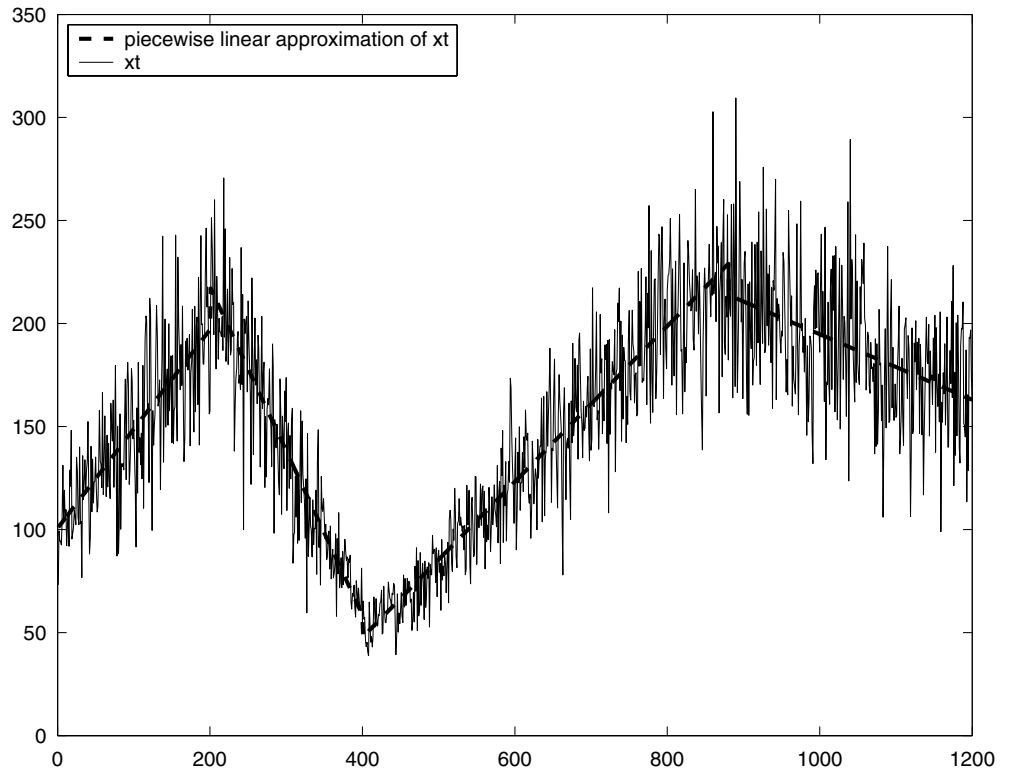
Table. 5a–c (for $N = 10$, 15, 20, respectively). It can be seen that the three N values used give approximately the same segmentations. Hence it appears that block-segmentation does not depend critically on block size N. Using the regression-by-lines, the optimal segmentations of orders $K = 1$, 2, ..., 10 are presented in Table. 6a–c (for $N = 10$, 15, 20, respectively). Again the three N values used give approximately the same segmentations. For all tables, the sequence of 10 segmentations is computed in times ranging from 0.66 to 2.31 s.

Fig. 14 Annual minimum level of the Nile river (*solid line*) and its regression-by-constants approximation (*dotted line*) (segmentation of order $K=25$)
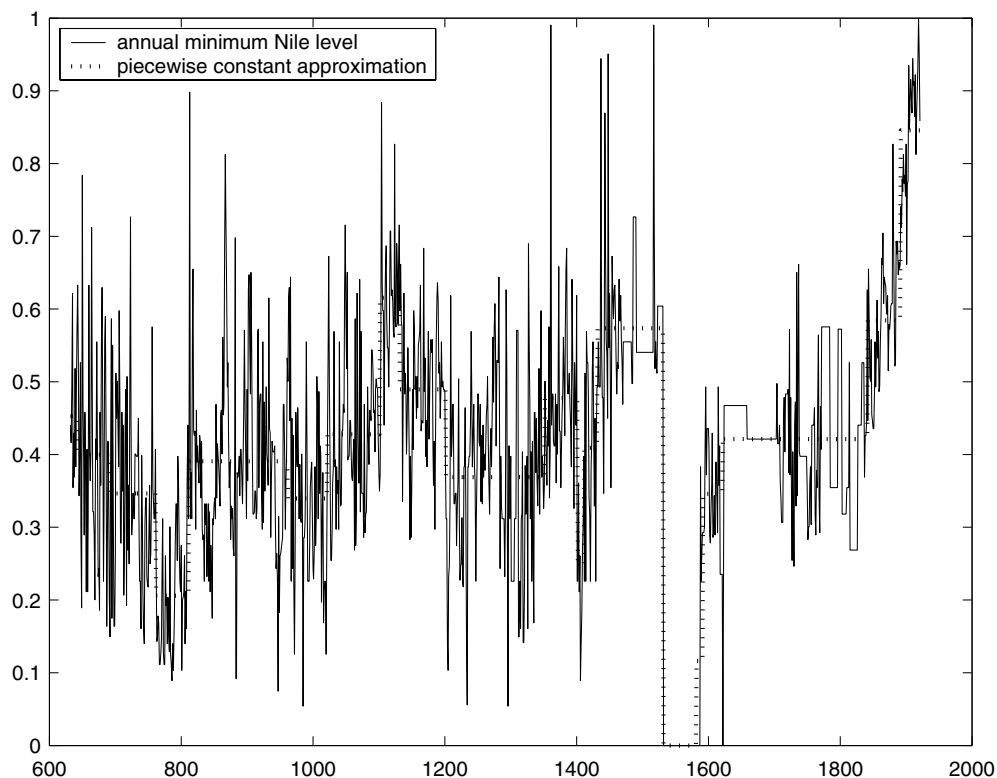
Table 5 Optimal segmentations of orders $K=1,2, ..., 10$ for the annual Nile minimum level time series, as obtained from regression-by-constants

$N=10$											
1	622	1921									
2	622	1852	1921								
3	622	1522	1582	1921							
4	622	1522	1582	1852	1921						
5	622	1422	1522	1582	1852	1921					
6	622	1012	1422	1522	1582	1852	1921				
7	622	1092	1192	1422	1522	1582	1852	1921			
8	622	1092	1192	1422	1522	1582	1832	1882	1921		
9	622	722	802	1092	1192	1422	1522	1582	1852	1921	
10	622	722	802	1092	1192	1422	1522	1582	1832	1882	1921
$N=15$											
1	622	1921									
2	622	1852	1921								
3	622	1522	1582	1921							
4	622	1522	1582	1852	1921						
5	622	1417	1522	1582	1852	1921					
6	622	1012	1432	1522	1582	1852	1921				
7	622	1087	1192	1417	1522	1582	1852	1921			
8	622	1087	1192	1417	1522	1582	1822	1882	1921		
9	622	727	802	1087	1192	1417	1522	1582	1852	1921	
10	622	727	802	1087	1192	1417	1522	1582	1822	1882	1921
$N=20$											
1	622	1921									
2	622	1842	1921								
3	622	1522	1582	1921							
4	622	1422	1522	1582	1921						
5	622	1422	1522	1582	1842	1921					
6	622	842	1422	1522	1582	1842	1921				
7	622	1082	1182	1422	1522	1582	1842	1921			
8	622	842	1082	1182	1422	1522	1582	1842	1921		
9	622	722	802	1082	1182	1422	1522	1582	1842	1921	
10	622	722	802	1082	1182	1422	1522	1582	1842	1882	1921

Note the re-appearance of some change points in segmentations for various K's and N's

Table 6 Optimal segmentations of orders $K = 1, 2, ..., 10$ for the annual Nile minimum level time series, as obtained from regression-by-lines

$N = 10$

1	622	1921									
2	622	1522	1921								
3	622	1522	1792	1921							
4	622	1522	1582	1792	1921						
5	622	1192	1522	1582	1792	1921					
6	622	802	1192	1522	1582	1792	1921				
7	622	802	1092	1192	1522	1582	1792	1921			
8	622	802	1092	1192	1422	1522	1582	1792	1921		
9	622	802	1092	1192	1342	1422	1522	1582	1792	1921	
10	622	802	1092	1192	1282	1392	1432	1522	1582	1792	1921

$N = 15$

1	622	1921									
2	622	1522	1921								
3	622	1522	1792	1921							
4	622	1522	1582	1792	1921						
5	622	1192	1522	1582	1792	1921					
6	622	802	1192	1522	1582	1792	1921				
7	622	802	1087	1312	1522	1582	1792	1921			
8	622	802	1087	1192	1432	1522	1582	1792	1921		
9	622	802	1087	1312	1387	1432	1522	1582	1792	1921	
10	622	802	1087	1192	1282	1387	1432	1522	1582	1792	1921

$N = 20$

1	622	1921									
2	622	1522	1921								
3	622	1522	1582	1921							
4	622	1522	1582	1802	1921						
5	622	802	1522	1582	1802	1921					
6	622	802	1182	1522	1582	1802	1921				
7	622	802	1082	1282	1522	1582	1802	1921			
8	622	802	1082	1342	1422	1522	1582	1802	1921		
9	622	802	1002	1122	1342	1422	1522	1582	1802	1921	
10	622	802	842	1002	1122	1342	1422	1522	1582	1802	1921

Note the re-appearance of some change points in segmentations for various K's and N's. Also compare with Table 5

There are two problems with the results presented in Table. 5 and 6. First, there is considerable discrepancy between the segmentations obtained with regression-by-constants and the ones obtained with regression-by-lines. Second, the BIC criterion fails by overestimating the number of segments, in the same way as in Sect. 4.4. However, from an intuitive point of view, Table. 5 and 6 yield some useful information regarding the structure of the time series. Some years appear in the early segmentations of all tables and persist all the way to high-order segmentations. Such years are, for example, 1522 and 1582. Hence it appears likely that some kind of hydrological change took place around these years, and it is captured by both types of regression and all block sizes. Other years appear in both types of regression but only in high-order segmentations (e.g. 802). These are "somewhat" likely to be genuine hydrological change points. We will further discuss this point in the next section.

Conclusion

We have presented a dynamic programming procedure which can be used for the segmentation of hydrological and environmental time series. Due to the use of recursive computation and block segmentation, the procedure can segment time series with over a thousand samples in a few seconds.

To automatically determine the "correct" number of segments, the procedure must be combined with a segmentation order selection criterion, such as BIC or Scheffe's criterion. However, these criteria do not always give a reliable estimate of the number of segments. Hence the user's judgment and knowledge of a particular domain may be a more useful criterion for the final selection of one of the optimal segmentations obtained by our procedure. User judgment may also be required to select an appropriate regression model (to obtain the function $d_{s,t}$).

Our DP procedure appears to be superior to Hubert's procedure in two respects. First, it is significantly faster and scales up better to long time series. Second, because of the fast execution, it is not necessary to employ pruning; hence, unlike Hubert's procedure, it is guaranteed to find the optimal segmentation (of every given order!). This is an advantage of the DP procedure as compared to the HMM procedure (Kehagias 2004) as well, since the HMM procedure is only guaranteed to find a *locally* optimal segmentation.

On the other hand, the HMM procedure has a "built-in" criterion (namely maximization of likelihood) for segmentation order selection. This is not the case for the

currently proposed DP procedure: automatic selection of segmentation order is a difficult problem for which we do not have a general solution. Note that this is not a problem with our procedure per se (for example, the same problem must be resolved for Hubert's procedure). As remarked in (Hubert 1997, 2000), Scheffe's criterion or BIC can be used to support the user's judgment. Hence it may be more appropriate to consider our DP procedure (as well as Hubert's procedure) as exploratory tools, which can be used to limit the set of candidate segmentations from the exponentially large set of *all* possible segmentations to the set of *optimal* segmentations (optimality being defined in terms of some precise criterion). An additional possibility is to use the DP procedure as an initialization step for the HMM procedure or for more sophisticated approaches, such as Fortin's Bayesian approach (Fortin et al. 2004).

Finally, it is worth emphasizing that our algorithm can be used to obtain an optimal *approximation* of a time series by a piecewise constant(or piecewise linear etc.) function. This becomes apparent by looking at Figs. 3, 4, 5, 6, 7, 8 etc. This is the point of view adopted in the *data mining* literature. In this case the problem of selecting the segmentation order becomes secondary.

Appendix 1

Recursive formulas for segment error

In this appendix we prove the validity of the recursive formulas (Eqs. 8, 9, 10, 11, 12) for the computation of segment error. Recall that $\mathbf{Q}_{s,t} = \mathbf{U}'_{s,t}\mathbf{U}_{s,t}, \mathbf{R}_{s,t} = \mathbf{U}'_{s,t}\mathbf{X}_{s,t}$. We have

$$\mathbf{Q}_{s,t+1} = \mathbf{U}'_{s,t+1}\mathbf{U}_{s,t+1} = \begin{bmatrix} \mathbf{U}'_{s,t} & | & \mathbf{u}'_{t+1} \end{bmatrix} \begin{bmatrix} \mathbf{U}'_{s,t} \\ -- \\ u_{t+1} \end{bmatrix}$$
$$= \mathbf{U}'_{s,t}\mathbf{U}_{s,t} + \mathbf{u}'_{t+1}t_{+1} = \mathbf{Q}_{s,t} + \mathbf{u}'_{t+1}\mathbf{u}_{t+1}.$$
$$(16)$$

Similarly, for $\mathbf{R}_{s,t}$ we obtain

$$\mathbf{R}_{s,t+1} = \mathbf{R}_{s,t} + \mathbf{u}'_{t+1}x_{t+1}. \qquad (17)$$

Furthermore, we have $d_{s,t} = | \mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}|^2$ and

$$d_{s,t+1} = \left|\mathbf{X}_{s,t+1} - \mathbf{U}_{s,t+1}\mathbf{A}_{s,t+1}\right|^2 \\ = \left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t+1}\right|^2 + \left|x_{t+1} - \mathbf{u}_{t+1}\mathbf{A}_{s,t+1}\right|^2. \qquad (18)$$

Now

$$\left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t+1}\right|^2 = \left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t} + \mathbf{U}_{s,t}\mathbf{A}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t+1}\right|^2 \\ = \left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right|^2 + \left|\mathbf{U}_{s,t}\left(\mathbf{A}_{s,t} - \mathbf{A}_{s,t+1}\right)\right|^2 \\ + 2\left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right)'\mathbf{U}_{s,t}\left(\mathbf{A}_{s,t} - \mathbf{A}_{s,t+1}\right)$$

Setting

$$\delta\mathbf{A} = \mathbf{A}_{s,t} - \mathbf{A}_{s,t+1}, \qquad (19)$$

we have

$$\left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t+1}\right|^2 \\ = \left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right|^2 + \left|\mathbf{U}_{s,t}\delta\mathbf{A}\right|^2 + 2\left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right)'\mathbf{U}_{s,t}\delta\mathbf{A}. \qquad (20)$$

Also, we have

$$d_{s,t} = \left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right|^2; \qquad (21)$$

and

$$\left|\mathbf{U}_{s,t}\delta\mathbf{A}\right|^2 \\ = \left(\mathbf{U}_{s,t}\delta\mathbf{A}\right)'\mathbf{U}_{s,t}\delta\mathbf{A} = \delta\mathbf{A}'\mathbf{U}'_{s,t}\mathbf{U}_{s,t}\delta\mathbf{A} = \delta\mathbf{A}'\mathbf{Q}_{s,t}\delta\mathbf{A}; \qquad (22)$$

and

$$\left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right)'\mathbf{U}_{s,t}\delta\mathbf{A} = \left(\mathbf{U}'_{s,t}\mathbf{X}_{s,t} - \mathbf{U}'_{s,t}\mathbf{U}_{s,t}\mathbf{A}_{s,t}\right)'\delta\mathbf{A} \\ = \left(\mathbf{U}'_{s,t}\mathbf{X}_{s,t} - \mathbf{U}'_{s,t}\mathbf{U}_{s,t}\left(\mathbf{U}'_{s,t}\mathbf{U}_{s,t}\right)^{-1}\mathbf{U}'_{s,t}\mathbf{X}_{s,t}\right)'\delta\mathbf{A} = 0. \qquad (23)$$

Replacing Eqs. 21, 22, 23 in Eq. 20 we get

$$\left|\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t+1}\right|^2 = d_{s,t} + \delta\mathbf{A}'\mathbf{Q}_{s,t}\delta\mathbf{A} \qquad (24)$$

and replacing Eq. 24 in Eq. 18 we get

$$d_{s,t+1} = d_{s,t} + \delta\mathbf{A}'\mathbf{Q}_{s,t}\delta\mathbf{A} + \left|x_{t+1} - \mathbf{u}_{t+1}\mathbf{A}_{s,t+1}\right|^2. \qquad (25)$$

Equations 16, 17, 19, 25 constitute the recursion presented in Sect. 3.2. The initialization $\mathbf{Q}_{s,s} = \mathbf{u}'_s\mathbf{u}_s$, $\mathbf{R}_{s,s} = \mathbf{u}'_s x_s$, $\mathbf{A}_{s,s} = \mathbf{Q}_{s,s}^{-1}\mathbf{R}_{s,s}$ follows from the definition of $\mathbf{Q}_{s,t}, \mathbf{R}_{s,t}, \mathbf{A}_{s,t}$.

Appendix 2

Block segmentation

In this appendix we present the modifications necessary for *block segmentation*. In block segmentation we assume that segment boundaries can only appear at constant multiples of a number N, which we call *block length*. For example, if $T = 1,000$ and $N = 10$, the only candidate boundaries are 0, 10, 20, ..., 1,000; this reduces computation by a factor of 100, at the cost of obtaining a "coarse" segmentation.

The definition of $d_{s,t}$ must be modified; now it is the error of segment $[(s-1)\cdot N + 1, t\cdot N]$. As far as computa-

tion is concerned, it is only required to change the input and output regression variables; $\mathbf{u}_t$ remains unchanged, but we now define

$$\mathbf{X}_{s,t} = \begin{bmatrix} x_{(s-1)\cdot N+1} \\ x_{(s-1)\cdot N+2} \\ \dots \\ x_{t\cdot N} \end{bmatrix}, \quad \widehat{\mathbf{X}}_{s,t} = \begin{bmatrix} \widehat{x}_{(s-1)\cdot N+1} \\ \widehat{x}_{(s-1)\cdot N+2} \\ \dots \\ \widehat{x}_{t\cdot N} \end{bmatrix},$$

$$\mathbf{U}_{s,t} = \begin{bmatrix} \mathbf{u}_{(s-1)\cdot N+1+1} \\ \mathbf{u}_{(s-1)\cdot N+1+2} \\ \dots \\ \mathbf{u}_{t\cdot N} \end{bmatrix}; \tag{26}$$

compare with Eq. 6. The formulas for $d_{s,t}$ and $\mathbf{A}_{s,t}$ also remain unchanged:

$$d_{s,t} = \left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right)'\left(\mathbf{X}_{s,t} - \mathbf{U}_{s,t}\mathbf{A}_{s,t}\right),$$
$$\mathbf{A}_{s,t} = \left(\mathbf{U}'_{s,t}\mathbf{U}_{s,t}\right)^{-1}\mathbf{U}'_{s,t}\mathbf{X}_{s,t}.$$

Finally, the recursive formula for $d_{s,t}$ is modified as follows

$$d_{s,t+1} = d_{s,t} + \delta\mathbf{A}'\mathbf{Q}_{s,t}\delta\mathbf{A} + \left|\mathbf{X}_{t,t+1} - \mathbf{U}_{t,t+1}\mathbf{A}_{s,t+1}\right|^2. \tag{27}$$

The segmentation algorithm of Sect. 3.2 remains unchanged, but now uses as input the newly defined *block* segment error $d_{s,t}$.

Appendix 3

Beeferman's segmentation metric P_k

Beeferman's segmentation metric $P_k(\mathbf{s}, \mathbf{t})$ measures the error of a proposed segmentation $\mathbf{s} = (0, s_1, \dots, s_{K-1}, t)$ with respect to a "true" segmentation $\mathbf{t} = (0, t_1, \dots, t_{L-1}, t)$. P_k first appeared in the *text* segmentation literature (Beeferman et al. 1999) but it can be applied to any segmentation problem. Here we will only give a short intuitive discussion of P_k. The interested reader can find more details in (Beeferman et al. 1999).

When is $\mathbf{s}$ identical to $\mathbf{t}$? The following two conditions are necessary and sufficient.

1. Each pair of integers (i, j) which is in the same segment under $\mathbf{t}$ is also in the same segment under $\mathbf{s}$.
2. Each pair (i, j) in a different segment under $\mathbf{t}$ is also in a different segment under $\mathbf{s}$.

It follows that the difference between $\mathbf{s}$ and $\mathbf{t}$ (i.e. the *error* of $\mathbf{s}$ with respect to $\mathbf{t}$) is the amount by which the above criteria are violated. This can happen in two ways:

1. Some pair (i, j) which is in the same segment under $\mathbf{t}$, is in a different segment under $\mathbf{s}$;
2. Or some pair (i, j) which is in the same segment under $\mathbf{t}$, is in a different segment under $\mathbf{s}$.

We can formalize the above description as follows. Define a function $\delta_{\mathbf{s}}(i, j)$ to be 1 when i and j are in the same segment under $\mathbf{s}$ and 0 otherwise; define $\delta_{\mathbf{t}}(i, j)$ similarly. Then we are interested in the quantity

$$\sum_{i=1}^{T}\sum_{j=1}^{T}|\delta_{\mathbf{s}}(i, j) - \delta_{\mathbf{t}}(i, j)|. \tag{28}$$

There are two problems with Eq. 28 and they both have to do with the range of the summations. First, considering *all* possible pairs (i, j) is too time consuming. Second, it yields an *undiscriminating* criterion, because even a grossly wrong $\mathbf{s}$ will agree with $\mathbf{t}$ on many pairs (for example most pairs $(i, i+1)$ will be placed in the same segment and most (i, j) pairs which are very far apart will be placed in different segments). Beeferman et al. propose to consider only pairs which are k steps apart $(i, i+k+1)$, where k is *half the average segment length*. It has been empirically validated that this choice of k works well; in (Beeferman et al. 1999) Beeferman et al. also discuss some theoretical justification for this choice. Hence P_k is defined as follows

$$P_k(\mathbf{s}, \mathbf{t}) = \sum_{i=1}^{T-k-1}|\delta_{\mathbf{s}}(i, i+k+1) - \delta_{\mathbf{t}}(i, i+k+1)| \tag{29}$$

and this is what we have used in Sect. 4.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Control 19:716–723

Auger IE, Lawrence CE (1989) Algorithms for the optimal identification of segment neighborhoods. Bull Math Biol 51:39–54

Basseville M, Nikiforov IV (1993) Detection of abrupt changes—theory and application. Prentice-Hall, Englewood Cliffs

Beeferman D, Berger A, Lafferty J (1999) Statistical models for text segmentation. Machine learning 34:177–210

Bellman R (1961) On the approximation of curves by line segments using dynamic programming. Commun ACM 4:284

Braun JV, Mueller H-G (1998) Statistical methods for DNA sequence segmentation. Stat Sci 13:142–162

Braun JV, Braun RK, Mueller H-G (2000) Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. Biometrika 87:301–314

Chen H-L (2002) Testing hydrologic time series for stationarity. J Hydr Eng 7:129–136

Chen J, Gupta AK (1998) Information criterion and change point problem for regular models. Tech. Rep. No. 98-05, Department of Math. and Stat., Bowling Green State U., Ohio

Fortin V, Perreault L, Salas JD (2004) Retrospective analysis and forecasting of streamflows using a shifting level model. J Hydrol 296:135–163

Fortin V, Perreault L, Salas JD (2004) Analyse retrospective et prevision des debits en presence de changement de regime. 57e Congres Annuel de l' Association Canadienne des Ressources Hydriques

Himberg J, Korpiaho K, Mannila H, Tikanmaki J, Toivonen HTT (2001) Time series segmentation for context recognition in mobile devices. Proc. of ICDM 2001, pp 203–210

Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems. Elsevier, Amsterdam

Hubert P (1997) Change points in meteorological analysis. In: Subba Rao T, Priestley MB, Lessi O (eds) Applications of time series analysis in astronomy and meteorology. Chapman and Hall, London

Hubert P (2000) The segmentation procedure as a tool for discrete modeling of hydrometeorogical regimes. Stoch Env Res Risk Ass 14:297–304

Jackson B et al (2005) An algorithm for optimal partitioning of data on an interval. IEEE Signal Proc Lett 12:105–108

Kay SM (1998) Fundamentals of statistical signal processing: detection theory. Prentice-Hall, Englewood Cliffs

Kearns M, Mansour Y, Ng AY, Ron D (1997) An experimental and theoretical comparison of model selection methods. Machine Learn 27:7–50

Kehagias Ath (2004) A hidden Markov model segmentation procedure for hydrological and environmental time series. Stoch Environ Res Risk Assess 18:117–130

Keogh E et al (2001) An online algorithm for segmenting time series. In: Proceedings of the 2001 IEEE International Conference on Data Mining, pp 289–296

Keogh E, Chu S, Hart D, Pazzani M (2003) Segmenting time series: a survey and novel approach, In: Data mining in time series databases. World Scientific Publishing Company, Singapore

Mann ME, Bradley RS, Hughes MK (1999) Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. Geophys Res Lett 26:759–762

Ninomiya Y (2003) Asymptotic theory for change-point model and its application to model selection. Tech. Report, Graduate School of Mathematics, Kyusho Univ., Japan

Pavlidis T, Horowitz SL (1974) Segmentation of Plane Curves. IEEE Trans Comput 23:860–870

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Sveinsson OGB, Salas JD, Boes DC, Pielke RA Sr (2003) Modeling the dynamics of long term variability of hydroclimatic processes. J Hydrometeorol 4:489–505

Yao Y-C (1988) Estimating the number of change-points via Schwarz' criterion. Stat Prob Lett 6:181–189