# A Hidden Markov Model Segmentation Procedure for Hydrological and Enviromental Time Series

Ath. Kehagias

Faculty of Engineering, Box 464

Division of Mathematics, Dept. of Math., Phys. and Comp. Sciences

Aristotle University of Thessaloniki, 54124 Thessaloniki, GREECE

email: kehagias@egnatia.ee.auth.gr

Phone no. 3031-995944, Fax no. 3031-414347

March 12, 2003

**Abstract**

In this paper we present a procedure for the segmentation of hydrological and enviromental time series. We consider the segmentation problem from a purely computational point of view which involves the minimization of Hubert's segmentation cost; in addition this *least squares* segmentation is equivalent to *Maximum Likelihood* segmentation. Our segmentation procedure maximizes Likelihood and minimizes Hubert's least squares criterion using a hidden Markov model (HMM) segmentation algorithm. This algorithm is guaranteed to achieve a local maximum of the Likelihood. We evaluate the segmentation procedure with numerical experiments which involve artificial, temperature and river discharge time series. In all experiments, the procedure actually achieves the global minimum of the Likelihood; furthermore execution time is only a few seconds, even for time series with over a thousand terms.

**Keywords:** Hidden Markov model, time series, segmentation, Maximum Likelihood, river discharge.

## 1 Introduction

In this paper we consider the following *time series segmentation* problem: a given time series must be divided into several *segments* (i.e. blocks of contiguous data) so that each segment is homogeneous,

while contiguous segments are heterogeneous (with homogeneity being defined in terms of some appropriate segment statistics). This problem falls within the framework of *change point detection and estimation*. Such problems appear often in hydrology and environmetrics. For example, in climate change studies it is sometimes required to test a river flow, rainfall or temperature time series for sudden changes of its mean value.

Our starting point is the use of a *hidden Markov model* (HMM) to formulate time series segmentation as a maximum likelihood (ML) problem. We use the HMM formulation to derive a procedure which yields the "optimal segmentation". Optimality can be understood in two senses. From the *probabilistic* point of view, our procedure yields the Maximum Likelihood (ML) segmentation. On the other hand, from a purely *numerical* point of view, our procedure minimizes the total square distance between within-segment samples and the corresponding segment means. This least-squares segmentation criterion has been previously used by Hubert in [17, 18].

Indeed, the major goal of the current paper is to improve (with respect to speed of execution and to the length of time series examined) Hubert's procedure for the segmentation of time series with multiple change points. Our HMM-based procedure is quite fast and can handle longer time series than the ones treated by Hubert (specifically, it segments time series with over a thousand samples in a few seconds). Let us stress at this point that our main concern is computational efficiency, *not* hydrological realism. In other words, the HMM should be understood as a computational aid, not as a physically plausible model.

Many types of hidden Markov models appear in the literature of pattern recognition, engineering, econometrics, biology and also in the hydrological literature. The particular type of HMM used in this paper is a pair of stochastic processes with the following properties.

1. The *unobservable* ("hidden") state process is Markovian and can take a finite number of values: 1, 2, ..., $K$.

2. At every time step, the state process can either remain unchanged or increase by one.

3. At every time step the *observable* process generates a sample from a normal distribution with mean value depending on the current state.

Now assume that the time series $x_1, x_2, ..., x_T$ is a realization of the observable process, and correspond one segment to every time interval during which the state process does not change value. Under

this correspondence, the segmentation problem is reduced to estimating the underlying state sequence $z_1, z_2, ..., z_T$.

Many variations of HMM's appear in the literature, but the abovementioned connection between state estimation and time series segmentation is always valid. Indeed, the first major application of HMM's was in speech recognition, where the goal was to divide a speech waveform into segments, each segment corresponding to a *phoneme*. Early papers such as [2, 3, 4, 28] deal with discrete valued time series; extensions to continuous valued time series have also been used (for instance [22, 23, 24]). An often cited review of HMM's is [36]; a more recent review is [5].

Let us now turn to the hydrological segmentation literature. An extensive introduction to *sequential* segmentation methods appears in [15, pp.655-733]. Regarding *nonsequential* methods, important early papers are [27], [9] and [7, 8]. Some examples of recent work include [16, 26] and [31, 40] and (from the Bayesian point of view) [32, 33, 34, 35] and [37] (these are just a few samples of the *very extensive* literature).

The references of the previous paragraph deal with a *single* change point. As already mentioned, a computational segmentation procedure which can handle *multiple* change points has been presented by Hubert [17, 18]. As far as we know there are few other *explicit* references to the multiple change point problem in the hydrological literature.

Many hydrologists have used HMM's as *realistic* models of hydrological processes. A recent and extensive review of stochastic models of climate time series is [41] which cites many papers that use HMM's. In fact, an early model of this type appears in [38]; this model can handle multiple change points, is very much related to our approach and has been used in a Bayesian setting in [13] (the possibility of using HMM's for hydrological time series segmentation was mentioned earlier in [35]). A related approach appears in [42, 43, 44]. Other HMM-based models appear in [1], [14], [19], [20, 21] and (in combination with AR models) in [29, 45]. According to the previously mentioned association of segments with HMM states, these papers are implicitly connected to multiple change point segmentation. However, the main goal of these papers is modelling the precipitation process, predicting rain levels etc. Hence their point of view is rather different from ours.

This paper is organized as follows. In Section 2 Hubert's formulation of the time series segmentation problem is reviewed. In Section 3 the segmentation problem is formulated in terms of hidden Markov models and Maximum Likelihood estimation. The segmentation procedure is presented in 4. Some

segmentation experiments are presented in Section 5. In Section 6 our results are summarized and some future research directions proposed. Finally, some technical points are presented in two Appendices.

## 2 Segmentation as Optimization

The formulation and notation introduced in this section is essentially the one used in [17, 18]. Given a time series $\mathbf{x} = (x_1, x_2, \ldots, x_T)$, a *segmentation* is a sequence of times $\mathbf{t} = (t_0, t_1, \ldots, t_K)$ which satisfy $0 = t_0 < t_1 < \ldots < t_{K-1} < t_K = T$. 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 $t_0, t_1, \ldots, t_K$ are called *change points* and $K$, the number of segments, is called the *order* of the segmentation. The *cost* of segmentation $\mathbf{t} = (t_0, ..., t_K)$ is defined by

$$D_K(\mathbf{t}) = \sum_{k=1}^{K} \sum_{t=t_{k-1}+1}^{t_k} (x_t - \widehat{\mu}_k)^2 . \tag{1}$$

where

$$T_k = t_k - t_{k-1}, \qquad \widehat{\mu}_k = \frac{\sum_{t=t_{k-1}+1}^{t_k} x_t}{T_k}, \quad k = 1, 2, ..., K. \tag{2}$$

When $D_K$ has a small value, the segments are homogeneous in the sense that the $x_t$'s are close to $\widehat{\mu}_k$ (for $k = 1, 2, ..., K$ and $t = t_{k-1} + 1, ..., t_k$ ). The minimum segmentation cost is denoted by $\widehat{D}_K = D_K(\widehat{\mathbf{t}})$, where $\widehat{\mathbf{t}}$ is the optimal $K$-th order segmentation. As observed in [17] we have

$$\widehat{D}_1 \geq \widehat{D}_2 \geq ... \geq \widehat{D}_T = 0 \tag{3}$$

(in fact there is only one segmentation of order $T$ , with every segment including a single time step).

In [17] it is noted that the number of possible segmentations grows exponentially with $T$. Hubert uses a *branch-and-bound* approach to search efficiently the set of all possible segmentations. In [18, 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.

# 3 Hidden Markov Models and Maximum Likelihood Segmentation

Now the time series segmentation problem will be formulated as a problem of *Maximum Likelihood (ML) estimation.* The connection of this to Hubert's approach will be discussed in Section 4.3.3.

We represent a time series with change points as a *hidden Markov model (HMM).* The term "hidden Markov model" denotes a broad class of stochastic processes; here a particular and somewhat restricted type of HMM is used, as illustrated by the following example.

The annual flow of a river is denoted by $X_t$. Assume that, for the years $t = 1, 2, ..., t_1$, $X_t$ is a normally distributed random variable with mean $\mu_1$ and standard deviation $\sigma$. In year $t_1$ a *transition* takes place and, for the years $t = t_1 + 1,\ t_1 + 2, ..., t_2$, $X_t$ is normally distributed with mean $\mu_2$ and standard deviation $\sigma$. This process continues with transitions taking place in years $t_2$, $t_3$, ... , $t_{K-1}$. This process is illustrated in Figure 1. The (unobservable) *states* of the river flow are indicated by circles and the possible transitions from state to state by arrows; the observable time series is indicated by the double arrows emanating from the states.

**Figure 1 to appear here**

The above mechanism (which can be applied not only to river flows, but to a variety of time series) can be described by a pair of stochastic processes $(Z_t, X_t)$ (with $t = 0, 1, 2, ...$) defined as follows.

1. The *state process* $Z_t$ is a finite state Markov chain with $K$ states; it has initial probability vector $\pi$ and transition probability matrix $P$. We assume that $Z_0 = 1$ with certainty. Hence, for any $T$, the joint probability function of $Z_1, Z_2, ..., Z_T$ is

$$\Pr(Z_1 = z_1, Z_2 = z_2, ..., Z_T = z_T) = \pi_{z_0} \cdot P_{z_0,z_1} \cdot P_{z_1,z_2} \cdot ... \cdot P_{z_{T-1},z_T}, \tag{4}$$

   where $\pi_1 = 1$, $\pi_k = 0$ for $k = 2, 3, ..., K$ and $P_{k,j} = 0$ for $k = 1, 2, ..., K$ and all $j$ other than $k$, $k + 1$. The parameters of this process are $K$ and $P$.

2. The *observation process* $X_t$ is a sequence of *conditionally independent,* normally distributed random variables with mean $\mu_{Z_t}$ and standard deviation $\sigma$. More precisely, for every $t$, the joint

probability density of $X_1, X_2, ..., X_t$ conditioned on $Z_1, Z_2, ..., Z_t$ is

$$f_{X_1,X_2,...,X_t|Z_1,Z_2,...,Z_t}(x_1, x_2, ..., x_t|z_1, z_2, ..., z_t) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^t} \prod_{i=1}^{t} e^{-\left(x_i-\mu_{z_i}\right)^2/2\sigma^2}. \tag{5}$$

The parameters of this process are $\mu_1$, $\mu_2$, ... , $\mu_K$ and $\sigma$. We will use the notation $\mathbf{M} = [\mu_1, \mu_2, ... , \mu_K ]$.

The $(Z_t, X_t)$ pair presented above and used throughout this paper is a *left-to-right continuous HMM* [36]. "Left-to-right" refers to the structure of state transitions (as depicted in Figure 1) and "continuous" refers to the fact that the observation process is continuous-valued. The model parameters are $K$, $P$, $\mathbf{M}$, $\sigma$.

There is a one-to-one correspondence between state sequences $\mathbf{z} = (z_1, z_2, ... , z_T)$ and segmentations $\mathbf{t} = (t_0, t_1, ..., t_{K'})$. Given a $\mathbf{z}$, the corresponding $\mathbf{t}$ is obtained by locating the times $t_k$ such that $z_{t_k} \neq z_{t_k+1}$, for $k = 1, 2, ..., K' - 1$ (and $t_0 = 0$, $t_{K'} = T$). The Markov chain described above allows only left-to-right transitions, hence $K' \leq K$, i.e. there will be *at most* $K$ segments, and every segment will be uniquely associated with a state.

Assuming that the observations $\mathbf{x} = (x_1, x_2, ... , x_T)$ are generated by a HMM (specified by $K$, $P$, $\mathbf{M}$, $\sigma$) the likelihood (i.e. the conditional probability) of every state sequence $\mathbf{z}$ (given $\mathbf{x}$) can be computed. In this manner, the ML state sequence $\widehat{\mathbf{z}}$ and the ML segmentation $\widehat{\mathbf{t}}$ can be obtained.

Some additional notation is required. The *conditional likelihood* of a state sequence $\mathbf{z}$ (given an observation sequence $\mathbf{x}$) is denoted by $L^1_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \sigma)$ or, equivalently, by

$$L^1_{K,T}(z_1, z_2, ..., z_T|x_1, x_2, ..., x_T; P, \mathbf{M}, \sigma) \tag{6}$$

and the *joint likelihood* of a state sequence $\mathbf{z}$ *and* an observation sequence $\mathbf{x}$ is denoted by $L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \sigma)$ or, equivalently, by

$$L^2_{K,T}(z_1, z_2, ..., z_T, x_1, x_2, ..., x_T; P, \mathbf{M}, \sigma). \tag{7}$$

$L^1_{K,T}$ and $L^2_{K,T}$ are understood as functions of $\mathbf{z} = (z_1, z_2, ..., z_T)$; the observations $\mathbf{x} = (x_1, x_2, ..., x_T)$, the number of segments $K$, and the length of the time series $T$, as well as the parameters $P$, $\mathbf{M}$, $\sigma$ are

assumed *fixed*. In place of $T$ any $t$ can be used; for instance

$$L^2_{K,t}(z_1, z_2, ..., z_t, x_1, x_2, ..., x_t; P, \mathbf{M}, \sigma). \tag{8}$$

denotes the joint likelihood of the subsequences $(z_1, z_2, ..., z_t)$, $(x_1, x_2, ..., x_t)$ etc. Finally note that, from (4), (5) follows

$$L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^T} \prod_{t=1}^{T} \left( P_{z_{t-1}, z_t} \cdot e^{-\left(x_t - \mu_{z_t}\right)^2 / 2\sigma^2} \right), \tag{9}$$

where $z_0 = 1$, according to the previously stated assumption.

Given a time series $\mathbf{x}$ , the ML state sequence $\widehat{\mathbf{z}}$ is the one which maximizes $L^1_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \sigma)$ as a function of $\mathbf{z}$. It is easy to see that

$$L^1_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \sigma) = \frac{L^2_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \sigma)}{A\left(\mathbf{x}, K, T, P, \mathbf{M}, \sigma\right)} \tag{10}$$

where $A\left(\mathbf{x}, K, T, P, \mathbf{M}, \sigma\right)$ is the marginal density of $x_1$, $x_2$, ... , $x_T$ and is independent of $\mathbf{z}$. Hence $\widehat{\mathbf{z}}$ also maximizes $L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \sigma)$ (as a function of $\mathbf{z}$ only!).

# 4 The Segmentation Procedure

To obtain the ML estimation of $\mathbf{z}$ discussed above, two problems must be solved. First, assuming that the parameters $K$, $P$, $\mathbf{M}$, $\sigma$ are known, an efficient algorithm is required to compute $\widehat{\mathbf{z}}$; this is the *state estimation* step. Second, since usually $K$, $P$, $\mathbf{M}$, $\sigma$ will be unknown, a method for their estimation is required; this is the *parameter estimation* step. The standard approach used in HMM problems is to perform a parameter estimation step followed by a state estimation step, and to repeat this process until convergence.

## 4.1 State Estimation

Assume first that the observations $\mathbf{x}$ and the parameters $K$, $P$, $\mathbf{M}$, $\sigma$ are fixed. Then the $\widehat{\mathbf{z}} = (\widehat{z}_1, \widehat{z}_2, ..., \widehat{z}_T)$ which maximizes $L^1_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \sigma)$ as a function of $\mathbf{z}$ can be found by the *Viterbi algo-*

*rithm* [12], a computationally efficient dynamic programming approach. From (10) follows

$$(\widehat{z}_1, \widehat{z}_2, ..., \widehat{z}_T) = \arg \max_{z_1, z_2, ..., z_T} L^2_{K,T}(z_1, z_2, ..., z_T, x_1, x_2, ..., x_T; P, \mathbf{M}, \sigma). \tag{11}$$

Define

$$q_{k,t} = \max_{z_1, z_2, ..., z_{t-1}} L^2_{K,t}(z_1, z_2, ..., z_{t-1}, k, x_1, x_2, ..., x_t; P, \mathbf{M}, \sigma), \quad t = 1, 2, ..., T, \quad k = 1, 2, ..., K \tag{12}$$

By standard dynamic programming arguments [6], both $\widehat{\mathbf{z}} = (\widehat{z}_1, \widehat{z}_2, ..., \widehat{z}_T)$ and the $q_{k,t}$'s of (12) can be computed recursively by the following algorithm; it takes as input the time series $x_1, x_2, ..., x_T$ and the parameters $K$, $P$, $\mathbf{M}$ and $\sigma$ and produces as output the optimal segmentation $\widehat{\mathbf{z}} = (\widehat{z}_1, \widehat{z}_2, ..., \widehat{z}_T)$ (with $x_1, x_2, ..., x_T$ and $K$, $P$, $\mathbf{M}$, $\sigma$ given).

---

**Viterbi Algorithm**

**Forward Recursion**

Set $q_{1,0} = 1$, $q_{2,0} = q_{3,0} = ... = q_{k,0} = 0$.

For $t = 1, 2, ..., T$

For $k = 1, 2, ..., K$

$$q_{k,t} = \max_{1 \le j \le K} \left( q_{j,t-1} \cdot P_{j,k} \cdot \frac{e^{-(x_t - \mu_k)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma} \right)$$

$$r_{k,t} = \arg \max_{1 \le j \le K} \left( q_{j,t-1} \cdot P_{j,k} \cdot \frac{e^{-(x_t - \mu_k)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma} \right).$$

End

End

**Backward Recursion**

$\widehat{L}^2_{K,T} = \max_{1 \le k \le K} (q_{k,T})$

$\widehat{z}_T = \arg\max_{1 \le k \le K} (q_{k,T})$.

For $t = T, T-1, ..., 2$

$$\widehat{z}_{t-1} = r_{\widehat{z}_t, t}.$$

End

---

$\widehat{L}^2_{K,T}$ (the maximum value of $L^2_{K,T}$, given $x_1, x_2, ..., x_T$ and $K$, $P$, $\mathbf{M}$ and $\sigma$) is obtained upon completion of the forward recursion; the state sequence $\widehat{\mathbf{z}}$ which achieves $\widehat{L}^2_{K,T}$ is obtained by the backward recursion. According to the previous remarks, $\widehat{\mathbf{z}}$ also maximizes $L^1_{K,T}$. The execution time of the algorithm has order $\mathrm{O}(T \cdot K^2)$, i.e. it is *linear* (rather than exponential) in the length of the time series $T$. This makes the algorithm computationally feasible even for long time series.

### 4.2 Parameter Estimation

Next, we turn to the problem of estimating the parameters $P$, $\mathbf{M}$, $\sigma$, assuming that a state sequence $\mathbf{z}$ (and the corresponding segmentation $\mathbf{t} = (t_0, t_1, ..., t_K)$) is given. A reasonable estimate for the components of $\mathbf{M} = [\mu_1, \mu_2, ..., \mu_K]$ is:

$$\widehat{\mu}_k = \frac{\sum_{t=t_{k-1}+1}^{t_k} x_t}{T_k}, \qquad k = 1, 2, ..., K. \tag{13}$$

One could estimate a separate $\sigma_k$ for each segment as follows:

$$\widehat{\sigma}_k = \sqrt{\frac{\sum_{t=t_{k-1}+1}^{t_k} (x_t - \widehat{\mu}_k)^2}{T_k - 1}}, \qquad k = 1, 2, ..., K. \tag{14}$$

A simpler approach (which maintains compatibility with Hubert's procedure) is to assume that

$$\sigma_1 = \sigma_2 = ... = \sigma_K = \sigma \tag{15}$$

and use the following estimate:

$$\widehat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T} (x_t - \widehat{\mu})^2}{T - 1}} = \sqrt{\frac{\sum_{k=1}^{K} \sum_{t=t_{k-1}+1}^{t_k} (x_t - \widehat{\mu})^2}{T - 1}}, \tag{16}$$

where

$$\widehat{\mu} = \frac{\sum_{t=1}^{T} x_t}{T}. \tag{17}$$

Note that the estimate (16) is *segmentation-independent*, while the estimate (13) is *segmentation-dependent*. The practical implication of this observation is that $\widehat{\sigma}$ is estimated once, while the $\widehat{\mu}_k$'s must be reestimated every time a new estimate $\widehat{\mathbf{z}}$ of the state sequence is computed.

We now turn to the estimation of the transition probability matrix $P$. The transition probabilities are important parameters of the HMM because they control the length of the segments. In a left-to-right HMM one has $P_{k,k+1} = 1 - P_{k,k}$ (for $k = 1, 2, ..., K-1$), $P_{K,K} = 1$ and $P_{k,j} = 0$ (for $j$ different from $k$ and $k+1$). Hence $P$ has $K-1$ free parameters, namely $P_{1,1}$, $P_{2,2}$, ... , $P_{K-1,K-1}$. Several estimates of the $P_{k,k}$ are possible.

1. The simplest possible approach is to assume equal transition probabilities:

$$P_{1,1} = P_{2,2} = ... = P_{K-1,K-1} = p \tag{18}$$

   and (*assuming that every state is traversed at least once*) estimate $p$ by

$$\widehat{p} = \frac{T-K}{T}. \tag{19}$$

   In other words, $\widehat{p}$ is the number of time steps during which $z_t$ did not change, divided by the total number of time steps.

2. A more sophisticated estimate of $p$ makes use of the so-called *forward* probabilities $a_{k,t}$ and the *backward* probabilities $\beta_{k,t}$ to obtain a *recursive* estimate $p^{(i+1)}$ in terms of a previous estimate $p^{(i)}$, as follows

$$p^{(i+1)} = \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{k,t} \cdot p^{(i)} \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2} \cdot \beta_{k,t+1}}{T}. \tag{20}$$

   The significance of the $\alpha$ and $\beta$ variables appearing in (20) is explained in Appendix A; they

depend on the previous estimate $p^{(i)}$ and are computed by

$$\alpha_{1,1} = 0, \quad \alpha_{k,1} = 0, \quad k = 2, 3, ..., K; \tag{21}$$

$$\alpha_{k,t+1} = \sum_{n=1}^{K} \alpha_{n,t} P_{n,k}^{(i)} \cdot \frac{e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma}, \qquad k = 1, 2, ..., K, \quad t = 1, 2, ..., T-1; \tag{22}$$

$$\beta_{k,t} = 1, \quad k = 1, 2, ..., K; \tag{23}$$

$$\beta_{k,t} = \sum_{n=1}^{K} P_{k,n}^{(i)} \cdot \frac{e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma} \cdot \beta_{n,t+1}, \qquad k = 1, 2, ..., K, \quad t = T-1, T-2, ..., 1; \tag{24}$$

(where $P_{k,k}^{(i)} = p^{(i)}$, $P_{k,k+1}^{(i)} = 1 - p^{(i)}$ and all the remaining $P_{k,n}^{(i)}$ equal zero).

3. It is possible to assume that every state has a different transition probability: $P_{k,k} = p_k$. In this case a simple estimate is

$$\widehat{p}_k = \frac{T_k - 1}{T_k}, \tag{25}$$

where $\mathbf{t} = (t_0, t_1, ..., t_K)$ is the segmentation associated with the state sequence $\mathbf{z}$ and $T_k = t_k - t_{k-1}$.

4. $\alpha, \beta$ estimates of the $p_k$'s can also be used, but we did not use this approach here.

At any rate, in the experiments of Section 5 we have found that using any one of the estimates (19), (20) and (25) gives practically identical segmentations. Note that (20) and (25) are segmentation-dependent, while (19) is segmentation-independent (*provided that the corresponding state sequence traverses all states*).

The important issue of selecting $K$ will be discussed in Section 4.4.

### 4.3 HMM Segmentation with Fixed Number of Segments

Now we combine the parameter and state estimation steps into a HMM segmentation algorithm. We first present a "basic" version of the HMM algorithm and then discuss some variants.

#### 4.3.1 The Basic Algorithm

In this algorithm we assume all transition probabilities to be equal (hence $P$ is determined by $p$). The algorithm works as follows. The input is the time series $\mathbf{x} = (x_1, x_2, ..., x_T)$, the number of segments

$K$ and a termination variable $\varepsilon$. An initial state sequence is chosen randomly and used to estimate the HMM parameters. Then, in every iteration of the algorithm, the HMM parameters are estimated from the current state sequence and a new state sequence is computed from the newly estimated HMM parameters. The algorithm converges to an *optimal* state sequence $\widehat{\mathbf{z}}$ (and a corresponding segmentation $\widehat{\mathbf{t}}$); convergence and optimality will be discussed more extensively in Section 4.3.3. The details of the algorithm are as follows.

---

**Basic HMM Segmentation Algorithm**

Choose randomly a state sequence $\mathbf{z}^{(0)} = (z_1^{(0)}, z_1^{(0)}, ..., z_T^{(0)})$.

Compute $\widehat{\sigma}$ from (16).

Compute $\widehat{p}$ (hence also $\widehat{P}$) from (19).

**Main**

For $i = 1, 2, ...$

**Parameter Estimation**

Compute $\mathbf{t}^{(i)}$ from $\mathbf{z}^{(i-1)}$.

Compute $\mathbf{M}^{(i)}$ from $\mathbf{t}^{(i)}$ and (13).

**State Estimation**

Compute $\mathbf{z}^{(i)}$ by the Viterbi algorithm using $\mathbf{x}$, $K$, $P$, $\mathbf{M}^{(i)}$ and $\widehat{\sigma}$.

**Termination Criterion**

If $\left| L^2_{K,T}(\mathbf{z}^{(i)}, \mathbf{x}; \widehat{P}, \mathbf{M}^{(i)}, \widehat{\sigma}) - L^2_{K,T}(\mathbf{z}^{(i-1)}, \mathbf{x}; \widehat{P}, \mathbf{M}^{(i-1)}, \widehat{\sigma}) \right| < \varepsilon$.

$\widehat{\mathbf{z}} = \mathbf{z}^{(i)}$.

$\widehat{\mathbf{M}} = \mathbf{M}^{(i)}$.

Exit the loop

EndIf

Compute $\widehat{\mathbf{t}}$ from $\widehat{\mathbf{z}}$.

End

---

One pass of the $i$ loop in the above algorithm has execution time $\mathrm{O}(T \cdot K^2)$. In all the examples of Section 5 the algorithm converges in very few iterations (typically 2, 3 or 4 iterations).

### 4.3.2 Variants of the Basic Algorithm

Several variants of the HMM segmentation algorithm can be used. For example, in place of (19) one can use the estimate of $p$ given by (20); or the $p_k$ estimates given by (25). Similarly, the $\widehat{\sigma}_k$ estimates of (14) could be used in place of (16).

Another family of variants is connected to the measure of homogeneity. The use of segment means follows from the assumption regarding the normal distribution of the observations. Other probability distributions can be used in (9). Also, an autoregressive model can be incorporated in the HMM segmentation algorithm. Assume that (for $k = 1, 2, ..., K$ and $t = t_{k-1} + 1, t_{k-1} + 2, ..., t_k$ ) we have

$$x_t = a_{0,k} + a_{1,k}x_{t-1} + a_{2,k}x_{t-2} + ... + a_{l,k}x_{t-l} + \epsilon_t, \tag{26}$$

where $\epsilon_t$ is a white noise term. Our segmentation algorithm can easily handle this assumption by a modification of the parameter estimation step. In other words, rather than reestimating $\mathbf{M}$, the algorithm reestimates the AR coefficients $a_{0,k}$, $a_{1,k}$, ... , $a_{l,k}$, using the data $x_{t_{k-1}+1}$, $x_{t_{k-1+2}}$, ... , $x_{t_k}$ and a least squares algorithm. This approach is used in Section 5.3 to fit a HMM autoregressive model to global temperature data. A similar approach can be used if it is assumed that the observations are generated by a polynomial regression of the form (for $t = t_{k-1} + 1, t_{k-1} + 2, ..., t_k$ and $k = 1, 2, ..., K$)

$$x_t = a_{0,k} + a_{1,k} \cdot (t - t_{k-1}) + ... + a_{l,k} \cdot (t - t_{k-1})^l + \epsilon_t \tag{27}$$

Again, the regression coefficients $a_{0,k}$, $a_{1,k}$, ..., $a_{l,k}$ can be computed by least squares fitting; in this case some constraints to enforce continuity across segments can also be used. In the case of first order polynomials it is only required to compute $a_{0,k}$, $a_{1,k}$, which are determined by the continuity constraints. This case may be of interest for detection of trends in the observable time series

### 4.3.3 Convergence and Optimality

We now discuss convergence and optimality issues. We only give the main results here; the details are presented in Appendix B. We start with the "basic algorithm" of Section 4.3.1. After some algebra,

the log-likelihood is found to be

$$\log L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \sigma) = -\sum_{t=1}^{T} \frac{(x_t - \mu_{z_t})^2}{2\sigma^2} - K \log\left(\frac{p}{1-p}\right) - T \log\left(\frac{\sqrt{2\pi}\sigma}{p}\right). \tag{28}$$

In a particular run of the algorithm $T, \mathbf{x}$ and $K, \widehat{P}, \widehat{\sigma}$ will be fixed. Hence, suppressing the dependence on these variables, we define

$$J_K(\mathbf{z}, \mathbf{M}) = -\log L^2_{K,T}(\mathbf{z}, \mathbf{x}; \widehat{P}, \mathbf{M}, \widehat{\sigma}). \tag{29}$$

In Appendix B we show that: if each of $\mathbf{z}^{(0)}$, $\mathbf{z}^{(1)}$, ... traverses all states, then we have

$$J_K(\mathbf{z}^{(0)}, \mathbf{M}^{(0)}) \geq J_K(\mathbf{z}^{(1)}, \mathbf{M}^{(1)}) \geq J_K(\mathbf{z}^{(2)}, \mathbf{M}^{(2)}) \geq ... \geq 0. \tag{30}$$

Hence, for $i = 0, 1, 2, ...$, the sequence $J_K(\mathbf{z}^{(i)}, \mathbf{M}^{(i)})$ converges to a value $\widehat{J}_K$ which is a *local* minimum of $J_K(\mathbf{z}, \mathbf{M})$. This is sufficient to ensure termination of the HMM segmentation algorithm. Upon termination the algorithm outputs a state estimate $\widehat{\mathbf{z}}$ and a means estimate $\widehat{\mathbf{M}}$ and we have $\widehat{J}_K = J_K(\widehat{\mathbf{z}}, \widehat{\mathbf{M}})$. It is easy to see that $\widehat{\mathbf{z}}, \widehat{\mathbf{M}}$ also maximize $L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \widehat{\sigma})$ and $L^1_{K,T}(\mathbf{z}|\mathbf{x}; P, \mathbf{M}, \widehat{\sigma})$ (both viewed as functions of $\mathbf{z}, \mathbf{M}$). Furthermore, if $J_K(\widehat{\mathbf{z}}, \widehat{\mathbf{M}})$ is a global minimum of $J_K(\mathbf{z}, \mathbf{M})$ and $\widehat{\mathbf{t}}$ is the segmentation associated with $\widehat{\mathbf{z}}$, then from (1) and (28) we see that $D_K(\widehat{\mathbf{t}}) = \widehat{D}_K$. However, note that $\widehat{D}_K \neq \widehat{J}_K$, since $J_K(\mathbf{z}, \mathbf{M})$ has additional terms depending on $T, K, P$ and $\sigma$.

### 4.4 The Full Segmentation Procedure – Selecting the Number of Segments

The last element required to complete our segmentation procedure is a method for choosing the best value of $K$. This is actually rather straightforward. From (28) we see that the log likelihood contains the term $-K \log\left(\frac{p}{1-p}\right)$. If $p > 0.5$, then $\log\left(\frac{p}{1-p}\right)$ is positive, hence segmentations with large $K$ are *penalized.* Consequently, it is reasonable and effective to take the "correct" value of $K$ to be the one which makes $L^2_{K,T}$ maximum[1].

[1]It is interesting to compare with Hubert's approach. Unlike $L^2_{K,T}$, Hubert's $\widehat{D}_K$ is a *decreasing* function of $K$. To avoid the trivial segmentation of $T$ segments, Hubert must penalize segmentations with a large number of segments (since $\widehat{D}_K$ is a decreasing function of $K$). This is achieved by the use of Scheffe's contrast criterion [17, 39] which checks for every $K$ the statistical significance of the optimal segmentation. Incidentally, the Scheffe criterion could also be used in

*In short, the full segmentation procedure consists in running the HMM algorithm of Section 4.3.1 (or one of the variants) for* $K = 1, 2, ..., K_{\max}$ *and selecting the segmentation which maximizes* $L^2_{K,T}$.

It is also possible to base the selection of $K$ on *human judgement*. Hubert [18, p.299] mentions the possibility of using his procedure with several values of $K$ and choosing the $K$ which looks "most reasonable". This type of "interactive segmentation" is even easier with our procedure, due to the short execution time of the HMM algorithm.

### 4.5 Some Additional Remarks

The HMM segmentation algorithm of Section 4.3.1 is an *approximate EM algorithm* [10]. It uses the basic EM idea of alternating *E*xpectation (parameter estimation) and *M*aximization (state estimation) steps. It is approximate because the $p, \mathbf{M}$ and $\sigma$ estimates we use are rather crude approximations of the *conditional expectations* of these quantities. For example, to estimate the means $\mu_k$ we use (13), while in a "complete" EM algorithm one would use $\widehat{\mu}_k = E(\mu_k|\mathbf{z})$. Note that the use of fixed $p$ and $\sigma$ values renders ML segmentation equivalent to Hubert's least-squares segmentation; this equivalence does not hold if $p$ and $\sigma$ are reestimated after every segmentation. As a practical matter, it will be seen in Section 5 that the segmentation algorithm is quite robust with respect to the exact value of $p$ (similar results have been obtained with respect to $\mathbf{M}$ and $\sigma$ but are omitted because of space limitations).

While the HMM algorithm (used for a specific value of $K$) is quite robust with respect to the $p$ value, $p$ is quite important for the overall segmentation procedure, in particular for the selection of the correct $K$ value. As mentioned previously, $K$ is selected so as to maximize likelihood; it can be seen from (28) that the likelihood depends on $K$ through the term $-K \log\left(\frac{p}{1-p}\right)$. Hence the choice of the correct $K$ depends on $p$. However, it will be seen in Section 5 that the simple estimate (19) gives very similar likelihood values to the ones obtained by using (20) and (25), hence the use of (19) again appears preferrable for reasons of simplicity.

---

conjunction with our HMM algorithm; this may be useful from the practical point of view but is not theoretically justified (we are grateful to the anonymous referee who pointed this out). In fact, even in the context of Hubert's procedure, the use of Scheffe's criterion is not entirely justified [18, p.300].

# 5 Segmentation Experiments

## 5.1 Annual Discharge of the Senegal River

This experiment uses the time series of the Senegal river annual discharge data, measured at the Bakel station for the years 1903-1988[2]. The length of the time series is 86. Hubert has applied his segmentation procedure on the same data set [17, 18] to find the segmentation which is optimal with respect to total deviation from segment means.

Three variants of the segmentation algorithm are run for $K = 2, 3, 4, 5, 6$. These variants only differ in the determination of the transition probability, according to the remarks of Section 4.3.2. I.e. the first variant uses the $p$ estimate (19); results for this variant are presented in Table 1. The second variant uses the $p_k$ estimates (25); results for this variant are presented in Table 2. The third variant uses the backward/forward $p$ estimate (20); results for this variant are presented in Table 3.

**Table 1 to appear here**

**Table 2 to appear here**

**Table 3 to appear here**

For every value of $K$ convergence is achieved by 2, 3 or, at most, 4 iterations of the algorithm. The maximum log likelihood achieved for each value of $K$ is listed in the last column of Tables 1, 2, 3. It can be seen that the optimal value of $K$ (i.e. the value which yields the maximum likelihood) is always 5. While the value of the log likelihood varies slightly in each table (reflecting its dependence on the exact value of the transition probabilities), the actual optimal segmentation is identical for all three variants, with segments [1903,1921], [1922,1936], [1937,1949], [1950,1967], [1967,1988]; these are the segments obtained by Hubert [17, 18] as well and indeed are the segments which yield the **globally** minimum total square error[3]. A plot of the time series, indicating the 5 segments and the respective means appears in Figure 2.

**Figure 2 to appear here**

[2]These data are available from Hubert's home page at `http://www.cig.ensmp.fr/~hubert`.

[3]The global optimality was checked with a dynamic programming algorithm (presented in [25]) which computes exactly the *globally* optimal segmentation (but it is slower than the HMM algorithm presented here).

The experiment was run with a MATLAB implementation of the HMM segmentation algorithm; the total execution time (i.e. obtaining the HMM segmentations of *all* orders) is between 0.72 and 1.20 sec (depending on the variant) on a Pentium III 1 GHz personal computer; it can be expected that a FORTRAN or C implementation would take about 10% to 20% of this time. For comparison purposes, Hubert reports execution time around 1 minute (but this is probably on a slower machine).

### 5.2 Annual Mean Global Temperature

This experiment uses 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 [30] and also at the Internet address `http://www.ngdc.noaa.gov/paleo/ei/ei_intro.html`. The length of the time series is 282.

The three segmentation variants mentioned in Section 4.3.2 are run for $K = 2, 3, 4, 5, 6$. Convergence is achieved in at most 4 iterations of the algorithm, for every value of $K$. The maximum log likelihood achieved for each value of $K$ is reported in the last column of Tables 4, 5, 6.

**Table 4 to appear here**

**Table 5 to appear here**

**Table 6 to appear here**

The optimal value of $K$ is always 4. While the actual value of the log likelihood varies slightly in each table (reflecting its dependence on the exact value of the transition probabilities), the actual optimal segmentation is *almost* identical for all three variants, with segments [1700,1720], [1721,1812], [1813,1930], [1931,1981] (the only difference appears in Table 5, where one segment boundary is 1718 rather than 1720). Once again it has been checked that these are the globally optimal segments. The total execution time for the experiment is is between 2.64 and 5.88 sec (depending on the variant). A plot of the time series, indicating the 4 segments and the respective means appears in Figure 3.

**Figure 3 to appear here**

### 5.3 Annual Mean Global Temperature with AR model

This experiment again uses the annual mean global temperature time series. The difference from the previous experiment is in the assumption regarding the data generation mechanism. Specifically, a

model of the following form is assumed:

$$x_t = a_{0,k} + a_{1,k}x_{t-1} + a_{2,k}x_{t-2} + a_{3,k}x_{t-3} + \epsilon_t, \tag{31}$$

for $k = 1, 2, ..., K$ and $t = t_{k-1} + 1, t_{k-1} + 2, ..., t_k$. The HMM segmentation algorithm can be modified to obtain the optimal segmentation with respect to the model of (31), as mentioned in Section 4.3.2. As previously, the three segmentation variants are run with $K = 2, 3, ..., 6$; results are presented in Tables 7, 8 and 9.

**Table 7 to appear here**

**Table 8 to appear here**

**Table 9 to appear here**

The optimal value of $K$ is always 4 and the three variants yield identical segmentations with segments [1700,1770], [1771,1835], [1836,1923], [1924,1981]. Once again it has been checked that these are the globally optimal segments. The total execution time for the experiment is between 2.80 and 6.17 sec (depending on the algorithm variant). A plot of the time series, indicating the 4 segments and the respective autoregressions appears in Figure 4.

**Figure 4 to appear here**

Recall that the means-based segmentation of the same time series yielded the segments [1700,1720], [1721,1812], [1813,1930], [1931,1981]. This seems in reasonable agreement with the AR-based segmentation, excepting the discrepancy of change points 1720 and 1770. From a numerical point of view, there is no a priori reason to expect that the AR-based segmentation and means-based segmentation should give the same results. The fact that the two segmentations are in relatively good agreement supports the hypothesis that actual climate changes have occurred approximately at the times indicated by both segmentation methods.

### 5.4 Artificial Time Series

Our next goal is to investigate the dependence of the segmentations on the value of the transition probabilities. To this end we perform several segmentation experiments where the value of $p$ is fixed. Hence in Table 10 we give segementations of the Senegal river time series with fixed $K = 5$ and various

values of $p$ (in every case $p$ is taken the same for all states and is held fixed). It can be seen that the same (optimal) segmentation is obtained for all $p$ values in the range $[0.7, 0.99]$. Similar results are presented for the global temperature time series in Table 11 (segmentation obtained by the means criterion) and Table 12 (segmentation obtained by the AR criterion). Hence it is reasonable to conclude that the segmentation algorithm is quite robust with respect to the $p$ value.

**Table 10 to appear here**

**Table 11 to appear here**

**Table 12 to appear here**

### 5.5 Artificial Time Series

The goal of the final experiment is to investigate the scaling properties of the algorithm, specifically the scaling of execution time with respect to time series length $T$ and the scaling of accuracy with respect to noise in the observations. To obtain better control over these factors, artificial time series are used, which have been generated by the following mechanism.

The time series are generated by a HMM with 5 states. Every time series is generated by running the HMM from state no.1 until state no.5. Hence, every time series involves 5 state transitions and, for the purposes of this experiment, this is assumed to be known a priori. On the other hand, it can be seen that the length of the time series is variable. With a slight change of notation, in this section $T$ will denote the *expected* length of the time series, which can be controlled by choice of the probability $p$. The values of $p$ were chosen to generate time series of average lengths 200, 250, 500, 750, 1000, 1250, 1500.

The observations are generated by a normal distribution with mean $\mu_k$ ($k$= 1, 2, ..., 5) and standard deviation $\sigma$. In all experiments the values $\mu_1 = \mu_3 = \mu_5 = 1$, $\mu_2 = \mu_4 = -1$ were used (hence the difference of two successive means is always 2). Several values of $\sigma$ were used, namely $\sigma$= 0.00, 0.10, 0.20, 0.30, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00.

For each combination of $T$ and $\sigma$, 20 time series were generated and the HMM segmentation algorithm was run on each one. For each run two quantities were computed: $c$, accuracy of segmentation, and $T_e$, execution time. Segmentation accuracy is computed by the formula

$$c = \frac{\sum_{t=1}^{T} \mathbf{1}(z_t = \widehat{z}_t)}{T} \tag{32}$$

where the indicator function $\mathbf{1}(z_t = \widehat{z}_t)$ is equal to 1 when $z_t = \widehat{z}_t$ and equal to 0 otherwise.

From these data two tables are compiled. Table 13 lists $T_e$ (in seconds) as a function of $T$ (i.e. $T_e$ is averaged over all time series of the same $T$). Table 14 lists average segmentation accuracy $c$ as a function of $T$ and $\sigma$ (i.e. $c$ is averaged over the 20 time series with the same $T$ and $\sigma$). As expected, segmentation accuracy is generally a decreasing function of $\sigma$, however it remains very close to 1 even for values $\sigma$ comparable to the difference between successive means $\mu_k$ and $\mu_{k+1}$ (e.g. for $\sigma = 1.25$).

**Table 13 to appear here**

**Table 14 to appear here**

# 6 Conclusion

In this paper our goal was to develop a computational procedure for the segmentation of multiple-change point time series. Using as starting points Hubert's pioneering work and some HMM methods (first used in the context of speech recognition) we have obtained a segmentation procedure which is faster and can handle longer time series than the one introduced by Hubert. Furthermore our procedure can incorporate various models of the observable time series (e.g. normal probability distribution with segment-specific means, autoregressive model with segment-specific AR coefficients etc.) and several different estimates of the important parameter $p$. We have demonstrated the local optimality of the $K$-th order segmentation and also given a simple method for the selection of the optimal $K$.

Our procedure also has some limitations. We have shown that the HMM segmentation algorithm produces a sequence of segmentations with increasing likelihood. This suffices to prove convergence to a *local* maximum of the likelihood but does not guarantee convergence to the global maximum. However, in all the experiments presented in this paper we have checked that the global maximum is actually achieved. Note that Hubert's procedure would always achieve the global optimum *if pruning were not used* – in the presence of pruning global optimality is not guaranteed.

Another possible drawback of our approach concerns the distribution of *state residence times*, i.e. the length of segments. Under the HMM, the residence time in any given state follows a geometric distribution. For example, assuming a common probability $p$ for all states, for every $k$ and $n$ we have

$$\Pr\left(t_k - t_{k-1} = n\right) = p^{n-1} \cdot (1-p)\,. \tag{33}$$

This is a drawback from the hydrological modelling point of view (since hydrological time series with change points do not appear to obey the above probability law). However, it is a peripheral issue when considered from our computational point of view, since it does not affect the *computational* efficiency of HMM's[4].

Indeed, throughout this paper we have adopted what we call a "computational approach". We have already stressed that our hidden Markov "model" is *not* intended as a realistic model of a hydrological or enviromental time series but as a computational tool.

This computational approach is often used by patter recognition practitioners and may have wider application to hydrological problems. Indeed, time series segmentation can be considered as a particular type of clustering, namely clustering under the constraint that clusters must respect the linear order of the samples. A vast number of clustering techniques has been reported in the Pattern Recognition literature. It would be interesting to apply some of these techniques to the problem of hydrological segmentation. The present paper can be understood as an example of this approach; an additional example is the dynamic programming segmentation algorithm which we have mentioned in footnote 3. Other examples of pattern recognition techniques which may be useful for hydrological time series segmentation are *hierarchical clustering, k-means clustering* etc.

# A Forward/BackWard Estimation of the Transition Probabilities

In Section 4.2 we have presented an iterative estimate of $p$ given by eqs.(20)–(24). In this Appendix we justify these formulas. The arguments are standard and can be found in [36].

The $i$-th estimate of $p$ is denoted by $p^{(i)}$. We also write $P^{(i)}$, keeping in mind that

$$P^{(i)}_{k,k} = p^{(i)}, \quad P^{(i)}_{k,k+1} = 1 - p^{(i)}, \quad P^{(i)}_{k,n} = 0 \text{ when } n \text{ is different from } k, k+1. \tag{34}$$

[4]The geometric distribution of HMM residence times has been discussed often in the engineering and pattern recognition literature [11] and has been accepted as a reasonable price to pay for the computational efficiency of HMM's.

Now, for $k = 1, 2, ..., K$ and $t = 1, 2, ..., T$ we define

$$\alpha_{k,t} = \Pr\left(X_1 = x_1, X_2 = x_2, ..., X_t = x_t, Z_t = k\right), \tag{35}$$

$$\beta_{k,t} = \Pr\left(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, ..., X_T = x_T | Z_t = k\right). \tag{36}$$

It is easily checked that the following recursive formulas hold

$$\alpha_{1,1} = 0, \quad \alpha_{k,1} = 0, \quad k = 2, 3, ..., K; \tag{37}$$

$$\alpha_{k,t+1} = \sum_{n=1}^{K} \alpha_{n,t} \cdot P_{n,k}^{(i)} \cdot \frac{e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma}, \qquad k = 1, 2, ..., K, \quad t = 1, 2, ..., T-1; \tag{38}$$

$$\beta_{k,t} = 1, \quad k = 1, 2, ..., K; \tag{39}$$

$$\beta_{k,t} = \sum_{n=1}^{K} P_{k,n}^{(i)} \cdot \frac{e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2}}{\sqrt{2\pi}\sigma} \cdot \beta_{n,t+1}, \qquad k = 1, 2, ..., K, \quad t = T-1, T-2, ..., 1; \tag{40}$$

In what follows $E(\cdot)$ denotes mathematical expectation and $\mathbf{1}(\cdot)$ is the *indicator function*. We denote the expected number of transitions from state $k$ to state $n$ at time $t$ by $\gamma_{t,k,n}$, i.e.

$$\gamma_{t,k,n} = E\left(\mathbf{1}\left(Z_t = k, Z_{t+1} = n\right)\right) = \Pr\left(X_1 = x_1, X_2 = x_2, ..., X_T = x_T, Z_t = k, Z_t = n\right). \tag{41}$$

Then we have

$$\gamma_{t,k,n} = \alpha_{k,t} \cdot P_{k,n}^{(i)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2} \cdot \beta_{n,t+1}. \tag{42}$$

Now, given an estimate $p^{(i)}$, a reasonable reestimate $p^{(i+1)}$ is

$$\begin{aligned}
p^{(i+1)} &= \frac{E\left(\text{“number of } k \text{ to } k \text{ transitions, for any } k \text{ and any } t\text{”}\right)}{T} \\
&= \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} E\left(\text{“number of } k \text{ to } k \text{ transitions at time } t\text{”}\right)}{T} \\
&= \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{t,k,k}}{T} \\
&= \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{k,t} \cdot P_{k,k}^{(i)} \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2} \cdot \beta_{k,t+1}}{T} \\
&= \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{k,t} \cdot p^{(i)} \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\left(x_{t+1} - \mu_{z_{t+1}}\right)^2 / 2\sigma^2} \cdot \beta_{k,t+1}}{T}
\end{aligned} \tag{43}$$

where the last equality follows from the fact that $P_{k,k}^{(i+1)} = p^{(i+1)}$.

## B Proof of Convergence

In this appendix we will justify the claim of eq.(30). We need some notation: we denote the set of all possible state sequences by $\Phi$, the set of all state sequences with $K$ transitions by $\Phi_K$ and the set of all $K$-dimensional real vectors by $\mathbf{R}^K$. We define $\phi(\mathbf{z})$ to be the number of transitions in the state sequence $\mathbf{z}$ (i.e. $\phi(\mathbf{z})$ = "number of times $z_{t-1} \neq z_t$") . If $\mathbf{z} \in \Phi_K$, then $\phi(\mathbf{z}) = K$.

Consider a single run of the basic HMM segmentation algorithm. During this run $T, K, \widehat{P}, \mathbf{M}$ and $\widehat{\sigma}$ are fixed. Taking the negative logarithm of (9) we obtain

$$\begin{aligned} -\log L^2_{K,T}(\mathbf{z}, \mathbf{x}; P, \mathbf{M}, \widehat{\sigma}) &= -\sum_{t=1}^{T} \log\left(P_{z_{t-1},z_t}\right) + \sum_{t=1}^{T} \frac{(x_t - \mu_{z_t})^2}{2\sigma^2} + T\log\left(\sqrt{2\pi}\sigma\right) \\ &= \sum_{t=1}^{T} \frac{(x_t - \mu_{z_t})^2}{2\sigma^2} - (T - \phi(\mathbf{z}))\log(p) - \phi(\mathbf{z})\log(1-p) + T\log\left(\sqrt{2\pi}\sigma\right) \\ &= \sum_{t=1}^{T} \frac{(x_t - \mu_{z_t})^2}{2\sigma^2} + \phi(\mathbf{z})\log\left(\frac{p}{1-p}\right) + T\log\left(\frac{\sqrt{2\pi}\sigma}{p}\right). \end{aligned} \tag{44}$$

Let us define

$$J_K(\mathbf{z}, \mathbf{M}) = \sum_{t=1}^{T} \frac{(x_t - \mu_{z_t})^2}{2\sigma^2} + K\log\left(\frac{p}{1-p}\right) + T\log\left(\frac{\sqrt{2\pi}\sigma}{p}\right). \tag{45}$$

Now suppose that *for* $i = 0, 1, 2, ...$ *the state sequence* $\mathbf{z}^{(i)}$ *traverses all states.* In other words, we assume that *for* $i = 0, 1, 2, ...$ *we have* $\phi\left(\mathbf{z}^{(i)}\right) = K$. Consider the sequence $J_K(\mathbf{z}^{(0)}, \mathbf{M}^{(0)})$, $J_K(\mathbf{z}^{(0)}, \mathbf{M}^{(0)})$, ... produced by a run of the basic HMM algorithm. Since $\mathbf{M}^{(i)}$ is estimated by (13), it follows that :

$$\forall \mathbf{M} \in \mathbf{R}^K : J(\mathbf{z}^{(i-1)}; \mathbf{M}) \geq J(\mathbf{z}^{(i-1)}; \mathbf{M}^{(i)}). \tag{46}$$

Also, since $\mathbf{z}^{(i)}$ (as computed by the Viterbi algorithm) yields the global maximum of the likelihood as a function of $\mathbf{z}$, we have:

$$\forall \mathbf{z} \in \Phi_K : J(\mathbf{z}; \mathbf{M}^{(i)}) \geq J(\mathbf{z}^{(i)}; \mathbf{M}^{(i)}). \tag{47}$$

Using first (46) and then (47) yields (for every $i$):

$$J(\mathbf{z}^{(i-1)}; \mathbf{M}^{(i-1)}) \geq J(\mathbf{z}^{(i-1)}; \mathbf{M}^{(i)}) \geq J(\mathbf{z}^{(i)}; \mathbf{M}^{(i)}). \tag{48}$$

Hence

$$J_K(\mathbf{z}^{(0)}, \mathbf{M}^{(0)}) \geq J_K(\mathbf{z}^{(1)}, \mathbf{M}^{(1)}) \geq J_K(\mathbf{z}^{(2)}, \mathbf{M}^{(2)}) \geq ... \geq 0. \quad (49)$$

This establishes eq.(30), under the condition that $\mathbf{z}^{(i)} \in \Phi_K$ for every $i$. This latter condition is easy to check and is usually satisfied. In fact, it can be *enforced* by choosing the parameter $p$ to be not too close to 1 (if $p \simeq 1$, then the high cost of state transitions may result to $\phi\left(\mathbf{z}^{(i)}\right) < K$ for some $i$).

## References


[1] A. Bardossy and E.J. Plate. "Modelling daily rainfall using a semi-Markov representation of circular pattern occurrence". *J. Hydrol.*, vol.122, pp.33-47, 1991.

[2] L.E. Baum and T.Petrie. "Statistical inference for probabilistic functions of finite state Markov chains". *Ann. of Math. Stat.*, 1966, vol.37, pp.1554-1563.

[3] L.E. Baum and J.A. Eagon. "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology". *Bull. Amer. Math. Soc.*, vol.73, pp.360–363, 1967.

[4] E. Baum et al. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains". *Ann. of Math. Stat.*, vol.41, pp.164–171, 1970.

[5] Y. Bengio. "Markovian models for sequential data". *Neural Comp. Surveys*,vol.2, pp.129-162, 1998.

[6] D. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice Hall, 1987.

[7] T.A. Buishand. "Some methods for testing the homogeneity of rainfall records". *J. Hydrol.*, vol.58, pp.11-27, 1982.

[8] T.A. Buishand. "Tests for detecting a shift in the mean of hydrological time series". *J. Hydrol.*, vol.75, pp.51-69, 1984.

[9] G. W. Cobb. "The problem of the Nile: Conditional solution to a changepoint problem". *Biometrika*, vol. 65, pp. 243- 252, 1978.

[10] A. P. Dempster, N. M. Laird and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *J. Roy. Statist. Soc. B*, vol.39, pp.1–38, 1977.

[11] G. Peng, B. Zhang, W.S-Y. Wang. "Performance of Mandarin connected digit recognizer with word duration modeling". In the *Proc. of ASR2000*, pp. 976-980, 2000.

[12] G. Forney. "The Viterbi algorithm". *Proceedings of the IEEE*, vol. 61, pp.268–278, 1973.

[13] V. Fortin, L. Perreault, J.C. Ondo and R.C. Evra. "Bayesian long-term forecasting of annual inflows with a shifting-level model". *ASCE Water Res. Planning on Managing the Extremes: Floods and Droughts*. Roanoke, Virginia, 2002.

[14] Y. Gyasi-Agyei and G. R. Willgoose. "Generalisation of a hybrid model for point rainfall". *J. of Hydrology*, vol. 219, pp. 218-224, 1999.

[15] A.I. Hipel and K.W. McLeod. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, 1994.

[16] H. Hoppe and G. Kiely. "Precipitation over Ireland – Observed changes since 1940". *Phys. Chem. Earth (B)*, vol.24, pp.91-96, 1999.

[17] P. Hubert. "Change points in meteorological analysis". In *Applications of Time Series Analysis in Astronomy and Meteorology*, T.Subba Rao, M.B. Priestley and O. Lessi (eds.). Chapman and Hall, 1997.

[18] P. Hubert. "The segmentation procedure as a tool for discrete modeling of hydrometeorogical regimes". *Stoch. Env. Res. and Risk Ass.*, vol. 14, pp.297-304, 2000.

[19] J.P. Hughes, P. Guttorp and S.P. Charles. "A non-homogeneous hidden Markov model for precipitation occurrence". *Appl. Stat.*, vol. 48, pp. 15-30, 1999.

[20] O.D. Jimoh and P. Webster. "The optimum order of Markov chain for daily rainfall in Nigeria". *J. Hydrol.* vol. 185, pp.45-69, 1996.

[21] O.D. Jimoh and P. Webster. "Stochastic modelling daily rainfall in Nigeria, intra-annual variation of model parameters". *J. Hydrol.* vol. 222, pp.1-17, 1999.

[22] B.H. Juang. "Maximum likelihood estimation for mixture multivariate stochastic observations of Markov chains". *ATT Tech. J.*, vol.64, pp.1235-1249, 1985.

[23] B.H. Juang and L.R. Rabiner. "Mixture autoregressive hidden Markov models for speech signals". *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, pp.1404-1412, 1985.

[24] P. Kenny, M. Lennig and P. Mermelstein. "A linear predictive HMM for vector-valued observations with applications to speech recognition". *IEEE Trans. on Sig.Proc.*, Vol.38, p.220-225, 1990 .

[25] Ath. Kehagias. "Hidden Markov model segmentation of hydrological and enviromental time series". `http://xxx.lanl.gov/abs/cs.OH/0206039.`

[26] G. Kiely, J.D. Albertson and M.B. Parlange. "Recent trends in diurnal variation of precipitation at Valentia on the west coast of Ireland". *J. Hydrol.*, vol.207, pp.270-279, 1998.

[27] A.S.F. Lee, and S.M. Heghinian. "A shift of the mean level in a sequence of independent normal random variables—a Bayesian approach". *Technometrics*, vol. 19, pp.503–506, 1977.

[28] S.E. Levinson, L. R. Rabiner and M.M. Sondhi. "An introduction to the application of the theory of probabilistic functions of a Markov chain", *The Bell Sys. Tech. J.*, vol. 62, pp.1035–1074, 1983.

[29] Z.Q. Lu and L.M. Berliner. "Markov switching time series models with application to a daily runoff series". *Water Resour. Res.*, vol. 35, pp.523-534, 1999.

[30] M.E. Mann, R.S. Bradley and M.K. Hughes. "Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations". *Geophys. Res. Lett.*, vol.26, pp.759-762, 1999.

[31] J.E. Paturel et al. "Climatic variability in humid Africa along the Gulf of Guinea. Part II: An integrated regional approach". *J. of Hydrol.*, vol.191, pp.16-36, 1997.

[32] L. Perreault, M. Hache, M. Slivitzky and B. Bobee. "Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach". *Stoch. Env. Res. and Risk Ass.*, vol. 13, pp.201-216, 1999.

[33] L. Perreault, E. Parent, J. Bernier, B. Bobee and M. Slivitzky. "Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences". *Stoch. Env. Res. and Risk Ass.*, vol. 14, pp.243-261, 2000.

[34] L. Perreault, J. Bernier, B. Bobee and E. Parent. "Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited". *J. Hydrol.*, vol. 235, pp.221-241, 2000.

[35] L. Perreault, J. Bernier, B. Bobee and E. Parent. "Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting". *J. Hydrol.*, vol. 235, pp.242-263, 2000.

[36] L.R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, vol. 77, pp.257-286, 1988.

[37] A. Ramachandra Rao and W. Tirtotjondro. "Investigation of changes in characteristics of hydrological time series by Bayesian methods". *Stoch. Hydrol. and Hydraulics*, vol. 10, pp.295-317, 1996.

[38] J.D. Salas and D.C. Boes. "Shifting level modelling of hydrologic time series". *Adv. in Water Res.*, vol.3, pp. 59-63, 1980

[39] M. Scheffe. *The Analysis of Variance*. Wiley, 1959.

[40] E. Servat et al. "Climatic variability in humid Africa along the Gulf of Guinea. Part I: detailed analysis of the phenomenon in Cote d' Ivoire". *J. Hydrol.,* vol.191, pp.1-15, 1997.

[41] R. Srikanthan and T.A. McMahon. "Stochastic generation of annual, monthly and daily climate data: a review". *Hydrology and Earth Sys. Sci.*, vol.5, pp.653-670, 2001.

[42] R. Srikanthan, M.A. Thyer, G. Kuczera and T.A. McMahon. "Modelling annual rainfall using a hidden state Markov model". In *EMS 2002, Integrated Assessment and Decision Support ,* June 2002, Lugano, Switzerland.

[43] M.A. Thyer and G. Kuczera. "Modelling long-term persistence in rainfall time series, Sydney rainfall case study". *Hydr. and Water Res. Symposium*, Inst. of Engineers, Australia, pp.550-555, 1999.

[44] M.A. Thyer and G. Kuczera. "Modelling long-term persistence in hydroclimatic time series using a hidden state Markov model". *Water Resour. Res.*, vol. 36, pp.3301-3310, 2000.

[45] W. Zucchini and P. Guttorp. "A hidden Markov model for space-time precipitation". *Water Resour. Res.,* vol. 27, pp.1917-1923, 1991.

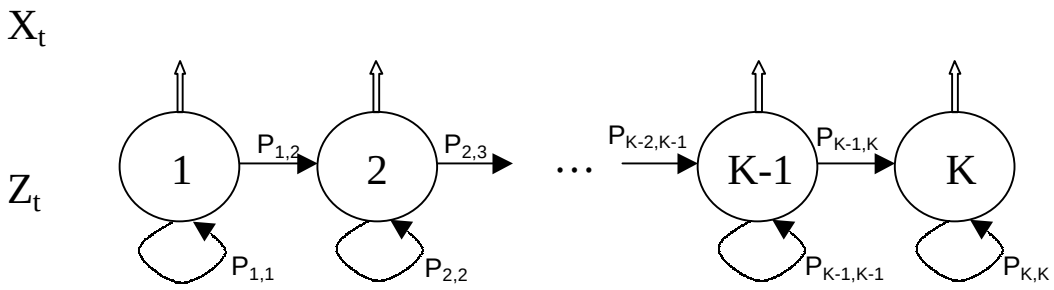


**Figure 1.** A diagrammatic representation of a hidden Markov model. Each circle denotes a state. Single line arrows denote state transitions; the transition from state i to state j has probability $P_{i,j}$. Double arrows denote the emission of state-dependent observations. The state process is $Z_t$ and the observation process is $X_t$.

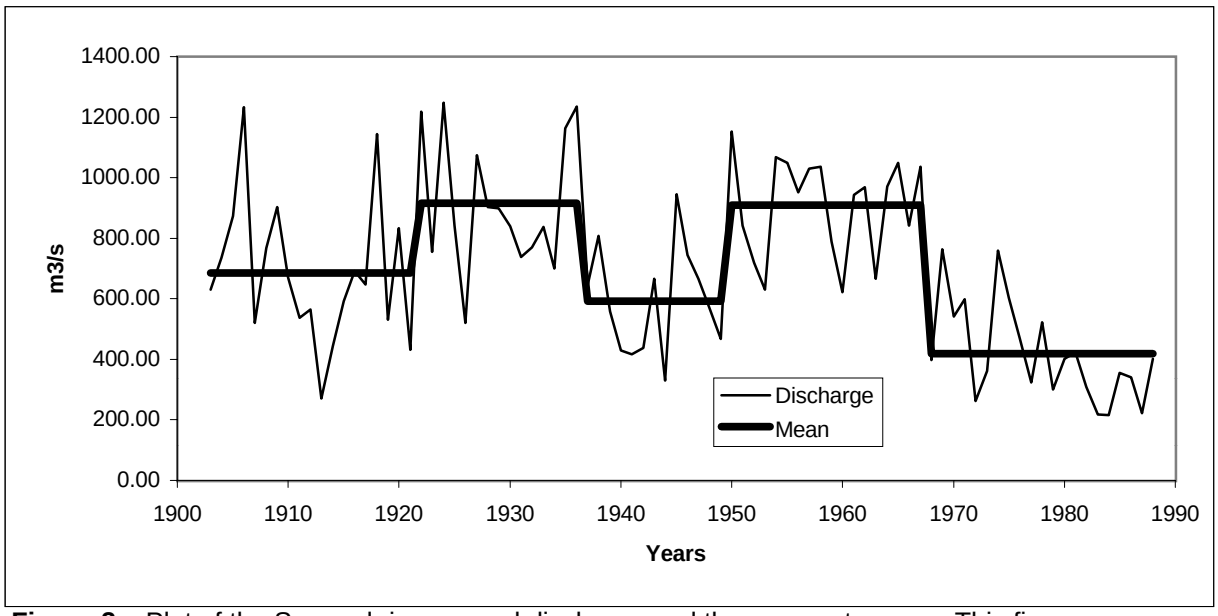


**Figure 2.** Plot of the Senegal river annual discharge and the segment means. This figure was obtained from the optimal 5-th order segmentation

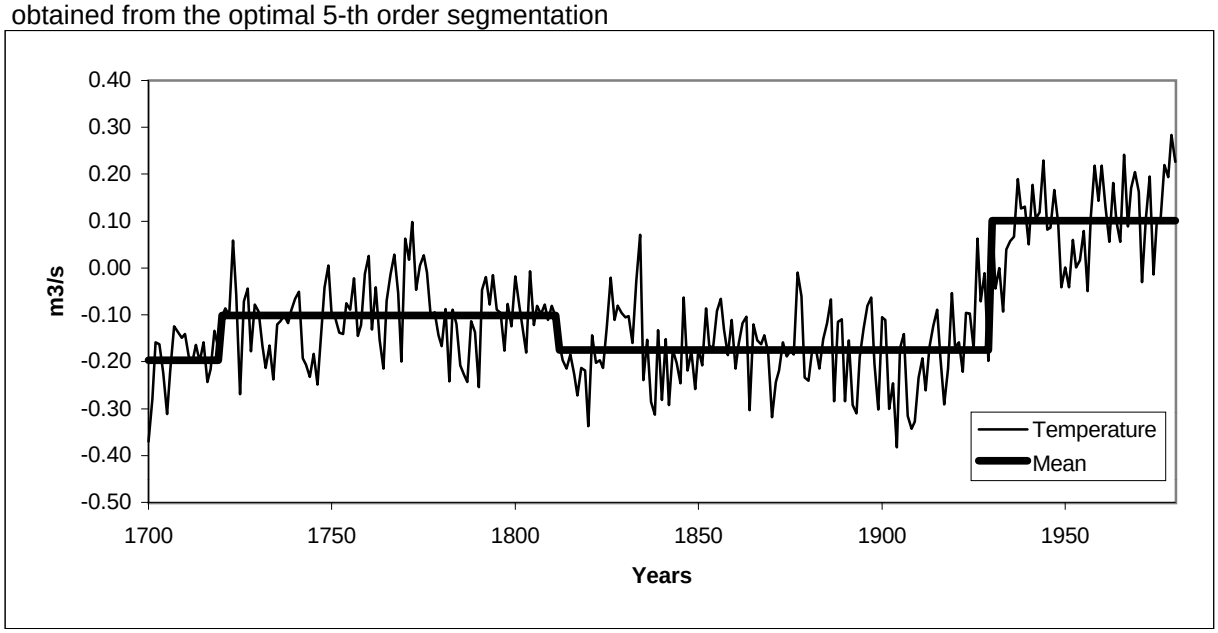


**Figure 3.** Plot of the annual mean global temperature and the segment means. This figure corresponds to the optimal fourth order segmentation.

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1902 | 1988 | | | | | | -41.500 |
| 2 | 1902 | 1967 | 1988 | | | | | -31.836 |
| 3 | 1902 | 1949 | 1967 | 1988 | | | | -31.839 |
| 4 | 1902 | 1917 | 1953 | 1967 | 1988 | | | -33.461 |
| **5** | **1902** | **1921** | **1936** | **1949** | **1967** | **1988** | | **-30.556** |
| 6 | 1902 | 1921 | 1936 | 1949 | 1967 | 1971 | 1988 | -31.668 |

**Table 1**: Optimal segmentation of the Senegal river annual discharge time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability p (common for all states) is estimated by (T--K)/T. Total execution time (for K=1,2,…,6) is 0.77 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1902 | 1988 | | | | | | -41.500 |
| 2 | 1902 | 1967 | 1988 | | | | | -31.728 |
| 3 | 1902 | 1949 | 1967 | 1988 | | | | -31.580 |
| 4 | 1902 | 1917 | 1953 | 1967 | 1988 | | | -33.172 |
| **5** | **1902** | **1921** | **1936** | **1949** | **1967** | **1988** | | **-30.501** |
| 6 | 1902 | 1921 | 1936 | 1949 | 1967 | 1969 | 1988 | -30.619 |

**Table 2**: Optimal segmentation of the Senegal river annual discharge time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability (of the k-th state) pk is estimated by (Tk--1)/Tk. Total execution time (for K=1,2,…,6) is 0.72 sec

| K | | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1902 | 1988 | | | | | | -41.500 |
| 2 | 1902 | 1967 | 1988 | | | | | -31.728 |
| 3 | 1902 | 1949 | 1967 | | | | | -31.818 |
| 4 | 1902 | 1917 | 1953 | 1967 | | | | -33.461 |
| **5** | **1902** | **1921** | **1936** | **1949** | **1967** | **1988** | | **-30.553** |
| 6 | 1902 | 1921 | 1936 | 1949 | 1967 | 1971 | 1988 | -31.669 |

**Table 3**: Optimal segmentation of the Senegal river annual discharge time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability p (common for all states) is estimated by the forward/backward formula (20). Total execution time (for K=1,2,…,6) is 1.20 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.000 |
| 2 | 1699 | 1930 | 1981 | | | | | -69.024 |
| 3 | 1699 | 1812 | 1930 | 1981 | | | | -67.732 |
| **4** | **1699** | **1720** | **1812** | **1930** | **1981** | | | **-66.834** |
| 5 | 1699 | 1748 | 1812 | 1890 | 1926 | 1981 | | -70.605 |
| 6 | 1699 | 1748 | 1777 | 1812 | 1890 | 1930 | 1981 | -71.621 |

**Table 4**: Optimal segmentation of the annual mean global temperature time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability p (common for all states) is estimated by (T--K)/T. Total execution time (for K=1,2,…,6) is 2.64 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.000 |
| 2 | 1699 | 1930 | 1981 | | | | | -68.877 |
| 3 | 1699 | 1812 | 1930 | 1981 | | | | -67.684 |
| **4** | **1699** | **1718** | **1812** | **1930** | **1981** | | | **-66.052** |
| 5 | 1699 | 1748 | 1812 | 1929 | 1930 | 1981 | | -66.724 |
| 6 | 1699 | 1748 | 1777 | 1812 | 1929 | 1930 | 1981 | -67.907 |

**Table 5**: Optimal segmentation of the the annual mean global temp. time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability (of the k-th state) pk is estimated by (Tk--1)/Tk. Total execution time (for K=1,2,…,6) is 2.85 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.000 |
| 2 | 1699 | 1930 | 1981 | | | | | -68.877 |
| 3 | 1699 | 1812 | 1930 | 1981 | | | | -67.684 |
| **4** | **1699** | **1720** | **1812** | **1930** | **1981** | | | **-66.821** |
| 5 | 1699 | 1748 | 1812 | 1890 | 1926 | 1981 | | -70.605 |
| 6 | 1699 | 1748 | 1777 | 1812 | 1890 | 1930 | 1981 | -71.621 |

**Table 6**: Optimal segmentation of the the annual mean global temp. time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on homogeneity of segment means. The transition probability p (common for all states) is estimated by the forward/backward formula (20). Total execution time (for K=1,2,…,6) is 5.88 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.500 |
| 2 | 1699 | 1930 | 1981 | | | | | -72.623 |
| 3 | 1699 | 1835 | 1926 | 1981 | | | | -65.651 |
| **4** | **1699** | **1770** | **1835** | **1923** | **1981** | | | **-64.105** |
| 5 | 1699 | 1747 | 1812 | 1869 | 1923 | 1981 | | -76.466 |
| 6 | 1699 | 1748 | 1774 | 1835 | 1893 | 1926 | 1981 | -76.907 |

**Table 7**: Optimal segmentation of the annual mean global temperature time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on AR error The transition probability p (common for all states) is estimated by (T--K)/T. Total execution time (for K=1,2,…,6) is 2.80 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.500 |
| 2 | 1699 | 1930 | 1981 | | | | | -72.475 |
| 3 | 1699 | 1835 | 1926 | 1981 | | | | -65.571 |
| **4** | **1699** | **1770** | **1835** | **1923** | **1981** | | | **-63.503** |
| 5 | 1699 | 1747 | 1812 | 1869 | 1923 | 1981 | | -76.435 |
| 6 | 1699 | 1748 | 1774 | 1835 | 1893 | 1923 | 1981 | -76.883 |

**Table 8**: Optimal segmentation of the the annual mean global temp. time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on AR error The transition probability (of the k-th state) pk is estimated by (Tk--1)/Tk. Total execution time (for K=1,2,…,6) is 2.97 sec

| K | Segment Boundaries (Change Points) | | | | | | | LogLikelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | 1699 | 1981 | | | | | | -139.500 |
| 2 | 1699 | 1930 | 1981 | | | | | -72.475 |
| 3 | 1699 | 1835 | 1926 | 1981 | | | | -65.612 |
| **4** | **1699** | **1770** | **1835** | **1923** | **1981** | | | **-64.103** |
| 5 | 1699 | 1747 | 1812 | 1869 | 1923 | 1981 | | -76.460 |
| 6 | 1699 | 1748 | 1774 | 1835 | 1893 | 1926 | 1981 | -76.907 |

**Table 9**: Optimal segmentation of the the annual mean global temp. time series for various values of K. The value which maximizes log likelihood is indicated in bold. Classification is based on AR error The transition probability p (common for all states) is estimated by the forward/backward formula (20). Total execution time (for K=1,2,…,6) is 6.17 sec

| p | Segment Boundaries (Change Points) | | | | | |
|---|---|---|---|---|---|---|
| 0.6000 | 1902 | 1921 | 1936 | 1949 | 1967 | |
| 0.7000 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.8000 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9000 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9100 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9200 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9300 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9400 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9500 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9600 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9700 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9800 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9900 | 1902 | 1921 | 1936 | 1949 | 1967 | 1988 |
| 0.9950 | 1902 | 1949 | 1967 | 1988 | | |
| 0.9990 | 1902 | 1988 | | | | |
| 0.9999 | 1902 | 1988 | | | | |

**Table 10**: Dependence of segmentation on the value of p. The above segmentations of the Senegal time series have been obtained in the same manner as the ones of Table 1, except that K is fixed at 5 and the value of p is the one indicated in the first column.

| p | Segment Boundaries (Change Points) | | | | |
|---|---|---|---|---|---|
| 0.70000 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.80000 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.90000 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.95000 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.99000 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.99900 | 1700 | 1720 | 1810 | 1930 | 1981 |
| 0.99990 | 1700 | 1812 | 1930 | 1981 | |
| 0.99999 | 1700 | 1930 | 1981 | | |

**Table 11**: Dependence of segmentation on the value of p. The above segmentations of the global temperature time series have been obtained in the same manner as the ones of Table 4, except that K is fixed at 4 and the value of p is the one indicated in the first column.

| p | Segment Boundaries (Change Points) | | | | |
|---|---|---|---|---|---|
| 0.70000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.80000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.90000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.95000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.97000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.99000 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.99900 | 1700 | 1770 | 1835 | 1923 | 1981 |
| 0.99950 | 1700 | 1835 | 1923 | 1981 | |
| 0.99990 | 1700 | 1923 | 1981 | | |

**Table 12**: Dependence of segmentation on the value of p. The above segmentations of the global temperature time series have been obtained in the same manner as the ones of Table 7, except that K is fixed at 4 and the value of p is the one indicated in the first column.

| **T** | 200 | 250 | 500 | 750 | 1000 | 1250 | 1500 |
|---|---|---|---|---|---|---|---|
| **$T_e$** | 0.193 | 0.249 | 0.585 | 1.024 | 1.845 | 3.026 | 4.600 |

**Table 13**: Scaling of execution time as a function of the time series length. In this table T is the length of the time series and $T_e$ is execution time.

| **T** | 200 | 250 | 500 | 750 | 1000 | 1250 | 1500 |
|---|---|---|---|---|---|---|---|
| **σ** | **c (Classification Accuracy)** | | | | | | |
| 0.00 | 1.0000 | 1.0000 | 1.0000 | 0.9692 | 1.0000 | 1.0000 | 0.9902 |
| 0.10 | 1.0000 | 1.0000 | 1.0000 | 0.9814 | 1.0000 | 1.0000 | 1.0000 |
| 0.20 | 1.0000 | 0.9806 | 1.0000 | 1.0000 | 1.0000 | 0.9716 | 1.0000 |
| 0.30 | 1.0000 | 1.0000 | 0.9999 | 0.9792 | 1.0000 | 0.9807 | 1.0000 |
| 0.50 | 0.9989 | 0.9993 | 0.9994 | 0.9997 | 1.0000 | 0.9997 | 1.0000 |
| 0.75 | 0.9945 | 0.9979 | 0.9663 | 0.9521 | 0.9988 | 0.9992 | 0.9991 |
| 1.00 | 0.9881 | 0.9880 | 0.9863 | 0.9974 | 0.9517 | 0.9981 | 0.9711 |
| 1.25 | 0.9778 | 0.9710 | 0.9762 | 0.9924 | 0.9965 | 0.9843 | 0.9781 |
| 1.50 | 0.9561 | 0.9701 | 0.9874 | 0.9341 | 0.9507 | 0.9362 | 0.9956 |
| 1.75 | 0.9337 | 0.8985 | 0.9494 | 0.9341 | 0.9708 | 0.9272 | 0.9942 |
| 2.00 | 0.8628 | 0.8617 | 0.8255 | 0.9141 | 0.8600 | 0.9523 | 0.8297 |

**Table 14**: Scaling of classification accuracy as a function of the time series length. In the above table T is the length of the time series and σ is the standard deviation of the observations.

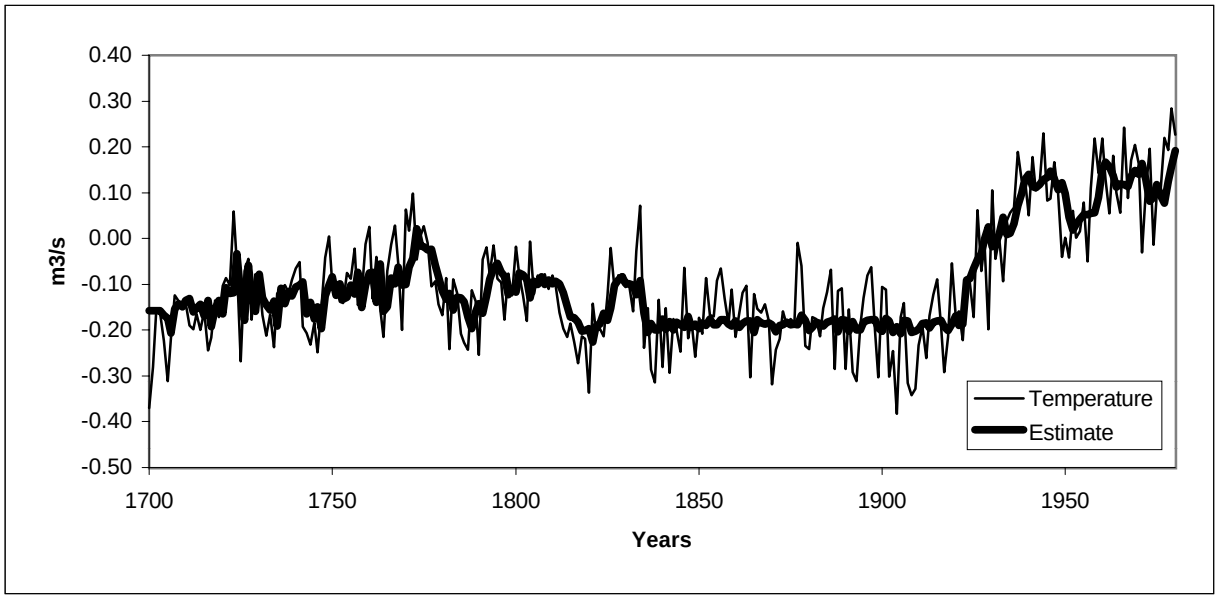


**Figure 4.** Plot of the annual mean global temperature and the AR estimate. This figure corrsponds to the optimal fourth order segmentation.