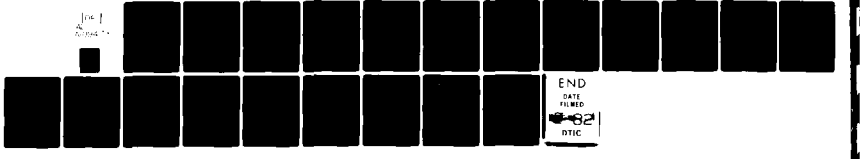
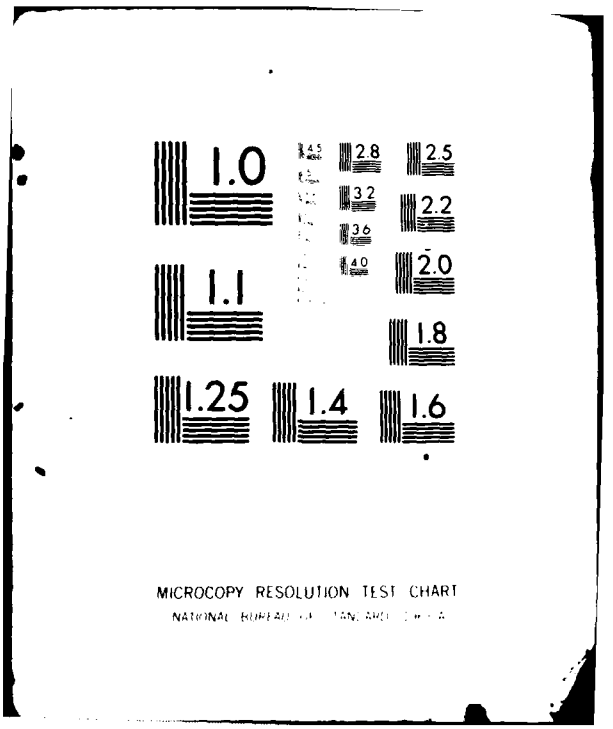
AD-A109 479 ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF MATHEMATICS F/G 12/1
ON SEGMENTATION OF TIME SERIES.(U)
NOV 81 S L SCLOVE N00014-80-C-0408
UNCLASSIFIED TR-81-1 NL



END
DATE FILMED
1-82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

(12)

ON SEGMENTATION OF TIME SERIES

LEVEL II

by

STANLEY L. SCLOVE
University of Illinois at Chicago Circle

Invited Paper, Special Session on Cluster Analysis, 789th Meeting,
American Mathematical Society, University of Massachusetts,
Amherst, MA, October 16-18, 1981

TECHNICAL REPORT NO. 81-1
November 30, 1981

PREPARED FOR THE
OFFICE OF NAVAL RESEARCH
UNDER
CONTRACT N00014-80-C-0408,
TASK NR042-443
with the University of Illinois at Chicago Circle

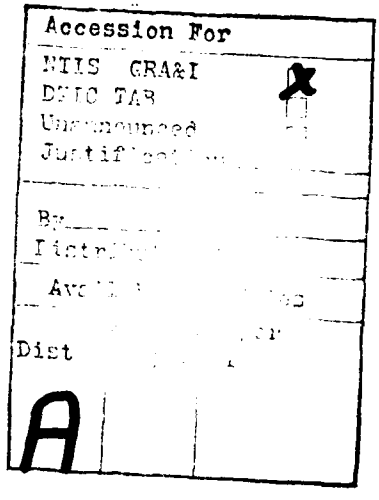
Principal Investigator: Stanley L. Sclove

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT CHICAGO CIRCLE
CHICAGO, ILLINOIS 60680

ON SEGMENTATION OF TIME SERIES*

Stanley L. Sclove
University of Illinois at Chicago Circle

CONTENTS



---

*Invited Paper, Special Session on Cluster Analysis, 789th Meeting, American Mathematical Society, University of Massachusetts, Amherst, MA, October 16-18, 1981

ON SEGMENTATION OF TIME SERIES*

Stanley L. Sclove
University of Illinois at Chicago Circle

ABSTRACT

The problem of partitioning a time-series into segments is considered. The segments fall into classes, which may correspond to phases of a cycle (recession, recovery, expansion in the business cycle) or to portions of a signal obtained by scanning (background/clutter, target, background/clutter again, another target, etc.; or normal tissue, tumor, normal tissue). Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov chain. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. A numerical example is given. Choice of the number of classes, using Akaike's automatic (model) identification criterion (AIC), is illustrated. Prediction is considered.

Key Words & Phrases: forecasting; prediction; signal analysis; isodata procedure; Markov chains; maximum likelihood; Akaike's automatic (model) identification criterion (AIC).

---

*Invited Paper, Special Session on Cluster Analysis, 789th Meeting, American Mathematical Society, University of Massachusetts, Amherst, MA, October 16-18, 1981

ON SEGMENTATION OF TIME SERIES*

Stanley L. Sclove
University of Illinois at Chicago Circle

1. Introduction

The problem of "segmentation" considered here is: Given a time series $\{x_t, t=1,2,\ldots,n\}$, partition the set of values of t into sub-series (segments, regimes) which are relatively homogeneous. The segments are assumed to fall into several classes. In processes which may be considered as cycles the classes are phases of the cycle.

Examples. (i) Segment an economic time series into periods of recession, recovery, and expansion. Here there are three classes of segment. (ii) Segment an electrocardiogram into rhythmic and arhythmic periods (two classes of segment). (iii) Segment an electroencephalogram of a sleeping person into periods of deep sleep and restless or fitful sleep (two classes of segment). (iv) Segment a received signal into segments of background, target, background again, another target, etc.

The observation X may be a scalar, vector, or matrix--any element of a linear space, for which the operations of addition and scalar multiplication are defined. (One needs to perform such operations as $x_t - cx_{t-1}$, where c is a scalar.)

In some applications the definition of the classes involves the values of the observed x; in others, their definition may be logically independent of the value-space of X. In the former case the classes may be viewed simply as a partition of the value-space of X.

---

*Invited Paper, Special Session on Cluster Analysis, 789th Meeting, American Mathematical Society, University of Massachusetts, Amherst, MA, October 16-18, 1981

2. The Model

One can imagine a series which is usually relatively smooth but occasionally rahter jumpy as being composed of sub-series which are first-order autoregressive [AR(1)], the autocorrelation coefficient ϕ being positive for the smooth segments and negative for the jumpy ones. In a simple case one might try fitting a segmentation with two classes given by $AR(1;\phi_1)$ and $AR(1;\phi_2)$, where one of the ϕ's is positive and the other is negative.

The mechanism generating the process changes from time to time, and these changes manifest themselves at some unknown time points (epochs) $\tau_1, \tau_2, \ldots, \tau_{m-1}$; that is, there are m segments. The integer m and the epochs τ_g, g=1,2,...,m-1, are unknown. Generally there will be fewer than m generating mechanisms. The number of mechanisms (classes) will be denoted by k; it will be assumed that k is at most m. In some situations, k is specified; in others, it is not. With the c-th class is associated a stochastic process, P_c, say. For example, above we spoke of a situation with k=2 classes, where, for c=1,2, the process P_c is $AR(1;\phi_c)$.

Now with the t-th observation (t=1,2,...,n) associate the label γ_t, which is equal to c if and only if x_t arose from class c, c=1,2,...,k. Each time-point t gives rise to a pair (x_t,γ_t), where x_t is observable and γ_t is not. The process $\{x_t\}$ is the observed time series, and $\{\gamma_t\}$ will be called the label process.

Define a segmentation, then, as a partition of the time index set $\{t: t=1,2,\ldots,n\}$ into subsets $S_1 = \{1,2,\ldots,t_1\}$, $S_2 = \{t_1+1,\ldots,t_2\}$, ..., $S_m = \{t_{m-1}+1, \ldots, n\}$, where $t_1<t_2<\ldots<t_{m-1}<t_m=n$. Each subset S_g is a segment. The integer m is not specified. In the context of this model, to segment the series is merely to estimate the γ's.

The idea underlying the development in the present paper is that of <u>transitions</u> between classes. The labels γ_t will be treated as random variables Γ_t with transition probabilities $Pr(\Gamma_t=d|\Gamma_{t-1}=c) = p_{cd}$, taken as stationary, i.e., independent of t. The matrix of transition probabilities will be denoted by $\underline{P}$, that is, $\underline{P} = [p_{cd}]_{\substack{c=1,2,\ldots,k \\ d=1,2,\ldots,k}}$.

If a process is to be strictly cyclic, like intake, compression, combustion for a combustion engine, or recession to recovery to expansion to recession, etc., in the business cycle, then this this condition can be imposed by using a transition probability matrix such as the following, with zeros in the appropriate places.

		Label at time t		
		1: Recession	2: Expansion	3: Recovery
Label at time t-1	1: Recession	p_{11}	p_{12}	$p_{13}=0$
	2: Recovery	$p_{21}=0$	p_{22}	p_{23}
	3: Expansion	p_{31}	$p_{32}=0$	p_{33}

Later we shall consider a matrix like this but with different restrictions; namely, we shall allow transitions only to adjacent states (classes). See Section 4.2.

Segmentation will involve the simultaneous estimation of the parameters of the stochastic processes P_c, c=1,2,...,k, the transition probability matrix $\underline{P}$, and the labels $\{\gamma_t, t=1,2,\ldots,n\}$.

A joint probability density function (p.d.f.) for $\{(X_t,\Gamma_t), t=1,2,\ldots,n\}$ is, using f as a generic symbol for any p.d.f., and successively conditioning each of Γ_1, X_1, Γ_2, X_2, Γ_3, X_3, ..., X_n on all preceding X'x and Γ's,

$$f(\gamma_1)f(x_1|\gamma_1)\prod_{t=2}^{n} f(\gamma_t|x_{t-1},\gamma_{t-1},\ldots,\gamma_1)f(x_t|\gamma_t,x_{t-1},\gamma_{t-1},\ldots,\gamma_1). \tag{2.1}$$

The working assumptions of this paper are the following.

(A.1) The γ_t are a first-order stationary Markov chain, independent of the x's:

$$f(\gamma_t|x_t,\gamma_{t-1},\ldots,x_1,\gamma_1) = p_{\gamma_{t-1}\gamma_t}. \tag{2.2}$$

(A.2) The random variable X_t depends upon the past only through its own label and through previous X's, not through previous labels:

$$f(x_t|\gamma_t,x_{t-1},\gamma_{t-1},\ldots,x_1,\gamma_1) = f(x_t|\gamma_t,x_{t-1},\ldots,x_1). \tag{2.3}$$

With these assumptions (2.1) becomes

$$f(\gamma_1)f(x_1|\gamma_1)\prod_{t=2}^{n} p_{\gamma_{t-1}\gamma_t} f(x_t|\gamma_t,x_{t-1},\ldots,x_1). \tag{2.4}$$

Note that this is

$$\prod_{c=1}^{k}\prod_{d=1}^{k} p_{cd}^{n_{cd}} f(\gamma_1)f(x_1|\gamma_1)\prod_{t=2}^{n} f(x_t|\gamma_t,x_{t-1},\ldots,x_1), \tag{2.5}$$

where n_{cd} = number of transitions from class c to class d (unobservable).

This model, with transition probabilities, has certain advantages over a model which uses only the epochs (change-points). The epochs are discrete parameters, and, even if the corresponding generalized likelihood ratio were asymptotically chi-square, the number of degrees of freedom would not be clear. On the other hand, the transition probabilities vary in an interval and it is clear that they constitute a set of k(k-1) free parameters.

Examples. (i) If each class-conditional process P_c is a first-order Markov process, then

$$f(x_t|\gamma_t,x_{t-1},\ldots,x_1) = f(x_t|\gamma_t,x_{t-1}). \tag{2.6}$$

(ii) If in addition the c-th class-conditional process is Gaussian first-order autoregressive with autoregression coefficient ϕ_c and constant term δ_c, with common variance σ^2, then (2.6) holds with

$$f(x_t|\gamma_t=c,x_{t-1}) = (2\pi\sigma^2)^{-1/2}\exp[-u_{tc}^2/(2\sigma^2)],$$

where

$$u_{tc} = x_t - (\phi_c x_{t-1} + \delta_c) .$$

E.g., the value of the likelihood for $\gamma_1 = 1 = \gamma_2 = \ldots = \gamma_m$ and $\gamma_{m+1} = 2 = \gamma_{m+2} = \ldots = \gamma_n$ is, for given x_1,

$$p_{11}^{m-1} p_{12} p_{22}^{n-m-2} (2\pi\sigma^2)^{-(n-1)/2} \exp[-q/(2\sigma^2)] ,$$

where

$$q = \sum_{t=2}^{n} [x_t - (\phi_1 x_{t-1} + \delta_1)]^2 + \sum_{t=m+1}^{n} [x_t - (\phi_2 x_{t-1} + \delta_2)]^2 .$$

In the simplest case the X's are (conditionally) independent, given the γ's. Then $f(x_t|\gamma_t, x_{t-1}, \ldots, x_1, \gamma_1) = f(x_t|\gamma_t)$. We shall pay special attention to this case in the present paper. The p.d.f.'s $f(x|\gamma_t=c)$, $c=1,2,\ldots,k$, are called the class-conditional densities. In the parametric case the class-conditional density takes the form

$$f(x_t|\gamma_t=c) = g(x_t;\beta_c), \tag{2.7}$$

where β is a parameter indexing a family of p.d.f.'s of form given by g.

3. An Algorithm

3.1. Development of the algorithm

The likelihood L is (2.4) or (2.5), considered as a function of the parameters, for fixed $\{x_t\}$. From (2.4), (2.5), and (2.7), the likelihood L can be written in the form

$$L = A(\{p_{cd}\})B(\{\gamma_t\},\{\beta_c\}). \tag{3.1}$$

Hence, for fixed values of the γ's and β's, L is maximized with respect to the p's by maximizing factor A. But $A = \prod_{c=1}^{k} \prod_{d=1}^{k} p_{cd}^{n_{cd}}$. The n_{cd} are specified by the γ's. So from the usual multinomial model, it follows that the maximum likelihood estimates of the p's, for fixed values of the other parameters, are given by

$$\hat{p}_{cd} = n_{cd}/n_c, \tag{3.2}$$

where

$$n_c = n_{c1} + n_{c2} + \ldots + n_{ck}.$$

Further, given the p's and γ's, the estimates of the distributional parameters --the β's--are easy to obtain. This suggests the following algorithm.

Step 0. Set the β's at initial trial values. Set the p's at intial trial values. Set $f(\gamma_1)$ at initial trial values, e.g., $f(\gamma_1) = 1/k$, for $\gamma_1=1,2,\ldots,k$.

Step 1. Estimate γ_1 by maximizing $f(\gamma_1)f(x_1|\gamma_1)$.

Step 2. For $t=2,3,\ldots,n$, estimate γ_t by maximizing

$$p_{\gamma_{t-1}\gamma_t} f(x_t|\gamma_t,x_{t-1},\ldots,x_1).$$

Step 3. Now, having labeled the observations, estimate the distributional parameters, and estimate the transition probabilities by (3.2).

Step 4. If no observations has changed labels from the previous iteration, stop. Otherwise, repeat the procedure from Step 1.

Step 2 is Bayesian classification of x_t, with prior probabilities $p_{\gamma_{t-1}\gamma_t}$. Hence all the techniques for classification in particular models are available (e.g., use of linear discriminant functions when the observations are multivariate normal with common covariance matrix).

3.2. The first iteration

When the k class-conditional processes consist of independent, identically distributed normally distributed random variables with common variance, one can start by choosing initial means and labelling the observations by a minimum-distance clustering procedure. [This is one iteration of ISODATA (Ball and Hall, 1967). One could iterate further at this stage.] From this clustering initial estiamtes of transition probabilities and the variance are obtained. This starting procedure could also be used for fitting AR models by taking the initial trial values of the autoregression coefficients as zero.

3.3. Restrictions on the transitions

As mentioned above, one might wish to place restrictions on the transitions, e.g., to allow transitions only to adjacent states. The model does permit restrictions on the transitions, as discussed above. The maximization is conducted, subject to the condition that the corresponding transition probabilities are zero. This is easily implemented in the algorithm. Once one sets a given transition probability at zero, the algorithm will fit no such transitions, and the corresponding transition probability will remain zero at every iteration.

4. An Example

Here, for a specific numerical example, the problems of fitting the model for a fixed k, choice of k, and prediction will be discussed.

Quarterly gross national product (GNP) in current (non-constant) dollars for the twenty years 1947 to 1966 was considered. (This makes a good size dataset for the current exposition.) Parameters were estimated from the first 19 years, the last four observations (1966) being saved to test the accuracy of predictions. (See Section 4.3.) The data and first differences are given in Table 1. The raw series is nonstationary, so the first differences (increases in quarterly GNP) were analyzed. The notation is

$$x_t = GNP_{t+1} - GNP_t, \quad t = 1,2,\ldots,79;$$

e.g., GNP_1 is the GNP at the end of the quarter 1947-1, GNP_2 is that at the end of 1947-2, and $x_1 = GNP_2 - GNP_1$ is the increase in GNP during the second quarter of 1947. (A negative value of an x indicates a decrease in GNP for the corresponding quarter.) A Gaussian model was used.

4.1. Fitting the model

In this section we discuss the fitting of a model with k=3 classes, discussion of the choice between alternative models being deferred to the next section. The three classes may be taken as corresponding to recession, recovery, and expansion, although some may prefer to think of the segments labeled as recovery as level periods corresponding to peaks and troughs. The approximate maximum likelihood solution found by the iterative procedure was $\hat{\mu}_1 = -1.3$, $\hat{\mu}_2 = 6.2$, $\hat{\mu}_3 = 12.3$, $\hat{\sigma} = 5.194^{1/2} = 2.28$ (the units are billions of current (non-constant) dollars) and

$$\hat{\underline{P}} = \begin{bmatrix} .625 & .250 & .125 \\ .156 & .625 & .219 \\ .039 & .269 & .692 \end{bmatrix} .$$

The estimated labels are given below; labels (r=recession, e=expansion) resulting from fitting k=2 classes (see below) are also given.

t:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
label, k=3:	2	2	3	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	1	2	3
label, k=2:	r	r	e	e	e	e	r	r	r	r	r	e	e	e	e	e	e	e	e	r	r	e	e

24	25	26	27	28	29	30	3]	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
2	2	1	1	1	1	2	2	3	2	2	2	2	2	2	2	2	2	2	1	1	2	3	3	3	3	1
e	r	r	r	r	r	r	e	e	e	e	e	r	r	r	e	e	r	e	r	r	r	e	e	e	e	r

51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
2	3	2	1	1	1	3	3	3	3	3	2	2	2	2	3	3	3	3	3	2	3	3	3	3
e	e	r	r	r	r	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e

The process was in state 1 for 21% of the time, in state 2 for 44% of the time, and state 3 for 35% of the time.

The conventional wisdom regarding recessions during the period of time covered by these data includes the following. [See, e.g., Mansfield (1974), pp. 209-211.] In 1948-1949 (t=4 to 11) there was a reduction of inventory

investment. In 1953-1954 (t=24 to 31) there was a reduction in government expenditures when the Korean conflict came to a close. In mid-1957 to late 1958 (t=42 to 45) an ongoing recession was aggravated by a drop in defense expenditures in late 1957. In 1960 (t=52 to 55) monetary and fiscal authorities had put on the brakes; interest rates had risen substantially during 1958 and 1959.

An interesting feature of the model and the algorithm is that, as the iterations proceed, some isolated labels change to conform to their neighbors. This should be the case when p_{cc} is large relative to p_{cd}, d=1,2,...,k, $d \neq c$.

It is customary to fit an ARI(1,1) model to such data. [See, e.g., Nelson (1973), pp. 64-65.] Hence AR(1)'s were fit within segments in a preliminary analysis of the data. One might expect that segmentation might absorb the autocorrelation. This was in fact found to be the case. The values of the estimated first-order autocorrelation coefficients were not significantly different from zero. Thus the model with conditional independence, given the labels, was used.

4.2. Choice of number of classes

Various values of k were tried, the results being compared by means of Akaike's Automatic (model) identification criterion (AIC). [See. e.g., Akaike (1981).] The AIC for a given model is

$$AIC = -2 \log_e \hat{L} + 2p,$$

where $\hat{L}$ is the maximized value of the likelihood and p is the number of parameters in the model. According to AIC, inclusion of an additional parameter in a model is appropriate if $\log_e \hat{L}$ increases by one unit or more, i.e., if $\hat{L}$ increases by a factor of e or more.

The model was fit with several values of k and unrestricted transition probabilities. Also, since it seems reasonable to restrict the transitions to those between adjacent states, these models were evaluated as weel. In the case of k=3, where the states might be considered as recession, expansion, and recovery, this means setting equal to zero the transition probabilities corresponding to the transitions, recession-to-expansion and expansion-to-recession. Also, by way of comparison, the ARI(1,1) model

$$x_t = \phi x_{t-1} + \delta + u_t, \quad x_t = GNP_{t+1} - GNP_t,$$

was fit. The IID model of independent and identically distributed Gaussian observations was fit also, just for comparison. The results are given in Table 2. The best segmentation model, as indicated by minimum AIC, is that with only two classes. [The AIC for AR(1,1) was even lower.] The AIC for the IID model was quite large, indicating a very poor fit, as would be expected.

The results for k=2 classes (which might be labeled recession, expansion) were $\hat{\mu}_1 = 0.43$, $\hat{\mu}_2 = 10.09$, $\hat{\sigma} = 3.306$, and

$$\underline{\hat{P}} = \begin{bmatrix} .667 & .333 \\ .170 & .830 \end{bmatrix}.$$

The process was in state 1 for 37% of the time and class 2 the other 63% of the time. The labels were given above.

A model with only two classes enjoys advantages relating to its relative simplicity.

4.3. Prediction

If there is feedback, in the sense that γ_t becomes known before x_{t+1} is to be predicted, then, given $\gamma_t = c$, one can give the prediction

$$\begin{aligned} \hat{x}_{t+1}|\gamma_t=c &= \hat{\mu}_1 \text{ with probability } \hat{p}_{c1} \\ &= \hat{\mu}_2 \text{ with probability } \hat{p}_{c2} \\ &\vdots \\ &= \hat{\mu}_k \text{ with probability } \hat{p}_{ck}. \end{aligned}$$

In this example this gives rise to a "recession probability," $\hat{p}_{c1}$, reminiscent of the "precipitation probability" of meteorology.

Similarly, one has

$$\begin{aligned} \hat{x}_{t+h}|\gamma_t=c &= \hat{\mu}_1 \text{ with probability } \hat{p}_{c1}^{(h)} \\ &= \hat{\mu}_2 \text{ with probability } \hat{p}_{c2}^{(h)} \\ &\vdots \\ &= \hat{\mu}_k \text{ with probability } \hat{p}_{ck}^{(h)}, \end{aligned}$$

where $\hat{p}_{cd}^{(h)}$ is the natural estimate of the k-step c-to-d transition probability, the c,d-th element of the h-th power of $\hat{\underline{P}}$.

These are vector estimates, with probabilities attached to the elements of the vector. A scalar estimate is given by $\sum_{d=1}^{k} \hat{p}_{cd}^{(h)} \hat{\mu}_d$, for any h = 1,2,... .

Now let us consider prediction based on the model with k=3 classes, fit in Section 4.1. We predict x_{76}, x_{77}, x_{78}, and x_{79}. Consider first the prediction of x_{76}. If, before it had to be predicted, one had been sure, due to the accumulation of information on various economic indicators, that the process had then been in an expansion (state 3), then the relevant estimated transition

probabilities would be .039, .269, and .692, for transitions from state 3 to states 1, 2, and 3, respectively. One would make the prediction

$$\hat{x}_{76}|\gamma_{75}=3 = -1.3 \pm 2.338 \text{ with probability .039}$$
$$= 6.2 \pm 2.318 \text{ with probability .269}$$
$$= 12.3 \pm 2.321 \text{ with probability .692,}$$

where the numbers after ± are approximate standard errors of prediction, namely, $[\sigma^2(1 + 1/n_c)]^{1/2}$, c = 1,2,3, or, since the numbers of observations assigned to the three groups were $n_1 = 19$, $n_2 = 29$, and $n_3 = 27$, $[5.194(1 + 1/19)]^{1/2} = 2.338$, $[5.194(1 + 1/29)]^{1/2} = 2.318$, and $[5.194(1 + 1/27)]^{1/2} = 2.321$. The actual value of x_{76} was 19.5, a very strong gain in GNP for that period, certainly consistent with a prediction of "expansion." The values of $\hat{x}_t|\gamma_{t-1}=3$, t = 77, 78, 79, are the same as those for $\hat{x}_{76}|\gamma_{75}=3$.

The difference 19.5 - 12.3 = 7.2 is rather large. However, the fitted ARI(1,1) model $x_{t+1} = 0.597x_t + 2.64$, with $\hat{\sigma} = 4.95$, also made a large error for this quarter. It gave a prediction of 13.9, with an approximate standard error of prediction of 4.95. Its successive forecasts for the last three quarters represented in the data set, in each case using the observation from the previous quarter, were 14.3, 10.9, and 10.2, each with standard error of prediction equal to 4.95, compared with actual results of 13.8, 12.6, and 14.8, respectively.

Now let us consider prediction more than one period ahead. Given information for t=75, we predict x_{77}, x_{78}, and x_{79}. Using the third row of the second, third and fourth powers of $\hat{\underline{P}}$, one finds

$$\hat{x}_{77}|_{Y_{75}=3} = -1.3 \text{ with prob. } .093$$
$$= 6.2 \text{ with prob. } .364$$
$$= 12.3 \text{ with prob. } .543,$$

$$\hat{x}_{78}|_{Y_{75}=3} = -1.3 \text{ with prob. } .136$$
$$= 6.2 \text{ with prob. } .397$$
$$= 12.3 \text{ with prob. } .467,$$

$$\hat{x}_{79}|_{Y_{75}=3} = -1.3 \text{ with prob. } .165$$
$$= 6.2 \text{ with prob. } .408$$
$$= 12.3 \text{ with prob. } .427.$$

We have

$$\lim_{h\to\infty} \hat{x}_{t+h}|_{Y_t=c} = -1.3 \text{ with prob. } .211$$
$$= 6.2 \text{ with prob. } .411 \qquad (4.1)$$
$$= 12.3 \text{ with prob. } .378,$$

independent of c and t, because (.211, .411, .378) is the estimated long-run distribution across the states. The predictions of ARI(1,1) are

$$\hat{x}_{77}|_{x_{75}} = 10.9, \text{ std. err. } = 5.76,$$
$$\hat{x}_{78}|_{x_{75}} = 9.2, \text{ std. err. } = 6.03,$$
$$\hat{x}_{79}|_{x_{75}} = 8.1, \text{ std. err. } = 6.12.$$

By way of comparison with (4.1), in the long run, these forecasts from ARI(1,1) tend to 6.55, the estimated mean of the process, with an estimated standard error of 6.167, the estimated standard deviation of the x's.

Acknowledgement. This research was supported by Office of Naval Research Contract N00014-80-C-0408, Task NR042-443, at the University of Illinois at Chicago Circle.

References

Akaike, Hirotugu (1981). "Likelihood of a Model and Information Criteria." Journal of Econometrics 16, 3-14.

Ball, Geoffrey H, and Hall, David J.(1967). "A Clustering Technique for Summarizing Multivariate Data." Behavioral Science 12, 153-155.

Mansfield, Edwin (1974). Economics: Principles, Problems and Decisions. W. W. Norton & Co., New York.

Nelson, Charles R. (1973). Applied Time Series Analysis for Managerial Forecasting. Holden-Day, Inc., San Francisco.

Table 1. GNP. Units: billions of current (non-constant) dollars (from Nelson (1973), pp. 100-101)

	Quarter							
	1	2	3	4	1	2	3	4
1947-48 GNP	224	228	232	242	248	256	263	264
ΔGNP		4.0	4.2	10.3	5.9	7.6	6.9	1.4
1949-50 GNP	259	255	257	255	266	275	293	305
ΔGNP		-3.3	1.9	-2.1	11.0	9.4	17.7	11.4
1951-52 GNP	318	326	333	337	340	339	346	358
ΔGNP		7.8	7.0	4.1	2.6	-0.4	6.5	12.1
1953-54 GNP	364	368	366	361	361	360	365	373
ΔGNP		3.3	-1.7	-5.0	-0.1	-0.3	4.3	8.7
1955-56 GNP	386	394	403	409	411	416	421	430
ΔGNP		8.2	8.1	6.3	1.8	5.6	4.4	8.9
1957-58 GNP	437	440	446	442	435	438	451	464
ΔGNP		3.0	6.4	-4.8	-6.8	3.6	13.1	13.0
1959-60 GNP	474	487	484	491	503	505	504	503
ΔGNP		12.9	-2.9	6.5	12.5	1.7	-0.5	-0.9
1961-62 GNP	504	515	524	538	548	557	564	572
ΔGNP		11.3	9.3	13.5	10.1	9.4	7.2	7.6
1963-64 GNP	577	584	595	606	618	628	639	645
ΔGNP		6.8	10.5	11.1	11.9	10.3	10.9	6.2
1965-66 GNP	663	676	691	710	730	743	756	771
		12.9	15.4	18.9	19.5	13.8	12.6	14.8

Table 2. Fitting models. (See Section 4.2.)

Model	AIC
Segmentation, 2 classes	481.4[a]
Segmentation, 3 classes, full trans. prob. matrix	483.6
Segmentation, 3 classes, sparse trans. prob. matrix[b]	488.5[-]
Segmentation, 4 classes, full trans. prob. matrix	507.1
Segmentation, 4 classes, sparse trans. prob. matrix[b]	486.8
Segmentation, 5 classes, full trans. prob. matrix	506.5[+]
Segmentation, 5 classes, sparse trans. prob. matrix[b]	stopped[c]
Segmentation, 6 classes, full trans. prob. matrix	stopped[c]
AR(1)[d]	453.2[e]
IID[f]	1721.4

a. Optimum, among segmentation models considered.

b. Allows transitions only to adjacent states.

c. Stopped, i.e., the algorithm reached an iteration where it allocated no observations to one of the classes.

d. AR(1) for the differences, i.e., ARI(1,1) for the original series.

e. Optimum, among all models considered.

f. Observations treated as a random sample from a normal distribution.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report 81-1 — 2. GOVT ACCESSION NO.: AD-A109479	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): On Segmentation of Time Series	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Stanley L. Sclove	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0408
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Mathematics, University of Illinois at Chicago Circle, Box 4348, Chicago, IL 60680	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: —	12. REPORT DATE: November 30, 1981
	13. NUMBER OF PAGES: 16 + ii
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Office of Naval Research, Statistics and Probability Branch, Arlington, VA 22217	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Unlimited distribution

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecasting, prediction; signal analysis; cluster analysis; classification; isodata procedure; Markov chains; maximum likelihood; Akaike's automatic (model) identification criterion (AIC)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of partitioning a time-series into segments is considered. The segments fall into classes, which may correspond to phases of a cycle (recession, recovery, expansion in the business cycle) or to portions of a signal obtained by scanning (background/clutter, target, background/clutter again, another target, etc.; or normal tissue, tumor, normal tissue).

(cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

(Abstract, continued)

Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov chain. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. A numerical example is given. Choice of the number of classes, using Akaike's automatic (model) identification criterion (AIC), is illustrated. Prediction is considered.