Some Experiences with Detection and Diagnosis of Model Parameter and Variance Changes

Theodor D. Popescu

Abstract—The problem of change detection and data segmentation has received considerable attention in a research context and appears to be the central issue in various application areas. The change detection and segmentation model used in this paper is the simplest extension of the linear regression models to data with abruptly changing properties. In the first part of the paper we give a general view on the main techniques used in change detection and segmentation: filtering techniques with a whiteness test and techniques based on sliding windows and distance measures. A new algorithm based on a likelihood technique, when sliding windows are used, for diagnosis of model parameter and variance changes is then presented. The results of some Monte-Carlo simulation for detection and diagnosis of model parameter and variance changes are included in the paper.

Keywords—Change detection, Diagnosis, Regression models, Decision making, Distance measure, Likelihood techniques.

I. INTRODUCTION

The problem of change detection or segmentation of data has received considerable attention during the last two decades in a research context and appears to be the central issue in various application areas.

The analysis of the behavior of such real data reveals the most of the changes that occur are either changes in the mean level, or changes in spectral characteristics. In this framework, the problem of segmentation between "homogenous" parts of the data (or detection of changes in the data) arises more or less explicitly.

The proposed problem formulation assumes the off-line or batch-wise data processing, although the solution is sequential in data and an on-line data processing can be used. The change detection and segmentation model is the simplest possible extension of linear regression models to data with abruptly changing properties. It is assumed that the data can be described by one linear regression model within each segment with distinct parameter vector and noise variance.

II. PROBLEM FORMULATION

The following problem is addressed: Let \( \{Y_0\} \) and \( \{Y_1\} \) two sets of stationary data, and one want to test the null hypothesis:

- \( H_0 \): \( \{Y_0\} \) and \( \{Y_1\} \) are generated by the same law.
- \( H_1 \): \( \{Y_0\} \) and \( \{Y_1\} \) are generated by different laws.

Concerning the data generating mechanism, it is assumed that under \( H_0 \), data sets \( \{Y_0\} \) and \( \{Y_1\} \) are generated by a linear regression process, whose parameters may jump at some unknown time, i.e.,

\[
y_t = \phi_0^T \theta + \epsilon_t, \quad E(\epsilon_t^2) = \sigma
\]

with different forms for \( \phi_0 \) and \( \theta \), depending on the model type (AR, ARX, ARMA, ARMAX, FIR, etc) and where

\[
\begin{align*}
\theta &= \theta_0, \quad \sigma = \sigma_0, \quad \text{for } t < t_0 \\
\theta &= \theta_1, \quad \sigma = \sigma_1, \quad \text{for } t \geq t_0
\end{align*}
\]

and \( \epsilon_t \) is the noise sequence.

This assumption is not too restrictive since many stationary processes encountered in practice can be closely approximated by such models. The advantage of this assumption consists in computational simplicity of the resulted test procedures.

The change detection problem consists in the sequential detection of the change, and the estimation of the change time, \( t_0 \), with few false alarms, short delay for detection and symmetrical detection (comparable performances when detecting a change from model (1) to model (2), or inverse).

There are different approaches to detect the changes in non-stationary signals. In this paper we give the conceptual description of some methods for sequential detection of changes in non-stationary data, based on filtering and a whiteness test, sliding windows and distance measures and for diagnosis of model parameter and variance changes, when a likelihood technique is used. The Monte-Carlo simulation results are presented only for the last approach.

III. CHANGE DETECTION BASED ON FILTERING

One useful approach for change detection consists in filtering of the observed data through a known or identified filter, and in looking for changes in the residual signal of innovations, \( \{\epsilon_t\} \). Actually, the use of cusum techniques based upon the innovations (one-step prediction errors), \( \{\epsilon_t\} \), or the squared innovations, \( \{\epsilon_t^2\} \), is a standard approach for change detection in linear regression models. Such a technique, using \( \{\epsilon_t^2\} \) is based upon the fact that, before the change \( E(\epsilon_t^2) = \sigma_0 \) and thus: \( E(\epsilon_t^2/\sigma_0 - 1) = 0 \).

To conclude, statistical whiteness tests can be used to test if the residuals are white noise as they should be if there is no change. Fig. 1 shows the basic structure, where the filter residuals are transformed to a distance measure, that measures...
the deviation from the no-change hypothesis. The stopping rule decides whether the deviation is significant or not. The most natural distances are listed below, [1]:

- Change in the mean. The residual itself is used in the stopping rule and \( s_t = \varepsilon_t \).
- Change in variance. The squared residual substraction by a known residual variance \( \sigma \) is used and \( s_t = \varepsilon_t^2 - \sigma \).
- Change in correlation. The correlation between the residual and past outputs and/or inputs are used and \( s_t = \varepsilon_t y_{t-k} \) or \( s_t = \varepsilon_t u_{t-k} \) for some \( k \).
- Change in sign correlation. For instance, one can use the fact that the white residuals should change sign every second sample in the average and use \( s_t = \text{sign}(\varepsilon_t\varepsilon_{t-1}) \).

The main problem in statistical change detection is now to decide what "large" are these distances. Many change detection algorithms can be recast into the problem of deciding on the following two hypotheses:

\[
H_0 : E(s_t) = 0, \\
H_1 : E(s_t) > 0,
\]

where \( s_t \) is a distance measure. A stopping rule is essentially achieved by low-pass filtering \( s_t \) and comparing this value to a threshold. Below, two such low-pass filters are given:

- The CUmulative SUM (CUSUM) test of Page, [2]:
  \[
  g_t = \max(g_{t-1} + s_t - \nu, 0), \quad \text{change if } g_t > h.
  \]
  The drift parameter \( \nu \) influences the low-pass effect, and the threshold \( h \) (and also \( \nu \)) influences the performance of the detector.

- The Geometric Moving Average (GMA) test in Roberts, [3].
  \[
  g_t = \lambda g_{t-1} + (1 - \lambda) s_t, \quad \text{change if } g_t > h.
  \]
  Here, the forgetting factor \( \lambda \) is used to tune the low-pass effect, and the threshold \( h \) is used to tune the performance of the detector. Using no forgetting at all (\( \lambda = 0 \)), corresponds to directly thresholding, which is one option.

It seems that classical approach consisting in testing how much the sequence of innovations, \( \{\varepsilon_t\} \) is far from hypothesis "zero-mean white noise" is not sufficient for change detection in practice.

IV. CHANGE DETECTION BASED ON SLIDING WINDOWS

The main idea underlying this approach consists in comparison of two models: a model (\( M_1 \)), based on data from a sliding window of size \( L \) (\( y_{t-L+1}, \ldots, y_t \)) is compared to a model (\( M_0 \)) based on all data or a substantially larger sliding window (\( y_1, y_2, \ldots, y_L \)), [4]. If the model based on the larger data window gives larger residuals

\[
\|\varepsilon_t^0\| > \|\varepsilon_t^1\| \tag{2}
\]

then a change is detected. The problem here is to choose a norm that corresponds to a relevant statistical measure. Some norms that have been proposed are:

- The Generalized Likelihood Ratio (GLR).
- The divergence test.
- Change in spectral distance. There are many methods to measure the distance between two spectra. One approach would be to compare the spectral distance of two models.

These criteria provide an \( s_t \) to be put into a stopping rule for instance, the CUSUM test. The choice of window size \( L \) is very critical here. On the one hand, a large value is needed to get an accurate model in the sliding window and, on the other hand, a small value is needed to get quick detection.

Concerning the distance functions presented above, we will give in the following their expressions.

In [5], two different test statistics for the case of two different models are given. A straightforward extension of the generalized likelihood ratio test leads to:

\[
d_{GLR} = L \log \frac{\sigma_0}{\sigma_1} + \frac{\| y_t - \phi_1^T \theta_0 \|^2}{\sigma_0} - \frac{\| y_t - \phi_1^T \theta_1 \|^2}{\sigma_1} \tag{3}
\]

This test statistic was as the same time proposed in Appel and Brandt, [6] and will be referred as Brandt’s GLR method.

To measure the distance between two models, any norm can be used. So, the Kullback discrimination information, [7] between two probability density functions \( p_1 \) and \( p_2 \) is defined as:

\[
I(1,2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx \geq 0 \tag{4}
\]

In the special case of Gaussian distribution, we get

\[
p_1(x) = N(\hat{\theta}_1, P_1) \Rightarrow I(1,2) = \frac{1}{2} \text{tr}(P_2^{-1}P_1 - I) + \frac{1}{2} \text{tr}(\hat{\theta}_1 - \hat{\theta}_2)^T P_2^{-1}(\hat{\theta}_1 - \hat{\theta}_2) - \frac{1}{2} \log \left( \frac{\det P_1}{\det P_2} \right)
\]

The Kullback information is not a norm (it is not symmetric) and is not suitable as a distance measure. Instead, Kullback divergence is used:

\[
V(1,2) = I(1,2) + I(2,1) \geq 0 \tag{5}
\]

From Kullback divergence, the divergence test can be derived and it equals:

\[
d_{DIV} = L \left( \frac{\sigma_0}{\sigma_1} - 1 \right) + \left( 1 + \frac{\sigma_0}{\sigma_1} \right) \frac{\| y_t - \phi_1^T \theta_0 \|^2}{\sigma_0} - \frac{2(y_t - \phi_1^T \theta_0)(y_t - \phi_1^T \theta_1)}{\sigma_1} \tag{6}
\]

The corresponding algorithm will be called the divergence test. \( d_{GLR} \) and \( d_{DIV} \) start to grow when a jump produced, and again the task of the stopping rule is to decide whether the growth is significant.
Concerning the parameter estimation of the models can be used the lattice implementation of the approximate least squares method, [8], for the long-term filter $M_0$, and the covariance method, [9], for the current filter $M_1$.

V. DIAGNOSIS OF PARAMETER AND VARIANCE CHANGES

A problem that frequently appears in practice, is to determine whether an abrupt change in the model parameters or noise variance has occurred at the change instant, $t_0$. For this the following signal model can be used:

$$y_t = \begin{cases} \phi_t^T \theta_0 + e_t, & E[e_t] = \lambda_0 \sigma_t \quad \text{if } t \leq t_0 \\ \phi_t^T \theta_1 + e_t, & E[e_t] = \lambda_1 \sigma_t \quad \text{if } t \geq t_0 \end{cases}$$  \hspace{1cm} (7)

where $\sigma_t$ is a known time-varying noise variance, and $\lambda$ is either a scaling of the noise variance or the variance itself ($\lambda t = 1$). Neither $\theta_0$, $\theta_1$, $\lambda_0$ or $\lambda_1$ are unknown. The following hypotheses are used:

- $H_0$: $\theta_0 = \theta_1$ and $\lambda_0 = \lambda_1$
- $H_1$: $\theta_0 \neq \theta_1$ and $\lambda_0 = \lambda_1$
- $H_2$: $\theta_0 = \theta_1$ and $\lambda_0 \neq \lambda_1$

The sufficient statistics from the filters are given bellow:

<table>
<thead>
<tr>
<th>Data</th>
<th>$y_1, y_2, \ldots, y_{t-L}, y_{t-L+1}, \ldots, y_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$M_0, M_1$</td>
</tr>
<tr>
<td>Time interval</td>
<td>$T_0, T_1$</td>
</tr>
<tr>
<td>RLS quantities</td>
<td>$\hat{\theta}_0, \hat{\theta}_1, P_0, P_1$</td>
</tr>
<tr>
<td>Loss function</td>
<td>$V_0, V_1$</td>
</tr>
<tr>
<td>Number of data</td>
<td>$n_0 = t - L, n_1 = L$</td>
</tr>
</tbody>
</table>

where $P_j, j = 0, 1$ denotes the covariance of the parameter estimate achieved from the RLS algorithm. The loss functions are defined by

$$V_j(\theta) = \sum_{k \in T_j} (y_k - \phi_k^T \theta)(\lambda_j \sigma_k)^{-1}(y_k - \phi_k^T \theta), \quad j = 0, 1.$$  \hspace{1cm} (10)

It makes sense to compute $V_1(\hat{\theta}_0)$ to test how the first model performs on the new data set. The maximum likelihood approach is stated in the slightly more general maximum a posteriori approach, where the prior probabilities $q_i$ for each hypothesis can be incorporated. The exact a posteriori probabilities

$$l_i = 2 \log p(H_i | y_1, y_2, \ldots, y_t), \quad i = 0, 1, 2$$  \hspace{1cm} (11)

are derived by Gustafsson [1]. So for the signal model (7) and hypothesis given in (8), the prior distribution for $\lambda$ taken as inverse Wishart $\lambda$ and the prior for the parameter vector

$\theta \in N(0, P_0)$, with the loss function (10) and the least squares estimation, the a posteriori probabilities are approximately given by:

$$l_0 \approx (n_0 + n_1 - 2 + m) \log \left( \frac{V_0(\hat{\theta}_0) + V_1(\hat{\theta}_1) + \sigma}{n_0 + n_1 - 4} \right)$$
+ \log \det(P_0^{-1} + P_1^{-1}) + 2 \log(q_0), \quad (12)

$$l_1 \approx (n_0 + n_1 - 2 + m) \log \left( \frac{V_0(\hat{\theta}_0) + V_1(\hat{\theta}_1) + \sigma}{n_0 + n_1 - 4} \right)$$
- \log \det(P_0 - \log \det P_0 + 2 \log(q_1), \quad (13)

$$l_2 \approx (n_0 - 2 + m) \log \left( \frac{V_0(\hat{\theta}_0) + \sigma}{n_0 - 4} \right)$$
+ (n_1 - 2 + m) \log \left( \frac{V_1(\hat{\theta}_0) + \sigma}{n_1 - 4} \right)
- 2 \log \det P_0 + 2 \log(q_2). \quad (14)

The last three equations are used in decision making concerning one of the three hypotheses presented above.

VI. EXPERIMENTAL RESULTS

The presented results are obtained by Monte Carlo simulation for a second order FIR model with 4 change times. After the first and fourth jump the model parameters change, while after the second and third jump the noise variance changes. Three experiments were performed. The following model structure was used:

$$y_t = \theta_1 * u_{t-1} + \theta_2 * u_{t-2} + \lambda * e_t$$  \hspace{1cm} (15)

where $u_t$ and $e_t$ are random sequences of zero mean and variance 1. The model parameters and $\lambda$ values for the first experiment are given in the Table I.

The signal resulted by simulation and the real change instants are given in Fig. 2. We present in Fig. 3 the model parameter and variance estimates obtained in a Monte Carlo simulation for 1000 noise realization. The resulted histogram are given in Fig. 4.

A similar experiment was performed for new values of noise variance given in Table II.

The signal resulted by simulation and the real change instants are given in Fig. 5, while in Fig. 6 the model parameter and variance estimates obtained in a Monte Carlo simulation for 1000 noise realization are presented. The resulted histogram is given in Fig. 7.
Fig. 2. The signal and the real change instants

Fig. 3. Model parameter and variance estimates obtained in Monte Carlo simulation

Fig. 4. The histogram for change detection instants in model parameters and variance of the noise

FIG. 5. The signal and the real change instants

FIG. 6. Model parameter and variance estimates obtained in Monte Carlo simulation

FIG. 7. The histogram for change detection instants in model parameters and variance of the noise

TABLE II
MODEL PARAMETERS AND λ VALUES

<table>
<thead>
<tr>
<th>Time</th>
<th>1-50</th>
<th>51-100</th>
<th>101-150</th>
<th>151-200</th>
<th>201-250</th>
</tr>
</thead>
<tbody>
<tr>
<td>θ₁</td>
<td>1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>λ</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
TABLE III

<table>
<thead>
<tr>
<th>Time</th>
<th>1-50</th>
<th>51-100</th>
<th>101-150</th>
<th>151-200</th>
<th>201-250</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta_1)</td>
<td>1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>-1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The last experiment in this case was performed for the following model parameter and noise variance values given in Table III.

The signal resulted by simulation and the real change instants are given in Fig. 8. We present in Fig. 9 the model parameter and variance estimates obtained in Monte Carlo simulation for 1000 noise realization. The resulted histogram is given in Fig. 10.

VII. CONCLUSIONS

The paper gives the conceptual description of some change detection and data segmentation methods based on filtering, sliding windows and likelihood techniques and presents some Monte-Carlo simulation for diagnosis of model parameter and variance changes of a second order FIR model. Based on the obtained results it can be noted that the performances of these methods are superior to the other methods.

REFERENCES


Theodor D. Popescu was born in Rosiori de Vede, Romania, on July 4, 1949. He received M.Sc. degree and Eng.Sc.D. (PhD) degree, both in Automatic Control, from "Politehnica" University of Bucharest in 1972 and 1983, respectively. He is with National Institute for Research and Development in Informatics, Bucharest, where is senior research scientist of first degree. The research activity was mainly oriented in the fields of system identification and parameter estimation, time series analysis, signal processing, change detection and diagnosis in signals and systems. He is author or co-author of more than 120 papers and of five books. He has been awarded with "Tudor Tanasescu" Prize of the Romanian Academy on 1995 for the works in the field of detection of changes in dynamics of signals and systems and General Association of Engineers in Romania Prize on 2000 in the field of Information Technology. He was granted by The Royal Society (1992), DAAD (1992, 1995, 1999), JSPS (1996, 1999), Australian Research Council(2002), CERGE-EI (2002). He is Senior Member of IEEE since 2002.