Abstract - We introduce in this article an optimal segmentation method of nonstationary random processes. Segmentation of a non stationary process consists in assuming piecewise stationarity and in detecting the instants of change. We consider here that all the data from all the sensors are available in a same time and perform a global segmentation. The bayesian fusion method we propose for the segmentation is based on the introduction of a joint prior model for the simultaneously segmentation and estimation of data coming from a set of sensors. We build a change process and define its prior distribution for the data fusion. That allows us to propose the MAP estimate as well as some minimum contrast estimate as a solution. We define, in the parametric processes distribution case, the expression and signification of all the segmentation’s parameters. We compare the performance of our detection method in the case of two or three sensor. Application to the fusion of wind data velocity and direction is proposed.

Keywords: Detection of changes, MAP estimate, Fusion.

1 Introduction

Data fusion is one of the active area of research in many applications such as nondestructive testing, geophysical signal processing, medical imaging, radio-astronomy, etc...This work is part of a global study on a surveillance system of the industrial reject. The dispersion of airborne material is mainly due to turbulent diffusion inherent in atmospheric motion. For realistic estimates of dispersion, it is therefore of primary importance to have an accurate description of atmospheric turbulence. The standard deviation of horizontal wind direction, and the mean direction and velocity of the vector wind give accurate estimates of the spread dispersion of airborne material. The fusion of this two signals during the segmentation gives better results. Our main object in this paper is not to focus on these application. Indeed, we want to show how we use classical probabilistic Bayesian approach to do data fusion.

We consider in this paper the segmentation problem of nonstationary processes. We assume that the processes are piecewise stationary and we want to detect the instants of changes as well as the statistics parameters of the processes. We consider here that all the data from all the sensors are available in a same time and perform a global segmentation. We shall assume that the distributions of the processes depend on parameters. Thus, the problem consists now of detecting changes of the parameters. Changes can affect, for example, the mean and the covariance structure of the process. The detection is off line and the criteria of good recovery will be only related to the detection errors.

We use a random change process that takes the value 1 at the change instants and is zero between two changes. The simultaneous segmentation and estimation of the process parameters are computed by maximising the a posteriori distribution of the instants of change, conditionaly to the parameters estimation [1]. The fusion of the processes is introduced in the prior probability definition of the change process. Then we propose a MAP estimate as well as some minimum contrast estimate as a solution of the segmentation.

In section 2 the bayesian model fusion is described. Section 3 defines the parameters we use for the changes detection in the mean and in the variance of parametrics processes models. The performance of the detection and numerical experiment on real data are presented in section 4.

2 The bayesian model fusion

Let $Z^i = \{Z^i_j\}_{j \in \mathbb{N}}$, $1 \leq i \leq p$, a set of non stationary real processes. Let $Z = \{Z^1, \ldots, Z^p\}$ a p-dimensional real process. We assume that the $Z^i$, $1 \leq i \leq p$ are piecewise stationary. Then there exists instants $(t^i_k, k \geq 0)$ such that $(Z_{t^i_k}, \ldots, Z_{t^i_{k+1}})$ is stationary for all $k \in \mathbb{N}$, and $1 \leq i \leq p$. The segmentation problem consists in detecting the changes in the distribution of the process $Z^i$. We assume that the distribution of the process depends on parameter $\theta^i$. Thus, the problem consists of detecting the changes of $\theta^i$. For example the changes can effect the mean and the covariance of the process. We assume that all the data is available, the detection is off line. The detection of change is seen here as the global segmentation of the vectors $Z^n_i = (Z^i_1, \ldots, Z^i_n)$, $1 \leq i \leq p$ in stationary pieces.

We shall introduce the random vectors $R^i = (R^i_1, \ldots, R^i_n)$, $1 \leq i \leq p$ that are defined by:

\[
\begin{align*}
R^i_j &= 1 \text{ if there exists } k \text{ such that } j = t^i_k \\
R^i_j &= 0 \text{ otherwise }
\end{align*}
\]

Let $S_p$, $1 \leq \rho \leq p$ be the number of $\rho$ simultaneously changes on the processes $\{Z^i\}_{1 \leq i \leq p}$ (for $\rho = 1$ $S_p$ is the...
number of nonsimultaneously changes). When a particular realisation $r^i$ of the vector $R_i$ is observed we define $\Pi(r^1, \ldots, r^p) = P_i(R_1 = r^1, \ldots, R_p = r^p)$ the prior probability of having the configuration $(r^1, \ldots, r^p)$. We define $\{R^i\}_{1 \leq i \leq p}$ as a sequence of independent Bernoulli variables. This probability will be described next.

The MAP estimate is obtained by maximising the conditional probability $P_i(R_1 = r^1, \ldots, R_p = r^p) = \frac{P_i(r^1, \ldots, r^p)}{P_i(Z^1, \ldots, Z^p)}$. In the case of parametric models, the distribution of the process $\{Z^i\}_{1 \leq i \leq p}$ depends on parameters whose values remain constant in each stationary pieces. For a given configuration $(r^1, \ldots, r^p)$ of $(K^1, \ldots, K^p)$ segments, let $\theta^i = (\theta^i_1, \ldots, \theta^i_{K^i})$, $1 \leq i \leq p$ be the sequences of parameters, such that $\theta^i_k$ is the parameter in the $k$th segment of the $i$th process. Then $\{\theta^i\}_{1 \leq i \leq p}$ and $\{R^i\}_{1 \leq i \leq p}$ can be estimated simultaneously, by maximising the a posteriori distribution of $\{R^i\}_{1 \leq i \leq p}$.

$$\hat{\pi} = \arg \max_{\theta^i} P_i(r^1, \ldots, r^p) \cdot h(z^1 \mid r^1, \theta^1) \ldots h(z^p \mid r^p, \theta^p) \cdot \Pi(r^1, \ldots, r^p)$$

where $\hat{\theta}$ is the estimation of $\theta^i$. In the previous expression we consider the data, coming from the sensors, independent. The probability $\Pi(r^1, \ldots, r^p)$ will join in the data fusion. Then the $\{\hat{\pi}\}_{1 \leq i \leq p}$ are obtained with the expression.

$$\hat{\pi} = \arg \max_{(r^1, \ldots, r^p) \in \{0,1\}} h(z^1 \mid r^1, \hat{\theta}(r^1)) \ldots h(z^p \mid r^p, \hat{\theta}(r^p)) \cdot \Pi(r^1, \ldots, r^p)$$

For a given configuration of change $r^i$, the maximum likelihood estimate of $\theta^i_k$ in the segment $k$ as :

$$\hat{\theta}_k^i(r^i) = \arg \max_{\theta_k} l(z^i_{t^i_k-1+1}, \ldots, z^i_{t^i_k}; \theta_k)$$

where $l(\ldots)$ is the log-likelihood. We assume that the different segments are independent, then the estimation of the vector $\theta^i$ on all the data is done by maximising the expression:

$$l(r^i, \theta^i) = \sum_{k=1}^{s^i} l(z^i_{t^i_k-1+1}, \ldots, z^i_{t^i_k}; \theta^i)$$

where $s^i$ is the number of changes in the process $Z^i$. All the $\{\hat{\pi}\}_{1 \leq i \leq p}$ are estimated in the same way and the expression 3 can be written as :

$$\hat{\pi} = \arg \max_{(r^1, \ldots, r^p) \in \{0,1\}} h(z^1 \mid r^1, \hat{\theta}(r^1)) \ldots h(z^p \mid r^p, \hat{\theta}(r^p)) \cdot \Pi(r^1, \ldots, r^p) - \ln(P_i(r^1, \ldots, r^p)).$$

The first term of this expression is related to the fit of the observation $\{z^i\}_{1 \leq i \leq p}$. Whereas the second term is related to the number of changes. The fusion consists in favouring the simultaneously changes on the signals. Then the problem leads in the definition of the joint probability $\Pi(r^1, \ldots, r^p)$ that binds the changes on the signals.

We define $\Pi(r^1, \ldots, r^p)$ as the probability : $\Pr(S_0$ changes) $\times \Pr(S_1$ simultaneous change and ... and $S_p$ simultaneous change/or $S_0$ changes), with $S_0 = \sum_{i=1}^{p} s^i$. This probability can be expressed as :

$$\Pi(r^1, \ldots, r^p) = \lambda^{S_0} \cdot (1 - \lambda)^{p - S_0} \cdot P_1 s^1 \times P_2 s^2 \times \ldots \times P_p s^p \cdot P_0 (\sum_{i=1}^{p} s^i),$$

(7)

where $\lambda$ is a real parameter between 0 and 1. $P_i s^i$ is the probability to have $S_i$ simultaneous change on the $p$ process. $P_0$ is the probability to have no change. Then the expression 6 is becoming :

$$\hat{\pi} = \arg \max_{(r^1, \ldots, r^p) \in \{0,1\}} h(z^1 \mid r^1, \hat{\theta}(r^1)) \ldots h(z^p \mid r^p, \hat{\theta}(r^p)) \cdot \Pi(r^1, \ldots, r^p) - \ln(P_i(r^1, \ldots, r^p)).$$

(8)

where $\beta = \ln(\lambda) / \lambda$ and $\beta_1 = \ln(\lambda^2) / \lambda$ ... and $\beta_p = \ln(\lambda^p)$. For the fusion of the processes we suppose that the probability to have $K$ simultaneous changes is greater than the probability to have $K-1$ simultaneous change. This supposition in term of probability will influence the segmentation. Then the determination of these parameters must be done under the constraints :

- To favour the simultaneously changes we must have :
  $$\beta_p < \beta_{p-1} < \ldots < \beta_1$$

- Normalisation :
  $$\sum_{i=0}^{p} P_i = 1$$

- When the expression 8 is minimum the augmentation of the number of simultaneously changes mustn’t decrease this minimum. Then we must have :
  $$\beta \geq \beta_i - \beta_{i-1} \forall i \in \{2, \ldots, p\}$$

- We must have $\beta > 0$ then $\lambda > \frac{1}{p}$.

We will show in the next section that for a parametric model, $\beta$ controls the resolution level of the segmentation and its expression depends on the position between two changes. We will show also the influence of the parameters $\{\beta_i\}_{1 \leq i \leq p}$ for the fusion.

3 Parametric model

We consider the following process :

$$Z^i_j = \mu_k + \varepsilon_j \quad t^i_{k-1} \leq j \leq t^i_k$$

and

$$1 \leq i \leq p,$$

where $\varepsilon$ is an additive noise. If $\varepsilon$ is a Gaussian white noise with variance $\sigma^2$ we can show that $\{z^i\}_{1 \leq i \leq p}$ is computed
by minimising:

\[ U_{\epsilon_1, \ldots, \epsilon_p}(p_1, \ldots, p_p) = V_{\epsilon_1, \ldots, \epsilon_p}(p_1, \ldots, p_p, \hat{\beta}(p_1), \ldots, \hat{\beta}(p_p)) \]

\[ + \beta \times S_0 + \sum_{i=1}^{p} \beta_i \times S_i. \]

(9)

where \([1]:

\[ V_{\epsilon_1, \ldots, \epsilon_p} [...] = \sum_{i=1}^{p} \sum_{k=1}^{N_{k}} n_k (p_k) \ln (\sigma^2 k (p_k)), \]

and

\[ \hat{\sigma}^2 k (p_k) = \frac{1}{n_k (p_k)} \sum_{j=t_{k-1}}^{t_k} (z_j - \hat{\mu}_k (p_k))^2 \]

with

\[ n_k (p_k) = \frac{t_k - t_{k-1}}, \]

\[ \hat{\mu}_k = \frac{1}{n_k (p_k)} \sum_{j=t_{k-1}}^{t_k} z_j. \]

Let \( \hat{\sigma}^2_k = \xi_k \sigma^2_k \) where \( \xi_k \) is the error of estimation. If we add a change at the \( j \)th position in the \( i \)th process. We assume that \( j \) belongs to the \( k \)th segment, then the variation \( \Delta V_{\epsilon_1, \ldots, \epsilon_p} [...] \) is easy to compute:

\[ \Delta V_{\epsilon_1, \ldots, \epsilon_p} [...] = n_k \ln (\sigma^2_k) - \left( \ln (\sigma^2 p_k) + (n_k - l) \ln (\sigma^2_{k+1}) \right) + \Delta V b_{\epsilon_1, \ldots, \epsilon_p} [...] \]

(12)

with

\[ \Delta V b_{\epsilon_1, \ldots, \epsilon_p} [...] = n_k \ln (\xi_k) - \ln (\xi_{kp}) - (n_k - l) \ln (\xi_{kp+1}) \]

let

\[ \Delta V [...] = \psi (n_k, l) + \Delta V b [...] \]

where \( l = j - t_{k-1} \) and \( \sigma^2 p_k \) and \( \sigma^2_{p+1} \) are the variance of \( Z^i \) calculated from each side of the \( j \)th position of the change. For a mean change in the \( k \)th segment of the process \( Z^i \), \( \psi (n_k, l) \) is expressed as:

\[ \psi (n_k, l) = \frac{1}{n_k} (\frac{1}{n_k} - 1) (\ln (1 + \frac{\sigma^2_{k+1}}{\sigma^2_k}) R^2) \]

(13)

where \( R = \frac{\mu_k - \mu_{k+1}}{\sigma_k} \) is positive and its minimum value is null. Every time we add a change, the value of \( V [...] \) decreases until it is null. This remark is systematically true if we don’t consider the error of estimation which are represented by the random variable \( \Delta V b \). The same kind of calculus and remark can be done for a change in the variance of the process.

The segmentation of \( Z \) with \( p=1 \) is done by the minimisation of \( U_{\epsilon_1} (r_1) \). The minimisation stop when \( \Delta V < \beta + \beta \) then we take \( \beta + \beta = \psi (n_k, l) \) (we suppose that \( \Delta V b = 0 \)). The value of \( \psi (n_k, l) \) will be defined by the user, it sets the resolution of the segmentation. For example, with the expression 13, \( \psi (n_k, l) \) is defined with the minimum signal to noise we want to detect. It’s easy to show that for a \( p \) dimensional process we have:

\[ \{ \beta^i = \psi^i - \beta_i \}_{1 \leq i \leq p}, \]

where \( \psi^i \) is defined by the user for the process \( Z^i \). Then the expression 9 is written as:

\[ U_{\epsilon_1, \ldots, \epsilon_p}(p_1, \ldots, p_p) = V_{\epsilon_1, \ldots, \epsilon_p}(p_1, \ldots, p_p, \hat{\beta}(p_1), \ldots, \hat{\beta}(p_p)) \]

\[ + \sum_{i=1}^{p} \beta_i \times S_i + \sum_{i=1}^{p} \beta_i \times S_i. \]

(15)

The parameters \( \beta^i \), set by the user, control the resolution level of the segmentation in the process \( Z^i \). Unfortunately the quality of the segmentation depends on the estimation error. We will show after that this quality can be improved with the fusion of the processes \( \{ Z^i \}_{1 \leq i \leq p} \).

The parameters \( \{ \beta_i \}_{1 \leq i \leq p} \) set by the user, control the fusion. The determination of their values is done under the constraints propose in the paragraph 2. This constraint can be rewritten as:

\[ \beta_1 = \psi - \beta \]

(16)

with \( \psi = \min_{i \in \{0, \ldots, p\}} \{ \psi^i \} \)

\[ 0 \leq \beta \leq \psi \]

\[ 0 \leq \beta_1 \leq \psi \]

\[ 0 \leq \psi - 2\beta \leq \beta_2 < \beta_1 \]

\[ 0 \leq \beta_2 - \beta \leq \beta_3 < \beta_2 \]

\[ \ldots \]

\[ 0 \leq \beta_{p-1} - \beta < \beta_p < \beta_{p-1} \].

We show on figure 1 the decision of detecting changes, \( f(\Delta V_{\epsilon_1}, \Delta V_{\epsilon_2}) \), calculate with the minimum value of \( U_{\epsilon_1, \epsilon_2}(p_1, p_2) \) for the fusion of two processes (we suppose that \( \Delta V b = 0 \)). Let:

- \( f(\Delta V_{\epsilon_1}, \Delta V_{\epsilon_2}) = 1 \) if we detect a simultaneous change on the two processes.
- \( f(\Delta V_{\epsilon_1}, \Delta V_{\epsilon_2}) = 2 \) if we detect a change on the first process.
- \( f(\Delta V_{\epsilon_1}, \Delta V_{\epsilon_2}) = 3 \) if we detect a change on the second process.
- \( f(\Delta V_{\epsilon_1}, \Delta V_{\epsilon_2}) = 0 \) if we don’t detect any change.

.We can see on this example that if one of the values \( \Delta V_{\epsilon_1} \) or \( \Delta V_{\epsilon_2} \) is superior to \( \beta_1 + \beta_1 \) and \( \beta_1 + \beta_2 \) respectively, we detect a simultaneous change if the other value of \( \Delta V_{\epsilon_1} \) or \( \Delta V_{\epsilon_2} \) is superior to \( (\beta_2 - \beta_1) + \beta_1 \) and \( (\beta_2 - \beta_1) + \beta_2 \) respectively \( ((\beta_2 - \beta_1) < 0) \). We detect also a simultaneous change if \( \Delta V_{\epsilon_1} + \Delta V_{\epsilon_2} > \beta_1 + \beta_2 + \beta_2 \). This is the effect of the fusion. For three sensors the values of \( \beta_3 \) and \( \beta_2 \) will respectively favour three and two simultaneous change in the same way.

In the real situation we must associate an error probability to the change detection. The estimation error, \( \Delta V b \), will influence the performance of the detection. We present in the next paragraph the performances comparison of the segmentation resulted with the fusion of two and three processes. Numerical experiments are also proposed on real and synthetic data.
4 Numerical experimentation

In the result we are presenting below, we show the performance of our segmentation technic and we present results on real and synthetic data. Let $p$ represent the probability of correct detection and correct localisation of a change. Let $q$ represent the probability of correct detection and bad localisation of a change. We show on figure 2.4 the experimental calcul of this probability for different values of the signal to noise ratio $R$. The signal is a synthetic gaussian process with a change in the mean of its statistic distribution.

On figure 2 the points marked with a * represent the probability calculate for a single process. The points marked with a o and a + represent the probability calculate from the fusion of two and three process respectively. This probability is calculated for a simultaneously change on the processes. The results show that the fusion improves the detection for all the value of the signal to noise ratio. The probability of correct detection and correct localisation increase with the number of fuse process. The probability of correct detection and bad localisation decreases in the same way.

On figure 4 the points marked with a * represent the probability calculate for a single process. The points marked with a o represent the probability calculate from the fusion of two process. The points marked with a + represent the probability calculate from the fusion of a process with a process which has a double signal to noise ratio. The results show that the performance increases with the signal to noise ratio of one of the process.

The wind velocity vector is described with an angle $\theta$ and a modul $\rho$. The value of this two parameters is available from a cup anemometer and a win vane. They are modelised by two independent random processes. In our application we have to detect changes in both the mean and the variance of the process. The changes of the process $\theta$ and $\rho$ depend on the atmospheric motion. The fusion of this two signals improves the segmentation results because $\rho$ has a better signal to noise ratio. The function we want to minimize for the segmentation can be decomposed into a sum of local potentials, then we use the iterative conditional mode (ICM) procedure to reach the optimal configuration of changes [2]. We show on figure 6, 7 the result of a segmentation on synthetic data. We show on figure 8, 9 the result of a segmentation on real data.
5 Conclusion

We present in this article an optimal segmentation of non-stationary random processes. The bayesian fusion method we propose for the segmentation is based on the introduction of a joined prior model for the simultaneously segmentation and estimation of data coming from a set of sensors. We determine the constraints for the definition of the segmentation parameters that will favorise the simultaneous changes in the processes fusion. We define in the parametric processes distribution case the expression and significance of all the segmentations parameters. In the numerical experimentation we show that the fusion of the processes gives better results for the segmentation, in terms of detection performance. We show also that these performances increase with the signal to noise ratio of one of the fused processes.

6 Acknowledgments

This work is supported by the French Minister of Education and Research. The authors thank Aloa Society and the Autonomous Port of Dunkerque for the wind measurements.

References


