IMPROVEMENT OF HIDDEN MARKOV MODEL EVALUATION OF THE MOBILE SATELLITE CHANNEL BY RESORTING TO A TRANSITION LOCALISATION METHOD

C. Alasseur, L. Husson

SUPELEC, Service Radioélectricité et Electronique, 3 rue Joliot Curie, Plateau de Moulon, 91 192 Gif-sur-Yvette, France
phone: +33 1 69851454, fax: +33 1 69851469, email: clemence.alasseur@supelec.fr, lionel.husson@supelec.fr

ABSTRACT

The mobile satellite channel has underlying Markovian properties and can then be represented by a Hidden Markov model (HMM). A challenging problem consists in estimating the model parameters from experimental data, especially when these parameters are not easily identifiable. In these cases, classification methods like k-means or scalable clustering, which are considered in this paper, show poor results when applied to the channel signal directly. We show that the detection of change-points of the signal, i.e. the detection of transitions between the model states, in a preliminary step, improves the estimation of the model parameters. We thus propose a method of model estimation including the detection of change-points that enables a better modelling of the satellite channel.

1. INTRODUCTION

Because of mobility of the receptor and/or of the satellite constellation that results in large variations of transmission conditions, the modelling of the satellite channel is a difficult issue. No a priori model, for example as [1] can efficiently model the channel in every context and there is a need for a method that modelled the channel without making priors so it can produce a model in any situation. Markovian models are characterized by their number of states, their corresponding distributions and the transition probabilities between states. At each time, the signal is considered to be the realization of one particular state. In this paper, we propose a method that enables to evaluate the model of the satellite channel signal when no knowledge about the signal is available before the beginning of the estimation process except the number of states of the model.

Because one observation of the satellite channel does not generally reveal the state the model currently is, we need to consider the signal by blocks of measures which we refer as the analysis window. One histogram is evaluated on every analysis window and linked to one state distribution of the HMM. The method used to estimate the HMM is the following: histograms corresponding to analysis windows are projected into a reduced subspace and their projections are then classified [2, 3]. Therefore, the parameters of the HMM can easily be estimated because the projection step has reduced the classification complexity by limiting the number of dimensions. The main classification difficulty is to find the correct length of the analysis window through which the signal is considered. In this paper, two classification methods are compared for which the size of the analysis window is either fixed or variable. When the length of the analysis window is variable, it means that the sampling of the channel signal follows the transition between states of the signal otherwise the channel signal is sampled every fixed analysis window length. The variable analysis window lengths are obtained by rupture between HMM states detection with a Monte Carlo Markov Chain (MCMC) method.

We resort to two classification methods that are detailed in the second part of this paper. The third part presents the rupture estimation method. The combination of the classification with the transition estimation for the HMM estimation is explained in part 4, followed by the experimental results.

2. CLASSIFICATION METHODS

The classification methods operate onto projections x of the histograms calculated over analysis windows of the signal. The projections of one particular state of the model are observed to cluster around a centre position y. That are these clusters that classification methods aim to identify.

2.1 K-means with the Mahalanobis distance

In the k-means method, the centres of the clusters are evaluated by successive iterations and the projected points are classified to the nearest cluster in term of Euclidean distance between them and the centres of the clusters. After several iterations, this preliminary rough classification converges and gives quite accurate centre m_i and covariance matrix C_i for each distribution. The projected points x are then once again classified to the nearest cluster but this time in term of the Mahalanobis distance D_i defined as follows: $D_i^2 = (x - m_i)' C_i^{-1} (x - m_i)$. One advantage of the Mahalanobis distance is to take into account the correlation of the projected points and makes it possible to have curved decision boundaries (unlike linear boundaries for the Euclidean distance).

2.2 Scalable Classification [4]

Considering a single cluster C of data, the contribution of each datum x to the centre y of the cluster, denoted by Py(x) is expressed by:

$$P_y(x) = \frac{e^{-\beta.e_y(x)}}{\sum_{x \in C} e^{-\beta.e_y(x)}} \quad (2.1)$$

where β is the scale factor
and $e_y(x)$ the distance measure between x and y.

The centre of the cluster minimises the free energy of the system and is satisfying the following relation:

$$y = \sum_{x \in C} \frac{x.e^{-\beta.e_y(x)}}{\sum_{x \in C} e^{-\beta.e_y(x)}} \quad (2.2)$$

This equation can not be solved analytically, however its solutions are equivalent to the fixed points of the following suite:

$$y \xrightarrow{f} y + \sum_{x \in C} \frac{(x-y).e^{-\beta.e_y(x)}}{\sum_{x \in C} e^{-\beta.e_y(x)}} \quad (2.3)$$

Initially a large value of the scale factor β is necessary in order to consider each data as a single cluster. Along the iteration process, the value of β decreases up to a value where every data is considered to belong to the same cluster.

Algorithm:

Initialization:
- Set β to β_{max} such as each datum is a cluster.

Iteration:
- evaluate the centre of each cluster by resolution of (2.3)
- identify the distinct centres. The clusters with identical centre are unified.
- increase iteration index
  decrease β: βi+1=βi/1.05

Iterate until only a unique cluster remain

The simplest metric, used in [4], is the Euclidian measure. However the projection subspace is characterized by the associated eigenvalues for each dimension. Injecting this consideration to the definition of the contribution Py(x) leads to the choice of the following weighted metric [2]:

$$e_y(x) = \frac{\sum_{k<N} \lambda_k (x_k - y_k)^2}{\sum_{k<N} \lambda_k} \quad (2.4)$$

where x_k and y_k are the coordinates of x and y on the N^{th} order subspace and λ_k the eigenvalue associated to the k^{th} dimension.

3. CHANGE-POINTS LOCALIZATION

The purpose of the change-points localization method is to provide an estimate of the change-points sequence $(r_t, n \geq t \geq 1)$, i.e. the transition time location between two states of the model. The method is based on a MCMC (Monte Carlo Markov Chain) principle.

3.1 Principle [5]

The purpose of MCMC method [6] is to generate an ergodic sequence (r) of associated distribution *f* without simulating (r) directly from *f*. For an arbitrary starting value r(0), an ergodic chain (r(t)) is generated using a transition kernel with stationary distribution *f*. For a large enough T, r(T) can be considered as distributed from *f* because of the ergodic theorem. The Hastings-Metropolis algorithm, chosen as the MCMC method, is based on a conditional density $q(x, x')$.

The change-points sequence (r_t) is null except when a transition between two states occurs for which r_t is equal to one. (r_t) is assumed to be a sequence of independent and identically distributed (iid) Bernouilly random variables with parameter λ.

Let (τ_k) be the sequence of change-points instants: τ_k is the instant *t* of the change-point *k* and $y = (y_t, t \geq 1)$ the observations sequence, i.e. the channel attenuation. Then a sequence of means (m_k) exits such that for any $\tau_{k-1} + 1 \leq t \leq \tau_k$, $y_t = m_{k,t} + \varepsilon_t$ with (ε) a sequence of zero-mean random process.

In theses conditions, the posterior distribution of *r* is shown to be [5]:

$$p(r / y; \theta) = C(y; \theta) e^{-\phi S_r - \gamma K_r} \quad (3.1)$$

where
y is the observation vector,
$C(y, \theta)$ is a normalizing constant ,
$\theta = (\lambda, V, \sigma^2)$ the set of hyper parameters with:

- λ the prior proportion of change-points
- V the variance of the Gaussian distribution for the vector of means (m_k)
- σ^2 the variance of the additive noise (ε)

$$\phi = \frac{V}{2\sigma^2(\sigma^2 + V)}, \quad \gamma = \frac{1}{2}\log\left(\frac{\sigma^2 + V}{\sigma^2}\right) + \log\left(\frac{1-\lambda}{\lambda}\right),$$

$$K_r = \sum_{t=1}^{n-1} r_t + 1 \text{ and } S_r = \sum_{k=1}^{K_r} \sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - \overline{y_k})^2$$

3.2 Hastings-Metropolis Algorithm

At iteration i, the Hastings-Metropolis algorithm performs in two stages:

- a candidate $\tilde{r}$ is proposed from the proposal kernel $q(r^i, \tilde{r})$
- $\tilde{r}$ is accepted as the new value for (r), $r^{i+1} = \tilde{r}$, with the probability

$$\alpha(r^i, \tilde{r}) = \min\left(1, \frac{p(\tilde{r} / y; \theta) q(\tilde{r}, r^i)}{p(r^i / y; \theta) q(r^i, \tilde{r})}\right) \quad (3.2)$$

At each iteration, candidate sequences $\tilde{r}$ are proposed from the following four kernels:

- *Independent* change-points sequence: $\tilde{r}$ is completely independent from r.
- *Creation* of a change-point: a new change-point position is chosen randomly among (r_t) such that $r_t = 0$.
- *Deletion* of an existing change-point: a new change-point position is chosen randomly such that $r_t = 1$.
- *Update* of one change-point position: two instant points (t, t') are randomly chosen such that $r_t = 1$ and $r_{t'} = 0$. All remaining instant points keep the same value except these two: $\tilde{r}_t = 1 - r_t$ and $\tilde{r}_{t'} = 1 - r_{t'}$.

For any $\tilde{r}$ acceptance, the set of hyper parameters is updated so that they maximise the complete likelihood $f(y, r, \theta)$: $\tilde{\lambda} = \frac{K_r - 1}{n - 1}$, $\tilde{\sigma}^2 = \frac{S_r}{n - K_r}$ and

$$\tilde{V} = \frac{1}{K_r}\left(\sum_{t=1}^{n}(y_t - \bar{y})^2 - S_r\right) - \sigma^2$$

Hastings Metropolis is run at a fixed and low temperature [5] in order to discriminate global and local maxima of the posterior distribution $p(r / y; \theta)$. Concretely, this induces the replacement of (ϕ, γ) by the couple $(\phi / T, \gamma / T)$. To run Hastings Metropolis algorithm at low temperature enables to lower the number of necessary iterations before convergence. This method can give perfect transition estimation but we would rather reduce the number of required iterations and obtain every existent change-points plus non-existent ones. The complexity is then reduced to 5.000 iterations, including 2.500 burn-in iterations, rather than 200.000 iterations, including 100.000 burn-in iterations, recommended in the method used in [7].

4. HMM ESTIMATION PROCEDURE

The states of the model are identified by their corresponding distribution. Histograms are calculated on intervals of the channel attenuation signal and classified to one particular state. To reduce the dimensionality of the classification problem, the histograms are projected into a 2-dimension space of the mean and variance (in dB). The distributions of tested signals are then assumed to be identifiable by their mean and variance. In the following, the signal is projected according to a fixed analysis window length [2, 3] or to a variable analysis window length linked to the transitions estimation.

4.1 Fixed analysis window length

- *Projection*: the size of the projection window is fixed. When the analysis window covers a transition between two states, the resultant projection is referred as an intermediate point. Intermediate point projections are located between two distributions centres because it is the result of two different states. They harden the classification, i.e. the clusters are less separated. An appropriate window length is a compromise between a short value necessary to detect every transitions, i.e. to reduce the number of intermediate points, and between a large value to obtain good mean and variance estimate.
- *Classification*: the projections are classified using one of the two proposed methods (k-means or scalable clustering).

4.2 Variable analysis window length

- *Projection*: every interval between two estimated change-points is projected onto the mean and variance (in dB) subspace in order to reduce the classification complexity. Contrary to the fixed length window, projections are calculated onto points that we are sure to belong to the same distribution. This implies that we dispose of a sequence of change-points location estimate obtained by the method developed in [3]. False detections which are obtained with this method are not a problem as they do not generate intermediate points: it only divides the signal in smaller windows that can be classified easily because every points of a particular window must belong to the same HMM state.
- *Classification* on *'safe'* projections: the classification method operates only on 'safe' projections, i.e. it selects only the projections that correspond to the longest time intervals. It avoids the presence of poor mean or variance estimates due to the short length of the analysis window and makes the classification easier.
- *Classification* of *every projection* points: the previous step gives quite accurate centre and covariance; every projection points (not only the 'safe' one) are then classified to the nearest distribution's centre according to the Mahalanobis distance. If the Euclidean distance is used, the final result is poorer: some short states are not detected.

5. SIMULATION RESULTS

The signals considered in this paper are generated according to the Lutz model [1] which is a well-known satellite propagation attenuation model for mobile satellite service context. The Lutz model corresponds to a satellite channel where the transmission encounters two main conditions: the LOS state (Line of Sight) when a direct link exits between the receiver and the satellite and the NLOS state (None Light of Sight) when the transmission path is obstructed. The NLOS state is represented by a Rayleigh distribution, the LOS by a Rice distribution. Simulations have been carried out on four different experimental signals whose parameters are presented in the following:

	Test 1	Test 2	Test 3	Test 4
$p_{transition}$	0.005	0.005	0.01	0.01
$\sigma_{Rayleigh}$	0.15	0.425	0.425	0.425
c_{Rice}	19.4 dB	9.6 dB	9.6 dB	4.42 dB

The value of $p_{transition}$ is given for a sampling time equal to one. The model estimation is getting harder from test 1 to test 4. In fact, the 2 distributions are closer in test 2 than in 1, transitions occur more rapidly in test 3 than in 2. And, in test 4, the distributions are also less separated than in 3.

Therefore, the classification is tougher for the conditions of test 4 than for these of test 1.

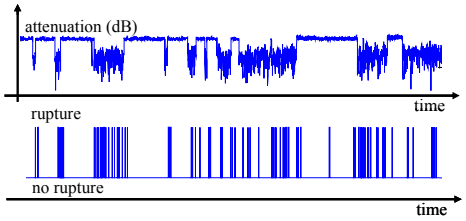


Figure 1: result of the rupture estimation sequence

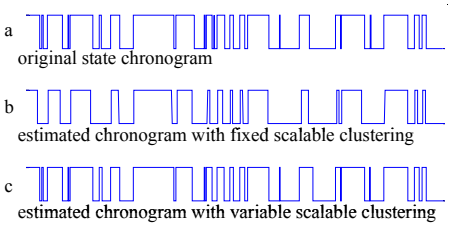


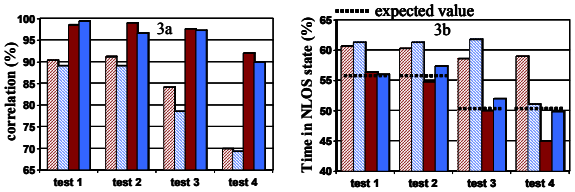
Figure 2: chronograms estimation

Figure 1 shows the satellite channel attenuation (in dB) obtained for test 3. It also presents the results of the transition detection: non-existent transitions are estimated but each relevant transition is also detected. Figure 2 presents examples of chronograms obtained for test 3: the original one (a), the estimated one with the scalable clustering/fixed window analysis (b), and the estimated one with the scalable clustering/variable window (c). It shows that when the transition estimation, i.e. the variable window analysis, is used that the chronogram is better estimated; all states are detected and their lengths are correct. Figure 3.a presents the matching measure, i.e. the correlation, between the original and the estimated chronograms. Fig. 3.b gives the percentage of time when the model is in the NLOS state. The final array, fig. 3.c, provides for the NLOS distribution, the chi-square distance between the empirical Rayleigh distribution and the estimated one. This distance is normalised by the minimum distance obtained over the four methods. Only results of the NLOS distribution are presented, because the LOS distributions are always well estimated, considering that intermediate states are, in practical, classified to the NLOS state and not to LOS.

The transition estimation enables to improve the model estimation for both classification methods, e.g. for test 3, the matching measure, which was only of 80% for the fixed window length, is around 97% for the variable window length. The estimated distributions are also better estimated: e.g. for test 3 with scalable-clustering, the NLOS distribution is 40 times closer when the variable window is used. The percentage of time in NLOS state is also more consistent with the expected value, e.g. for test 3, it is around 50% for the variable window, that corresponds to the empirical measure, and around 60% for the fixed window.

6. CONCLUSIONS

In this paper, we study the performance of classification



3c RAYLEIGH	Test 1	Test 2	Test 3	Test 4
K-means, fixed length	24.3	11	24.6	2.2
Scalable-clustering, fixed length	31.2	15	40.7	2.2
K-means, variable length	1.5	1	1	1
Scalable-clustering, variable length	1	2.4	1.1	1.7

Figure 3: comparison between the different methods

methods with or without transition detection for the evaluation of the HMM for the mobile satellite channel. When transition locations are detected, we show that better performances are possible for the estimation of the model. We thus proposed a procedure which is to add a preliminary step, i.e. the detection of change-points before the classification which is performed either by a k-means or scalable clustering method. An advantage of this method is also to avoid cases where the analysis window is considered while a transition between two states of the model occurs that degrade the classification performances. With this method, the satellite channel is easily modelled and the computational complexity is reduced.

7. REFERENCES

[1] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky and W. Papke, "The land mobile satellite communication channel", IEEE trans. on Vehic. Technol., vol. 40, no. 2, pp. 375-386, May 1991

[2] L. Husson and al. "Extraction of empirical models from experimental measurements in the L-band for mobile satellite transmissions", 19th AIAA Int. Com. Satellite Systems Conf., 2001, Toulouse, France

[3] L. Husson, J.C. Dany, S. Chambon, K. Berradi, A. Beffani. "Modelling of Mobile Satellite Channels by Scalable Clustering Algorithm", VTC 2002 spring, 6-10 May 2002, Birmingham, USA

[4] Y. Wong, E.C. Posner, "A new clustering algorithm applicable to multispectral and polarimetric SAR images", IEEE trans. on geoscience and remote sensing, vol. 31, no. 3, pp. 634-644, May 1993

[5] M. Lavielle and E. Lebarbier, "An Application of MCMC methods for the multiple change-points problem", Signal Processing 81, pp. 39-53, 2001

[6] C. Robert, "Méthodes de Monte Carlo par chaînes de Markov", Statistique Mathématique et Probbilité, ECONOMICA

[7] S. Richardson and P.J. Green, "On bayesian analysis of mixtures with an unknown number of components", J. R. Statist. Soc. Ser. B, vol 59, n. 4, pp. 731-792, 1997.