

# Modelling dengue epidemics with autoregressive switching Markov models (AR-HMM)\*

Madalina Olteanu<sup>1</sup>, Esther García-Garaluz<sup>2</sup>, Miguel Atencia<sup>3</sup>, and Gonzalo Joya<sup>2</sup>

<sup>1</sup> SAMOS - CES, Université de Paris I (France)

<sup>2</sup> Departamento de Tecnología Electrónica

<sup>3</sup> Departamento de Matemática Aplicada  
Universidad de Málaga (Spain)

Campus de Teatinos, 29071 Málaga, Spain  
madalina.olteanu@univ-paris1.fr

**Abstract.** In this work, autoregressive switching-Markov models (AR-HMM) are applied to the dengue fever epidemics (DF) in La Havana (Cuba). This technique allows to model time series which are controlled by some unobserved process and finite time lags. A first experiment with real data of dengue is performed in order to obtain the characterization of different stages of the epidemics. The aim of this work is to present a method which can give valuable information about how an efficient control strategy can be performed.

## 1 Introduction

Since the eighteen century, mathematical models of infectious diseases have been developed and used for providing information to health authorities [1]. During the 60s, the improvements in the prevention programs turned the research interests in the industrialized countries to other diseases. However, during the last few years dengue fever (DF) has become a major health problem affecting tropical and sub-tropical regions around the world, especially urban and suburban areas [2]. Thus, the question of modelling infectious diseases arose again. For health authorities, a well fitted mathematical model may help in figuring out new strategies for an efficient fight against the epidemics.

In this paper, a method based on hidden Markov models is described. For vector-borne diseases, this is a very suitable technique since it enables characterization of unknown processes. The method presented herein allows to take into account finite-time feedback as well. This is an interesting feature since incubation and transmission rates may be considered.

\* This work has been partially supported by the Agencia Española de Cooperación Internacional para el Desarrollo (AECID), project no. D/017218/08. Thanks are due to Dr. Hector de Arazoza for his useful suggestions and, in particular, for providing the model of Dengue epidemics. The authors are grateful to Dr. Joseph Rynkiewicz for the REGRESS software.



The rest of the paper is organized as follows: in section 2, the autoregressive switching-Markov model (AR-HMM) is described. Section 3 presents the features of dengue fever epidemics. In section 4, AR-HMM is applied to real data of dengue fever epidemics. Finally, in section 5, conclusions and future work are proposed.

## 2 Autoregressive Switching-Markov models

In some cases, time series may be explained not only by their past values, but also by some unknown, random process. Typical examples are economic cycles (recession and growth periods) or DNA identification (coding and non-coding regions). For these examples, the observed series will depend on the “state” or the “regime” of the unknown, hidden process. Usually, the hidden process is supposed to have a finite and fixed number of regimes, corresponding to the expected states of the phenomena to be modelled (for example, recession versus growth periods)

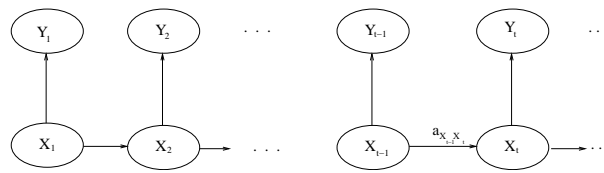
### 2.1 Notations and definitions

Let us start by introducing hidden Markov chains (HMC) and recall the definition.

A bivariate process  $(X_t, Y_t)_{t \in \mathbb{Z}}$  is a hidden Markov chain if:

- $(X_t)$  is a homogeneous Markov chain valued in a finite state space  $E = \{e_1, \dots, e_N\}$ ,  $N \in \mathbb{N}^*$ , with transition matrix  $A = (a_{ij})_{i,j=1,\dots,N}$ ,
- $$a_{ij} = P(X_{t+1} = e_j \mid X_t = e_i)$$
- $(Y_t)$  valued in  $R^d$ ,  $d \geq 1$ , conditionally independent w.r.t.  $(X_t)$  and  $\mathcal{L}(Y_t \mid X_t = e_i) \equiv \mathcal{L}_i$ .

The hidden Markov chain is described in Figure 1.



**Fig. 1.** Representation of Hidden Markov chains (HMC)

Hidden Markov models were introduced by [3] in 1966. They soon became essential tools in speech-recognition ([4], [6]), bioinformatics and analysis of biological sequences ([5]). Hidden Markov models have been already used for modelling epidemics data, with promising results ([8]).

Hidden Markov chains may be generalized by letting  $Y_t$  depend on its past values and on  $X_t$ . This new class of models may be defined as an autoregressive switching-Markov model (AR-HMM). AR-HMM were initially proposed in [7] and were extensively used in the sequel for modelling financial and economic time series. However, to our knowledge, there are no current applications for epidemics data.

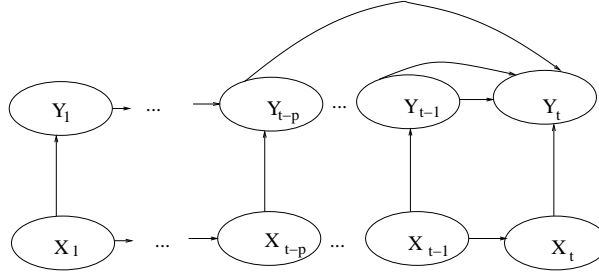
An autoregressive switching-Markov process is a bivariate process  $(X_t, Y_t)_{t \in \mathbb{Z}}$  such that:

- $(X_t)$  is a homogeneous Markov chain valued in  $E = \{e_1, \dots, e_N\}$ ,  $N \in \mathbb{N}^*$ , with transition matrix  $A = (a_{ij})_{i,j=1,\dots,N}$ ,  $a_{ij} = P(X_{t+1} = e_j \mid X_t = e_i)$
- $Y_t = F_{X_t}(Y_{t-1}, \dots, Y_{t-p}) + \sigma_{X_t} \varepsilon_t$
- $F_{e_i}$  is a parametric function (linear or nonlinear) depending on some  $\theta_i$
- $\sigma_{e_i} \in \{\sigma_1, \dots, \sigma_N\} \subset (R_+^*)^N$  and  $(\varepsilon_t)$  is an iid sequence

We will suppose that the noise  $\varepsilon_t$  is a standard Gaussian. The parameter to be estimated contains the transition probabilities of the hidden Markov chain, the parameters in the regression functions and the standard deviations for the noise:

$$\Theta = (a_{ij}, \theta_i, \sigma_i, i, j = 1, \dots, q)$$

The size of the parameter  $\Theta$  is  $q(q-1) + dq + q = q^2 + dq$ .



**Fig. 2.** Autoregressive switching-Hidden Markov Model (AR-HMM)

An autoregressive switching-Markov model is represented in Figure 2.

The estimation procedure may be summarized as follows:

1. First, the user has to set the number of regimes,  $q$ , the number of time lags,  $p$  ( $p$  may be chosen using some AIC/BIC information criteria or by cross-validation) and the type of regression functions (linear, MLP, ...).
2. The parameter  $\Theta$  is estimated using the EM-algorithm (see [9] for more details). A first characterization of the regimes (stability, influence of the past values,...) will be done using these results.
3. With the estimated values of the parameters, one may compute the posterior probabilities of belonging to each of the regimes, at any moment. These

probabilities will allow the a posteriori identification and characterization of the regimes.

### 3 Dengue Epidemics

Dengue Fever (DF) and Dengue Haemorrhagic Fever (DHF) are infectious diseases transmitted to humans by the bites of *Aedes* mosquitoes. If a mosquito bites a Dengue Fever infected, it gets the disease and, after an incubation period, is capable of transmitting the infection to other people. That kind of disease, where an infectious organism that does not cause diseases itself transmits infection by contact among different hosts, is called a vector-borne disease [2] and *Aedes* mosquito is the vector.

The disease of Dengue fever shows a resurgent pattern worldwide in the last years, becoming one of the most important health problems in subtropical countries. WHO estimates [2] there may be 50 million dengue infections every year in the world, and two fifths of the world population are now at risk of Dengue. In the Americas more than 890 000 cases of Dengue fever were reported, of which 26 000 was of Haemorrhagic Fever.

There is not vaccine against dengue fever. Besides, people who suffer dengue are immune against that dengue strain. For these reasons, Dengue fever is frequently described as an epidemic SIR model for a short or medium-term model (under the assumption that only one strain of the virus exists). In this model the whole population is classified in three groups: susceptible, infected and recovered people. Susceptible people are those who are not immunized against the disease and do not present symptoms. Infected people are who present symptoms of the disease. Recovered are those who got the disease in the past and are immunized against it.

Several models have been proposed for dengue fever which depend on the vector population [10], [11]. This becomes an important limitation since generally it is not possible to have enough real information about it. On the other hand, dengue fever, as most biological processes, has a feedback which affects the system in a finite time way. This is the case of human and mosquito incubation periods or recovering rate.

The ability to model systems in which the influence of an unknown process appears and with a finite-time feedback make an autoregressive switching-Markov model specially suitable for dengue fever characterization.

## 4 A first attempt to model Dengue epidemics using HMM-AR

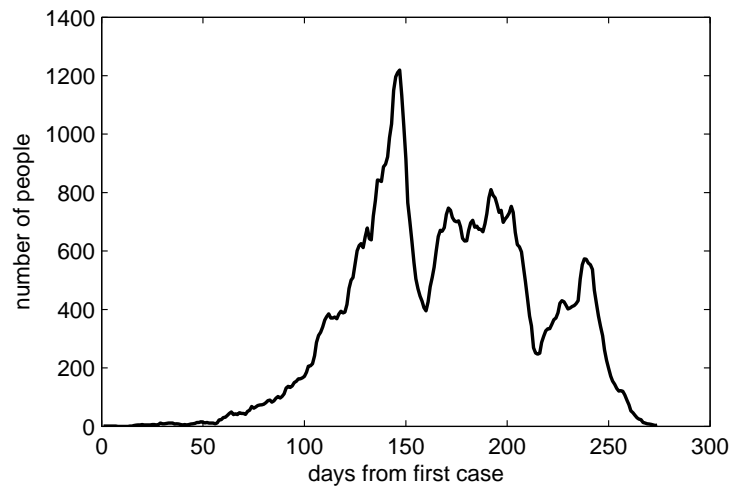
### 4.1 A first toy model

In order to perform the experiment, a 274-day register of new infections for Dengue Fever in La Havana is available. From that database and the known recovering rate of the disease, the time series for the global number of infected

people was obtained (Figure 3). In it, the decrease periods are related with the fumigation programs performed by the health authorities.

One of the important considerations is the time lags in the experiment. Because of the incubation rate of human and mosquito lags between 8 and 15 days are considered.

For these epidemic data, we shall test three regimes: epidemics, non-epidemics and decrease periods. In all cases, each regime will be characterized a posteriori using these estimated probabilities.



**Fig. 3.** Number of Dengue infected individuals in Havana (Cuba)

In this case, the estimated transition matrix is :

$$P = \begin{pmatrix} 0.953 & 0 & 0.044 \\ 0 & 0.960 & 0.031 \\ 0.467 & 0.040 & 0.925 \end{pmatrix}$$

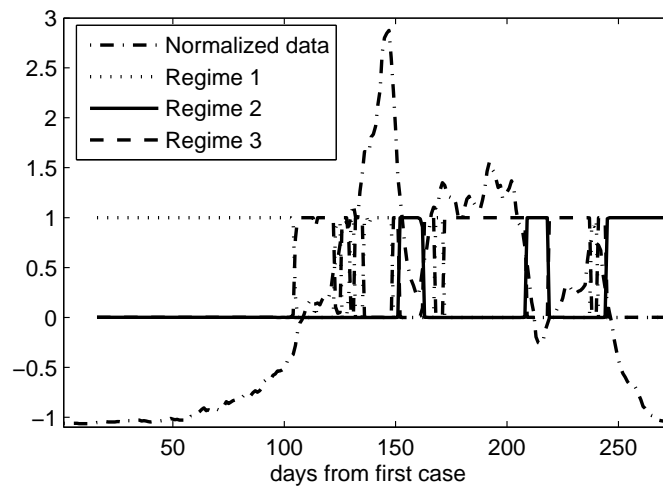
From this transition matrix, some aspects should be highlighted

- All three regimes are very stable.
- There's no transition between the first and the second regime.

Besides the transition matrix, a model for each regime was obtained with the form:

$$Y_t = aY_{t-8} - bY_{t-9} + \dots + hY_{t-7} + i + \epsilon_t$$

Nevertheless, we are not showing the results for every coefficient for simplicity.



**Fig. 4.** Number of Dengue infected individuals in Havana (Cuba)

In the Figure 4 the probability of every regime is plotted.

From the whole results the obtained regimes can be characterized. The first and third regimes represent the non-epidemic and the epidemic periods respectively. Besides, it is necessary that the epidemics reach a high enough incidence to be removed. The decrease period is represented by the second regime.

## 5 Conclusion and future works

In this work the autoregressive-switching-Markov model is described. This modelling technique is applied to an epidemiological model of dengue fever. Autoregressive switching-Markov model is specially intended for model it because of its ability to consider both finite feedback and the influence of hidden process. An experiment with a real data-base of the dengue epidemics in La Havana (Cuba) has been made. The results enable the characterization of three different regimes: the epidemic, the no-epidemic and the decrease period. Some ideas can be inferred from the results, among them, that a high enough value of the epidemic prevalence is necessary in order to reach the non-epidemic state.

Some future works can be proposed. The next step is to test other configurations with different lags of time and different regression functions. The consideration of vertical transmission of dengue, that is, the transmission from the female to the eggs would be interesting as well. On the other hand, another future work line is extracting information from the available data with this technique in order to develop more efficient fumigation strategies. It would allow to perform prevention measures while the prevalence of the disease is still reduced.

## References

1. Blower, S. and D. Bernoulli (2004), An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it: 1766. *Reviews in Medical Virology* 14: 275-288.
2. World Health Organization, <http://www.who.int>
3. Baum L., Petrie T. (1966), Statistical inference for probabilistic functions of finite state Markov chains, *Ann. Math. Statist.*, **37**, p.1554-1563
4. Rabiner R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of IEEE*, **77(2)**, p.257-286
5. Durbin R., Eddy S.R., Krogh A., Mitchison G. (1998) *Biological sequence analyses: probabilistic models of proteins and nucleic acids*, Cambridge University Press
6. Bourlard H.A., Morgan N. (1994), *Connectionist speech recognition: a hybrid approach*, Kluwer academic publ.
7. Hamilton J. D. (1988), A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, **57(2)**, 357-384
8. Rath T.M., Carreras M., Sebastiani P. (2008) Automated Detection of Influenza Epidemics with Hidden Markov Models, *Advances in Intelligent Data Analysis V*, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, p.521-532
9. Rynkiewicz J. (2002) , Estimation of Hybrid HMM/MLP models, *Proceedings of ESANN'2001* p. 383-390
10. Esteva, L., Vargas, C.: A Model for Dengue Disease with Variable Human Population. *J. Math. Biol.* 38, 220–240 (1999)
11. Coutinho, F.A.B., Burattini, M.N., Lopez, L.F., Massad, E.: Threshold Conditions for a Non-Autonomous Epidemic System Describing the Population Dynamics of Dengue. *Bulletin of Mathematical Biology* 68, 2263–2282 (2006)