



## Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis



Claudia Torres<sup>a,\*</sup>, Samier Barguil<sup>a</sup>, Miguel Melgarejo<sup>a</sup>, Andrés Olarte<sup>b</sup>

<sup>a</sup> Laboratorio de Automática e Inteligencia Computacional, Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Carrera 7 No. 40-53, Bogotá, Colombia

<sup>b</sup> Grupo de Modelamiento y Control de Sistemas Biológicos, Facultad de Ingeniería, Universidad Nacional de Colombia, Carrera 45 No. 26-85, Bogotá, Colombia

### ARTICLE INFO

#### Article history:

Received 3 May 2013

Received in revised form

29 November 2013

Accepted 30 November 2013

#### Keywords:

Takagi–Sugeno model

Fuzzy clustering

Wavelet transforms

Dengue

### ABSTRACT

**Objective:** This article presents a model of a dengue and severe dengue epidemic in Colombia based on the cases reported between 1995 and 2011.

**Methodology:** We present a methodological approach that combines multiresolution analysis and fuzzy systems to represent cases of dengue and severe dengue in Colombia. The performance of this proposal was compared with that obtained by applying traditional fuzzy modeling techniques on the same data set. This comparison was obtained by two performance measures that evaluate the similarity between the original data and the approximate signal: the mean square error and the variance accounted for. Finally, the predictive ability of the proposed technique was evaluated to forecast the number of dengue and severe dengue cases in a horizon of three years (2012–2015). These estimates were validated with a data set that was not included into the training stage of the model.

**Results:** The proposed technique allowed the creation of a model that adequately represented the dynamic of a dengue and severe dengue epidemic in Colombia. This technique achieves a significantly superior performance to that obtained with traditional fuzzy modeling techniques: the similarity between the original data and the approximate signal increases from 21.13% to 90.06% and from 18.90% to 76.83% in the case of dengue and severe dengue, respectively. Finally, the developed models generate plausible predictions that resemble validation data. The difference between the cumulative cases reported from January 2012 until July 2013 and those predicted by the model for the same period was 24.99% for dengue and only 4.22% for severe dengue.

**Conclusions:** The fuzzy model identification technique based on multiresolution analysis produced a proper representation of dengue and severe dengue cases for Colombia despite the complexity and uncertainty that characterize this biological system. Additionally, the obtained models generate plausible predictions that can be used by surveillance authorities to support decision-making oriented to designing and developing control strategies.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Dengue is a viral disease transmitted between susceptible and infected individuals through a vector (the *Aedes aegypti* mosquito). The dengue virus belongs to the Flaviviridae family and four distinct serotypes are known. When a person recovers from an infection by one of the virus serotypes s/he will acquire lifelong immunity against that serotype and temporary immunity against others [1–3]. The dengue infection may cause mild symptoms similar to those of flu (fever, nausea, muscle pain, etc.), but may sometimes

develop into severe dengue, a potentially deadly disease characterized by severe bleeding and organ damage [4].

This disease imposes a significant social and economic burden on affected populations and represents an international public health issue. According to the World Health Organization, about 2.5 billion people live in countries where dengue is endemic and between 50 and 100 million dengue infections occur every year, of which 24,000 result in death [2].

In recent years, the incidence and severity of dengue have significantly increased because of migratory movements, weather conditions and a lack of preventative measures. This is evidenced by the figures reported: dengue-endemic countries increased from nine in the year 1970 to 100 in 2013 and 40% of the global population is at risk of contracting dengue [2]. According to the Pan American Health Organization, between 2001 and 2007, Colombia

\* Corresponding author. Tel.: +57 1 3010585; fax: +57 1 3165241.

E-mail addresses: [crtorresm@correo.udistrital.edu.co](mailto:crtorresm@correo.udistrital.edu.co), [claudiatrrs190@gmail.com](mailto:claudiatrrs190@gmail.com) (C. Torres).

had the highest number of cases (81%) and dengue deaths (73%) of the Andean countries (subregion comprising Bolivia, Colombia, Ecuador, Peru and Venezuela). This subregion contributed 19% of cases of dengue in the Americas and has the largest number of reported cases of severe dengue [3].

For these reasons it has become necessary to create quantitative models that explain the dynamics of dengue infection. A model that represents the dynamics of the disease with sufficient accuracy could be used as a tool for decision-making and the distribution of resources to tackle epidemics. Such models would allow the authorities to anticipate outbreaks of the disease, facilitating the design of control strategies or optimizing them [5,6].

Because of its importance, several models of dengue have been proposed, mainly based on differential equations with different levels of complexity. For instance, models including two or more serotypes of the virus have been developed [7,8], others have divided human populations according to age and contagion risk [9–12] and models analyzing the effect of the inclusion of control strategies have also been proposed [13–15]. The development of models with differential equations, however, requires explicit knowledge of the existing relationship between system variables, which could generate constraints in the case of complex adaptive problems characterized by high levels of uncertainty, such as those found in epidemiology [16–18].

Particularly, the modeling of a dengue epidemic from data involves dealing with uncertainty largely owed to the many unreported cases, misdiagnosis and environmental factors [2]. For example, the diagnosis and confirmation of each case evaluated as a dengue infection should be subject to an immunological assay. This test is included by law in health plans in Colombia, but its application is only mandatory for between 10% and 20% of the cases that are treated as dengue infections [19]. As a consequence a source of uncertainty is introduced into the data set.

On the other hand, it is well known that fuzzy systems are appropriate for modeling systems with uncertainty and nonlinearity [17,20], which motivated the authors to build a dynamic model of a dengue and severe dengue epidemic in Colombia based on real data and using a fuzzy approach. To achieve this, a methodological proposal was presented, where the fuzzy modeling identification (FMID) technique and multiresolution analysis (MRA) were combined for the first time in epidemiological analysis. This approach breaks down and analyzes the signal in several time and frequency scales, followed by an approximation by Takagi–Sugeno (TS) models tuned with fuzzy clustering.

To the best of the authors' knowledge, there are no published studies reporting the modeling of epidemiological dynamics such as dengue, which are based on data and use the FMID technique. Other biological processes, however, have been represented with fuzzy clustering, for instance: estimation of respiratory parameters [21], temperature behavior in the composting process [22], the conversion process of G penicillin [23], the process of pH neutralization in a tank [24] and waste-water treatment [25].

The approach proposed in this paper was compared with the FMID technique, which performance was evaluated through root mean square error (RMSE) and the variance between the original and the approximated signal, known as variance accounted for (VAF). Results showed the low performance of the classic fuzzy modeling technique in terms of representing the dynamics of a dengue and severe dengue epidemic, whereas the inclusion of MRA generated a significant improvement in the performance of the models.

The organization of this paper is as follows: Section 2 presents a description of the database, algorithms and techniques used. Section 3 outlines the experimental methodology, the results obtained and the comparison between the performance of the classic fuzzy modeling technique and the multiresolution

analysis hybrid technique. Finally, Sections 4 and 5 provide a discussion of the results, draw conclusions and offer suggestions for future work.

## 2. Materials and methods

This section describes the data set used to generate the model, as well as the general aspects of fuzzy modeling and the proposal of a hybrid modeling technique using MRA. Finally, the metrics used to evaluate the performance of the obtained models are presented.

### 2.1. Reported cases of dengue and severe dengue

The database used is the result of a weekly compilation generated by the System of Public Health Surveillance from the Colombian National Institute of Health (NIH) for reported cases of dengue and severe dengue in the periods from 1995 to 2011 and 1997 to 2011, respectively (Fig. 1). These data sets represent national consolidation of the cases reported in each of the departments and main cities of the country.

Analyzing the national consolidation of dengue and severe dengue, one could argue that both cases present time-variant behavior with significant changes from one year to the next. Among these changes, the behavior observed between the years 2009 and 2011 in the reported dengue cases stands out particularly because it is far removed from the dynamics that the disease presented in previous years.

### 2.2. Fuzzy model identification

This technique allows the construction of fuzzy models from data so that it is not necessary to characterize the system internally. In general, this task is achieved through the decomposition of nonlinear dynamics in locally linear sub-systems, which reduces the complexity of the problem studied [26].

The structure of the nonlinear system identification method is presented in Fig. 2. The purpose of each step is described below. Data collection is the first step in the identification process and involves determining the inputs and outputs of the system. Results obtained depend on the amount of information included in the selected data.

In the next step, the identification of nonlinear dynamic systems with an input  $\mathbf{u}$  and output  $\mathbf{y}$  is defined as a regression problem in which a regression vector  $\mathbf{x}(k)$  is mapped towards the predicted output  $y(k+1)$ . This process can be done by using the nonlinear auto-regressive model with exogenous input (NARX), which regressors consist of a finite number of past inputs and outputs of the system, whereas the regressand is the predicted output [24,26]. Such a model is defined in the below equation:

$$y(k+1) = F(y(k), \dots, y(k-n_y-1), u(k), \dots, u(k-n_u-1)) \quad (1)$$

where  $k$  symbolizes the value of the current sample and  $n_u$  and  $n_y$  refer to the number of delays of the input and the output, respectively and  $F$  is the fuzzy system. Thus, the identification problem of a nonlinear system consists in an optimization problem where the objective function  $F$  minimizes the error of the existing data and the obtained approximation. For this reason, it is important to select the regressors and the order of the model carefully, since adoption of only a few regressors can generate inaccurate predictions, and a higher number of them can over-parameterize the model [24].

#### 2.2.1. Fuzzy clustering

Since the identification method proposed is based on the decomposition of a nonlinear system into locally linear sub-problems, it is necessary to use fuzzy clustering algorithms on the predefined regression surface. These algorithms allow a data set to be

Download English Version:

<https://daneshyari.com/en/article/377650>

Download Persian Version:

<https://daneshyari.com/article/377650>

[Daneshyari.com](https://daneshyari.com)