

Research Article

Capturing the Data Uncertainty Change in the Cocaine Consumption in Spain Using an Epidemiologically Based Model

Christopher Anaya, Clara Burgos, Juan-Carlos Cortés, and Rafael-J. Villanueva

Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Building 8G, Access C, 2nd Floor, Camino de Vera s/n, 46022 València, Spain

Correspondence should be addressed to Juan-Carlos Cortés; jccortes@mat.upv.es

Received 12 September 2016; Revised 31 October 2016; Accepted 8 November 2016

Academic Editor: Gongnan Xie

Copyright © 2016 Christopher Anaya et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A probabilistic model is proposed to study the transmission dynamics of the cocaine consumption in Spain during the period of 1995–2011. Using the so-called probabilistic fitting technique, we study if the model is able to capture the data uncertainty coming from surveys. The proposed model is formulated through a nonlinear system of difference equations whose coefficients are treated as stochastic processes. A discussion regarding the usefulness and limitations of probabilistic fitting technique in order to capture the data uncertainty of the proposed model is presented.

1. Introduction

In [1], the authors presented an epidemiologically based model to study the transmission dynamics of the cocaine consumption in Spain. This model worked well in the presented scenario, 1995–2005, where the population of cocaine consumers had an increasing trend. However, in 2008, the economic crisis began and the Spanish Health Ministry started to apply the National Drug Plan. As a consequence, the nonconsumer population increased and the consumer populations started to decrease. This sudden change could not be captured by the model as it was shown in [2].

In this work, we propose to use a variation of the model stated in [1] and assume that the model parameters are variable over the time in order to see if the variable model parameters are able to adapt and help the model to capture the observed changes in the cocaine consumption.

With this aim, we use the probabilistic fitting technique recently developed in [3, 4] by some of the authors. With this technique, we expect to capture the data uncertainty (data coming from surveys), despite the change of trend, and provide an estimation to the probability distribution of the model parameters. The latter will help us study if the observed data variation also produces variations in some or all model parameters as well as their quantification.

These types of mathematical models also have been used in the study of epidemics and other drug addictions, such as alcohol, tobacco, ecstasy, or heroin addiction [5–9], and in the approach to other sociological topics that are spread by social contact as obesity or extreme ideological behaviors [10, 11].

The work is organized as follows. In Section 2, we present the data we are going to use about cocaine consumers in Spain. In Section 3, we build a type-epidemiological model describing the cocaine consumption dynamics in Spain. Also, the model is scaled in order to match data and model magnitudes. In Section 4, we recall and adapt the *probabilistic fitting* technique introduced in [3]. This technique is applied to the model in order to obtain both data estimation with uncertainty and an estimation of the probability distribution of the model parameters. In Section 5, we present and discuss the results. In Section 6, conclusions are drawn.

2. Data

In this study, we use data from the Survey on Alcohol and Drugs in Spain (EDADES), part of the Spanish National Drug Plan [12]. This survey is published every two years. Specifically, we have focused on the following key question quoted in this survey: *how often have you consumed cocaine?* The available responses were as follows: “never consumed

TABLE 1: Survey results. % of people who belong to nonconsumers, occasional consumers, regular consumers, or habitual consumers of cocaine in Spain from 1995 to 2011.

Survey dates	Nonconsumer	Occasional consumer	Regular consumer	Habitual consumer
$t_1 = 1995$	0.944	0.034	0.018	0.004
$t_2 = 1997$	0.948	0.032	0.015	0.005
$t_3 = 1999$	0.948	0.031	0.015	0.006
$t_4 = 2001$	0.911	0.049	0.026	0.014
$t_5 = 2003$	0.903	0.059	0.027	0.011
$t_6 = 2005$	0.884	0.070	0.030	0.016
$t_7 = 2007$	0.874	0.080	0.030	0.016
$t_8 = 2009$	0.860	0.102	0.026	0.012
$t_9 = 2011$	0.879	0.088	0.022	0.011

(nonconsumers),” “at least once in your life (occasional consumers),” “at least once in the past year (regular consumers),” and “at least once in the past 30 days (habitual consumers).” This survey was launched in 1995 and we have data until 2011. The data are collected in Table 1.

In order to simplify the notation, we are going to define the set of time instants where survey data are available:

$$\begin{aligned} \Omega &= \{t_1 = 1995, t_2 = 1997, t_3 = 1999, t_4 = 2001, t_5 \\ &= 2003, t_6 = 2005, t_7 = 2007, t_8 = 2009, t_9 \\ &= 2011\}. \end{aligned} \quad (1)$$

3. Model Building

From the available survey data, we then introduce an epidemiological model to describe the dynamics of the four subpopulations over the time. While a traditional epidemiological study considers the transmission of a biological contagion, for our study, we will apply a *social contagion* approach. The underlying assumption is the following: we can consider cocaine consumption to be a contagion insofar as the probability that one individual becomes a cocaine consumer depends on the interaction with people who already exhibit this behavior, that is, fellow cocaine consumers [1, 13]. Let us assume *homogeneous population mixing*; that is, each individual can contact with any other individual [14, 15]. With this assumption, we can now build a system of difference equations based on the model introduced in [1] to further describe the transmission dynamics.

As with the survey results, we divide the Spanish 15–65-year-old population (the age group considered in the surveys) into four subpopulations:

- (1) N_t , the amount of people who have never consumed cocaine at year t
- (2) O_t , the amount of people who occasionally consume cocaine (at least once in their lives) at year t
- (3) R_t , the amount of people who regularly consume cocaine (at least once in the past year) at year t

- (4) H_t , the amount of people who habitually consume cocaine (at least once in the past month) at year t

Furthermore, we have the total population

$$T_t = N_t + O_t + R_t + H_t, \quad (2)$$

which is not going to be a constant value over the time.

The transitions between the different subpopulations are determined as follows:

- (i) Let us consider that the newly recruited 15-year-old individuals become members of N_t subpopulation at rate μ ; that is, we consider that they have never consumed cocaine before.
- (ii) An individual in the nonconsumer population can become an occasional consumer through contact with occasional, regular, or habitual consumers. Then, this social contagion term can be modeled with $\beta_t(N_t/T_t)(O_t + R_t + H_t)$.
- (iii) An occasional consumer can increase their consumption and become a regular consumer. This transition term can be modeled by $\gamma_t O_t$.
- (iv) A regular consumer can increase their consumption and become a habitual consumer. We assume that this can be modeled by the term $\sigma_t R_t$.
- (v) A habitual consumer can decide to go to drug-treatment therapy. If he/she stays in therapy for 6 months, according to experts, he/she can be considered a nonconsumer [16]. This term is modeled by $\epsilon_t H_t$.

Using the above assumptions, a dynamic cocaine consumption model for the 15–65-year-old Spanish population is given by the following nonlinear system of difference equations:

$$\begin{aligned} N_{t+1} &= N_t + \mu T_t - d_N N_t - \beta_t \frac{N_t}{T_t} (O_t + R_t + H_t) \\ &\quad + \epsilon_t H_t, \end{aligned} \quad (3)$$

$$O_{t+1} = O_t - d_O O_t + \beta_t \frac{N_t}{T_t} (O_t + R_t + H_t) - \gamma_t O_t, \quad (4)$$

$$R_{t+1} = R_t - d_R R_t + \gamma_t O_t - \sigma_t R_t, \quad (5)$$

$$H_{t+1} = H_t - d_H H_t + \sigma_t R_t - \epsilon_t H_t, \quad (6)$$

where

- (i) μ is the birth rate (the mortality between 0 years old and 14 years old is so small that it can be discarded),
- (ii) d_N is the death rate for nonconsumers,
- (iii) d_O is the death rate for occasional consumers,
- (iv) d_R is the death rate for regular consumers,
- (v) d_H is the death rate for habitual consumers,

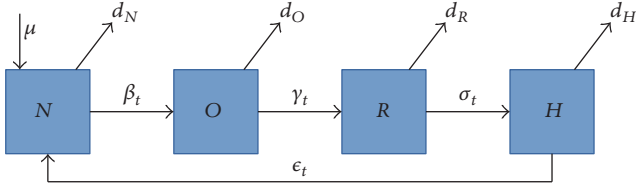


FIGURE 1: Flow diagram of the mathematical model (3)–(6) for the dynamics of cocaine consumption in Spain. The boxes represent the subpopulations and the arrows represent the transitions between the subpopulations. Arrows are labeled by the parameters of the model.

- (vi) β_t is the time-dependent transmission rate for cocaine consumption,
- (vii) γ_t is the time-dependent transition rate between occasional and regular cocaine consumers,
- (viii) σ_t is the time-dependent transition rate between regular and habitual cocaine consumers,
- (ix) ϵ_t is the time-dependent rate at which habitual consumers enter and complete drug-treatment therapy.

Figure 1 shows the flow diagram of the dynamic cocaine consumption model. The boxes represent the subpopulations and the arrows represent the transitions between the subpopulations. Arrows are labeled by their corresponding parameters of the model.

Following the idea proposed in [1], we are going to determine bounds for the model parameters. The birth rate $\mu \in [0.009168052, 0.011275930]$; both ends of this interval represent the minimum and maximum values for the birth rate during the period 1995–2011 [17]. Analogously, the death rate $d_N \in [0.008166102, 0.009228115]$. Taking into account the fact that the death rate of the consumers may be up to 6.8% higher than d_N [12] and the fact that the more an individual consumes drugs the higher probability to die because of drug abuse is, we have that $d_N < d_O < d_R < d_H \leq 0.009228115 \times 1.068$.

Also, we have the time-dependent parameters β_t , γ_t , σ_t , and ϵ_t for $t \in \Omega$ for which we consider that all of them lie between 0 and the double of the values given in [1] (maximum likelihood). Then, $\beta_t \in [0, 2 \times 0.09614]$, $\gamma_t \in [0, 2 \times 0.0596]$, $\sigma_t \in [0, 2 \times 0.0579]$, and $\epsilon_t \in [0, 2 \times 0.0000456]$, for $t \in \Omega$.

The time step for simulations will be of two years according to the time instants in Ω . Then, the total number of model parameters is 1 birth rate, 4 death rates, $9\beta_t$, $9\gamma_t$, $9\sigma_t$, and $9\epsilon_t$, that is, 41 model parameters.

3.1. Model Scaling. Despite the fact that (3)–(6) describe the cocaine consumption dynamics in Spain, there is a mismatch between the input for these difference equations, which is formulated for total population counts, and the survey data available, which is stated in percentages. We therefore need to scale the equations. To do this, we add together (3)–(6). Then, taking into account (2), the left-hand side yields

$$N_{t+1} + O_{t+1} + R_{t+1} + H_{t+1} = T_{t+1}, \quad (7)$$

while the right-hand side yields

$$T_t + \mu T_t - d_N N_t - d_O O_t - d_R R_t - d_H H_t. \quad (8)$$

Now, we shall define

$$\begin{aligned} n_t &= \frac{N_t}{T_t}, \\ o_t &= \frac{O_t}{T_t}, \\ r_t &= \frac{R_t}{T_t}, \\ h_t &= \frac{H_t}{T_t}, \end{aligned} \quad (9)$$

from which we get $n_t + o_t + r_t + h_t = 1$. From here, we will scale the equations, starting with (3). Here, using (8)–(9), one obtains

$$\begin{aligned} n_{t+1} &= \frac{N_{t+1}}{T_{t+1}} \\ &= \frac{N_t + \mu T_t - d_N N_t - \beta_t (N_t/T_t)(O_t + R_t + H_t) + \epsilon_t H_t}{T_t + \mu T_t - d_N N_t - d_O O_t - d_R R_t - d_H H_t}, \end{aligned} \quad (10)$$

and by dividing the numerator and denominator by T_t , one gets

$$n_{t+1} = \frac{n_t + \mu - d_N n_t - \beta_t n_t (o_t + r_t + h_t) + \epsilon_t h_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}. \quad (11)$$

Applying the same method to the other equations, the following expressions are obtained:

$$\begin{aligned} o_{t+1} &= \frac{O_{t+1}}{T_{t+1}} \\ &= \frac{O_t - d_O O_t + \beta_t (N_t/T_t)(O_t + R_t + H_t) - \gamma_t O_t}{T_t + \mu T_t - d_N N_t - d_O O_t - d_R R_t - d_H H_t} \\ &= \frac{o_t - d_O o_t + \beta_t n_t (o_t + r_t + h_t) - \gamma_t o_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}, \\ r_{t+1} &= \frac{R_{t+1}}{T_{t+1}} \\ &= \frac{R_t - d_R R_t + \gamma_t O_t - \sigma_t R_t}{T_t + \mu T_t - d_N N_t - d_O O_t - d_R R_t - d_H H_t} \\ &= \frac{r_t - d_R r_t + \gamma_t o_t - \sigma_t r_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}, \end{aligned}$$

$$\begin{aligned}
 h_{t+1} &= \frac{H_{t+1}}{T_{t+1}} \\
 &= \frac{H_t - d_H H_t + \sigma_t R_t - \epsilon_t H_t}{T_t + \mu T_t - d_N N_t - d_O O_t - d_R R_t - d_H H_t} \\
 &= \frac{h_t - d_H h_t + \sigma_t r_t - \epsilon_t h_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}.
 \end{aligned}
 \tag{12}$$

And the scaled equations can be written in the following form:

$$\begin{aligned}
 n_{t+1} &= \frac{n_t + \mu - d_N n_t - \beta_t n_t (o_t + r_t + h_t) + \epsilon_t h_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}, \\
 o_{t+1} &= \frac{o_t - d_O o_t + \beta_t n_t (o_t + r_t + h_t) - \gamma_t o_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}, \\
 r_{t+1} &= \frac{r_t - d_R r_t + \gamma_t o_t - \sigma_t r_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}, \\
 h_{t+1} &= \frac{h_t - d_H h_t + \sigma_t r_t - \epsilon_t h_t}{1 + \mu - d_N n_t - d_O o_t - d_R r_t - d_H h_t}.
 \end{aligned}
 \tag{13}$$

4. Probabilistic Fitting

This technique consists of using information of the data survey to assign probability distributions to the data. Then, we sample data values from these probability distributions and fit the model to the sampled data. Thus, we find model parameters that fit not only the data but also the uncertainty contained into the intrinsic survey error. Thus, these model parameters will allow the model to capture the data uncertainty (with 95% confidence intervals).

4.1. Data 95% Confidence Intervals (95% CI). Data in Table 1 correspond to the mean percentage obtained from the EDADES survey conducted between 1995 and 2011 every two years. In the technical specifications of each survey, we can see sample sizes of 15000, 15000, 15000, 15000, 15000, 27934, 23715, 20109, and 22128 interviews, respectively.

Taking into account the fact that the sample is not the same for each survey, let us assume that the survey outputs are independent. For each one of the 9 available surveys, let us denote by $X^j = (X_1^j, X_2^j, X_3^j, X_4^j)$, $0 \leq X_i^j \leq n_j$, $i = 1, 2, 3, 4$, and $j = 1, \dots, 9$, a random vector whose entries are $X_1^j = \#$ nonconsumers, $X_2^j = \#$ occasional consumers, $X_3^j = \#$ regular consumers, and $X_4^j = \#$ habitual consumers and $n_1 = 15000$, $n_2 = 15000$, $n_3 = 15000$, $n_4 = 15000$, $n_5 = 15000$, $n_6 = 27934$, $n_7 = 23715$, $n_8 = 20109$, and $n_9 = 22128$ are the sample sizes of surveys. These components represent exclusive selections (events) with probabilities

$$\begin{aligned}
 P^j(X_1^j = x_1) &= \theta_1^j, \\
 P^j(X_2^j = x_2) &= \theta_2^j,
 \end{aligned}$$

$$\begin{aligned}
 P^j(X_3^j = x_3) &= \theta_3^j, \\
 P^j(X_4^j = x_4) &= \theta_4^j, \\
 & j = 1, \dots, 9,
 \end{aligned}
 \tag{14}$$

where $\theta_1^j, \theta_2^j, \theta_3^j$, and θ_4^j are the percentages collected in Table 1 for each survey j : $j = 1, \dots, 9$. Thus, each random vector X^j follows a multinomial probability distribution. Therefore, the probability that X_1^j occurs x_1 times, X_2^j occurs x_2 times, X_3^j occurs x_3 times, and X_4^j occurs x_4 times is given by

$$\begin{aligned}
 P_{n_j}^j(x_1, x_2, x_3, x_4) &= \frac{n_j!}{x_1! x_2! x_3! x_4!} (\theta_1^j)^{x_1} (\theta_2^j)^{x_2} (\theta_3^j)^{x_3} (\theta_4^j)^{x_4}, \\
 & j = 1, \dots, 9,
 \end{aligned}
 \tag{15}$$

where $x_1 + x_2 + x_3 + x_4 = n_j$. The resulting multinomials for each EDADES survey can be seen in Table 2.

Now, we compute the quantiles 2.5 and 97.5 (95% CI) of each one of the joint multinomial distributions in Table 2: $j = 1, 2, \dots, 9$. The obtained 95% CI are collected in Table 3.

4.2. Probabilistic Estimation. Let $M(t; \alpha)$ be a short representation of the scaled model (13), where α denotes the list of 41 model parameters; that is,

$$\alpha = (\mu, d_N, d_O, d_R, d_H, \beta_{t_1}, \dots, \beta_{t_9}, \gamma_{t_1}, \dots, \gamma_{t_9}, \sigma_{t_1}, \dots, \sigma_{t_9}, \epsilon_{t_1}, \dots, \epsilon_{t_9}).
 \tag{16}$$

Also, we have the data 95% confidence intervals (95% CI) of Table 3 obtained by sampling the joint probability distributions of Table 2 and calculating the percentiles 2.5 and 97.5.

Given the probability distributions P^j , $i = 1, \dots, 9$, in Table 2, we take a sample $d_{i,j}^*$, $i = 1, \dots, 9$ and $j = 1, \dots, 4$, and we look for the values to the parameters α^* so that

$$\left\| \begin{bmatrix} M(t_1; \alpha^*) \\ M(t_2; \alpha^*) \\ \vdots \\ M(t_9; \alpha^*) \end{bmatrix} - \begin{bmatrix} d_{t_1,1}^*, d_{t_1,2}^*, d_{t_1,3}^*, d_{t_1,4}^* \\ d_{t_2,1}^*, d_{t_2,2}^*, d_{t_2,3}^*, d_{t_2,4}^* \\ \vdots \\ d_{t_9,1}^*, d_{t_9,2}^*, d_{t_9,3}^*, d_{t_9,4}^* \end{bmatrix} \right\|_F
 \tag{17}$$

is as minimum as possible, with $\| \cdot \|_F$ being the Frobenius norm [18]:

$$\| [a_{ij}] \|_F = \sqrt{\sum_i \sum_j a_{ij}^2}.
 \tag{18}$$

This procedure is a classic optimization problem that can be carried out through genetic algorithms, PSO, Nelder-Mead, and so forth [19, 20] but having the following key

TABLE 2: Date and data joint multinomial probability function of each survey.

Survey dates	Joint multinomial probability functions
$t_1 = 1995$	$P^1 = P_{15000}^1(x_1, x_2, x_3, x_4) = \frac{15000!}{x_1!x_2!x_3!x_4!} 0.944^{x_1} 0.034^{x_2} 0.018^{x_3} 0.004^{x_4}$
$t_2 = 1997$	$P^2 = P_{15000}^2(x_1, x_2, x_3, x_4) = \frac{15000!}{x_1!x_2!x_3!x_4!} 0.948^{x_1} 0.032^{x_2} 0.015^{x_3} 0.005^{x_4}$
$t_3 = 1999$	$P^3 = P_{15000}^3(x_1, x_2, x_3, x_4) = \frac{15000!}{x_1!x_2!x_3!x_4!} 0.948^{x_1} 0.031^{x_2} 0.015^{x_3} 0.006^{x_4}$
$t_4 = 2001$	$P^4 = P_{15000}^4(x_1, x_2, x_3, x_4) = \frac{15000!}{x_1!x_2!x_3!x_4!} 0.911^{x_1} 0.049^{x_2} 0.026^{x_3} 0.014^{x_4}$
$t_5 = 2003$	$P^5 = P_{15000}^5(x_1, x_2, x_3, x_4) = \frac{15000!}{x_1!x_2!x_3!x_4!} 0.903^{x_1} 0.059^{x_2} 0.027^{x_3} 0.011^{x_4}$
$t_6 = 2005$	$P^6 = P_{27934}^6(x_1, x_2, x_3, x_4) = \frac{27934!}{x_1!x_2!x_3!x_4!} 0.884^{x_1} 0.07^{x_2} 0.03^{x_3} 0.016^{x_4}$
$t_7 = 2007$	$P^7 = P_{23715}^7(x_1, x_2, x_3, x_4) = \frac{23715!}{x_1!x_2!x_3!x_4!} 0.874^{x_1} 0.08^{x_2} 0.03^{x_3} 0.016^{x_4}$
$t_8 = 2009$	$P^8 = P_{20109}^8(x_1, x_2, x_3, x_4) = \frac{20109!}{x_1!x_2!x_3!x_4!} 0.86^{x_1} 0.102^{x_2} 0.026^{x_3} 0.012^{x_4}$
$t_9 = 2011$	$P^9 = P_{22128}^9(x_1, x_2, x_3, x_4) = \frac{22128!}{x_1!x_2!x_3!x_4!} 0.879^{x_1} 0.088^{x_2} 0.022^{x_3} 0.011^{x_4}$

TABLE 3: 95% CI of the EDADES surveys data using the joint multinomial probability function of each survey.

Survey dates	Nonconsumer	Occasional consumer	Regular consumer	Habitual consumer
$t_1 = 1995$	[0.940, 0.947]	[0.031, 0.037]	[0.016, 0.020]	[0.003, 0.005]
$t_2 = 1997$	[0.944, 0.952]	[0.029, 0.035]	[0.013, 0.017]	[0.004, 0.006]
$t_3 = 1999$	[0.944, 0.952]	[0.028, 0.034]	[0.013, 0.017]	[0.005, 0.007]
$t_4 = 2001$	[0.906, 0.916]	[0.046, 0.053]	[0.023, 0.029]	[0.012, 0.016]
$t_5 = 2003$	[0.898, 0.908]	[0.055, 0.063]	[0.024, 0.030]	[0.009, 0.013]
$t_6 = 2005$	[0.880, 0.888]	[0.067, 0.073]	[0.028, 0.032]	[0.015, 0.018]
$t_7 = 2007$	[0.870, 0.878]	[0.076, 0.083]	[0.028, 0.032]	[0.014, 0.018]
$t_8 = 2009$	[0.855, 0.865]	[0.098, 0.106]	[0.024, 0.028]	[0.011, 0.014]
$t_9 = 2011$	[0.875, 0.883]	[0.084, 0.092]	[0.020, 0.024]	[0.010, 0.012]

TABLE 4: Model fitting to N samples of the data's probability distributions.

Error	Parameters	Model
e_1^*	α^1	$M(t; \alpha^1)$
e_2^*	α^2	$M(t; \alpha^2)$
\vdots	\vdots	\vdots
e_N^*	α^N	$M(t; \alpha^N)$

difference: now we will fit the data sampled from probability distributions instead of the raw data. We shall perform the fitting N times (N being a large number), storing both the parameter values α^* and the calculated errors e^* , ordered from smallest to largest errors. The result of this procedure is a list of model parameters fitted to a sample of the data with their corresponding errors represented in Table 4.

The value N should be a large number in order to capture as much data uncertainty as possible during the sampling process and this uncertainty could be fitted by the model. In our case, we take $N = 25000$.

Now, we take $M(t; \alpha^1)$ and $M(t; \alpha^2)$ in Table 4 and calculate the outputs for times t_1, \dots, t_9 in Ω , the time instants where data are available. For each time instant, we shall calculate percentiles 2.5 and 97.5, one for each one of the 4 subpopulations. Hence, we will name m_2 the sum of the following:

- (i) The Frobenius norm of the difference between the percentiles 2.5 from the model output and from the data percentiles 2.5 in Table 3
- (ii) The Frobenius norm of the difference between the percentiles 97.5 from the model output and from the data percentiles 97.5 in Table 3

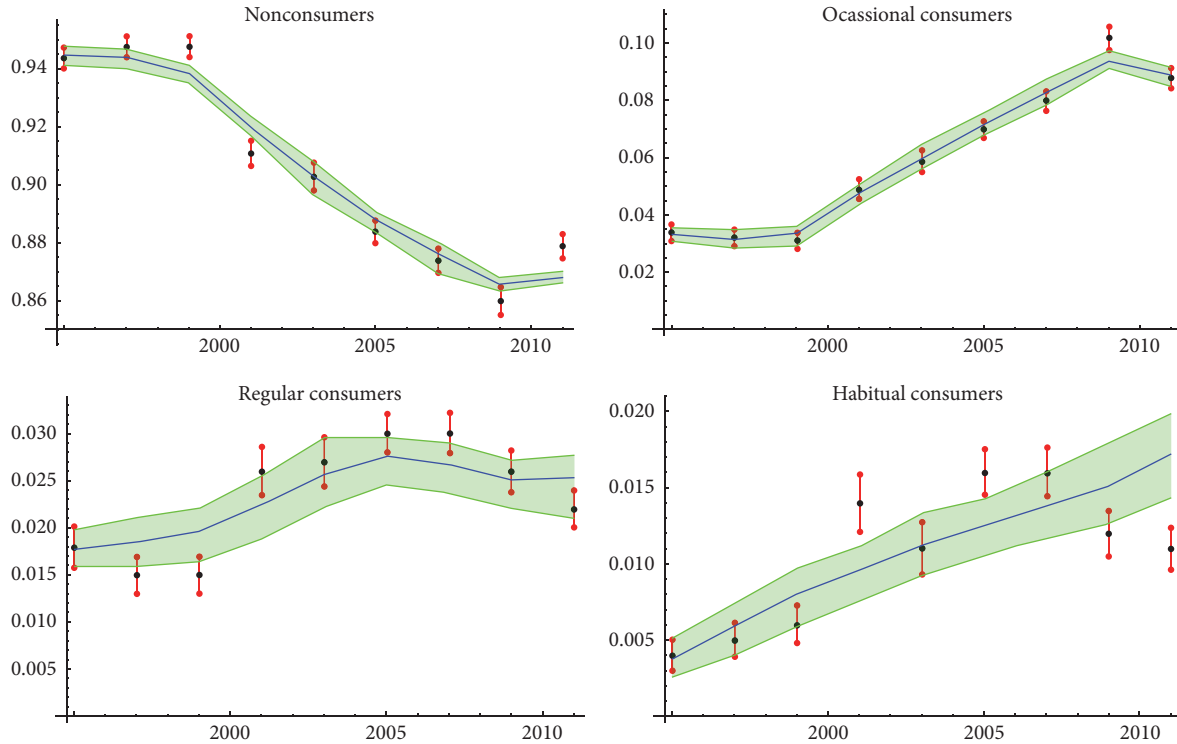


FIGURE 2: Result of the probabilistic fitting. The green band corresponds to 95% CI model output and the blue line its mean. The red points are the 95% CI data given in Table 3 and the black points their means. We can see how the model tries to capture the data uncertainty, but the goal is not achieved completely.

We will repeat the above process with the outputs from $M(t; \alpha^1)$, $M(t; \alpha^2)$, and $M(t; \alpha^3)$, obtaining m_3 , the measure between the confidence bands from the outputs and the data. The same shall be done for m_4 with $M(t; \alpha^1)$, $M(t; \alpha^2)$, $M(t; \alpha^3)$, and $M(t; \alpha^4)$ and so on, until $M(t; \alpha^1)$, $M(t; \alpha^2)$, \dots , $M(t; \alpha^N)$, obtaining m_N as the measure between the confidence bands from the outputs and the data.

Taking $m_k = \min\{m_1, \dots, m_N\}$, we ensure that the 95% confidence band of the outputs from $M(t; \alpha^1)$, $M(t; \alpha^2)$, \dots , $M(t; \alpha^k)$ is the closest to the 95% confidence band from the data, with which our model will capture the maximum uncertainty of the data from the output of the models $M(t; \alpha^1)$, $M(t; \alpha^2)$, \dots , $M(t; \alpha^k)$.

5. Results

The probabilistic fitting procedure returns, as the best fitting, the one giving for $k = 40$ with an error of $m_k = 0.0502666$. A graphical representation of this result can be seen In Figure 2.

The model 95% CI band captures almost all data uncertainty except 1999 and 2011 in nonconsumers, 2009 in occasional consumers, and 2001 and 2011 in habitual consumers (there is no intersection between the model output band and these data 95% CI).

The decrease in the nonconsumer subpopulation during the period of 1997–2001 is not properly captured by the model. With regard to the economic crisis and the National

Drug Plan proposed in 2008 by the Spanish government, their effect does not appear until 2011, where an increase in nonconsumers and a decrease in occasional consumers arise. As can be seen in Figure 2, for these two subpopulations, the model is unable to capture the data uncertainty. Also, in the habitual consumers' subpopulation, the model still does not realize the trend change.

All these comments lead us to say that the proposed model was fairly good to explain what happened when the number of consumers was increasing, mainly in occasional and regular consumers; nevertheless when the trend changed, it could not adapt and explain entirely what was happening.

With respect to the model parameters, in Figure 3, we can see the estimation of the probability density function (PDF) of β_t , γ_t , σ_t , and ϵ_t for every $t \in \Omega$.

Figure 3 has been built taking $k = 40$ sets of model parameters selected by the probabilistic fitting procedure, in particular corresponding to β_t , γ_t , σ_t , and ϵ_t for $t \in \Omega$, and generate a kernel PDF in each time instant in Ω using the *Mathematica* [21] command `KernelMixtureDistribution []`.

In Figures 3 and 4, we can see the variability of the parameters over the time, trying to adapt to the cocaine consumption behavior changes. Note that ϵ_t has a mean and 95% confidence interval fairly stable, which means that the flow of people going to therapy is almost invariable over the time with independence of the changes in the consumption behavior. Therefore, we could make $\epsilon_t = \epsilon$ for all $t \in \Omega$ and transform this random process into a random variable.

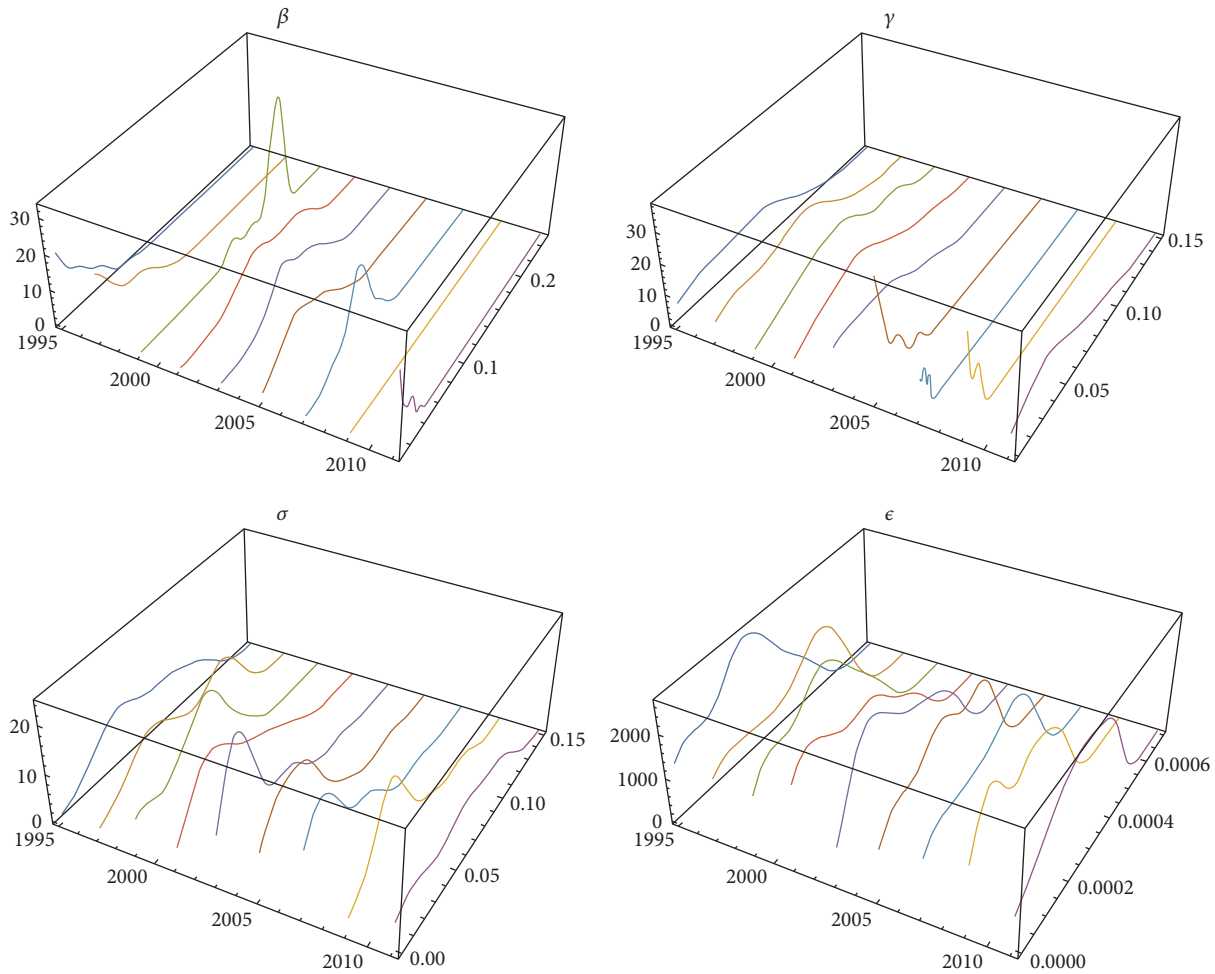


FIGURE 3: Estimation of the random processes $\beta_t, \gamma_t, \sigma_t,$ and ϵ_t using the estimated PDF of the parameters in the time instants in Ω .

The use of variable parameters may allow us to identify changes in the studied phenomenon which have not been considered in the hypotheses or which could arise unexpectedly, as it seems to happen in this regard to the economic crisis and the National Drug Plan, in this case. Accordingly, we can see in Figure 4 how parameters β_t and γ_t , and with lower changes parameter σ_t , try to adapt in the rightmost part of the graphs to data changes, but, unfortunately, they do not change enough to capture the data uncertainty when the consumption trend varies. This leads us to think that the model may not be appropriate to explain the decreasing trends in cocaine consumption.

6. Conclusion

In this work, we present a model formulated through a system of nonlinear difference equations, where the model parameters are variable over the time. This model intends to describe the cocaine consumption dynamics in Spain in the period of 1995–2011. The model follows fairly well the dynamics until 2009, where the effect of two nonconsidered facts, the effect of the economic crisis and the implantation of the National Drug Plan, introduced a change in the behavior of

the individuals, especially in the young people who are the group that is increasing the nonconsumer subpopulation.

We expected that the introduction of variable parameters over the time would allow us to capture the changes in the cocaine consumption dynamics; however, it could not. This indicates that the proposed model may not be appropriate when a decreasing trend in cocaine consumption emerges.

With regard to the parameters, we must say that it was not necessary to consider the parameter ϵ_t as variable, because it is fairly stable over the time. The remaining parameters, $\beta_t, \gamma_t,$ and $\sigma_t,$ are trying to adapt to the data trend changes, but, in nonconsumer and habitual consumer subpopulations, they fail to do it accurately.

Finally, the three sides that this study, a model with variable parameters using the probabilistic fitting technique, provides should be pointed out:

- (1) Consider the model parameters by default as random processes (a family of random variables indexed by time), which is more versatile than considering them as random variables, and then we can detect if they really vary over the time.

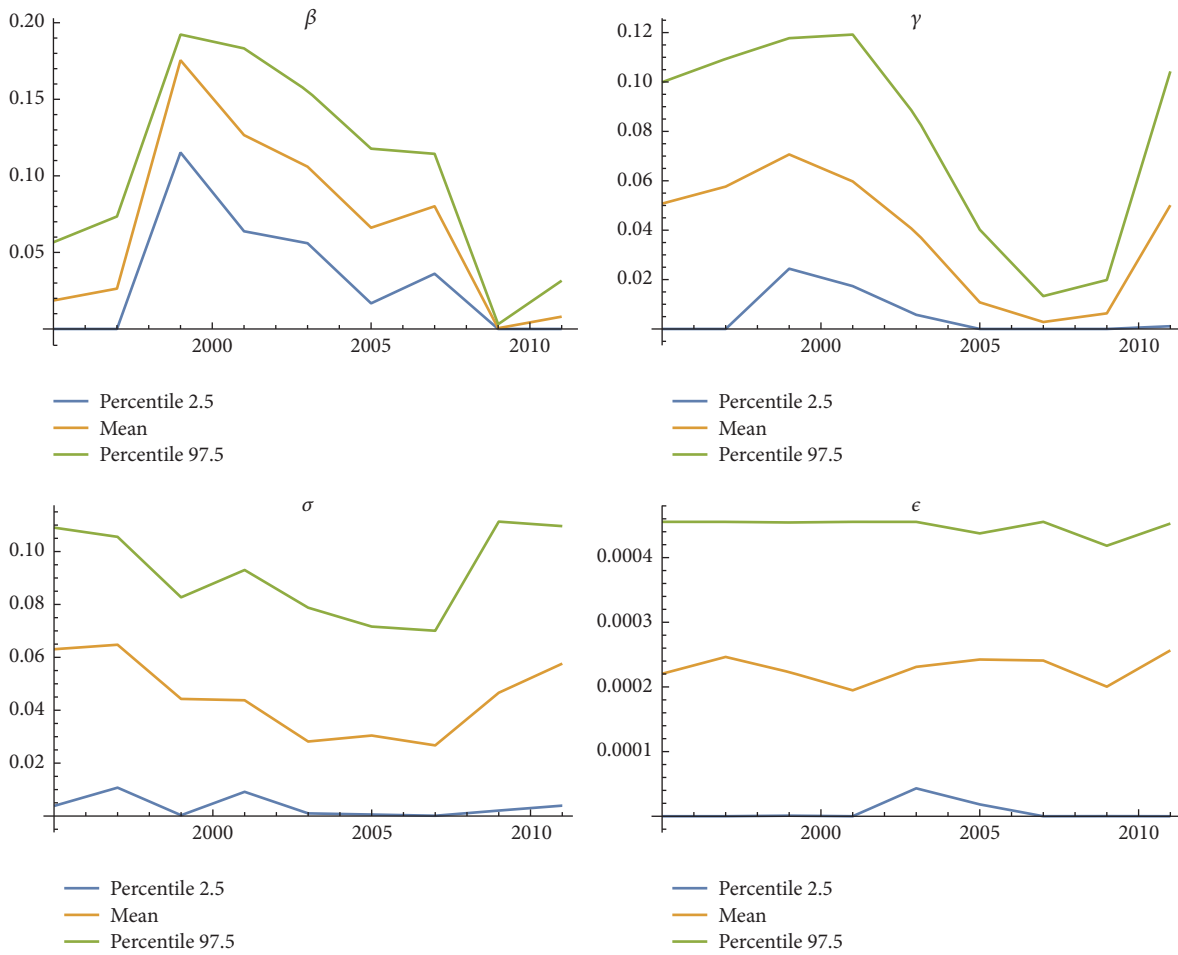


FIGURE 4: Estimated 95% confidence intervals (band) and means of the random processes β_t , γ_t , σ_t , and ϵ_t plotted in Figure 3.

- (2) Study the estimation of the parameter random processes and obtain estimations for the means, the confidence intervals, the PDFs, and hence other higher statistical moments.
- (3) Determine the usefulness of the model in several realistic situations or when data changes because unexpected scenarios emerge.

Although it was not mentioned before, a great computational effort has been done to obtain the presented results and this is one of the major drawbacks we want to improve in future works.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this article.

Acknowledgments

This work has been partially supported by the Ministerio de Economía y Competitividad (Grant MTM2013-41765-P).

References

- [1] E. Sánchez, R.-J. Villanueva, F.-J. Santonja, and M. Rubio, "Predicting cocaine consumption in Spain: a mathematical modelling approach," *Drugs: Education, Prevention and Policy*, vol. 18, no. 2, pp. 108–115, 2011.
- [2] M. Rubio, *Uncertainty quantification in dynamical models. An application to cocaine consumption in Spain [Ph.D. dissertation]*, Universitat Politècnica de València, Valencia, Spain, 2015, <https://riunet.upv.es/handle/10251/55844>.
- [3] J.-C. Cortés, F.-J. Santonja, A.-C. Tarazona, R.-J. Villanueva, and J. Villanueva-Oller, "A probabilistic estimation and prediction technique for dynamic continuous social science models: the evolution of the attitude of the Basque Country population towards ETA as a case study," *Applied Mathematics and Computation*, vol. 264, pp. 13–20, 2015.
- [4] J.-C. Cortés, F. Sánchez, F.-J. Santonja, and R.-J. Villanueva, "A probabilistic analysis to quantify the effect of March 11, 2004, attacks in Madrid on the March 14 elections in Spain: a dynamic modelling approach," *Abstract and Applied Analysis*, vol. 2015, Article ID 387839, 8 pages, 2015.
- [5] F.-J. Santonja, I.-C. Lombana, M. Rubio, E. Sánchez, and J. Villanueva, "A network model for the short-term prediction of

- the evolution of cocaine consumption in Spain,” *Mathematical and Computer Modelling*, vol. 52, no. 7-8, pp. 1023–1029, 2010.
- [6] D. M. Gorman, J. Mezic, I. Mezic, and P. J. Gruenewald, “Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice,” *American Journal of Public Health*, vol. 96, no. 11, pp. 2055–2060, 2006.
- [7] O. Sharomi and A. B. Gumel, “Curtailling smoking dynamics: a mathematical modeling approach,” *Applied Mathematics and Computation*, vol. 195, no. 2, pp. 475–499, 2008.
- [8] B. Song, M. Castillo-Garsow, K. R. Ríos-Soto, M. Mejran, L. Henso, and C. Castillo-Chavez, “Raves, clubs and ecstasy: the impact of peer pressure,” *Mathematical Biosciences and Engineering*, vol. 3, no. 1, pp. 249–266, 2006.
- [9] E. White and C. Comiskey, “Heroin epidemics, treatment and ODE modelling,” *Mathematical Biosciences*, vol. 208, no. 1, pp. 312–324, 2007.
- [10] L. Jódar, F. J. Santonja, and G. González-Parra, “Modeling dynamics of infant obesity in the region of Valencia, Spain,” *Computers & Mathematics with Applications*, vol. 56, no. 3, pp. 679–689, 2008.
- [11] F. J. Santonja, A. C. Tarazona, and R. J. Villanueva, “A mathematical model of the pressure of an extreme ideology on a society,” *Computer and Mathematics with Applications*, vol. 56, no. 3, pp. 836–846, 2008.
- [12] Ministerio de Sanidad, Servicios Sociales e Igualdad (Ministry of Health, Social Services, and Equality), Observatorio Español Sobre la Droga y las Toxicomanas, Informe 2015 (Observatory in Spain Concerning Drugs and Addiction, 2015 Report), http://www.pnsd.mssi.gob.es/profesionales/sistemasInformacion/informesEstadisticas/pdf/INFORME_2015.pdf.
- [13] N. A. Christakis and J. H. Fowler, *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, Brown and Company, Hachette Book Group, 2009.
- [14] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Springer Science & Business Media, 2001.
- [15] J. D. Murray, *Mathematical Biology*, Springer, New York, NY, USA, 2002.
- [16] L. Dutra, G. Stathopoulou, S. L. Basden, T. M. Leyro, M. B. Powers, and M. W. Otto, “A meta-analytic review of psychosocial interventions for substance use disorders,” *The American Journal of Psychiatry*, vol. 165, no. 2, pp. 179–187, 2008.
- [17] INE, Spanish Statistic Institute, <http://www.ine.es>.
- [18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 4th edition, 2012.
- [19] C. Jacob and N. Khemka, “Particle Swarm Optimization in Mathematica an exploration kit for evolutionary optimization,” in *Proceedings of the 6th International Mathematica Symposium (IMS '04)*, Banff, Canada, 2004.
- [20] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [21] Mathematica, <http://www.wolfram.com/mathematica/>.