

GENERATING ARTIFICIAL POPULATIONS USING A MULTI-LEVEL FUZZY INFERENCE ENGINE

Carlos R. García-Alonso
Gabriel M. Pérez-Alcalá

Escritor Castilla Aguayo, 4
ETEA, Business Administration Faculty
Córdoba, 14004, SPAIN

ABSTRACT

The design of complex artificial populations is the first step in simulating evolution during the time span of socio-economic variables as the family income. In this paper, a new hybrid model based on Monte-Carlo simulation and fuzzy inference is described to design environmental conditions, the basic socio-economic structure and to determine the causes for mortality in an artificial population. The model is based on three main databases that describe the characteristics of the environment, individuals and mechanisms (mortality). These expert-based characteristics guide the simulation model which has a fuzzy inference engine to evaluate fuzzy dependence relationships. These relationships have been formulated to automatically determine complex environmental and individual characteristics as well as mechanism parameters, and they are based on expert knowledge. An artificial population has been designed with satisfactory results when critical design factors are carefully adjusted.

1 INTRODUCTION

In order to analyze evolution throughout the time span of social, demographic or economic variables, researchers have built models that usually describe only a specific situation (Stern 1997). These models are used to evaluate political alternatives or to suggest different courses of action (Kydland 2006). But the nature of their specifications render them useless in other contexts, when the original situation changes, due to difficulties in determining relations among covariates or, even worse, because information is incomplete and must be estimated.

Monte-Carlo simulation has been widely used to model and design artificial data series based on some fixed prerequisites (statistical distributions) defined by experts depending on the purpose of the study (Fishman 1996, Stern 1997). This methodology can be used to model complex events but, in order to avoid absolutely blind models, it is absolutely necessary to manage expert knowledge (always

scarce) to define, if possible, the variable behavior (represented by mathematical functions, statistical distributions, etc.), on one hand, and the relationships among them (many, complex and sometimes unknown), on the other. In these cases, variable and parameter values in the model depend on the values of other variables and/or parameters (of the model or external) and, sometimes, on their own values. There are many approaches for modeling and evaluating these dependence relationships (Setnes et al. 1998, Cox 2005), but when it is not possible to define algebraic expressions, only artificial intelligence approaches can be used, if expert knowledge can be explicitly stated. Fuzzy logic can be considered as an appropriate tool when expert-based rules can be defined (Cordon et al. 2002, Cordon et al. 2003, Cox 2005). The huge amount of these expert rules usually needed to model a framework under study can be a serious problem because experts cannot accurately define all of them (Wang et al. 2005) and, in addition, their evaluation is very computer demanding (Gegov 2007).

In order to deal with the rule number problem (Güven and Passino 2001, Pal et al. 2002), in this paper we propose a new procedure to make rule design and evaluation automatic. This procedure can only be used in structured systems where experts can identify which variables (outputs) depend on others (inputs) as well as some additional parameters to use as rule-design guidelines. Rather than defining and stating all the rules explicitly in a rule base (Xiong and Litz 2002), this procedure designs, instantiates and evaluates only the appropriate rules and when needed.

The aim of this paper is to design and develop a Monte-Carlo simulation model that incorporates a multi-level fuzzy inference engine to generate and understanding the evolution of complex artificial populations of individuals based on expert knowledge.

A population is a set of individuals. Each individual is born in a specific socio-economic environment and dies due to different causes that can be algebraically described by mortality functions. These sets of functions are called mechanisms (i.e. mortality). Once the environmental characteristics are known, those associated to each specific in-

dividual can be determined at the moment of his/her birth. From this starting point, the individual characteristics evolve the function of his/her age according to complex relationships among individual and socio-economic covariates. Finally the individual dies, again due to complex reasons.

The structure of the socio-economic environment has been defined by a set of complex variables. Basically, there are two different variable types: non-conditioned and conditioned. The non-conditioned variable (inputs) values are determined first by a standard Monte-Carlo model (Fishman 1996) taking into consideration their appropriate expert-based statistical distributions. Conditioned variables (outputs) are related to non-conditioned or other conditioned variables (previously evaluated and considered inputs). Each conditioned variable or output defines a dependence relationship (DR). These DRs can be defined by algebraic functions, statistical distributions or, finally, fuzzy rules. Depending on the DR type chosen and the values of non-conditioned or conditioned variables, output values can be finally evaluated.

Individual characteristics can also be classified using the same categories: non-conditioned (inputs) and conditioned (inputs/outputs). Their values evolve through time (the age of the individual) according to both the environmental socio-economic characteristics and his own. Non-conditioned characteristics, like individual gender and his/her age, are strange in complex systems. The rest, conditioned, depend on both environmental (per capita income, ...) and individual (age, gender, family income, ...) characteristics. These DRs can be also modeled by algebraic functions, statistical distributions or fuzzy rules to be evaluated by the appropriate algorithms.

Our proposed Monte-Carlo Model has two sections. First, the socio-economic variables that describe the evolution of the artificial environment are determined based on an expert-driven framework. In the second, individuals are born (birth rate), live (they learn, generate income and so on) and, finally, die. In both sections non-conditioned variables (inputs) are determined by the Monte-Carlo model according to specific and expert-based probability distributions. The rest can be evaluated by an algebraic solver, by Monte-Carlo simulation or by our fuzzy inference engine, according to the predefined framework.

In order to check our proposed model, an artificial population has been designed based on the characteristics of a medium-large sized town in a developed country. DRs were mostly defined by fuzzy rules because there was not enough evidence to define specific algebraic functions for all the possible inputs and outputs combinations. Depending on individual characteristics, four different mortality causes are evaluated. The mortality mechanisms are (Carnes et al. 1996): child (including maternal and infant) mortality, non-specific base mortality, accidental and disease mortality (which can include wars and natural disas-

ters) and, finally, deterioration mortality. All of them have been designed as functions of a time whose results can be evaluated directly or by applying conditioned probabilities according to standard probability functions.

Results verify that the behavior of the model and the structure of the resulting population matches expectations.

This paper is structured as follows. First, the Monte-Carlo model is defined in section two. The third section introduces the structure of a real example. Some relevant results obtained using the predefined structure are described in section four. Finally, a brief discussion of relevant issues concludes this paper.

2 MODEL STRUCTURE

The Monte-Carlo model is based on four main elements: i) databases; ii) dependence relationships, iii) the fuzzy inference engine; and, iv) the simulation model.

2.1 Databases

Our model is structured on three main databases: i) environment; ii) individuals; and iii) mechanisms.

2.1.1 Environmental Characteristic Database

Socio-economic characteristics, like per capita income and social security expenditures, are essential (Wood et al. 1980, Carnes et al. 1996, Murray et al. 1996, Olshansky et al. 1997) for the design of artificial populations because they always condition individual behavior. Experts know that these DRs exist, but their structure is complex and, on many occasions, unknown.

Each environmental characteristic is identified and can be described by experts. Its structure has many parameters that define its behavior in the Monte-Carlo model. It can be non-conditioned (input) or conditioned (input/output) and, in both cases, it can be relevant during the whole life of the individual or only at a/some specific age ranges. Its value can be determined by: experts (parameter), an algebraic function, statistical distribution or, finally, a DR (fuzzy rules). In these three final cases, its value can depend on all or some (including its own previous values) environmental characteristic values. Finally, an environmental variable evolves through the time: increasing, decreasing, reaching peaks or having a conditioned probability.

2.1.2 Individual Characteristic Database

Each specific individual in the artificial population is defined by a set of characteristics (variables) that determines his/her life. These characteristics like, for example, gender, family income and so on, are also chosen by experts. Each individual characteristic or variable has a structure similar to that which was described for environmental ones. It can

also be non-conditioned or conditioned, can be relevant in the whole life of the individual or only in a/some age ranges and, finally, can have, or not, minimum and maximum admissible values. The variable value can be determined by the Monte-Carlo model when is fitted to a statistical distribution (with conditional probabilities or not), but it can also be calculated by an algebraic function or by a DR. Finally, it evolves through time and its value can be more or less constant, increase, decrease or can fluctuate.

2.1.3 Mechanism Parameter Database

A mechanism is a process, defined by a set of algebraic functions, that must be evaluated to be able to design a reliable artificial population of individuals. Some examples of mechanisms are: mortality, employment, etc. Mechanisms are evaluated on individuals, for example: individuals die, can work and so on; and evolve through time depending on the individual and environmental characteristics.

Parameter values in mechanism functions are defined by DRs once the function is instantiated. This means that there is enough expert knowledge to define the global structure of mechanism functions through time, but experts cannot accurately define their parameter values for each specific situation. These parameters are always related to individual and environmental characteristics and can be evaluated by the fuzzy inference engine. As mentioned above, parameters are determined when the corresponding function is instantiated. This procedure means that the parameter values are different for each individual (the function structure is identical) and are inherited during his/her life. Nevertheless and depending on environmental and/or individual variable values, these parameter values can be distorted for a specific year/age according to conditioned probabilities.

In this paper, only one mechanism has been described: mortality. It was defined by four mortality functions: child (decreasing exponential. Lynch 2006), non-specific base (degenerated normal. LeClere 2000), accident and disease (displaced Weibull. Murray, et al. 1996) and deterioration (increasing exponential. Gavrilov et al. 2003). In all of them, parameters were determined by the fuzzy inference engine using DRs where the output was the corresponding parameter and the inputs, individual and environmental characteristics.

2.2 Dependence Relationship Structure

In our model, many variables as well as mechanism parameters depend on individual or environmental characteristics. These relationships are defined by fuzzy rule sets where individual and/or environmental characteristics are the inputs and other individual or environmental characteristics and mechanism parameters are the outputs. The set of

rules that relates one output with some inputs is called a dependence relationship (DR) and its basic structure must include:

- Inputs/Output. For each output, inputs must be defined. Inputs can be ranged according to expert knowledge.
- Time ranges. DRs can be applicable only at specific time/age ranges (i.e. child mortality).
- Function value ranges. Output values (results) can be ranged.
- Conditioned probabilities. Output values can be fitted to specific probability functions.

2.3 The Fuzzy Inference Engine

For each DR, inputs and outputs are expressed in linguistic terms using semantic labels (membership functions or fuzzy sets in Fuzzy Logic). For example: the family income is *high* or the national health care system is *poor*. Semantic labels define fuzzy sets that allow us to evaluate fuzzy relationships using *IF* input1 *IS* label₁ *and* input2 *IS* label₂ *and* ... *THEN* output *IS* label_o, rules (Cordon 2002). Once the real value for a specific population characteristic is known, we can fuzzify it determining its membership functions (MF) –semantic labels- and their corresponding membership degrees (MF_i , $\mu_{MF_i}(x_i)$) ($i=1,2, \dots,m$), m being the number of inputs. The maximum number of rules that the system should be capable of managing, r , is an exponential function (Cordon et al. 2002, Cordon et al. 2003) of the number of inputs m , membership functions p and, finally, the number of DRs d in the form:

$$r = dp^m \quad (1)$$

Depending on d , p and m values, r in (1) can be enormous making it impossible for the experts' explicit definition one per one (Dubois et al. 1997). This means that the Monte-Carlo model must automatically design and evaluate the corresponding rules (depending on the input values) when they are needed.

This procedure can be completely automatic and needs a simple rule base that should include:

- DR description: inputs and outputs, type and intensity (described in detail in the next section).
- The relative weight of each input on the output defined by the characteristics (range, mean, standard deviation ...) of a probability distribution chosen by the experts.
- The influence (positive or negative) of each input on the output.

In order to design and evaluate automatically fuzzy rules, an inference engine has been included in the Monte-Carlo simulation model. Fortunately, population design is a structured system where experts can identify which inputs, to a certain extent, are related to specific outputs: the DR basic structure. In this phase of knowledge acquisition, sys-

tem designers can also request experts to identify a DR type. A DR type is a category, defined by system designers and comprehensible for experts, that describes how specific combinations of input MFs identify the output MF in a fuzzy inference engine. For example, five types of DR can be defined as follows:

- Very negative (*VN*) type. Given the MFs for the DR inputs (identified in a fuzzification process) the corresponding ones for the outputs are shifted far to the left.
- Negative (*N*). Given the MFs for the DR inputs, the corresponding ones for the outputs are shifted slightly to the left.
- Neutral (*X*). Given the MFs for the DR inputs, the corresponding ones for the outputs are located in the same relative position. It is the default selection.
- Positive (*P*). Given the MFs for the DR inputs, the corresponding ones for the outputs shift slightly to the right.
- Very positive (*VP*). Given the MFs for the DR inputs, the corresponding ones for the outputs are shifted far to the right.

This procedure makes the complete (but simple) automation of rule design and evaluation possible and it does not imply that the MF number for each input and output should be identical. Each input can have an influence on the output that is positive (the greater the value –or MF– of the input, the greater the value –or MF– of the output) or negative (the greater the value –or MF– of the input, the lower the value –or MF– of the output). Experts must define the influence of each input as well as its relative weight (relevance) with respect to the output. In a Monte-Carlo simulation model, these weights can be modeled selecting appropriate probability distributions.

Once the input weights w_i ($i=1,2, \dots,m$) are determined at random (according to their probability distributions, Monte-Carlo) for a specific simulation and their corresponding input MF_i are identified (the fuzzification process determines all standard grades of membership $\mu_{MF_i}(x_i) \in [0,1]$), the MF for a specific output MF_o is:

$$MF_o = \text{round} \left[\left(\frac{\sum_{i=1}^m w_i MF_i}{\sum_{i=1}^m w_i} \right) + Q \right]$$

Q being an integer chosen according to p ($|Q| \leq (p-1)/2$) and the DR type selected by the experts. Q can be positive (positive DR types: *P* and *VP*), zero (neutral) or negative (negative DR types: *N* and *VN*). If $Q > 0$, then neutral MF_o will be biased to the right and vice versa.

The selection of both the number of MFs p and the DR type determines the design (now automatic) and the behavior of the fuzzy rules in the fuzzy inference engine. This process automatically determines MF_o whatever the values of the inputs might be. This means that the rules are always

true and their fulfillment does not admit nuances. Usually, experts cannot agree with this fuzzy inference engine behavior. The experts are capable of identifying fuzzy ranges for each input where the fulfillment degree fd of the corresponding fuzzy rules (automatically designed) can be more or less arguable (vagueness and ambiguity). When this situation occurs, new easy-to-understand categories, called DR intensities, can be defined by system designers in order to describe the inner ambiguity of each specific DR. For example, if five intensities are chosen, they would be:

- Neutral (*N*). No intensity is evaluated. It is the default selection.
- Increasing (*I*) intensity. The greater the MF_i ($i=1,2, \dots,m$), the greater the rule fulfillment degree.
- Decreasing (*D*) intensity. The greater the MF_i , the lower the rule fulfillment degree.
- Centered-increasing (*CI*) intensity. The more centered the MF_i , the greater the fulfillment degree.
- Centered-decreasing (*CD*) intensity. The less centered (in the extremes) the MF_i , the greater the fulfillment degree.

For each rule and in each simulation, the determination of the fulfillment degree ($0 \leq fd \leq 1$) can be considered as a standard fuzzy problem that can also be solved by a fuzzy inference engine once MF_i ($i=1,2, \dots,m$) are known. When a pseudo-random number generated by the computer is lower than fd , the fulfillment of the rule is arguable and MF_o is determined by:

$$MF_o = \text{round} \left[\left(\frac{\sum_{i=1}^m w_i MF_i}{\sum_{i=1}^m w_i} \right) + Q + D \right]$$

D being the distortion, an integer $|Q| + |D| \leq (p-1)/2$ that shifts the original MF_o to the right or to the left. The D value depends on the fd value. D is selected at random within a range (composed by negative and positive integers, zero is also included), the lower the fd , the wider the range.

The fuzzy inference engine, specifically designed to evaluate mechanism parameters –outputs, has finally the following structure:

- Input values –characteristics of a population group– x_i ($i=1,2, \dots,m$) are fuzzified ($MF_i, \mu_{MF_i}(x_i)$).
- Input weights w_i ($i=1,2, \dots,m$) are determined at random (Monte-Carlo) according to their expert-defined probability distributions.
- For each MF_i ($i=1,2, \dots,m$) combination, the fuzzy inference engine determines the fulfillment probability fd according to the intensity selected by the experts for the corresponding DR [3,4].
- If fd is lower than a pseudo-random number generated by the computer, then the distortion D is determined at random within an automatically generated range of integers $[-v,v]$ where v depends

on fd , that is: the lower the fd , the greater the v (being $|v| \leq (p-1)/2$).

- Again for each MF_i ($i=1,2, \dots,m$) combination, the fuzzy inference engine determines MF_o as if the neutral DR type had been selected.
- According to the DR type chosen and the MF_o previously determined, Q is then calculated (if the neutral type was selected $Q=0$). The easiest way to evaluate Q for a DR type or category is adding or subtracting 1 to the corresponding Q value of the previous category. For example, if in the neutral (X) type $Q=0$, then in the positive (P) type $Q=1$ and in the very positive (VP) category $Q=2$ and the contrary holds true for negative types.
- Taking into consideration D and Q values, the final MF_o is calculated for each MF_i ($i=1,2, \dots,m$) combination.

Finally, some relevant parameters must be defined by experts:

- The number of categories for DR types and intensities (in our example five each).
- The number of MFs for fuzzifying/defuzzifying (precision and accuracy) the input and output values.
- The number of MFs for evaluating the degree of fulfillment .

The structure proposed to model DRs can be easily formulated in a standard relational database and admits, in a very compact format, a huge amount of DR and, consequently, rules.

2.4 The Simulation Model

The Monte-Carlo simulation model is structured in two sections: the former determines the environment structure in each period of time (by default, the year is the time unit) and the latter designs the artificial population. Table 1 shows the model pseudo-code.

In the first section (steps 2 to 4, Figure 1) the evolution of environmental socio-economic variables is determined using the process that was described in section 2.1.1.

In the first loop of the second section (step 5 in Figure 1), The number of individuals in each *year* is defined by the birth rate (a new generation) which was fitted to an expert-based uniform statistical distribution and evaluated by the Monte-Carlo model.

In the second loop of the second section (step 7 in Figure 1), a new individual in the generation is born and starts living. For each individual age (third loop of the second section, step 8 in Figure 1) his/her characteristics are evaluated depending on, if it was established before, environmental and/or other individual characteristics. Then

(step 10 in Figure 1), the mechanism parameters are determined by the fuzzy inference engine according to the specific characteristics of the new individual. Finally (step 11 in Figure 1), the new individual continues alive until a mortality cause is fulfilled.

Table 1: Model Pseudo-Code. ⁽¹⁾number of years to be analyzed, ⁽²⁾number of individuals (birth rate is a predefined parameter) and ⁽³⁾maximum age that an individual can reach (i.e. 100 years old).

Step	Description
1	Determination of MF limits for all the inputs and outputs
2	For <i>year</i> =1 to <i>years</i> ¹ do
3	Structure of the socio-economic environment: Monte-Carlo simulation and fuzzy inference
4	End for <i>years</i>
5	For <i>year</i> =1 to <i>years</i> ¹ do
6	Number of children that will be born <i>nindividuals</i> ² : Monte-Carlo simulation
7	For <i>individual</i> =1 to <i>nindividuals</i> do
8	For <i>age</i> =1 to <i>agemax</i> ³ do
9	Characteristics of the <i>individual</i> of <i>age</i> years old: Monte-Carlo simulation and fuzzy inference
10	Mechanism parameter determination according to the <i>individual</i> and environmental characteristics: Monte-Carlo simulation (conditioned probabilities) and fuzzy inference
11	Determination of the mortality causes for the <i>individual</i> of <i>age</i> years old
12	Evaluation of the mortality causes
13	If <i>individual</i> =dies
14	The <i>age</i> loop finishes
15	End if
16	End for <i>age</i>
17	The life and the mortality cause of the <i>individual</i> are saved
18	End for <i>individual</i>
19	End for <i>years</i>

Once an individual is born and for each year of his/her life, four mortality causes (except if an age range is defined, section 2.1.3) are evaluated (step 12 in Figure 1) and if a pseudo-random number is lower than one or more of them, then the individual dies. The life (years) and the mortality cause are saved in the result database (step 17 in Figure 1).

Individuals have a maximum expecting life (selected by experts), if they are still alive at this moment they die (no individual can be older than the predefined expecting life). Once the population is older than the individual expecting life, it can be considered stable.

3 AN ILLUSTRATIVE EXAMPLE

To check the model, five socio-economic variables (Wood et al. 1980, Carnes et al. 1996, Murray et al. 1996, Olshansky et al. 1997) were chosen to describe the population environment: national per capita income (*PCI*), public health expenditures (*PHE*), social security expenditures (*SSE*), annual investment in public infrastructure (*IPI*) and, finally, insecurity and accident rate (*IAR*). All of their values were standardized in a global range [0,100]. Their structures are defined in Table 2.

Table 2: Structure of the socio-economic variables in the environment database.

	Conditioned	Function	Range	Evolution
<i>PCI</i>	No	Uniform	[40,75]	Increasing
<i>PHE</i>	Yes	Fuzzy	[50,90]	None
<i>SSE</i>	Yes	Fuzzy	[50,80]	None
<i>IPI</i>	Yes	Fuzzy	[50,90]	None
<i>IAR</i>	No	Uniform	[5,25]	Constant

Taken into account their statistical distributions (Table 2), non-fuzzy variables *PCI* and *IAR* were calculated at random by the Monte-Carlo model in the first year. Their evolution from year 2 (step 3 in Table 1) is *increasing* for *PCI* with a probability of 0.85 and *constant* for *IAR* with a probability of 0.5. Their values were finally obtained using uniform statistical distributions based on their corresponding evolutions.

PHE, *SSE* as well as *IPI* variables were evaluated using the fuzzy inference engine. They defined three DRs (described in Table 3) where their values are determined depending on *PCI* and *IAR* values. Taken into account 9 MFs, 243 rules can be potentially evaluated.

Table 3: Structure of the DRs for evaluating environmental variables: DR type, Probability of rule fulfillment –the rule is instantiated, DR intensity, intensity minimum probability, *PCI* weight and influence (+: positive; -: negative), *IAR* weight and influence (+,-).

	Structure
<i>PHE</i>	(<i>X</i> ,0.98, <i>I</i> ,0.25,0.85 <i>PCI</i> +,0.15 <i>IAR</i> -)
<i>SSE</i>	(<i>X</i> ,0.90, <i>I</i> ,0.25,0.90 <i>PCI</i> +,0.10 <i>IAR</i> -)
<i>IPI</i>	(<i>X</i> ,0.80, <i>I</i> ,0.25,0.75 <i>PCI</i> +,0.25 <i>IAR</i> -)

Eleven population characteristics (standardized within [0,100]) have also been considered: Gender (*IG*), family income through his/her childhood (*FI*), sanitary environment through his/her childhood (*SE*), parent experience (*PE*), childhood schooling (*CH*), location –rural/urban- of the his/her home in childhood (*LH*), youth schooling (*YS*), parent educational level in his/her childhood (*FE*), adulthood educational level (*AE*), town size where his home is (*TS*) and personal income (*PI*). Their structure are shown in Table 4.

Table 4: Structure of the characteristics for describing individuals in the artificial population. ⁽¹⁾Conditioned; ⁽²⁾Value ranges; ⁽³⁾Age ranges; ⁽⁴⁾To the end of his/her life.

	Cond ¹	Function	VR ²	AR ³	Evolution
<i>IG</i>	No	Binary	[0,0]		None
<i>FI</i>	No	Gamma	[25,75]	[0,18]	Increasing
<i>SE</i>	Yes	Fuzzy	[40,80]	[0,18]	None
<i>PE</i>	No	Normal	[5,50]	[0,5]	Increasing
<i>CH</i>	Yes	Fuzzy	[10,20]	[3,8]	None
<i>LH</i>	No	Uniform	[20,90]	[0,18]	Constant
<i>YS</i>	Yes	Fuzzy	[20,40]	[9,18]	Constant
<i>FE</i>	Yes	Fuzzy	[20,90]	[0,18]	None
<i>AE</i>	Yes	<i>f(CH)</i>	[40,90]	[19,- ⁴]	Increasing
<i>TS</i>	No	Uniform	[30,90]	[19,-]	Constant
<i>PI</i>	Yes	Fuzzy	[20,85]	[19,-]	None

Family income was initially modeled as a gamma (2,18) and fuzzy relationships are described in Table 5.

Table 5: DR structure in the individual characteristic database: DR type, Probability of rule fulfillment, DR intensity, intensity minimum probability, *characteristic* weight and influence (+: positive; -: negative).

	Structure
<i>SE</i>	(<i>X</i> ,0.98, <i>I</i> ,0.25,0.9 <i>FI</i> +,0.1 <i>LH</i> +))
<i>CH</i>	(<i>X</i> ,0.98, <i>NI</i> ,0,0.8 <i>FI</i> +,0.1 <i>FE</i> +,0.05 <i>PCI</i> +,0.05 <i>IPI</i> +))
<i>YS</i>	(<i>X</i> ,0.98, <i>NI</i> ,0,0.8 <i>FI</i> +,0.1 <i>FE</i> +,0.05 <i>PCI</i> +,0.05 <i>IPI</i> +))
<i>FE</i>	(<i>X</i> ,0.98, <i>I</i> ,0.25,0.7 <i>FI</i> +,0.15 <i>PCI</i> +,0.15 <i>IPI</i> +))
<i>PI</i>	(<i>X</i> ,0.98, <i>I</i> ,0.25,0.7 <i>AE</i> +,0.15 <i>TS</i> +,0.1 <i>PCI</i> +,0.05 <i>IPI</i> +))

All variables involved in fuzzy rules were adjusted to nine MFs in the fuzzy inference engine.

Four mortality functions were considered and designed for the mortality mechanism. The child mortality rate (*CMR*) was designed as a decreasing exponential of the time *t* (years) $CMR = be^{-at}$, *a* and *b* being the parameters, $t \in [0,15]$ and its results were conditioned by a normal probability function in the form: $N(CMR + C_{CMR}, 0.1CMR)$, $C_{CMR} \in \mathbb{R}$ being a coefficient that depends on the individual characteristics at the moment (age) *t* when the function is evaluated. The unexpected base mortality rate (*BMR*) was designed as a degenerated normal of the time in the form:

$$BMR = c \frac{e^{\frac{[(t-\bar{x})/\sigma_x]^2}{-2d}}}{\sqrt{2\pi\sigma_x}}$$

where *c* and *d* are the parameters that were calculated once the average \bar{x} and the standard deviation σ_x are defined by the experts. *BMR* results were also conditioned by a normal probability function $N(BMR + C_{BMR}, 0.05BMR)$, $C_{BMR} \in \mathbb{R}$ being a coefficient that depends on individual

characteristic values in t . The accident and disease mortality rate (AMR) was designed as a displaced in time Weibull probability distribution in the form:

$$AMR = \alpha\beta(t-I)^{(\beta-1)} e^{-\alpha\beta(t-I)}$$

α and β being the mechanism parameters, I being the initial moment (age. In this case $I=35$) when the formula goes into effect (a uniform probability distribution $U[14,16]$) because the function is time ranged $t \in [14,100]$. Finally, its results were also conditioned by a normal probability distribution $N(AMR + C_{AMR}, 0.15AMR)$, $C_{AMR} \in \mathbb{R}$ being the distortion coefficient that depends on the individual characteristics in t .

Finally, the fourth function was the deterioration mortality rate (DMR). It was an in time increasing and displaced exponential $DMR = he^{k(t-I)}$, h and k being the parameters, I' being the moment (age) when the formula goes into effect (fitted to a $U[28,35]$) and, finally, $t \in [28,100]$. Its results were also conditioned to a normal probability function in the form $N(DMR + C_{DMR}, 0.1DMR)$, C_{DMR} being a distortion coefficient that again depends on individual characteristics in t . The initial and basic characteristics of all the mechanism parameters are shown in Table 6 and an example of non-distorted mechanism functions are introduced in Figure 1.

Table 6: Membership function (MF) basic structure: range, minimum value (MF min), central value (MF central) and maximum value (MF max), for all the mechanism parameters.

	Range	MF min	MF central	MF max
A	[0.15,0.25]	0.15	0.2	0.25
B	[10E ⁻⁵ ,20E ⁻⁵]	10E ⁻⁵	15E ⁻⁵	20E ⁻⁵
\bar{X}	[20,40]	20.5	30	39.5
σ_x	[2,4]	2	2.8	4
C	[0.35,0.55]	0.35	0.4	0.55
D	[4,6]	4	5	6
I	[14,16]	14	15	16
α	[0.36,0.42]	0.36	0.38	0.42
β	[0.9,1.2]	0.9	1	1.2
H	[8E ⁻³ ,12E ⁻³]	8E ⁻³	9.5E ⁻³	12E ⁻³
K	[0.06,0.07]	0.06	0.065	0.07
I'	[28,35]	28	31	35

All the population characteristics in Table 4 were considered as inputs and the mechanism parameters (except \bar{X} and σ_x) as outputs to be evaluated by the fuzzy inference engine. DRs are described in Table 7.

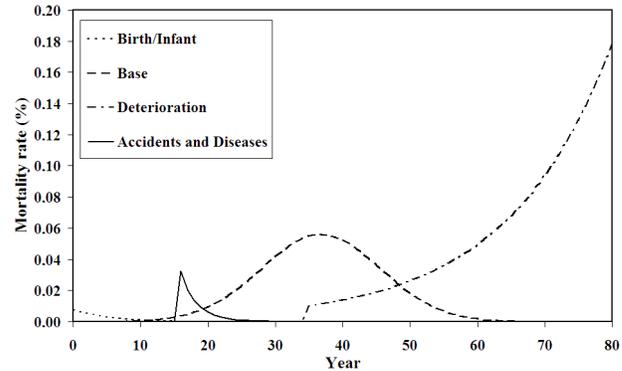


Figure 1: Mechanism (mortality) functions, a detail to highlight the lower mortality rates.

Table 7: DR structure, all of them ($X,0.98,I,0.25$), for determining the mechanism parameter values: *characteristic weight and influence (+, -)*. Gender was considered an input that defines the final structure of the rule: ⁽¹⁾Females; ⁽²⁾Males.

Structure		
a	F ¹	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
b	F	(0.65FI-,0.15SE-,0.1PE-,0.1SSE-)
\bar{X}	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
σ_x	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
c	F	(0.65FI-,0.1SE-,0.15PE-,0.15SSE-)
d	F	(0.7FI-,0.2SE-,0.05PE-,0.05SSE-)
α	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
β	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
I	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
h	F	(0.7FI-,0.1SE-,0.1PE-,0.1SSE-)
k	F	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
I'	F	(0.65FI-,0.15SE-,0.1PE-,0.1SSE-)
a	M ²	(0.65FI-,0.15SE-,0.1PE-,0.1SSE-)
b	M	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
\bar{X}	M	(0.5FI-,0.25SE-,0.15PE-,0.1SSE-)
σ_x	M	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
c	M	(0.65FI-,0.15SE-,0.1PE-,0.1SSE-)
d	M	(0.7FI-,0.2SE-,0.05PE-,0.05SSE-)
α	M	(0.5FI-,0.2SE-,0.15PE-,0.15SSE-)
β	M	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
I	M	(0.6FI-,0.2SE-,0.1PE-,0.1SSE-)
h	M	(0.65FI-,0.15SE-,0.1PE-,0.1SSE-)
k	M	(0.7FI-,0.1SE-,0.1PE-,0.1SSE-)
I'	M	(0.5FI-,0.3SE-,0.1PE-,0.1SSE-)

The birth rate was defined by a uniform probability distribution $U[3.25,3.5]$. The maximum individual age was established at 100 years old and the model was stopped when the 300th generation was created. The fuzzy inference engine uses the product-sum-gravity with superposition method to evaluate fuzzy rules (Cox 2005).

4 RESULTS

The artificial population generated by the Monte-Carlo model reached a stable situation in the 100th generation. Due to the model's complexity, it was parallelized in a multiple processor cluster (5 processors) following the strategy of dividing the expected number of individuals in the final population by 5. The simulation model designed and evaluated successfully the appropriate fuzzy rules, but their results were sometimes unstable depending on their fulfillment probabilities. Figure 2 shows, as an example of the results obtained using the fuzzy inference engine, the evolution through time of *PHE* considered as an output of a DR defined in Table 3.

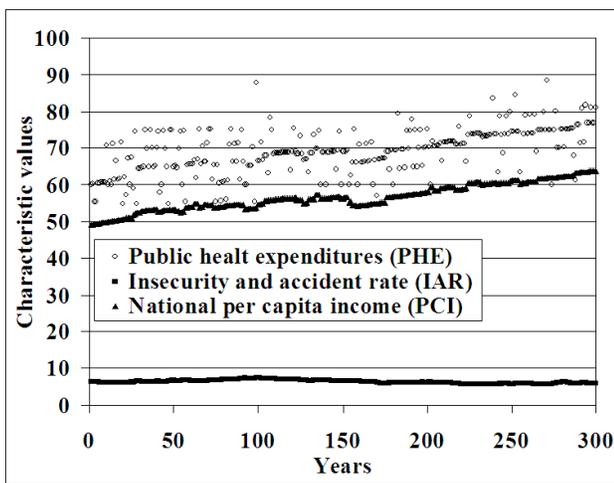


Figure 2: Evolution of public health expenditures based on the national per capita income and insecurity and accident rate.

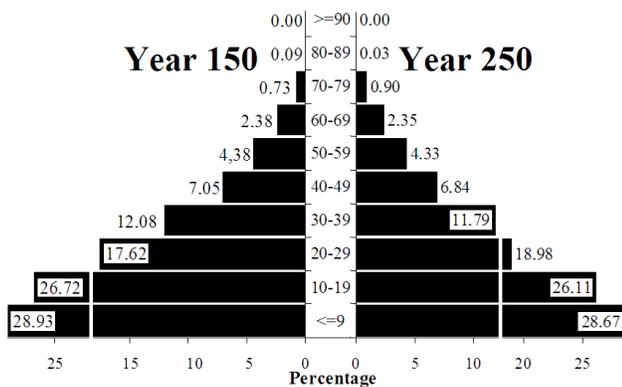


Figure 3: Population structure (age intervals) in two different years once it was stable.

The structure of the population remains constant in relative terms once a stable situation is reached. As it is shown in Figure 3, the population can be considered young (over 50% under 20 years of age) and the relative weights of different age intervals are practically constant if the

150th and the 250th generations are compared. Mann-Whitney U non-parametric test confirmed that neither structure can be considered to come from different populations.

Once a stable situation is reached, the evolving population can be used to evaluate other mechanisms related to family structure and income evolution

5 DISCUSSION AND CONCLUSIONS

In this paper, we have introduced a hybrid system (Monte-Carlo simulation and fuzzy logic) to design artificial populations when complex relationships among environmental/individual characteristics and mechanism parameters should be taken into account. Expert knowledge was incorporated into the Monte-Carlo model by identifying the appropriate algebraic functions, statistical distributions and by defining the structure of DRs.

Fuzzy inference was chosen to formalize the vagueness of the expert's reasoning involved, when DRs must be defined among environmental and individual characteristics as well as mechanism parameters, and to simplify the design process in this large scale but structured system. In the artificial population design, there is enough expert knowledge to define DR structure, but it is not possible to design all the rules needed. In this paper, we have proposed a new procedure to make the design and evaluation of fuzzy rules automatic. This procedure requires: i) identifying the related inputs and outputs (DRs); ii) selecting a type and a rule fulfillment probability for each DR; iii) assuming, if necessary, an intensity and an intensity minimum probability also for each DR; iv) assigning a relative relevance level (weight) for all the inputs in each DR, where a specific probability distribution will determine the behavior of these weights in the Monte-Carlo model; and, finally, v) identifying the influence (positive or negative) of each input on the corresponding output.

Using this simplified way of designing rule bases, a fuzzy inference engine was designed to automatically evaluate the appropriate rules when input values have been determined by the simulation model. This fuzzy inference engine first evaluates the fulfillment degree of the rule to admit nuances in the inference process. These nuances are based on expert knowledge and selected by a specific intensity.

In a second step, the fuzzy inference engine determines if the resulting MF output should be biased or not according to the DR type and intensity selected. Using this procedure a Monte-Carlo simulation model has been designed and developed to design artificial populations. This large scale but, fortunately, structured problem was successfully modeled and results confirm that we are capable of generating stable populations based on expert knowledge.

Fuzzy rules have been designed, instantiated and evaluated correctly by the Monte-Carlo model but their results were sometimes unstable throughout time. There are some critical factors in the design process that must be adjusted very carefully to avoid unexpected results: input/output ranges (time, age and values), DR probabilities (fulfillment and intensity) and, finally, value tendencies (increasing, decreasing, etc.).

Basically, due to the number of generations needed to reach stable but complex populations with many different individuals, the Monte-Carlo model designed is very computer demanding. Nevertheless, the characteristics of the problem make it easy to parallel following different strategies (i.e. number of individuals, gender groups, etc.).

ACKNOWLEDGEMENTS

The author gratefully acknowledges the financial support provided by the Spanish Department of Research of the Ministry of Education and Science under the TIN2005-08386-C05-02 projects. FEDER also provided additional funding.

REFERENCES

- Carnes, B.A., S.J. Olshansky and D. Grahm. 1996. Continuing the Search for a Law of Mortality, *Population and Development Review*, 22(2): 231:264.
- Cordon, O., F. Herrera and I. Zwir. 2002. Linguistic Modeling by Hierarchical Systems of Linguistic Rules. *IEEE Transactions on Fuzzy Systems* 10(1):2-20.
- Cordon, O., F. Herrera and I. Zwir. 2003. Fuzzy Modeling by Hierarchically Built Fuzzy Rule Bases. *International Journal of Approximate Reasoning* 27:61-93.
- Cox, E. 2005. *Fuzzy Modeling and Genetic Algorithms for Data-Mining and Exploration*. Elsevier, San Francisco.
- Dubois, D., H. Prade and L. Ughetto. 1997. Checking the Coherence and Redundancy of Fuzzy Knowledge Rule Bases. *IEEE Transactions on Fuzzy Systems* 5(3):398-417.
- Fishman, G.S. 1996. *Monte Carlo: concepts, algorithms and applications*. Springer.
- Gavrilov, L.A. and N.S. Gavrilova. 2003. The quest for a general theory of aging and longevity. *Science's SAGE KE*. (<http://sageke.sciencemag.org/cgi/>)
- Gegov, A. 2007. *Complexity Management in Fuzzy Systems: a Rule Base Compression Approach*, Springer, Berlin Heidelberg.
- Gems, D. and J.J. McElwee. 2003. Ageing: Microarraying mortality. *Nature* 424: 259-261.
- Guven, M. and K. Passino. 2001. Avoiding exponential parameter growth in fuzzy systems. *IEEE Transactions on Fuzzy Systems* 9(1):194-199.
- Kydland, F. 2006. Quantitative Aggregate Economics. *The American Economic Review* 96(5):1373-1383.
- Lynch, S.M. 2006. Explaining life course and cohort variation in the relationship between education and health: the role of income. *Journal of Health and Social Behavior* 47(4):324-338
- LeClere, F.B. and M.J. Soobader. 2000. The effects of income inequality on the health of selected US demographic groups. *American Journal of Public Health* 90 (12):1892-7.
- Murray, C.J.L. and A. López (eds.), 1996. *The Global Burden of Disease: A comprehensive assessment of mortality from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge, Massachusetts: Harvard University Press.
- Olshansky, S.J. and B.A. Carnes. 1997. Ever since Gompertz. *Demography*, 34(1):1-15
- Pal, N., V. Eluri and G. Mandal. 2002. Fuzzy Logic Approaches to Structure Preserving Dimensionality Reduction. *IEEE Transactions on Fuzzy Systems* 10(3):277-286.
- Setnes, M., R. Babuska and H. Verbruggen. 1998. Rule-Base Modeling: Precision and Transparency. *IEEE Transactions on Systems, Man and Cybernetics* 28(1):165-169.
- Stern, S. 1997. Simulation-Based Estimation. *Journal of Economic Literature* 25:2006-2039.
- Wang, F., H. Shang, L. Wang and Y. Sun. 2005. How to Determine the Minimum Number of Fuzzy Rules to Achieve a Given Accuracy: a Computational Geometric Approach to SISO Case. *Fuzzy Sets and Systems* 150:199-209.
- Wood, Ch.H. and P.A. Lovell, Indirect measures of child mortality: Overview and application to Brazil, 1980. *Social Indicator Research* 23(3):247-267.
- Xiong, N. and L. Litz. 2002. Reduction of Fuzzy Control Rules by Means of Premise Learning: Method and Case Study. *Fuzzy Sets and Systems* 132:217-231.

AUTHOR BIOGRAPHIES

CARLOS R. GARCÍA-ALONSO is a professor in the Department of Quantitative Methods at the University of Córdoba, Spain. He received a degree in engineering and a PhD in Agricultural Engineering from the University Politécnica of Madrid, Spain. His current research and teaching interests are in operational research (simulation and Data Envelopment Analysis), artificial intelligence (fuzzy logic and evolutionary algorithms) and information systems in order to design simulation-artificial intelligence-based models in economics. He has published four papers in refereed journals as well as many book chapters. He is a member of EURO and INFORMS.

GABRIEL M. PÉREZ-ALCALÁ is a professor in the Department of Economy at the University of Córdoba, Spain. He received a degree in economics from the Comillas University of Madrid, Spain, and his PhD in Economics from the University of Córdoba, Spain. His current research and teaching interests are in economy, family income distribution and labor market evolution.