Adaptive Neuro Fuzzy Inference System Used to Build Models with Uncertain Data: Study Case for Rainfed Maize in the State of Puebla (Mexico)

Anäis Vermonden, Carlos Gay-García and Iván Paz-Ortiz

Abstract A model was built using Adaptive Neuro Fuzzy Inference System (ANFIS) to determine the relationship between the natural suitability index of rainfed maize and yield per hectare and percentage of production area lost for the state of Puebla. The data used to build the model presented inconsistencies. The data of the INEGIs land use map presented more municipalities without rainfed maize agriculture than the database of SAGARPA. Also the SAGARPA data, in terms of the percentage of production area lost, do not mark any distinctions of the loss. Even with data inconsistencies ANFIS produced a coherent output reviewed by experts and local studies. The model shows that higher the percentage of production area lost and high yields, the higher the suitability index is. According to local studies this is due to the high degradation of the soils and confirmed with the second model built adding soil degradation information.

Keywords Fuzzy logic · Agriculture · Adaptive neuro fuzzy inference system · ANFIS · Uncertain data · Mexico

1 Introduction

The municipality of Tehuacan in the state of Puebla, is considered to be one of the places where maize originated, since archaeologists have found some of the oldest maize fossil dating 5,500 y.b.p. [9]. This case of study for the state of Puebla was

A. Vermonden (✉) · C. Gay-García · I. Paz-Ortiz
Programa de Investigación En Cambio Climático, Universidad Nacional Autónoma de México, Edificio de Programas Universitarios, Circuito Exterior CU, 04510 Coyoacan, Mexico
e-mail: anaisvt@gmail.com

C. Gay-García
e-mail: cgay@servidor.unam.mx

I. Paz-Ortiz
e-mail: ivnpaz@gmail.com

M.S. Obaidat et al. (eds.), *Simulation and Modeling Methodologies, Technologies and Applications*, Advances in Intelligent Systems and Computing 319,
DOI 10.1007/978-3-319-11457-6_10

developed to understand the relationship between the natural suitability index for rain-fed maize, the yield per hectare and the percentage of the area lost. Today 62 % of the cultivated land in the state of Puebla is destined to grow maize, 70 % is rainfed, and up to 80 % are varieties of native maize [15].

The natural suitability index for rainfed maize was calculated using the mean temperature, mean precipitation, soil depth and slope. According to the experts, for an area to be suitable the four variables must be suitable, even if one of the variables is in another class then the index will belong to that class. The index of suitability for rainfed maize was calculated with a fuzzy model based on expert knowledge [14] and on the previous work of Monterroso [11]. The Index had a resolution of 1 km by 1 km. The natural suitability index helps determine which lands are best to practice rainfed agriculture. These can also be projected under different climate change scenarios to determine areas that will be more vulnerable as the suitability index changes. How is the natural suitability index related to other variables, and could these also be projected into a future made more uncertain due to climate change?

At state level the information was provided by two sources, the Secretaría de Agricultura, Ganadería, Desarrollo Social, Pesca y Alimentación (SAGARPA) and the Instituto Nacional de Estadística y Geografía (INEGI). The information from both government bodies came at different scales: SAGARPAs is at municipality level; INEGIs was obtained at 1 km × 1 km, making it coincide with the calculated natural suitability index for rainfed maize.

When examining the data from INEGI and SAGARPA, inconsistencies started to show. According to the data from SAGARPA [12] only three municipalities in the state of Puebla have no rainfed maize agriculture, Altepexi, Atzala and Zinacatepec, while the map of land use from INEGI [6] shows 16 other municipalities laking rainfed agriculture, see Fig. 2. Due to the scale these small areas do not show on the map.

The data record available from SAGARPA at municipality level goes from 2003 to 2012. In previous years the municipalities were grouped in Rural Development Districts, but how they are grouped is not published.

In the study developed with only two variables, soon it became apparent a third variable was needed to explain the results, as higher suitability index had higher yields and higher percentage in area lost. Further research claimed soil degradation was a big factor in loss of rainfed maize production in the state. As it can be observed in Fig. 7, in the state of Puebla nearly 60 % of the land used for rainfed agriculture shows soil degradation.

2 Method

Adaptive Neuro Fuzzy Inference System (ANFIS), represents a Sugeno-type neuro-fuzzy system. By integrating both neural networks and fuzzy logic principles, it has the potential to capture the benefits of both in a single framework. It has the ability to construct sets of fuzzy if-then rules to approximate nonlinear functions. ANFIS

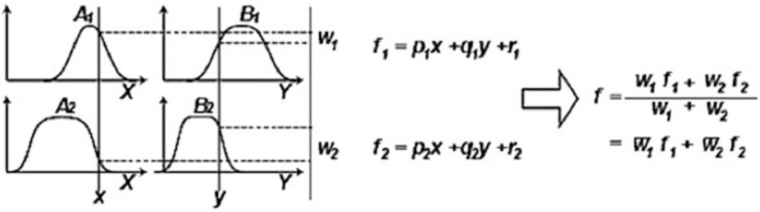

Fig. 1 Diagram of a Sugeno model (evaluation of two fuzzy rules with two input variables, i.e. *A* and *B*)

can also build appropriate membership functions to generate the stipulated input-output pairs to be used in the model [8]. The Neuro-adaptive learning techniques provide a method for building a fuzzy model from the information contained in a data set. The fuzzy system enables flexibility in the variables and the representation of incomplete data, as membership to a fuzzy set is denoted by the degree of membership to the set. Since the ANFIS can deduce relations between the inputs/outputs, ANFIS forms an input output mapping based both on human knowledge (based on fuzzy if-then rules) and generated input/output data pairs by using a hybrid algorithm that is the combination of the gradient descent and least square estimates [7]. The main characteristic of the Sugeno inference system is that the consequent, or output of the fuzzy rules is a function, as shown in Eq. 1.

$$\begin{aligned} &R1 : \mathit{If}\ A\ \mathit{is}\ A1\ \mathit{and}\ B\ \mathit{is}\ B1\ \mathit{the} f1 = p1 * a + q1 * b + r1 \\ &R2 : \mathit{If}\ A\ \mathit{is}\ A2\ \mathit{and}\ B\ \mathit{is}\ B2\ \mathit{the} f2 = p2 * a + q2 * b + r2 \end{aligned} \quad (1)$$

Figure 1 graphically describes the inference process of a Sugeno model composed by the two rules described in Eq. 1.

The first step combines a given input tuple (Fig. 1), x and y, through antecedent rules by determining the degree to which each input belongs to the corresponding fuzzy set. The min operator is used to obtain the weight of each rule, which is later used in the final output computation, f. Sugeno has two differentiated set of parameters, the first set corresponds to the input variable and the second to the output function of each rule, i.e. p_i, q_i, and r_i. ANFIS uses two optimization algorithms to automatically adjust the two sets of parameters. Back-propagation (gradient descendent) to learn the parameters of the antecedents (membership functions) and least square estimation is used to determine the coefficients of the linear combinations in the rules' consequents.

3 Data

The examination of the data obtained from SAGARPA and the INEGI showed inconsistencies. According to the data from SAGARPA only three municipalities in the state of Puebla have no rainfed maize agriculture, Altepexi, Atzala and

Fig. 2 Map of the state of Puebla showing the municipalities division and in green land use for rainfed agriculture

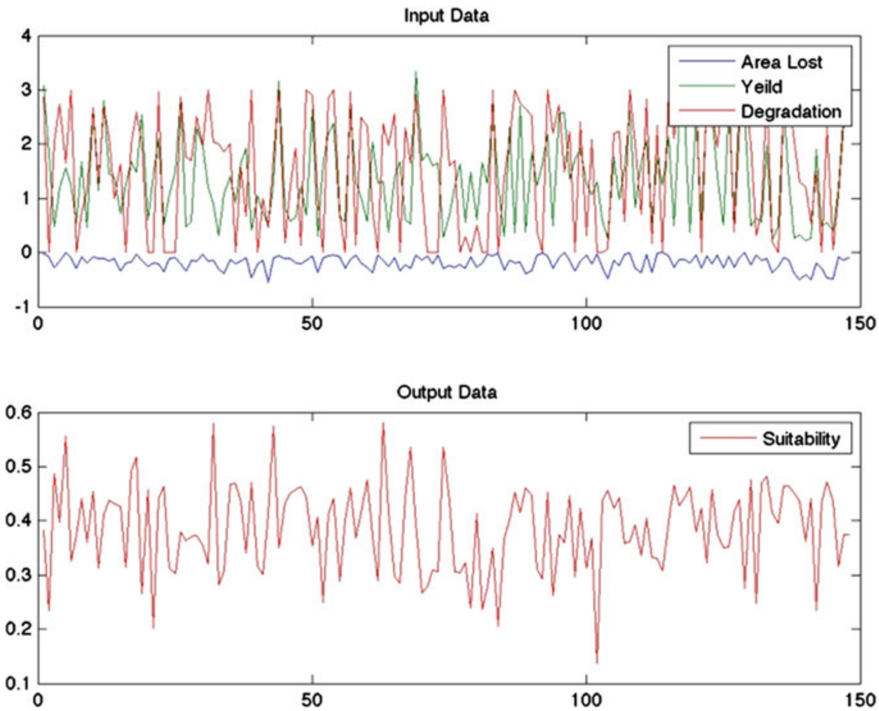

Fig. 3 Input and output data for the first model. Input data on top, percentage of area lost (*blue*) and yield (*green*). Output data on the bottom natural suitability index (*red*)

Zinacatepec, while the map of land use from INEGI [6] show another 16 municipalities without rainfed agriculture, see Fig. 2. Another assumption made for the land use map for rainfed agriculture is that in all the area presented maize is being cultivated. Maize is the most important cereal in the Mexican diet and 92 % of the farmers in Mexico that own between 0 and 5 ha produce 56.4 % of the countrys maize by the rainfed farming practice. In the state of Puebla 60 % of the land used for rainfed agriculture is used to cultivate maize; but how much of the land used for this crop varies between areas, in some of them, for example the municipality of Cohetzala, it is as high as 95 %. Therefore the assumption is: the areas presented as rainfed agriculture are rainfed maize agriculture.

Since the data from SAGARPA is at the municipality level for the period of 2003–2012, the map of land uses published by INEGI was used as a mask to extract the data of the suitability index, since it would correspond to the area marked where rainfed maize agriculture was being produced. An average was calculated to obtain the natural suitability index for rainfed maize to an equivalent scale of the following two variables. Figure 3 shows the values of percentage of production area lost and yield per hectare as well as the natural suitability index for rainfed maize. The land degradation map was obtained from INEGI, published in 2002. The mask from the land use map was once again used to calculate an average to obtain data at the same scale. In Fig. 6 a graphic representation of the values for the input data in addition the level of soil degradation is shown.

4 Results

To determine the relationship between the three variables, a subtractive clustering algorithm [1] was used to generate a fuzzy system. This algorithm allows to estimate the number of clusters and their centers, to later build the membership functions and the relations between the variables. First, the centers are established through subtractive clustering methods [2], once the centers are calculated their radius of influence are determined. For each data of the set a potential measure is calculated to check the center of the cluster using the density of surrounding data. This is done to identify natural groupings of data from a large set, allowing concise representation of embedded relationships. In this case four clusters were calculated, thus reducing the complexity of the sets and the analysis of the relationship between the variables. The clusters calculated were used in the training of the model.

The clusters were used to generate the if-then rules and membership functions. The information was added to the genfis2 function [10] and 75 % of the data set was used to train and generate a fuzzy inference system (FIS) Sugeno type [13]. First, with the information obtained from the subclustering method, it determines the number rules, antecedent membership functions and uses the least square estimation to determine each rules consequent equations; then returning a FIS structure that contains a set of fuzzy rules to cover the featured space.

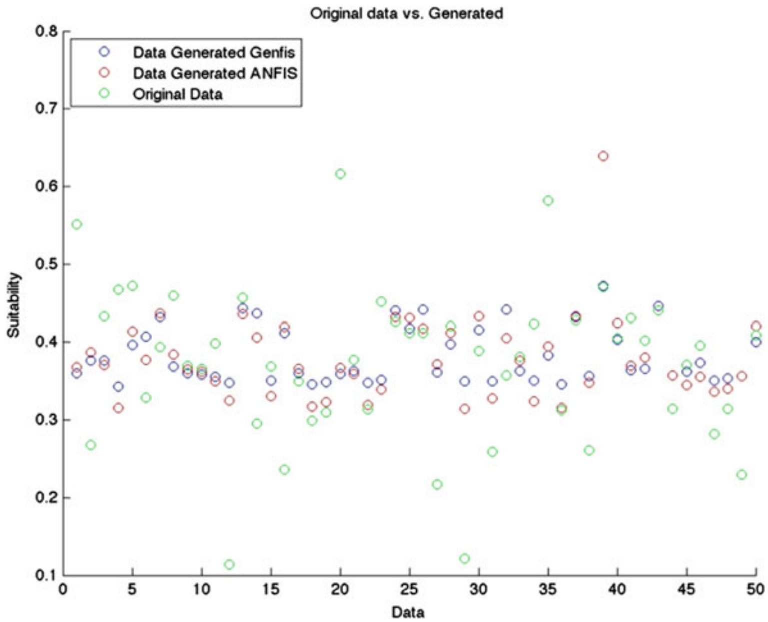

Fig. 4 Graph showing the 25 % of the data used to verify the model, in *green circles* the original data, in *blue circles* the data generated by the first model with genfis2 and in *red circles* the data generated with ANFIS

The remaining 25 % of the data set is then used to verify the model. To verify the model the root mean square error of the system generated by the training data was calculated to be 0.0766. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0953. Both very close to zero.

ANFIS is used to improve the capacity of the FIS to model data. Again 75 % of the data is used to train the neuro-adaptive network. In this case the hybrid optimization method was used, which combines gradient descent and the least squares method. The gradient descent is used on the premise parameters that define the membership functions; for the consequent parameters that define the coefficients of each output equations the least squares method is used. A hundred Epochs were used and the training error tolerance was set to zero. Stability of the training was achieved before Epoch 30. To verify the model the root mean square error of the system generated by the training data was calculated to be 0.0745. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0937, improving the previous FIS generated by genfis2. See Fig. 4.

The fuzzy surface of the rules generated with the data (Fig. 5) show that the areas with the highest suitability index have the highest percentage of production area lost (Fig. 6).

A study of the municipality of Molcaxac [3], which has a high suitability index for the period of 2002–2003, informed that 35 % of the total production of the cereal was lost due to soil degradation. A third variable would further explain the

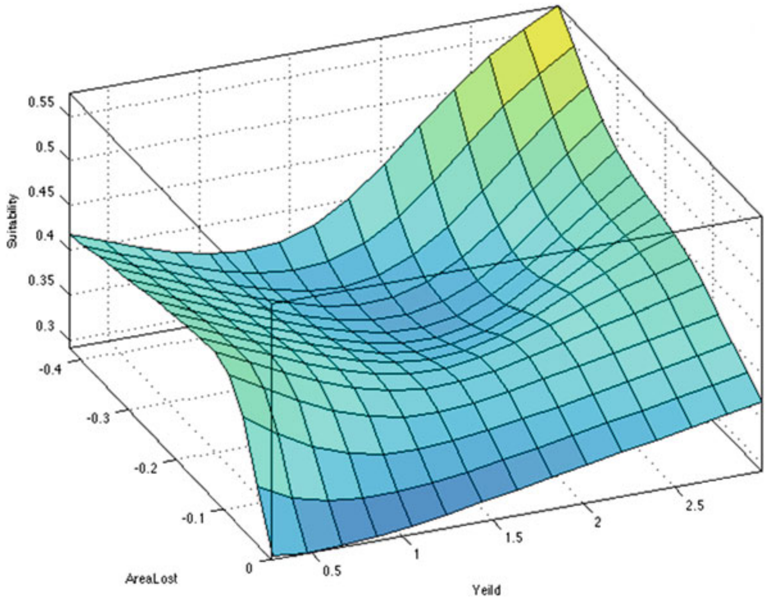

Fig. 5 Graph of the surface created by the rules of the FIS, using two variables: yields and percentage of area lost

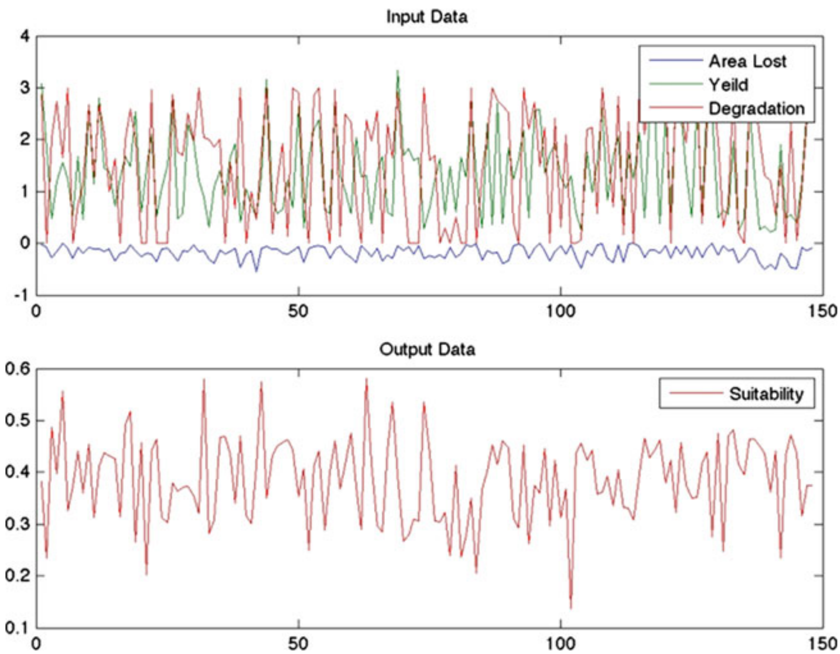

Fig. 6 Input and output data for the second model. Input data on top, percentage of area lost (*blue*), yield (*green*) and level of soil degradation (*red*). Output data on the bottom natural suitability index (*red*)

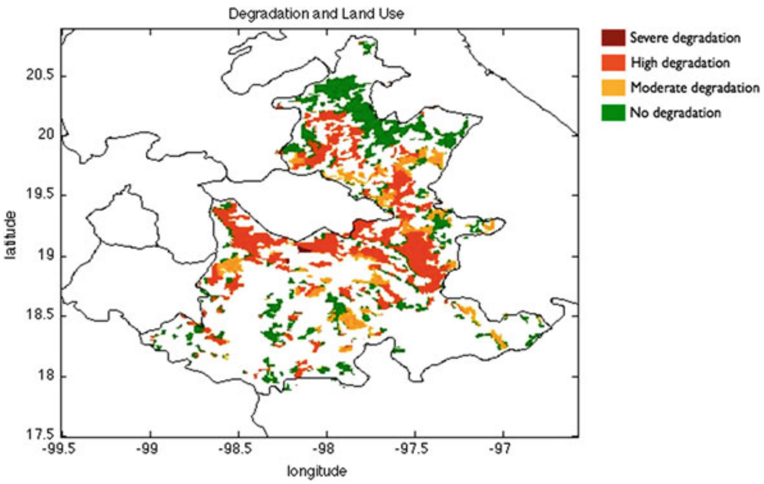

Fig. 7 Map of the state of Puebla showing level of degradation on land use for rainfed agriculture. With data from [5]

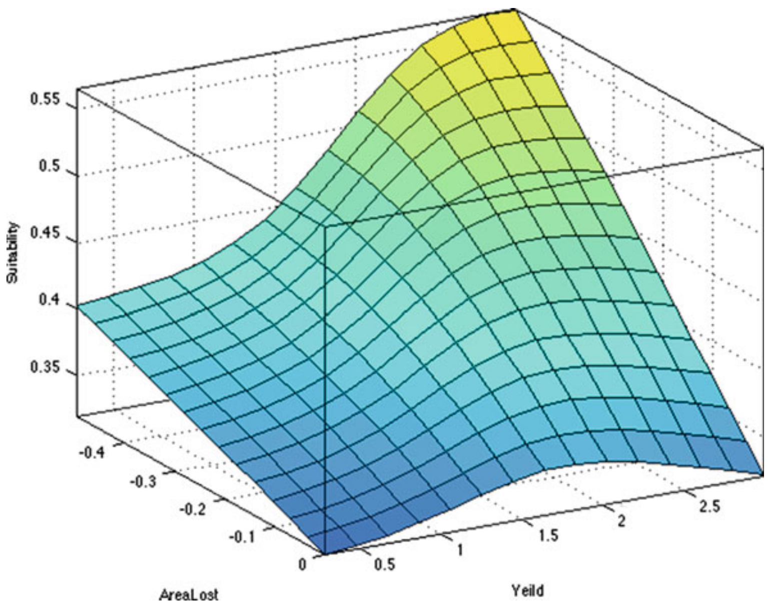

Fig. 8 Graph of the surface generated by the rules using three variables: yields, percentage of area lost and soil degradation

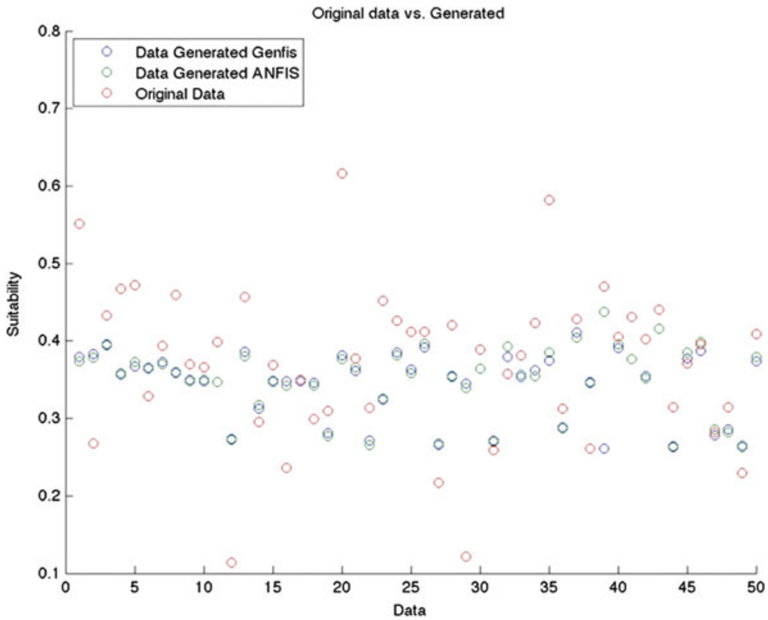

Fig. 9 Graph showing the 25 % of the data used to verify the model with the three variables, in *green circles* the original data, in *blue circles* the data generated by the first model with genfis2 and in *red circles* the data generated with ANFIS

results obtained. Figure 7 shows how 60 % of the land used for rainfed agriculture presents some level of soil degradation. The two major types affecting these areas are chemical degradation and soil erosion. Both are directly linked to poor agricultural practices.

Repeating the procedures described above, but this time adding the soil degradation variable, confirmed that the areas of higher suitability, high yields and higher percentage of area lost were affected by soil degradation, as shown in Fig. 8. The signal of degradation was set to the max, showing the behavior obtained with the model with only two variables.

This model was also improved with the use of ANFIS since the root mean square error of the system generated by the training data was calculated to be 0.0757, and became 0.0752. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0868, it improved slightly to 0.0814. The improvements can be seen on the graph in Fig. 9.

This model shows the importance of tackling soil degradation as a measure of adaptation to climate change, since today the most suitable areas are losing their productive capacity.

5 Conclusions

The state of Puebla is known for the origin of cultivated maize. The methodology used was the subtractive clustering analysis and ANFIS, to establish the relationships between the suitability index for rainfed maize and the other variables. This preliminary model shows that where suitability is higher the area lost is higher, soil degradation being one of the reasons. This has a very logic explanation as the most suitable areas have been used over the years to produce rainfed maize under poor agricultural practices. This can be verified, one of the major degradation types found over these lands is chemical degradation, which reduces or eliminates the biological productivity of the soil. The methodology used allows users to build models even when the data presents uncertainties. These models can always be improved with better information, expert knowledge and field validation. This first approach presents an important advance for case studies in regions were there is little information and with obvious restrictions. In this case the first restriction was the scale at which the two sources presented the data. It was important to use both to obtain an average of the suitability index at municipality level. If the land use mask had not been used the suitability index would have become insignificant. The other source of uncertainty was to assume that all the areas marked on the land use map for rainfed agriculture were maize. But the municipalities marked on the land use map had rainfed maize production according to the data obtained from SAGARPA, as many of the subsistence farmers in Mexico will grow maize in association with other crops, mainly beans and zuchinis.

SAGARPA data had other restrictions. The earliest data published at municipality level is from the year 2003. Older data are presented as Rural Development Districts, which groups together municipalities, but how they are grouped is unknown as this information is not published. The data used for this study is from 2003 to 2012. Also, the lost production area has no distinctions on why this happened; it could be climate, pests, even that farmers did not plant the area they had first declared or did not harvest all the area planted (which occurs when maize prices fall and no longer compensate the harvesting cost, this way they can claim insurance).

A case study at the municipality of Cohetzala, in the south west part of the state of Puebla [4], showed 98.5 % of the 1,308 ha were destined to grow maize and 95 % is under rainfed method. This municipality has a high suitability index and low degradation. This study found that most of the farmers in this municipality used traditional practices to grow maize: native seeds, association and crop rotation, water and soil conservation techniques, and use of manure as fertilizer (as they are more cost effective), instead of the recommendations made by the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) a subdivision of SAGARPA that promotes the use of fertilizers, pesticides, herbicides and hybrid seeds. The yields per hectare are low, but Huato et al. argue that what marked the biggest difference in yield was the use of irrigation (one of the reasons can be that the area has regosols soils), and the comparison between the farmers using traditional practices to the ones using INIFAP recommendations indicate that the differences are insignificant. This kind of information could be added to the model and create distinctions between traditional practices and more modern practices to determine how it affects the yields, degradation and the percentage of area lost.

These models show that agriculture is a complex human system, but this first approach is showing there is an alarming trend in the state of Puebla due to soil degradation. Degradation must be a a priority in the adaptation policies to climate change, since, in the present day scenario, it causes very high losses.

This kind of models facilitate decision making processes even in situations with uncertain data.

Acknowledgments The present work was developed with the support of the Programa de Investigación en Cambio Climático (PINCC) of the Universidad Nacional Autónoma de México (UNAM) and the Consejo Nacional de Ciencia y Tecnología (Conacyt). We would like to thank Dr. Cecilia Conde, Dr. Alejandro Monterroso and M. Guillermo Rosales for their valuable inputs and serving as the experts to validate the model.

References

1. Chiu SL (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2 (3):267–278
2. Dubois D, Prade H (1980) Fuzzy sets and systems: theory and applications. Academic Press, New York
3. Gaspar Angeles E, Ortiz Torres E et al (2010) Caracterización y rendimiento de poblaciones de maíz nativas de Molcaxac, Puebla. Revista fitotecnica mexicana 33(4)
4. Huato MAD, Cruz Leon A et al (2012) Management of corn at Cohetzala, Puebla, Mexico: between the local and the global. Estudios Sociales XX(40)
5. INEGI (2002) Degradación de suelos. Degradación del medio ambiente, SEMARNAT and Colegio de Postgraduados
6. INEGI (2005) Uso del Suelo y Vegetación
7. Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685
8. Jang J-SR, Sun C-T (1997) Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice Hall, Upper Saddle River
9. Long A, Benz BF et al (1989) First direct AMS dates on early maize from Tehuacan, Mexico. Radiocarbon 31(3):1035–1040
10. MATLAB (2009) Fuzzy logic toolbox. I. The MathWorks
11. Monterroso AI, Conde C et al (2009) Assessing current and potential rainfed maize suitability under climate change scenarios in Mexico. Atmósfera 24(1):53–57
12. SAGARPA, SIAP (2003–2012). Producción anual de maíz del estado de Puebla
13. Sugeno M (1977) Fuzzy measures and fuzzy integrals: a survey. North-Holland, NY
14. Vermonden A (2012) Modelo Difuso para la Evaluacón de la Aptitud Actual y Potencial del Maíz de Temporal en México con Cambio Climático. Mexico, Universidad Nacional Autonoma de Mexico. Mastería, Posgrado de Ciencias de la Tierra
15. Viveros Flores CE (2010) Estudio de la Dinámica de Aprovechamiento del Maíz en las Unidades de Producción Familiar en el Valle de Puebla, México. Postgrado de estrategias para el desarrollo agrícola regional. Puebla, Colegio de Postgraduados. PhD, p 119