SUPPORTING PAPERS
ISBN 979-97243-9-2

USING FUZZY LOGIC TO SIMULATE THE COMPOSITION OF FARMING SYSTEMS IN THE MEKONG DELTA

Bosma R.H.[1], Le Thanh Phong[2], J van de Berg[3], U. Kaymak[3], H.M.J. Udo[1], J.A.J. Verreth[1].

[1] Animal Science Group, Wageningen University, Netherlands; [2] Can Tho University, Viet Nam; [3] Computer Science Department, Erasmus University, Rotterdam, Netherlands.

ABSTRACT

Though specialisation is the global trend in agriculture, integrated agriculture aquaculture farming systems expanded in Vietnam during the past 20 years. We interviewed 24 farmers of three categories of well-being in three hamlets in the Mekong Delta and analysed their motivations for the diversification. Subsequently we conceived a three level hierarchical tree of fuzzy logic inference systems to simulate the farm composition. The model simulated with 80 to 98% of accuracy the distribution of the 'hard to change' farm components: rice field, orchard, and fishpond. The presence of the 'easy to change' livestock components was less well simulated: an accuracy of 30 to 70 %, either because of their lower frequency or because these components can be stopped or started from one day to the other.

Key words: Vietnam, Mixed-Farming, Decision-Making, Simulation, Fuzzy-Sets

INTRODUCTION

Though specialisation is the global trend in agriculture, integrated agriculture aquaculture farming systems (IAAS) spread over Vietnam during the past 20 years. IAAS attracted our interest because mixed farming is thought to be more sustainable and its' expansion on a global scale desirable. Using a.o. the capital assets framework we showed that integration is beneficial but that farmers motives to increase income from agriculture are divers (Bosma *et al.*, 2006b). However, most simulation models consider that farmers are driven by economic utility maximisation only. Wanting to elucidate farmers motives further, we opted for fuzzy logic to include qualitative data in the simulation model (Zadeh, 1978 in Jang *et al.*, 1997).

METHODOLOGY

Fuzzy systems use dual or multi-valued logic, the way humans usually argue, to control machines or to model decision-making. The mathematical principles of fuzzy logic are demonstrated in figure 1. The building of our fuzzy is reconstructed in nine steps: data collection; input and output variables; description of the architecture for the FISs (Fuzzy Inference Systems); linguistic sets and membership functions (MF) of the

ISBN 979-97243-9-2

variables; composition of the FISs' rule bases; programming; calibration/fine-tuning; and validation.

Data collection

To assess farmers' motives and drivers to practise a particular component we interviewed 72 farmers in three hamlets from the fresh water alluvial zone of the MD through semi-open interviews. In each hamlet 24 farmers were selected through stratified random sampling based on wealth rankings of poor, intermediate and well-off households (Bosma *et al*, 2006b). The interviews started with mapping and characterising the farms' physical resources, and subsequently data were collected on the family composition, the present farming components, and the financial results of the past year. The open part of the interview insisted on past changes in farm composition, the causes, reasons or conditions under which farmers implement a change or innovation, and if applicable the farmers' motives for not applying other components. We classified children contributing to farm activities as youngsters (10–18 years) and their grandparents still working on the farm as elders. For the calculation of labour availability, elders not participating in work and young children were both classified as non-working. We asked the farmers to rank on a scale of 1 to 5 their preference of having their own rice-field for food security and their know-how on the various farming activities.

Input variables of the fuzzy inference systems

During the semi-open interviews the farmers repeatedly mentioned the following drives for innovation: improving income and diversifying the diet, both mainly for the well-being of children. The analysis with the livelihood capital asset framework showed that other factors guiding their decisions were: the availability of labour, water, land, capital, market-price and know-how (Bosma *et al*, 2006b).

The household labour availability was derived from the weighted number of family members in the age categories: *adult* –0.25 × *non-working* + 0.5 × *youngster* + 0.75 × *elder*. For the MD nine sources of water were distinguished: river, primary and secondary canal, natural source, seasonal river, rainwater reservoir, permanent well, deep-well or bore-hole, and shallow well (ranked for diminishing availability). The soils were classified in a Land Quality Index (LQI) of 10 categories (Bosma *et al*, 2006a). The land most Vietnamese farms use is dispersed and each plot has its' own characteristics relating to e.g. soil quality and water availability, and thus supports different types of activities. We considered this variation by using 3 categories of land: homestead, irrigated and upland. The orchards constructed as ditch-dike systems were included in the homestead. If a plot of upland was situated next to the homestead the two were considered to constitute one plot of homestead. Land that flooded seasonally was classified as irrigated; flood level and length were collected individually. Rain period and rain level were derived from national statistics and applied uniformly to all cases.

The collateral value of land with a red certificate, attributing owner rights, was twice as high than for land with a green user certificate, attributing user rights and conferring obligations (Bosma *et al*, 2006a). Six categories of increasing risk behaviour

ISBN 979-97243-9-2

were used, based on the source of credit people used and the activity it was used for: none, relatives, bank, input providers, private money lenders and high risk credit.

Tabel 1 *Product' dummy prices* applied (x1000VND, per kg or head for livestock (except pig)*

	Rice	Fruit	Fish	Veg.	Duck	Chicken	Egg	Pig	Piglet	Lrum	Goat
year 2003	2.1	4	8	6	15	18	0.8	10000	800	2000	200

Crop= other crop than rice; Veg.= vegetables; Lrum = Large ruminants (cattle and buffalo).

The category of well-being applied as factor for the availability of capital for investments and for the capacity to hire labour.

The market price of the products applied in the model was a dummy equal for all farmers (Table 1). We applied the average of the farm gate prices per kg of product category. Though those do not reflect the net margin of the component, farmers were aware of the price level that resulted in a break-even or a profit, or which price level caused financial losses.

Output variables of the model

The output variables considered for the model were the most frequent farm components: rice-field, fruit-orchard, fish-pond, vegetable-garden, upland-crops, large ruminants (buffalo and cattle), goats, pigs, chickens and ducks. The garden component was included because of its' growing interest and to be able to use the model on other datasets. However, the estimations for the number of farmers having a vegetable garden could not be calibrated nor validated because their benefits had been included in those of the irrigated land and thus mixed with rice-field in the delta. In the available dataset the financial outputs for goats and large ruminants were cumulated due to their low frequency; the model included separate estimations of both for future use.

Architecture and FISs

The knowledge acquisition identified 45 variables affecting farmers' decision making (Figure 2). We structured those in a hierarchical decision-making tree. Hereafter we describe each subset of the hierarchical structure in four sections: the primary production factors, the product opportunities, the product options and the FIS of the final output layer.

The FIS of the homestead contained four input variables: its' area, its' soil quality, the length of the rainy season(s) and the level of rainfall; the linguistic value increasing with larger, better, longer, and higher values of the respective variables. Next to the variables applied for the homestead, the upland FIS contained also the distance to the homestead; the longer the distance the lower the linguistic value. The FIS of the irrigated land also five variables as for the upland FIS, but the two factors related to the rain were replaced by the length of the flooding and the flood level, both restricting the period the land could be used.

The water availability depended on five variables: the length and level of both the rain season and the flooding, and on the source of the water. The capital availability depended on three variables: the collateral value of the owned land, the rank of risk

SUPPORTING PAPERS
ISBN 979-97243-9-2

behaviour, and the rank of well-being. The availability of labour depended on two variables: the household labour and the capacity to hire labour, which was determined by the rank of well-being.

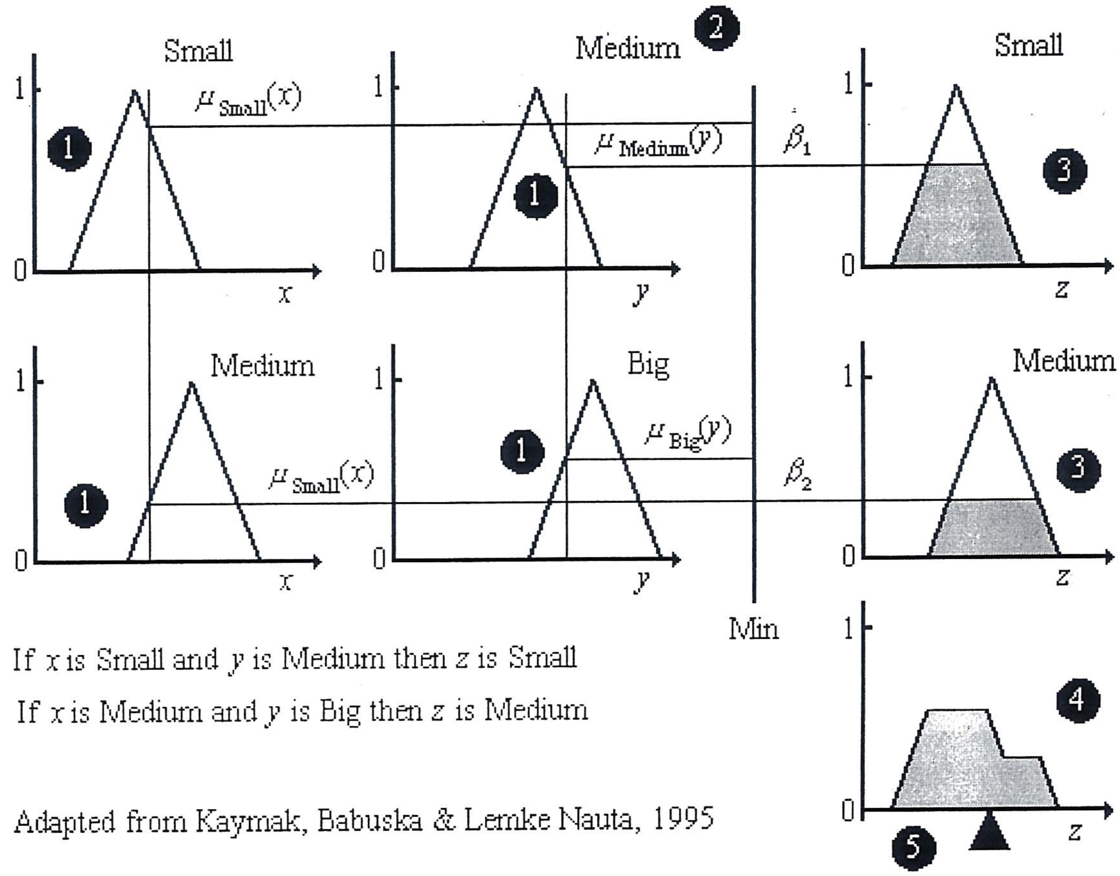

Figure 1. The reasoning engine of the fuzzy inference for 2 rules (from (Bosma et al., 2006a).

For most activities, the FISs for the opportunities to practice a product included its' general market price, the individual farmers' know-how on the specific product and the average distance between the farm and the product market. The FISs for the opportunity to raise pigs, ducks and chickens related to two types of product and the know-how and prices were represented by both specialisations: fattening and reproduction (eggs). A high price for eggs was always a positive incentive for raising both ducks or chickens. A price of piglets could either be positive if the farmer's know-how on breeding was high or negative if this know-how was low.

The farmer may decide to do an activity if the value of (one of) the three land types, the availability of water, capital and labour are no constraint and if the product opportunity is acceptable or good. Those seven variables were included in the second layer of FISs that represent the farmers' option to raise ducks, chickens, pigs, goats,

ISBN 979-97243-9-2

cattle, or fish, or to have an orchard or a garden. Though for efficiency reasons all those variables with their MFs were included in each FIS they were not all applied in the rule-base: e.g. in the FIS 'option to raise ducks' the value of upland did not appear in any of the rules. The FIS 'option to do a rice field' also included the ranking of the importance the farmers gave to rice as a key to food security, scaled from 1 to 5.

Whether a farmer opts for one or several components depends on the relation between the components and on the individual family farm household reference frames. Some details on the relations between the components will be given below in the section dealing with the rule-base. The farmer's reference frame (FRF) was separated in two distinct FIS: one related to the factors decisive for the integration of various components and the other to the motives and drives for the diversification of the farming systems. The FRF for integration of farm components depended on six variables: the distance between the fields and the homestead, the area of homestead and irrigated land, the farmers' education level, and his/her tendency to integration. To represent the farmers' tendency to integration we used the index for the integration of farm components that we assessed for all farms (Bosma *et al*, 2006b). The diversification was driven by the socio-economic context and by individual farmers' motives. The socio-economic context was already assessed through the market prices included in the FISs for product opportunities. Three individual farmers' motives composed the FRF for diversification: the phase in the household life cycle, the age of the household head and the number of children and young in the household (ibid.).

Linguistic sets of the fuzzy systems

The number of linguistic value sets for a variable and the related memberships functions (MF) for each variable were determined by the normality of the data and the need to distinguish intermediate outputs. If the distribution of the input data is skewed more than two values are needed. The more linguistic values one variable has the more complex will be the rule base and therefore preferably only two values were distinguished.

Table 2 Example of a rule-base (farmers' opportunity to make profit from raising cattle) with a 'don't car statement' and the possibility to trim the remaining rules to 7 (see text).

1='if pricattle is low, then proficattle is bad';
2='if market is far and know-how is low and pricattle is acceptable, then proficattle is bad';
3='if market is far and know-how is low and pricattle is high, then proficattle is fine'; 3,7
4='if market is far and know-how is high and pricattle is acceptable, then proficattle is fine';
5='if market is far and know-how is high and pricattle is high, then proficattle is good'; 5,9
6='if market is close and know-how is low and pricattle is acceptable, then proficattle is fine';
7='if market is close and know-how is low and pricattle is high, then proficattle is fine'; 3,7
8='if market is close and know-how is high and pricattle is acceptable, then proficattle is good';
9='if market is close and know-how is high and pricattle is high, then proficattle is good'; 5,9

See the related text for an explanation of numbers; Observe that the applied rule base was different.

Rule definition

SUPPORTING PAPERS
ISBN 979-97243-9-2

The fuzzy 'if-then' rules were composed according to the farmer's statements and motives and to the results of the empirical and statistical analysis of the data (Bosma *et al.* 2006b). Whenever possible we reduced the rule explosion by straightforward definition of a "don't care" statement. In the FIS of Table 2 a 'don't care' is "If *Pricattle* (price of cattle) is low then *Proficattle* (profit from cattle) is bad"; so, if the value for *Pricattle* is low the value of the other 2 variables does not affect the decision anymore, which reduces this rule base dramatically. For the remaining sections of the rule base the rules with the same consequence for all the values of one variable can be trimmed; e.g. the 2 sets of four rules in Table 2 can be trimmed down to 7 by eliminating all but one of the marked rules, while deleting the section of the variable with different values ('market is close/far' Know-how is high/low'). The rule reduction led to a different number of rules for each FIS; e.g. the FIS 'Option to do rice-field' had close to 100 while the one for cattle was trimmed down to 8 rules.

If in the third layer an input with three linguistic values was used, a simple antecedent rule was composed with the high linguistic value and a positive consequence for the product. In most other rules of the third layer one or both of the FRFs needed to be good if a component was to be applied. Next to the FRFs, the rules for the decision to integrate a component included products relations. We give two examples. If raising cattle and goats are both good options, the farmers will do only cattle. Even if the opportunity to raise cattle is high the farmer will only carry out this option if he has fields of which he can collect feed, so if he has also a positive option for rice or an upland crop.

Programming the model

The model was programmed in Matlab 6.1 and the fuzzy logic toolbox of Matlab 6.1, release 13.1 (Anonymous, 2002) was used for the fuzzy logic inference procedures. We used Mamdani fuzzy sets with the minimum operator to determine the degree of fulfilment of the fuzzy rules (Jang *et al*, 1997). After aggregation, the outputs of the first and second layer FISs were defuzzified (decoded) according to the center-of-gravity method. The resulting centroïd value was directly used as an input in the second or third layer FIS, respectively.

Calibration and fine-tuning

The numerical household and farm data, descriptive information, and farmers' ranking provided the inputs for the database. For the calibration of the parameters with a training dataset we randomly sampled 48 cases from the 72 farmers of the dataset. The model was calibrated by comparing the overall number of positive outputs with the observed situation of the training dataset. For the distribution at the aggregated level we referred to two thresholds: the lower was the number of farmers earning cash income from the components and the upper was the number of all farmers practising the component; the difference constituting the households' that consume themselves all produce of the component, or that did not sell a ruminant during the period considered. The fit at the individual level was checked for both thresholds separately.

To obtain optimal fit between the model estimates and the expected tendency in distribution, three steps were needed: adjustment of the rules; shifting the parameters of

SUPPORTING PAPERS
ISBN 979-97243-9-2

the MFs of the basic variables (of the market price mainly); adjusting the parameters of the MFs for the inputs of the second or third layer to the distribution and level of the values of the output from the corresponding variable in the first and second layer respectively. Though initially the MFs and parameters of the linguistic values of the intermediate outputs and the corresponding inputs for the next layer were similar, most of the parameters for the intermediate inputs were adjusted during calibration and manual fine-tuning. Individual fine-tuning was done for those output variables of which the estimated number of farmers practising at an aggregated level did not fit in the range of farmers practising the component for cash and of all practising farmers, e.g. rice fields and ruminants.

Validation and analysis

During fine-tuning we checked the face validity (Law&McComas, 2001) and for result validity we applied the preliminary model to the 24 remaining delta farms. Again we compared the overall number of positive outputs with the observed situation for both thresholds. For further analysis of the simulation model we calculated the individual classification rate (ICR) only for the positive cases in Excel. A case is called positive if the simulation confirmed that the farmer practices the activity.

SUPPORTING PAPERS
ISBN 979-97243-9-2

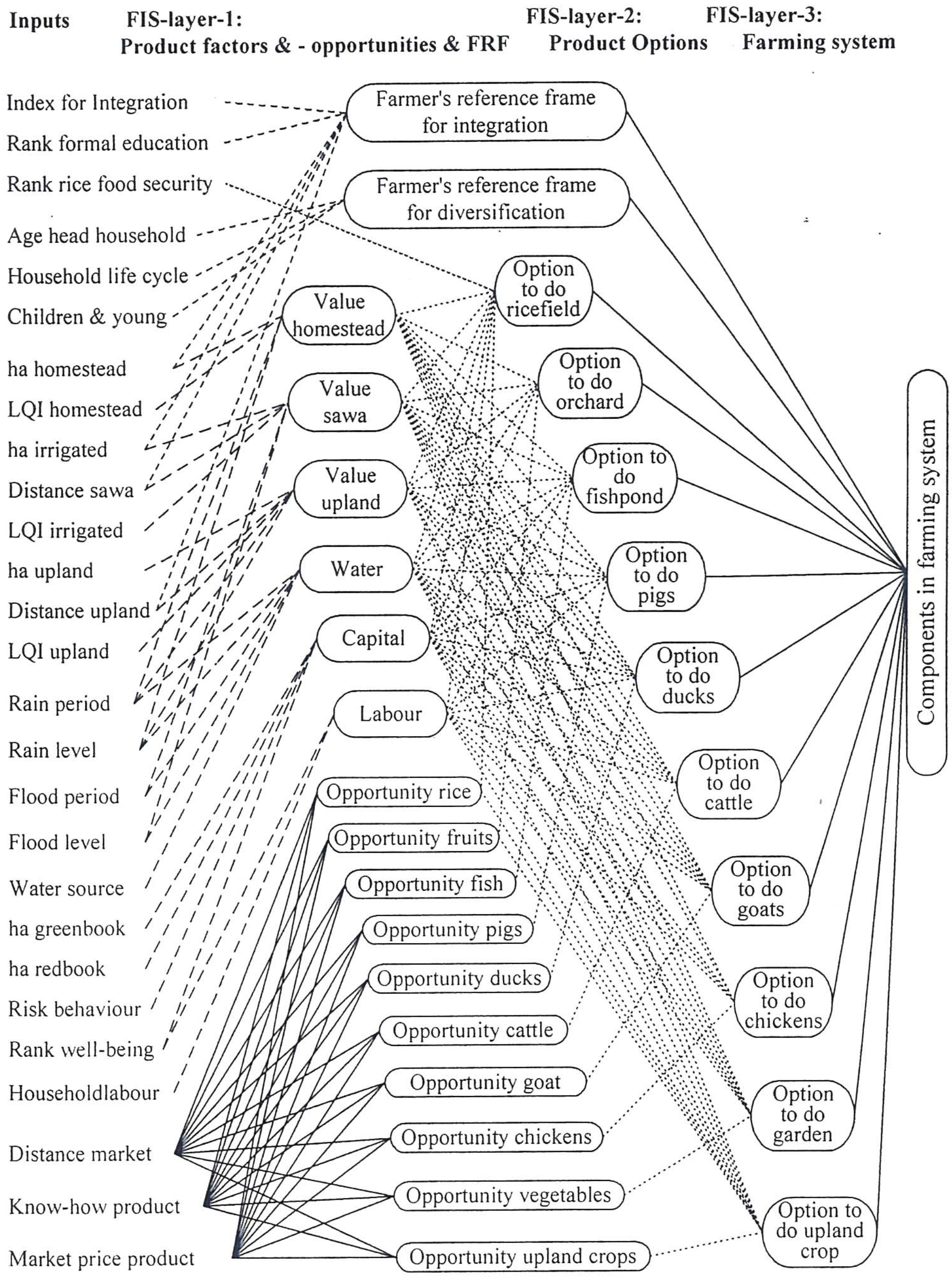

Figure 2. Structure of the hierarchical fuzzy model simulating farmers' decision-making on their farm composition: at the left the inputs for 18 first layer Fuzzy Inference Systems (FIS) generating centroïd values for the 10 FISs of the 2nd layer, that aggregate the option for a component considering the resources and the opportunity to make a profit from each. The centroïd outputs of these last 10 FISs are infered with the 2 FISs for the farmers' reference frame in the third layer FIS at the right hand to simulate the composition of the farming system. (FRF = farmer reference frame; ha = hectare; LQI = land quality index; input variables for the farmers' references frames = short hyphened line, for the production factors = long hyphened line, for opportunity to make a profit =continuous line).

SUPPORTING PAPERS
ISBN 979-97243-9-2

RESULTS

For both the calibration and the validation set the estimated number of farmers practising a component was within the range of all practising farmers and those doing the activity for cash also, the difference being those practising the component for home-consumption only (Figure 3). The number of farmers raising ruminants was overestimated in both the calibration and validation set. The slight overestimation of the number having orchards or raising ducks for cash in the calibration set was compensated by an underestimation in the validation set. In total, the number of farmers estimated to have a rice field was close to the total number, the estimated number of farmers having an orchard, ducks or chickens was close to the real number practising the activity also for cash, while for fish and pigs the number estimated was intermediate between the two extremes.

The individual classification rates were lowest for duck, close to 33%, and highest for rice, above 90% (Table 2). The average individual classification rate for doing rice, fruit or fish for cash was above 80%, while it was close to 50% for the animal components. For some products the individual classification rate was different whether estimating all cases of practising farmers or estimating those doing it for cash also.

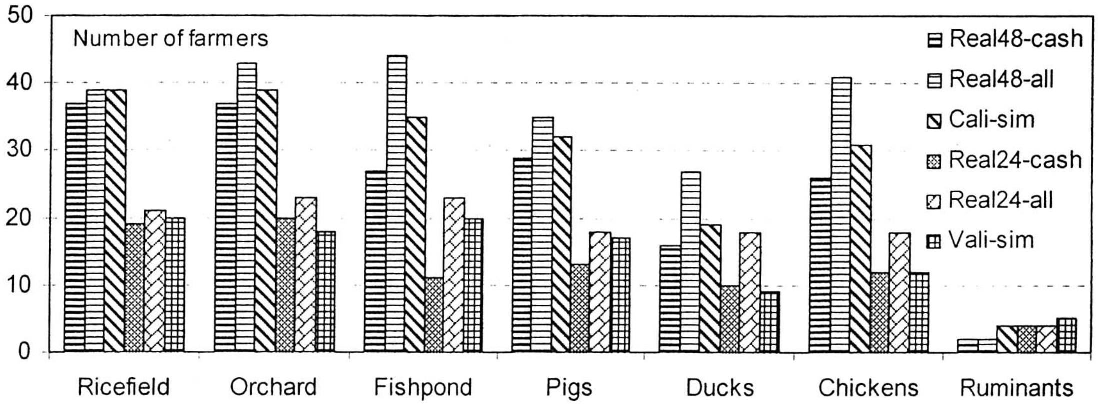

Figure 3: The simulated number of farmers doing the activity (sim) compared to the real numbers of all farmers doing the activity (all) and those making cash income also (cash), for the calibration set (-cal) of 48 farmers, and the validation set (-val) of 24 other farmers

SUPPORTING PAPERS
ISBN 979-97243-9-2

Table 2. The individual classification rate of the positives, i.e. the fit of the individual simulation of the farmers practising a specific activity, either for cash and home consumption or for home consumption.

Dataset		Rice	Fruit	Fish	Pigs	Ducks	Chickens	Ruminants
Training	Cash	95	84	81	72	31	62	50
	All practising	95	84	75	69	33	63	50
Validation	Cash	98	80	91	69	40	58	25
	All practising	90	78	87	67	33	50	25

DISCUSSION AND CONCLUSION

For the land-based activities we obtained an individual classification rate (ICR) between 75 and 98%. These rates are higher than the simulation of the spatial dynamics of land use in the Philippines and Malaysia: 65% to 85% (Verburg *et al.*, 2002). The ICR seems lower for those activities with a smaller number of farmers practising the activity to generate cash: ruminants and ducks in the delta. In general the individual classification rates for the land-based activities (rice, fruit or fish) are higher compared to those for the livestock activities relying less on the availability of private land. The livestock components are relatively easy to start or to stop and have been called 'easy to change' in contrast to the 'hard to change' land-based activities. The rice field, fruit orchard and fish pond are long term investments a farmer can not just stop from on day to the other, but all livestock can be sold from one day on the other. This decision to stop or continue with livestock is related to various factors: household composition, willingness to deal with risk, need for cash, and know-how, which are mostly individual motives and may be subject to rapid change.

The use of only the products' market price, except for piglets, might be considered a weakness of the model. Nevertheless, the farmers' knowledge of the breakeven price in a particular year with a particular technology and a set of input and output prices remains valid. However for a model to have a general applicability it should include changes in the general production context. The inclusion of the piglet price shows the feasibility to apply input prices and the application of farmers' know-how can be considered to stand for the technological level determining the productivity of the component.

Models can become valid decision support tools for planning at the operational level if they 1/ integrate social and natural drivers, 2/ are flexible, i.e. allow to change the system composition and the inter-component relationship, 3/ can address changing issues at farm level as well as a changing context, and 4/ are "tools to think with" rather than "to learn from" (McIntosh *et al.*, 2005). We demonstrated that physical drives and social motives can be integrated in a model using fuzzy logic, and discussed the feasibility to address the changing market context. In the present architecture of the model the system composition is flexible but most inter-component relationships are defined in the rule bases and are not accessible through an interface. E.g. for a model to be interactive an interface should allow to make choices like "use more family labour on-farm". Our efforts show that constructing fuzzy expert models is feasible. This could

ISBN 979-97243-9-2

allow to associate farmers to the modelling and to prepare decision support tools for farmers and extension services. However issues like changing inter-component relationships and changing issues at farm level need to be addressed to make it a "tool to think with" for the end-users.

ACKNOWLEDGEMENTS

We wish to thank the farmers, Can Tho University and the Village People Committees of An Phu, An Thanh, Cai Be, O Mon and Tam Binh for their kind collaboration, and last but not least, Mr. Cao Quoc Nam and Mr. Phan Minh Duc for assisting in the fieldwork and interviewing the farmers. The field-work of this study was funded by the INREF-POND programme from the Interdisciplinary Research and Education Fund (INREF) of Wageningen University. Part of the time spend on modelling and editing the paper was funded by the European Commission through the related POND-LIVE project (ICA4-CT2001-10026).

REFERENCES

Anonymous, 2002. MATLAB, The Language of Technical Computing, MathWorks, 7 ed. The Mathworks Inc., Natick, USA.

Bosma, R.H., L.T. Phong, U. Kaymak, J.v.d. Berg, H.M. Udo, M.E.F.v. Mensvoort & L.Q. Tri, 2006a. Assessing and modelling farmers' decision-making on integrating aquaculture into agriculture in the Mekong Delta Farmer. *Netherlands Journal of Agricultural Sciences* 53: 281-300.

Bosma, R.H., C.Q. Nam, H.M. Udo, J.A. Verreth & L. Visser, 2006b. Farmers' Motives for Agriculture Diversification and its' Contribution to Livelihoods in the Mekong Delta. *Asian Journal of Agricultural Development,* accepted.

Jang, J.-S.R., C. Sun & E. Mizutani, 1997. Neuro-Fuzzy and Soft Computing. A computional Approach to Learning and Machine Intelligence. Prentice Hal, 614 pp.

Law, A.M. & M.G. McComas, 2001. How to build valid and credible simulation models. In: B.A. Peters, *et al.* (Eds.), 2001 Winter Simulation Conference. IEEE, pp. 22-29.

McIntosh, B.S., P. Jeffrey, M. Lemon & N. Winder, 2005. On the Design of Computer-Based Models for Integrated Environmental Science. *Environmental Management* 35: 741-752.

Verburg, P.H., W. Soepboer, A. Veldkamp, R. Limpiada, V. Espaldon & S.A.A. Mastura, 2002. Modellling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management* 30: 391-405.