Official URL
DOI : https://doi.org/10.1007/978-3-030-22270-3_5

PASHAMAMA: An Agricultural Process-Driven Agent-Based Model of the Ecuadorian Amazon

Doryan Kaced[1(✉)], Romain Mejean[2(✉)], Aurélien Richa[3], Benoit Gaudou[1,4,5], and Mehdi Saqalli[2]

[1] UMR 5505 IRIT, CNRS, University Toulouse 1 Capitole, Toulouse, France
doryan.kaced@ut-capitole.fr
[2] UMR 5602 GEODE, CNRS, University Toulouse 2 Jean Jaurès, Toulouse, France
romain.mejean@univ-tlse2.fr
[3] UMR 5505 IRIT, CNRS, University Paul Sabatier, Toulouse, France
[4] Sorbonne University, IRD, UMMISCO, 93143 Bondy, France
[5] USTH - ICTLab, Hanoi, Vietnam

Abstract. This article presents the PASHAMAMA model that aims at studying the situation in the northern part of the Amazonian region of Ecuador in which the intensive oil extraction has induced a high rise of population, pollution, agricultural work and deforestation. It simulates these dynamics impacts on both environment and population by examining exposure and demography over time thanks to a retro-prospective and spatially explicit agent-based approach. Based on a previous work that has introduced roads, immigration and pollution (induced by the oil industry) dynamics, we focus here on the agricultural and the oil salaried work sides of the model. Unlike many models that are highly focused on the use of quantitative data, we choose a process-based approach and rest on qualitative data extracted from interviews with the local population: farmers are not represented by highly cognitive agents, but only attempt to fulfill their local objectives by fulfilling sequentially their constraints (*e.g.* eating before earning money). We also introduce a new evaluation method based on satellite pictures that compares simulated to "real" data on a thematic division of the environment.

Keywords: Agent-based model · Socio-ecological systems · Colonization · Ecuadorian Amazon · Deforestation

1 Introduction

The Northern Ecuadorian Amazon, the region of Ecuador called "Oriente", carries stigmas of a spontaneous agricultural colonization. Encouraged by the State, since the 1970s and 1980s, and then reinforced and facilitated by oil prospecting and exploitation, which opened roads for settlement [3], the region has been the target of a huge migration of inhabitants; this process that can be referred as an agricultural colonization has profoundly altered the landscape.

P. Davidsson and H. Verhagen (Eds.): MABS 2018, LNAI 11463, pp. 59–74, 2019.
https://doi.org/10.1007/978-3-030-22270-3_5

This work takes place in the MONOIL project that aims at developing "a prospective of future dynamics combining contamination exposure, demographics, production activities, with oil but also agriculture, and public policies and their impacts altogether". For this goal, we have developed a spatially explicit agent-based model, named PASHAMAMA [4], integrating the oil leaks and spreading in the environment, and the colonization by families, their settlements and their activities. It is developed on three parishes of the Oriente (Dayuma, Pacayacu and Joya de Los Sachas), but we limit the presentation to Dayuma in this article due to space limitations. This article is focused on the presentation of the socio-economic and demographic part of the model: it aims at tackling the question of the interaction and co-evolution of the colonization and demography with the agri-cultural submodel.

The challenges to develop such a model are multiple and this article provides contributions to face two of them. The first one lies on the shortage of data available on the area and more specifically of data describing human activities: most of the data are available at the global scale. This makes the initialization, dynamics and evaluation on a spatial-explicit model much harder, but it is a common issue to face when building such a large-scale model. The evaluation of such a model in particular is extremely complicated given this lack of data at the proper scale. In particular, among available data, the demography is an input data of the model and can thus not be used to evaluate it. It is possible to get the agricultural production at the scale of Dayuma only, which prevents us to use them to a spatial evaluation. We thus had no choice but to use a proxy to evaluate the spatial accuracy of our model. We thus use land cover maps, based on satellite images classifications and carried out by Ecuadorian government services[1], on which are identified the deforested area. We propose a thematic way to evaluate the model based on a meaningful division of the space.

The second contribution is related to the agent behavior architecture. We argue that, given the lack of data we can gather and in a context of bounded rationality, it is irrelevant to model agent decision-making process with global optimization over all its possible alternatives. On the contrary, we choose an extended KIDS-like [5] approach in which we theorized that agents behavior is highly constraint and they do every task they have to do in the best way they can. They thus perform multiples optimization in a sequential way, which allows to have a good and easy way to implement risk management. The model is thus built from qualitative ground survey results and observations of the people behavioral process. The model had at its core, the process by which the locals make decision about their installation and their land management. This model is process-based, which means that we try to understand and implement the behavior of the locals at the best we can. The behaviour we try to model is the installation of settlers in Dayuma and how they manage their land. We use the pattern of deforestation we get as a proxy value which help us to calibrate and validate the model.

[1] *Mapa de cobertura y uso de la tierra del Ecuador continenal año 1990*, Ministerio del Ambiente, 2014.

This paper is organized as follows. Section 2 presents the context of the study and a brief state of the art of related works. Sections 3, 4 and 5 are focused on the model presentation, using the standard O.D.D. protocol [9]. Section 6 shows the preliminary results obtained on the parish of Dayuma in the province of Orellana. Finally Sect. 7 concludes and highlights perspectives.

2 Context

2.1 Historical Context

From the first petrol discovery in 1967, until the beginning of the conflictual and trial era (1990–2000) and the emblematic Aguinda vs. Texaco, Inc. court case in the 1990s [11], our study site (the Northern Ecuadorian Amazon, the region called "Oriente") lived mainly controlled by the Texaco Inc. company that played a central role in local governance, exploiting the petrol resource with a though policy that had a disastrous impact on the environment and human communities.

During this era, the North Oriente was also the object of a colonization plan supported by the various central governments of that time in an effort to relocate the surplus of peasants of the mountain (*Sierra*) and coastal (*Costa*) areas, most of whom lacked land tenure. The plan was supported by two laws (1964 and 1973) that led to the creation of the IERAC (Ecuadorian Institute of Agrarian Reform and Colonization, which organized this colonization) but also to the opening of gravel roads into the forest and connecting oil wells by the petrol company following an agreement with the government. This thus means that the current spatial structure of the colonization reflects the organization of the geological resources more than the potentialities of the surface. Primary forests were therefore exploited for their wood and colonized along and around these roads and tracks: each family of *colonos* received a *finca* (farm) of approximately 50 hectares and had to clear at least half of it for agricultural purposes. The forest lost territory while indigenous communities regressed, either changing their way of life or disappearing. However, progressively effective applications of the law of *comunas* (1937) leased parts of indigenous territories to natives (both locals and those coming from southern provinces, such as some Shuar families), offering some protection thanks to collective land tenure.

Prior to establishing national law to control oil exploitation in 1990, including waste disposal, oil companies did not undertake measures to protect the environment. It is why we stopped our simulation in 1990 with the prospect of adding a political module in the future.

2.2 Related Works

The biophysical processes are based on a previous work presented in [4] which focus on oil hazards in the area and their impact on inhabitants. It has been developed using the generic agent-based modeling and simulation GAMA platform [8]. This first model lacked farmers' agricultural behavior model to manage their exploitation and make their decision in terms of cropping.

The use of Agent-Based Models (ABMs) to study Socio-environmental Systems has now widely spread in the modeling community [17] and the description of agents representing human beings (*e.g.* farmers) with decision to make is still a really challenging task. According to [1], many approaches, derived from economic and social sciences theories, have been used to model farmers' behaviour in relation with natural systems in ABMs. Firstly, the micro-economic approach, which consists of agents maximizing an utility function based on revenue or profit, rely on the assumption of rational agents (homo economicus) (*e.g.* [16] among many others). Then and more, a psychological and cognitive approach integrating more abilities for agents and their aspirations, beliefs and intentions as well as social structures effects like social norms or social reproduction. In this way, [6] used a specific framework for their agents, that integrate behavioural drivers, in their study of the use of pesticides in Colombia. Similarly, [7] used Belief Function Theory to model yearly decision-making process of cropping plan in the South-East of France. Other models, from a participatory approach, directly involve stakeholders in the modelling process [18]. Also, KIDS models are driven by empirical data [10]: [15] have developed an ABM on the Ecuadorian Amazon used to simulate land use change on farms, based on empirical rules from a socio-economic and demographic survey; [2] have used a cluster analysis on empirical data to distinguish four types of farmers' agents in a spatially explicit model based on cadastral information. Recently, [12] reviewed ABMs for agricultural policy evaluation.

The next three sections are dedicated to the model description using the O.D.D. protocol (Overview, Design Concepts, Details) [9].

3 Overview

3.1 Purpose

The aim of the model is to reproduce the evolution of the parish (*parroquia*) of Dayuma under the impact of the migration of Escuadorian farmers in the Amazonian forest induced by the petrol exploitation of the area and their settlement.

We aims at reproducing the migration and settlement of Ecuadorian farmers in the Amazonian forest induced by the petrol exploitation of the area and at observing its consequences in terms of deforestation (for agricultural purpose).

3.2 Entities, State Variables, and Scales

Scales. As detailed in the input data presented in the Sect. 5.2, the modeled system is an area of approximately 87 km by 58 km. In this area, we chose as smallest spatial unit, the plot, that is the agricultural unit in our model. A plot is defined as a square with an edge of 90 m (for the high resolution Digital Elevation Model).

The simulations are launched from the 1st of January 1960 and run until the 31st of December 1990, with a simulation step lasting 1 month.

Entities. As presented previously, our main interest is to study the colonization and the anthropization of the parish of Dayuma over a long period of time (30 years). This process should take into account arrivals of new settlers and their allocation of a new `finca` (driven by the evolution of the `road` network), but also the development of the agricultural activity and its effect on the landscape. We thus made the hypothesis that the `family` is the key entity in this process (rather than the individual). We have introduced individual `person` and `couple` to simulate the demography, *i.e.* aging, wedding and new individual birth processes. The evolution of the number of inhabitants has an influence on the food needs and working force of each family and thus on the agricultural process, and finally the territory land cover. Agriculture relies on two kinds of entities: human activities and decision-making and physical support. The family think its agriculture activities at the scale of the `finca` with individual implementations at `plot` scale. We made the choice for modularity purpose to introduce the `finca_manager` agent that manages the `activity` set for the whole *finca* and in particular their `activity_state` that evolve step by step. We also made the choice to split the set of activities (coffee, market gardening...) into `subsistence_activity`, necessary to feed the family, and `payout_activity`, that provides financial incomes.

To make its decision about new activities, the `family` via the `finca_manager` needs to know the history of its own production, the `market` price and via the `path_manager` the price and time to send it to the market.

These entities, their main attributes and relations are summarized in the class diagram presented in Fig. 1. Only the main attributes and operations are presented in the diagram (in particular attributes used only for internal computations are not displayed).

Environment Variables. The environment variables are mostly used to initialize the family state: initial capital (in dollars), the number of working day per month and of working hour per day. We also have the price of every production and the demography (number of inhabitants, birth rate ...) at each time step.

We initialized the capital of each family at 150 dollars to explicit that they don't have many resources. We have put as a hypothesis the amount of work that a human can do in a month. We attribute to every adult 20.5 Human Day per month and to every child or elderly half this amount. During every day, a human can work 8 h. We also have the price of every production at each time step, the demography, the legal context in order to be able to change them in futures studies and use them to test the effect of different politics.

3.3 Process Overview and Scheduling

At each simulation step, the processes are scheduled sequentially in order to avoid unexpected side effects of possible interactions. First the demography is applied: it updates the family population and installs the new people on *fincas*. Then the bio-physical environment is updated: the biomass of each plot, the roads and markets are updated given preloaded data. The path manager gives for every plot, the time and the price for every activity to go from one plot to the

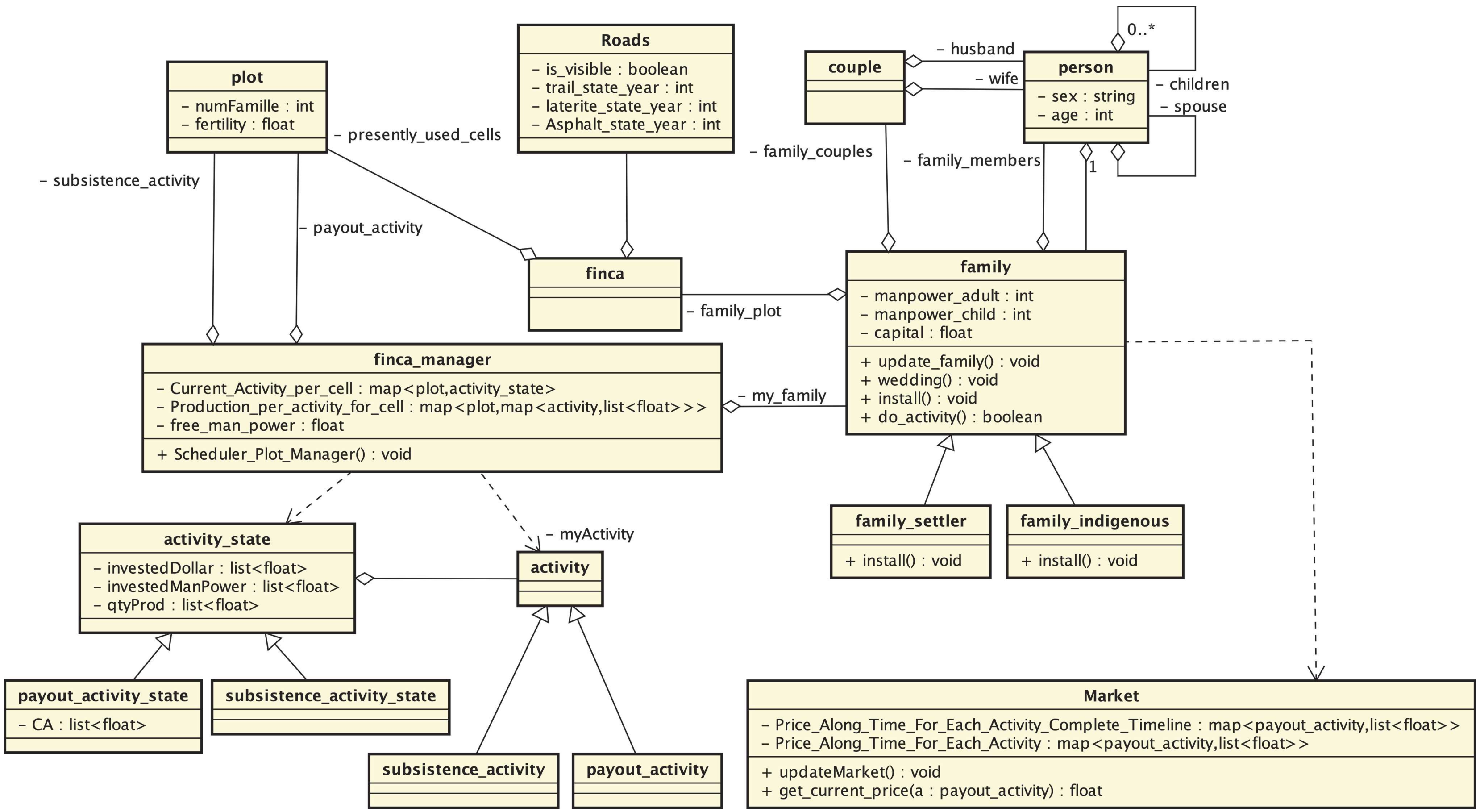

Fig. 1. UML class diagram

market of the map. Finally, families can start managing their *finca*, *i.e.* exploiting it with the constraints of first trying to feed the family members and only in case of remaining working time, earn money.

4 Design Concepts

As recommended by the authors of the O.D.D. protocol [9], only the relevant Design Concepts are presented below.

Objectives. The main objectives of the families are first to produce enough for their needs (through subsistence crops) and then to maximize their incomes with wage-earning and crash crops taking into account their current amount of money and their available working force. The self-feeding is divided in two objectives. The first one is the need to produce enough proteins for the entire family and the second one is to produce enough calories for everyone in the family. Families will produce enough proteins and its associated calories and will try to reach the total amount of needed calories through carbohydrates (carbohydrate here).

We do not take into account vitamins because the gardening tends to produce enough of it.

Learning. We make the hypothesis that families have a poor knowledge on their ground fertility but a good knowledge on their past productions depending on the activity and the plot. Each family will thus remember its production plot each crop was made on, to better know their ground and thus better predict their future incomes.

This limited knowledge prevent them to predict their future generated incomes by each crop. But over the simulation they will observe this production and learn to have a better knowledge over their ground. The learning revolve around a perfect memory for previous events. No fog of war is used. Every family will remember its production and the fertility of the plot each crop was made on.

Interaction. In addition to the use of learning, families try to improve their knowledge by asking neighbor families an estimation of their production if they don't already have their own information. Families choose to only take into account the information given by the family which is in the closest situation about to the use of one type of production on one plot.

Families can share information. When a family has no knowledge on the production it can expect, it will ask an estimation to the neighbouring family. Each family with knowledge on this matter will then provide information about their productions. The asking family will only take the information given by the knowing neighbour which had this activity on a plot with the closest estimated fertility to the one family want to use.

Sensing. To make their decisions, families have different information coming from their sensing: they have a perfect perception of the price markets and their ground fertility. Families have knowledge over the state of the transport network: they have access to the time and price, and manpower needed to send their productions from their farm to the marker.

Adaptation. The simulations are run over 30 years with a step of 1 month; during this period of time, both the bio-physical (route network or soil fertility) and socio-economic (activity prices or demography) environments evolve. The families have to adapt to theses changes by abandoning less-profitable activities to settle new ones that can bring more incomes.

Prediction. At several steps in the family behavior, it needs to make some choice depending on the production an activity can give on a plot. It can be the case to predict whether the family is self-sufficient (*i.e.* it produces enough food to feed all the family[2]) or to choose which activity is the best to install on a given plot (in particular because both the turnover and the expenses are computed from the production). In order to predict the production of a plot for a given activity, the family extracts from its knowledge all the time it had a production of this activity. It then looks at the production made by the plot with the closest fertility and use it as the predicted value and thus an estimator of the production on the given plot.

At different moment, the family have to make choice depending on the production an activity can give on a plot. In order to succeed in this task, it will extract from its knowledge all the time it produce crops of this activity and the fertility of the relief it was made on. It will then look at the production made by the plot with the closest fertility and use it as a predicted value.

The family have to predict if it will be self-sufficient (if it will produce enough food to feed all the family). In order to succeed in this task, it must know of all the macronutrients it needs and an estimation of the macronutrients produced by the subsistence crops. The needed macronutrients are calculated according to [13]. The produced macronutrients are calculated by predicting the total amount of each crop we produced and summing the associated macronutrients. If all the predicted produced macronutrients are superior or equal to the needed macronutrients, the family is self-sufficient.

The family needs to predict the gross incomes gained by putting an activity on a plot, in order to choose which activity is the best. In order to do this, the family will have to predict the turnover and the expenses. Both needs to predict the production. The turnover is calculated by multiplying the price per kilogram of a crop by the number of crops in kilograms produced. The expenses are a sum of the taxes on the turnover and the cost of sending the production to the market.

Stochasticity. The main stochastic part of the model is the evolution of the population: in particular, the initial number of individuals in each settler family is random. We use the demographic data to get the new number of people that will get in Dayuma, but the existing families update themselves randomly given the birth and mortality rate. Then we create a number of families which is equal to the number of new people coming divided by the average number of people in

[2] The macronutrients are calculated given an activity production based on [13].

a family. In the family, the number of children is between 2 and 5, the number of adult between 2 and 4 and the number of old people between 0 and 2. The chief is the oldest person in the family.

There is also an element of stochasticity on the production of every activity. The maximal production has a top and bottom bound and the actual production is calculated according to the quality of the plot and the calculated maximal production. The work in the oil industry is stochastic too: family members only have a certain chance of getting a one-month job, at each time step. The salary is set between 300 and 400 dollars per month.

Collectives. The model contains two main collectives: the family and the *finca*. The family is the entity that takes all the decisions in terms of agriculture development, *i.e. finca* management (through the proxy entity `finca_manager`). It is composed of individuals and couples: these two kinds of entities are integrated in the model only to manage the demography part of the model. They are also used to compute the family working force and its food need. The *finca* is only the administrative property of the family and gather all the plots it contains, but all the dynamics are implemented at the plot scale.

Observation. During the simulation, we observe activities productions and the money amount of families. We also observe the final map of forest/deforestation (produced in the model given the plot biomass) with the intention to compare it with actual satellite image, in order to evaluate our model. This final indicator is the one that is used to compare the results of the simulation to the real deforestation rate.

5 Details

5.1 Initialization

- We initialize the plots from the Digital Elevation Model of Dayuma. Every plot has an initial surface, fertility and biomass. Its biomass is calculated according to the local biomass.
- The position of the market is initialized at the place that is extracted from the real-life data. And the *fincas* and *communas* are created from input data.
- For the entire duration of the simulation, the same set of activity is used. We have 7 of them, 3 being annuity farming (coffee, cacao and breeding) and 4 subsistence farming (plantain banana, market gardening, corn/manioc and small breeding).
- We initialize the family capital to 150 dollars, the working time of an adult to 20.5 Human days per month and half of it for a child or an elderly and a day to 8 h.

5.2 Input Data

First, the input data contain multiple spatial data: the shapefile of Dayuma bounds, the shapefile of main roads (with, for each road, its construction date, its state (trail, laterite or asphalt) and its mean transport speed), a shapefile of pedology (containing in particular fertility data), a cadastre file (a shapefile or the *fincas* and *communas*) and the Digital Elevation Model of the area (with a resolution of $90\,\text{m} \times 90\,\text{m}$).

In addition, the input data includes a tabular file with the demography for every month after 1960: it contains the total population, its evolution, the birthrate and the migration. Finally, the simulation needs a file that contains every subsistence or crop activity. For each activity, we can extract every data we need to do those activities, including the transport cost: the cost in man power or money for installation and maintenance, the maximum or minimum production, the lowest fertility on which we can produce crops with this activity, the questioning frequency, the necessary surface to do it and the max number of plots we can put this activity on. For the subsistence activities, we have the percentage of carbohydrates or proteins contained in the production. Finally, we have a file that provides selling prices for each cash crop and for every year.

5.3 Submodels

The Colonization Submodel. At every time step, this model will simulate the demography of Dayuma. Families will be created and install themselves. 51% of the created families are settlers and the 49% remaining ones are indigenous. The indigenouses install themselves in the *communa*, with other indigenous families and are assigned a farm on which they have the operating rights. The settlers install themselves near the roads and in the *finca* that is the closest to the market.

The main stochastic part of the model is the evolution of the population: in particular, the initial number of individuals in each settler family is chosen randomly. We use the demographic data to get the new number of people that will get in Dayuma. We make the existing family update themselves with the birth and mortality rate. Then we create a number of families which is equal to the number of new people coming divided by the average number of people in a family. In the family, the number of children is between 2 and 5, the number of adult between 2 and 4 and the number of old people between 0 and 2. The chief is the oldest person in the family.

The Market Submodel. The market find for every cash crop the price at which it is sold for the current year in the simulation and gives it to the family.

The Farming Submodel. The family has two objectives (in the following priority order): to produce enough food to feed all the family members and maximizing its money incomes. At every step, the family behavior can be described by the following steps:

1. **Prepare each plot for its activity.** In order to install an activity on a plot or to maintain already installed activities, the plot must have enough free space. The family ensures that this is the case for every active plot, otherwise it prepares it (*i.e.* and thus destroy the exceeding biomass). Preparation has

a cost, both in time and money, that is conditioned by the activity. This cost will be in money and in man power because it takes time and the family might have to rent a chainsaw and gasoline to cut the excess in biomass. This action is made for every plot, even the plot that are producing, because the family must maintain its crops.

2. **Do subsistence farming.** The subsistence farming has a cost in man power and it is done before the cash crops because it is more important to feed the family than to get money. From this activity, family members will get the needed proteins and carbohydrates.
3. **Work for the oil industry.** This activity costs 20.5 Human Day per month and earns money in order to create exploitations. The chance of getting a job follows a linear function which is at 10% in 1960 and 1% in 1990. They do this work before the cash crops only if they have less than 1,000 dollars in capital. If not, they prefer to add new cash crop activities and postpone trying to work in industry to the end of the step, if remaining working time is available.
4. **Do cash crops.** At this step, the family will do its activity of cash crops in order to get money. It is at this step that it will produce, send and sell its production.
5. **Question the activities and remove the useless ones if necessary.** At every step, the family checks if its active plots still have enough fertility. Furthermore, every activity must be questioned at some interval. Cash and subsistence activities are not questioned with the same criteria. In order to know if a subsistence activity must be stopped, the family checks if it produces too much food (30% more here) and if it is the case, it drops one activity while maximizing the food diversity. In the case of crops activity, it will check if it produces enough food and if it does not, it might have to delete the less profitable crop activity in order to free some manpower.
6. **Put in fallow the plots with less than 30% of fertility.**
7. **Add a new activity on a plot if possible.** The choosing of the plot on which the activity is done with the following Criterion of Selection.

$$CS(x,t) = \frac{p(t) + Fertility(x,t)}{(D(x)+1).(l(x)+1)}$$

with $D(x)$ the distance to the closest exiting plot of the *paroquia* of the plot x; $p(t)$: a random number in order to shuffle the plot with the closest score; $l(x)$ the risk of flood on the plot x. At this point, we have the choice between adding a cash crops activity or a subsistence farming activity. First, the family tries to predict if it is self-sufficient in food. If it is not, it adds a subsistence activity. If it is, it adds a cash activity. The adding of a subsistence activity is done according to the principle of plurality of the crops. Every family will try to diversify its food intake. So the family will try to have the same number of plot on each subsistence activity. When multiple activities don't have the same amount of occurrences as the most done activity of subsistence, we rank them accordingly to the inverse of the manpower needed to do the activity and install the best. If we want to add a cash crop, the family will estimate for every possible activity the gross income it could generate. The best activity will be installed on the plot. This process is explained in the prediction part of the design concept.

6 First Results

Due to lack of precise data, we cannot evaluate the simulation outputs directly with the real corresponding data, such as the spatial distribution of crops. As a proxy, we use the deforestation as the only indicator we can evaluate. As reference data, we choose to use a map extracted from satellite data by identifying deforested plots. In order to evaluate our model, we thus compare the level of deforestation in our simulation outputs[3] and the one recorded in the region of Dayuma in 1990, extracted from land use and land cover maps[4].

Due to our lack of data, we opted for a grid-based method to evaluate the outputs, rather than pixel to pixel methods (statistical indices like Cohen's kappa or confusion matrices), because we had no interest in reproducing pixel-precise dynamics for a process-based model. The pixel-to-pixel evaluation method consists in computing the number of pairs of pixels from the real data and simulated data that are different[5]. This distance is thus not that interesting because we want to know where the settlers came; and how they impacted the field is just a consequence of his/her behavior. From our point of view, there is no difference between using one plot or the one that is just next to it because it is still in his farm. We want to see if our agents have proportionally the same amount of effect on their land than the settler who was in the same situation.

Therefore, in order to compute the grid-based error indicator, we created a grid with a mesh of five squared kilometres covering the entirety of our study area, for which we calculated a proportion of deforested pixels (on the one hand from satellite image classification data and on the other hand from simulation data) by mesh in a geographic information system (GIS). Then, the layers are subtracted from each other (deforestation rate in simulation data - deforestation rate in satellite image classification data), to obtain a deviations map (Fig. 2) which indicates over-estimations and under-estimations of deforestation by the model. We can see on it that the model over-estimates forest clearing for the most part, especially in the northern and in the southern parts of the study area (up to 24.4% of over-estimation in some meshes). However, there is a bias related to the presence of a side effect, which we have tried to limit (without losing too much information) by eliminating from the analysis the border meshes of which less than half of the pixels were included in the study area. This evaluation method just give us an evaluation at a larger scale than the pixel-to-pixel method but doesn't allow us to test our assumptions. Here our assumption are that settler

[3] Deforested pixels are plots where at least 50% of the biomass is missing.

[4] We consider as "deforested" pixels belonging to the categories "populated areas" and "agricultural land" of level 1 of *Mapa de cobertura y uso de la tierra del Ecuador continenal año 1990*, Ministerio del Ambiente (MAE), 2014.

[5] It is thus close to the classical Hamming distance [14].

want to be able to travel as fast as possible to the city, so they choose well-connected farms and the oil industry has an impact on the amount of money that the farmers have and thus on the amount of work they can do, which has a big impact of the land. Moreover, farmers in well-connected roads have to pay less to make theirs products travel to the market and have more money at the end of the month to invest in their farm.

A similar but more thematic and spatial approach to evaluate the simulation outputs of the model was used using the road network (Fig. 3). Indeed, given the role of the road network in the colonization process, it seemed interesting to take it as a reference point for the evaluation of simulation data. Thus, four zones, corresponding to different buffers around the roads (from one kilometre to ten kilometres and more), were used to calculate the same deforestation rates as previously. This thematic method allows us to spatially evaluate our results: over-estimations appear lower (less than 3% from one to five kilometres around roads) and underestimations are negligible (less than 1% beyond ten kilometres around roads).

We propose a global method of evaluating spatial results according to buffers that are made to test initial assumptions. It allows us to test the model and check if our assumptions can generate the expected results. Two type of error can be seen here:

- First there are error of modeling because we can see that in places that are not at all well-connected, there are still a few places that are deforested, and thus there are people who came here. We cannot explain why. Our assumptions still are coherent with the majority of the settlers, but are not complete.
- Second, there might be fine-tuning errors (or still modeling errors) because we are overestimating deforestation in well-connected areas.

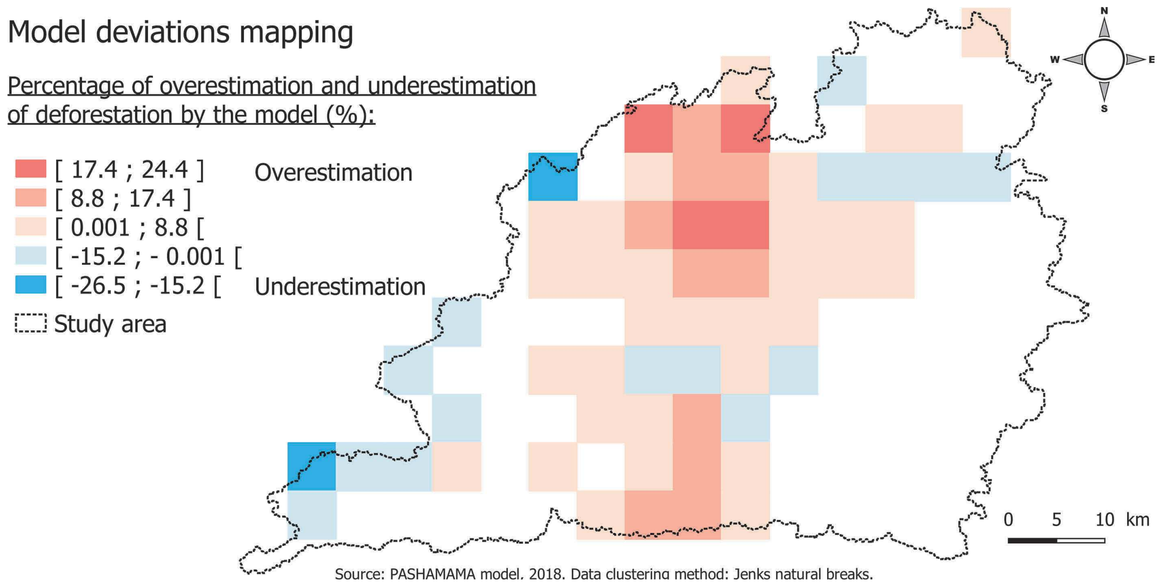

Fig. 2. Deforestation error by grid (in 1990)

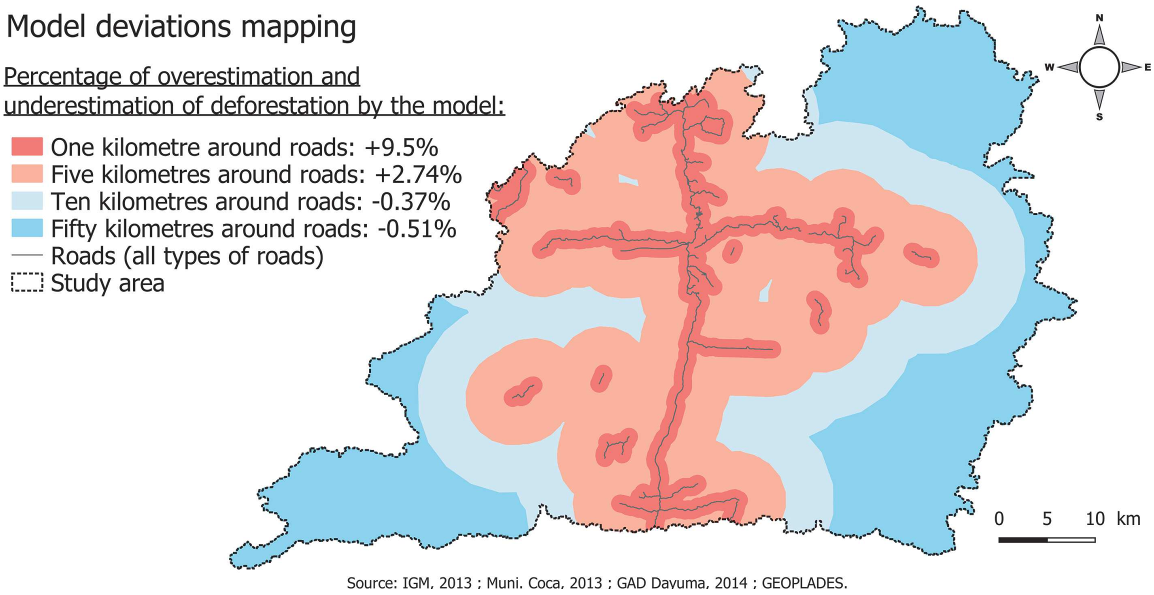

Fig. 3. Deforestation error by buffer (in 1990)

The separation of space according to our assumptions allows us to see easily what went wrong with the modeling process and how we can improve our model, either by fine-tuning the parameters or changing our assumptions. Here, we see that most errors are concentrated on the zone around one kilometer from the roads. It could mean that our understanding of the phenomenon in this area in particular is not that good. It allows us to confront our expectation of the spatial organization for the studied phenomenon and can help us understand the localized effect of our assumptions and thus what we must change.

7 Conclusion and Perspectives

In this article, we have presented the PASHAMAMA model and how we used qualitative data in order to design process-driven low-level cognitive agents. This approach based on multiple sequential optimizations is used to keep the process as descriptive as possible and generic enough to be applied to many case studies. Simulation results are displayed in Figs. 2 and 3. On the buffer map, we have better overall results than in the grid map, which means that the model is able to well reproduce the dynamics in the areas of interest. It shows us that the hypothesis of people installing themselves as close as possible to the roads might be correct. The first people who installed themselves are those who will cause the most deforestation because they have higher resources due to the high probability of getting a job. The results show some divergences between the reality and the model output. We argue that over-estimations are due to the excess in money earned by the agents: it seems to be due to the lack of data we have on the chances of getting a job with the oil industry. We have overestimated them. For the under-estimation, they are two kind. First, we cannot explain why people settle near isolated areas. The second kind is about the places with high under-estimation (near 25%) where there are settlers, but they can not earn enough money to have an efficient enough farm (and thus deforest).

Here, we can see that our model over-estimate the impact and efficiency of the first settlers that are close to the roads and might under-estimate that of the newcomers.

The work presented here is only a first step. In particular, the agricultural module still needs tuning on some parameters: after preliminary studies, the most important of them is the probability of getting a work in the oil industry. We need to design a function computing the hiring rate given the oil production and the cost of the oil barrel. This model needs to be extended to help us to understand more dynamics in Dayuma. We will consider the addition of a policy module in which the institutions will be able to put some rules for the agricultural production, oil production, or for the installation of health institutions.

Acknowledgment. This study was partly funded by the French research funding agency ANR (Agence Nationale de la Recherche), two French research institutions, the "Institut des Amériques" and "Maison des Sciences de l'Homme et de la Société de Toulouse" (USR CNRS 3414), ECOLAB Funds, Occitanie region and university of Toulouse. The team received logistical support from the IRD.

References

1. An, L.: Modeling human decisions in coupled human and natural systems: review of agent-based models. Ecol. Model. **229**, 25–36 (2012)
2. Bakker, M.M., van Doorn, A.M.: Farmer-specific relationships between land use change and landscape factors: introducing agents in empirical land use modelling. Land Use Policy **26**(3), 809–817 (2009)
3. Bilsborrow, R.E., Barbieri, A.F., Pan, W.: Changes in population and land use over time in the Ecuadorian Amazon. Acta Amaz. **34**(4), 635–647 (2004)
4. Chapotat, W., et al.: An agent-based model of the Amazonian forest colonisation and oil exploitation: the Oriente study case. In: Sauvage, S., Sanchez Perez, J.-M., Rizzoli, A.E. (eds.) iEMSs, vol. 5, pp. 1335. iEMSs (2016)
5. Edmonds, B., Moss, S.: From KISS to KIDS – *an 'anti-simplistic' modelling approach*. In: Davidsson, P., Logan, B., Takadama, K. (eds.) MABS 2004. LNCS (LNAI), vol. 3415, pp. 130–144. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32243-6_11
6. Feola, G., Binder, C.R.: Towards an improved understanding of farmers' behaviour: the integrative agent-centred (IAC) framework. Ecol. Econ. **69**(12), 2323–2333 (2010)
7. Gaudou, B., et al.: The MAELIA multi-agent platform for integrated analysis of interactions between agricultural land-use and low-water management strategies. In: Alam, S.J., Parunak, H.V.D. (eds.) MABS 2013. LNCS (LNAI), vol. 8235, pp. 85–100. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54783-6_6
8. Grignard, A., Taillandier, P., Gaudou, B., Vo, D.A., Huynh, N.Q., Drogoul, A.: GAMA 1.6: advancing the art of complex agent-based modeling and simulation. In: Boella, G., Elkind, E., Savarimuthu, B.T.R., Dignum, F., Purvis, M.K. (eds.) PRIMA 2013. LNCS (LNAI), vol. 8291, pp. 117–131. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-44927-7_9

9. Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F.: The ODD protocol: a review and first update. Ecol. Model. **221**, 2760–2768 (2010)
10. Janssen, M.A., Ostrom, E.: Empirically based, agent-based models. Ecol. Soc. **12**(2) (2006)
11. Juteau-Martineau, G., Becerra, S., Maurice, L.: Ambiente, petróleo y vulnerabilidad política en el oriente ecuatoriano: ¿hacia nuevas formas de gobernanza energética? América Lat. Hoy **67**, 119–137 (2014)
12. Kremmydas, D., Athanasiadis, I.N., Rozakis, S.: A review of agent based modeling for agricultural policy evaluation. Agric. Syst. **164**, 95–106 (2018)
13. Laure, J.: Nutrition et population en vue de la planification alimentaire. ORSTOM, PARIS (1983)
14. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error-Correcting Codes. Elsevier, Amsterdam (1977)
15. Mena, C.F., Walsh, S.J., Frizzelle, B.G., Xiaozheng, Y., Malanson, G.P.: Land use change on household farms in the Ecuadorian Amazon: design and implementation of an agent-based model. Appl. Geogr. **31**(1), 210–222 (2011)
16. Saqalli, M., Bielders, C.L., Gerard, B., Defourny, P.: Simulating rural environmentally and socio-economically constrained multi-activity and multi-decision societies in a low-data context: a challenge through empirical agent-based modeling. J. Artif. Soc. Soc. Simul. **13**(2), 1 (2010)
17. Schulze, J., Müller, B., Groeneveld, J., Grimm, V.: Agent-based modelling of social-ecological systems: achievements, challenges, and a way forward. J. Artif. Soc. Soc. Simul. **20**(2) (2017)
18. Voinov, A., Bousquet, F.: Modelling with stakeholders. Environ. Model. Softw. **25**(11), 1268–1281 (2010)