Simulating the Spatial Distribution of Pollutant Loads from Pig Farming using an Agent-based Modeling Approach

An The Ngo (ngothean@gmail.com)
Vietnam National University of Agriculture https://orcid.org/0000-0002-0518-318X
Giang Thi Huong Nguyen
Vietnam National University of Agriculture
Duong Huu Nong
Vietnam National University of Agriculture
Linda See
IIASA: International Institute for Applied Systems Analysis

Research Article

Keywords: Agent-based model, pollutant load, pig farming, planned behavior, water pollution, livestock waste

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-572047/v1

Simulating the spatial distribution of pollutant loads from pig farming using an agent-based modeling approach

An The Ngo [1,*], Giang Thi Huong Nguyen [1], Duong Huu Nong[1] and Linda See [2,*]

[1] Faculty of Environment, Vietnam National University of Agriculture; Trau Quy, Gia Lam, Hanoi, VIETNAM
[2] Advancing Systems Analysis Program, International Institute for Applied Systems Analysis (IIASA); Schlossplatz 1 - A-2361 Laxenburg, AUSTRIA
* Correspondence: ntan@vnua.edu.vn; Tel.: +84-9122-54886 (A.T.N.) and see@iiasa.ac.at; Tel.: +34-2236-807-423 (L.S.)

Abstract

This research developed an agent-based model (ABM) for simulating pollutant loads from pig farming. The behavior of farmer agents was captured using concepts from the Theory of Planned Behavior. The ABM has three basic components: the household or farmer agent, the land patches and global parameters that capture the environmental context. The model was evaluated using a sensitivity analysis, and then validated using data from a household survey, which showed that the predictive ability of the model was good. The ABM was then used in three scenarios: a baseline scenario, a positive scenario in which the number of pigs was assumed to remain stable but supporting policies for environmental management were increased, and a negative scenario, which assumed the number of pigs increases but management measures did not improve relative to the baseline. The positive scenario showed reductions in the discharged loads for many sub-basins of the study area while the negative scenario indicated that increased loads will be discharged to the environment. The scenario results suggest that to maintain the development of pig production while ensuring environmental protection for the district, financial and technical support must be provided to the pig producers. The experience and education level of the farmers were significant factors influencing behaviors related to the manure reuse and treatment, so awareness raising through environmental communication is needed in addition to technical measures.

Keywords: Agent-based model, pollutant load, pig farming, planned behavior, water pollution, livestock waste

1 Introduction

In recent years, surface water quality in Vietnam has been severely degraded due to increased environmental pressures from domestic, industrial and agricultural activities (MONRE, 2014). In the agricultural sector, pig production is the largest source of pollution, accounting for 51% of the total nitrogen load discharged into rural water basins (An et al., 2020). Many local water pollution problems are attributed to pig production (Nguyen et al., 2021, Vu et al., 2007). Accordingly, the government and authorities have issued a waste management program aimed at minimizing sources of pollution and environmental problems around pig farming areas (Nguyen et al., 2021).

The individual's responsive behavior is an important factor in environmental policy implementation. However, this factor is often overlooked in policy and planning processes due to lack of information at the individual level and is often treated as a "black box" in policy assessment (Berkes et al., 2000). In fact, agricultural production in general and pig production in particular are not only influenced by external factors (e.g., policy), but they are also impacted significantly by internal factors existing within the farm (Happe et al., 2011). Therefore, farm management behaviors are heterogeneous in nature, depending on the complex interactions between their own specific conditions and the external context. For example, some farmers may be willing to fully comply with regulations on waste treatment while there are also many farmers who do not respond to the policy as expected. Examples of negative or non-compliant behavior could be the discharge of untreated manure in prohibited areas or the rejection of mitigation technologies recommended by the government (Zheng et al., 2013). The difference in the behavior and effectiveness of policy implementation is also due to the heterogeneity of economic conditions and different perceptions of farmers (Feng and Heerink, 2008, Ngo et al., 2012).

Pro-environmental behavioral studies have been carried out for many years, typically based on Ajzen's studies (Ajzen, 1991). Based on this approach, researchers have conducted studies on waste management behaviors and tested the hypotheses that the attitudes, subjective norms and perceived behavioral controls have a significant influence on recycling and composting behaviors (Taylor and Todd, 1995b, Taylor and Todd, 1997). These factors have been proven as the determinants of recycling behaviors in research undertake in Taiwan (Chu and Chiu, 2003) and Scotland (Edgerton et al., 2009). Most studies on the Theory of Planned Behavior (TPB) have been based on a large survey data set to determine and validate the influencing factors. Some studies based on statistical analysis have also attempted to extend the research focus to farm characteristics and social context factors (Tonglet et al., 2004, Supaporn et al., 2013, Ittiravivongs, 2011, Nsimbe et al., 2018). However, the findings were focused mainly on the deterministic interaction without including contextual factors that might influence collective individual behaviors stochastically. Therefore, these results are not capable of predicting behavior in the context of heterogeneity, especially perceptions of farmers as mentioned above.

To simulate such aggregated behavior within a heterogeneous and interacting environmental system, a novel modeling approach is needed. An agent-based model (ABM), which integrates spatial analysis, has become a particularly promising approach for such a purpose. In an ABM, a system is modeled as a collection of autonomous decision-making entities called agents, each of which can individually assess their situation and make decisions on the basis of a set of rules (Bonabeau, 2002). By modeling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviors, which together give rise to the behavior of the system as a whole, can be observed (Macal and North, 2010). ABMs can connect the decision making agent at a microcosmic level together with the macroscopic phenomena of the system, which can help us study complex adaptive systems (Wooldridge et al., 2000). Current available spatial analysis software (e.g., ArcGIS, BASINS) has been augmented with built-in terrain and hydrological analysis toolkits. These software tools become extremely useful in supporting the environmental decision-making process, especially when they are integrated with an ABM approach.

ABMs have been applied previously in waste management research (Meng et al., 2018, Scalco et al., 2017b). Regarding waste from pig production, ABMs have been initially tested in some select places (Karmakar et al., 2007, Zheng et al., 2013, Van der Straeten et al., 2010). The most relevant study on the application of ABM in animal waste management was conducted in China (Zheng et al., 2013). In particular, the authors developed an ABM called the Nutrient Emission Model (ANEM) to integrate the decision-making process of individuals into an environmental impact assessment. Decisions related to the conversion of farm-scale, manure collection technology and manure handling patterns were formulated in the ANEM through the underlying function of the cost, benefit and risk related to the performance of the farms. In addition to the conversion rate at the livestock scale, this model could also forecast the nutrient emissions for the study area. The limitation of this model was the simplified behaviors of the waste collection and handling. As the authors acknowledged in their paper, important factors, such as education level, socioeconomic status and innovation, were not included in the simulation. However, these factors are very important determinants of behavior that should be included (Ajzen and Fishbein, 2005, Chu and Chiu, 2003, Edgerton et al., 2009, Taylor and Todd, 1997).

ABM applications in the field of TPB have also been developed (Scalco et al., 2017a, Scalco et al., 2018). However, since ABMs are still a relatively new approach and the TPB is inherently complex, the existing models have mostly been developed and tested in specific areas, and they have been unable to represent general behaviors. The integration of contextual factors into the TPB model still requires considerable validation, especially the definition of these contextual factors in spatial terms within a complex landscape. Applied research on the TPB is particularly limited in Vietnam (Si et al., 2019), while ABMs applied to TPB are even rarer. Yet, environmental controlling behavior is often considered a decisive factor in the implementation of local environmental management policies (Edgerton et al., 2009). In particular, waste recycling and composting play an important role in promoting circular economic development in the pig production sector, which several sustainable development policies aim for in the coming years.

This study aims to develop an ABM to simulate the spatial distribution of pollutant loads generated from pig farming, thereby supporting waste management planning based on the actual load accumulated in the sub-basins. The model systematically analyzes the interactions between the agents (i.e., the farmers) and their surrounding environment through aggregated environmental performance based on the internal attributes of the farms and the external policy context at the national level. The model applies the TPB approach to systematically predict farmers' behaviors within Vietnamese conditions. Therefore, this study contributes to broadening the application of TPB to systems that have dynamic individual perceptions within a heterogeneous context. At the same time, the model also integrates spatial analysis in an integrated model, in particular the analysis of waste accumulation zoning and the interaction of the location and distance between the farms, which has not been the focus of previous research.

2 Study area

The study area is Yen Dung district, which is a semi-mountainous district including 19 communes and two towns in Bac Giang province in northern Vietnam. Yen Dung is surrounded by the Cau, Thuong and Luc Nam Rivers. The western part of the district has a high mountain range running through four communes: Noi Hoang, Yen Lu, Nham Son, and Leo Town. Other communes have low sloping terrain, interleaved with lowland for water retention before spilling into the local river systems. The hydrological flows are highly dispersed on the surface terrain, with the major direction of flow from the northwest to the southeast. The study area is shown in Fig. 1.

This study focuses mainly on pig farming at both medium and small scales. In Yen Dung district, pigs are mostly raised on a small scale in households. There are a few farms that have more than 1000 pigs, but no large-scale farms were found in the district. In 2020, the district had 4,274 pig farms with a total pig population of 82,313 (Bac Giang department of Animal husbandry and Veterinary, 2020). However, environmental issues have arisen from the management of the pig waste. Although some of the pig manure is reused through composting or selling, a significant volume of waste is treated by biogas digesters or discharged into the environment if the farm has no waste treatment system, which leads to water pollution.

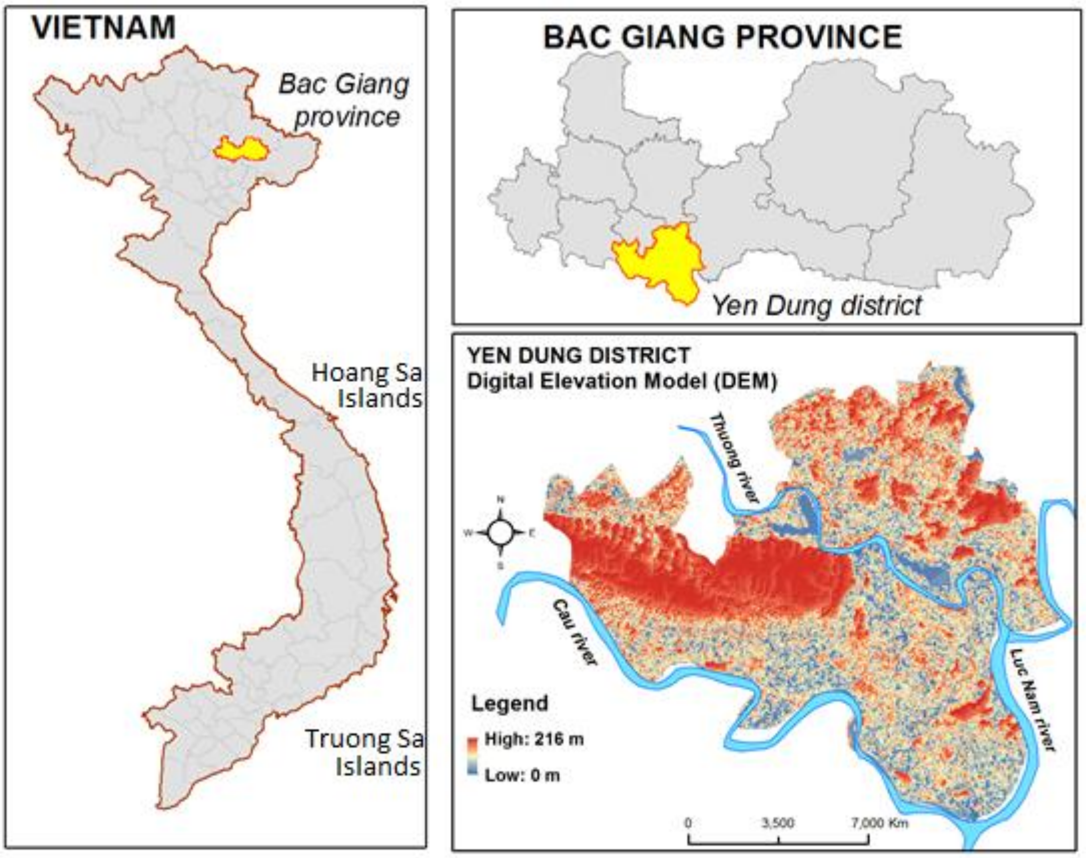

Fig. 1: The location of Yen Dung district and a DEM of the district. Source of the DEM: SRTM.

3 Materials and methods

3.1 Modelling approach

The ABM for predicting pollutant discharge loads from pig farming was developed based on the TPB approach (Ajzen, 1991), with an extension to include the influence of background factors according to the Theory of Reasoned Action (Ajzen and Fishbein, 2005). Many researchers have applied these theories to explain and predict intended human behavior with a special focus on waste management (Scalco et al., 2017a, Chu and Chiu, 2003, Taylor and Todd, 1997).

There are two important behaviors related to the pollutant load that need to be simulated: (i) recycling and (ii) treatment of pig manure. According to the TPB, each behavior will be determined by the intention to perform that behavior (BI). BI, in turn, is formed by (i) attitude (AT), which reflects feelings of favorableness towards a behavior; (ii) subjective norms related to environmental protection (SN), which reflects perception that significant influences (i.e., family, society) desire the individual to perform a behavior or refrain from it; and (iii) perceived behavioral control (PBC), which reflects beliefs regarding control over factors that may facilitate or inhibit the implementation of a behavior.

The behavioral intention (BI) of a farmer is formulated as follows (Ajzen, 1991):

$$BI = w_A AT + w_{SN} SN + w_{PBC} PBC \quad [1]$$

where w is the weight factor. The relationship between BI and the actual ability to perform the behavior (B) is defined as:

$$B = w_{BI} BI + w_{PBC} PBC \quad [2]$$

Each of the determinants of intention (i.e., AT, SN, and PBC) for a particular behavior is, in turn, determined by the strength of belief (b) associated with each possible outcome (i) when performing the behavior as well as the subjective evaluation of an individual (e) regarding the desirable benefits from the outcome. Stated formally, AT is proportional to the sum of the attitudinal belief (b_i), multiplied by an evaluation of the desirability of that outcome (e_i):

$$AT \propto \sum b_i e_i \quad [3]$$

According to previous research (Chu and Chiu, 2003, Taylor and Todd, 1995b), two essential attributes that determine attitudes towards waste recycling or composting were the relative perception of personal benefits (PRB) and social benefits (SRB). PRB, in this study, was defined similarly to the study by Taylor and Todd (1997), which includes the cost or reward that livestock owners think they would receive when performing the behavior, which could also simply be the feeling of being recognized by others as a good person. The SRB includes aspects such as protecting natural resources and reducing the overall social cost to the community (e.g., reducing water pollution). Thus, equation [3] above can be written as:

$$A \propto \sum (b_{PRB} e_{PRB} + b_{SRB} e_{SRB}) \quad [4]$$

The SN is determined by the normative belief (n) that an agent should perform a particular behavior based on the views of relevant social references, and the motivations for the individual to comply with those references (m):

$$SN \propto \sum n_i m_i \quad [5]$$

As suggested in previous studies (Oskamp et al., 1991, Chu and Chiu, 2003, Edgerton et al., 2009), one important predictor of recycling behavior was having friends and neighbors who recycled. The SN's determinants were divided into two groups based on peer influences: primary normative belief (PNB), which include beliefs influenced by family, friends, and neighbors (Taylor and Todd, 1995a); and secondary normative belief (SNB), which includes beliefs influenced by environmental groups and governments (Edgerton et al., 2009), which can, thus, be defined as:

$$SN \propto \sum (n_{PNB} m_{PNB} + n_{SNB} m_{SNB}) \quad [6]$$

Perceived behavioral control (PBC) is determined by control belief (*c*), which reflects an individual's perceived difficulty or ease in performing a particular task; and by the perceived facilitation (*p*) to perform that behavior:

$$PBC \propto \sum c_i p_i \quad [7]$$

Based on previous studies (Ajzen, 1991, Chu and Chiu, 2003, Taylor and Todd, 1995b), PBC was divided into two components: self-efficacy (SE,), which is related to the perceived knowledge and effectiveness of recycling; and facilitating conditions (FC), which refer to factors such as time, household space, and the convenience of recycling. Hence, Equation [7] can be elaborated as follows:

$$PBC \propto \sum (c_{SE} p_{SE} + c_{FC} p_{FC}) \quad [8]$$

The framework in Fig. 2 shows that AT, SN and PBC are all influenced by household background factors such as education, experience, labor, income, social relations, land, treatment infrastructure, animal, interventions from the outside, etc. This formulation is compatible with findings from the literature (Supaporn et al., 2013, Nsimbe et al., 2018, Mustafa-Msukwa et al., 2011, Ajzen and Fishbein, 2005). In addition, environmental conditions such as the pollution status of the community, the distance to the neighboring households, and the treatment status of the neighboring households also have a great influence on people's intentions (Scalco et al., 2017b, Oskamp et al., 1991).

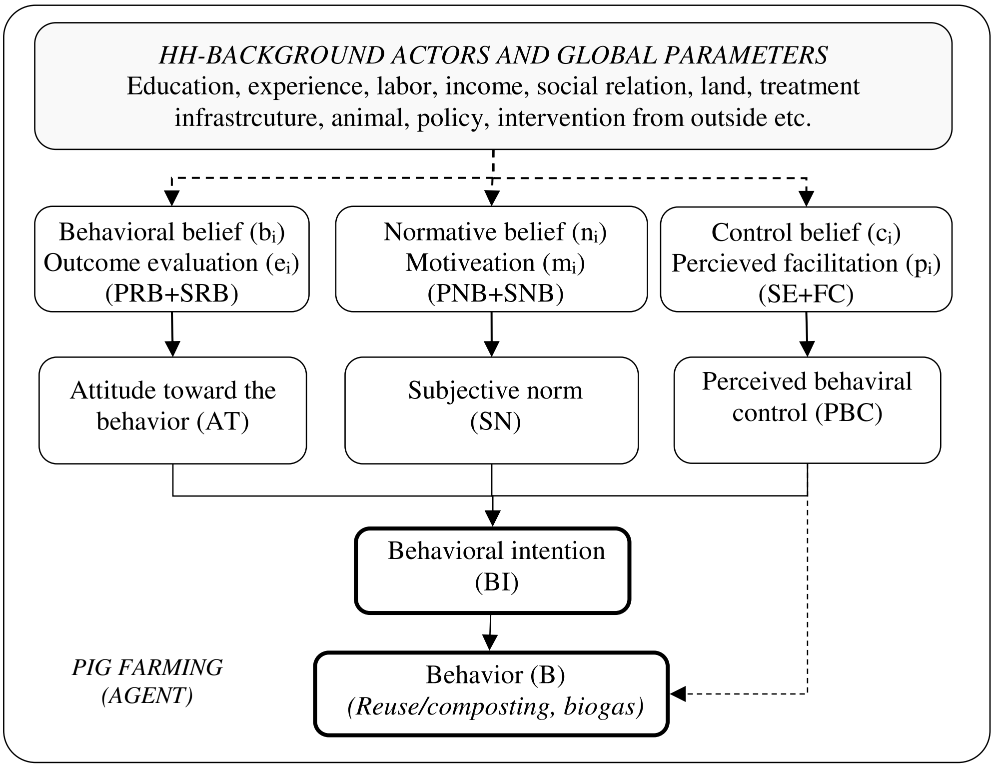

Fig. 2: A framework for predicting pig manure recycling behavior. Adapted from Azjen and Fishbein (2005).

The role of global parameters such as policy interventions (e.g., financial support, technical support, environmental pollution control) by local environmental management organizations were confirmed to have an impact on the decision-making process of environmental management (Edgerton et al., 2009, Supaporn et al., 2013, Ajzen and Fishbein, 2005, Scalco et al., 2018). However, household/farm context and global parameters may only have an indirect effect on AT, SN and PBC (Ajzen and Fishbein, 2005).

According to the theories mentioned above, when BI becomes actual behavior (B) will depend once again on the PBC of an individual. Both BI and PBC play an important role in predicting behavior, but, in practice, it is possible to use only one of them (Ajzen and Fishbein, 2005).

The approach described above is for predicting a behavioral intention based on attitudes, subjective norms and belief behavioral control and finally actual behavior. This approach has much better predictive ability than the direct prediction of a specific behavior (Ajzen, 1991). This approach is applied to predict the behavior of manure reuse and is embedded in the ABM as described below.

3.2 The agent-based conceptual model

The pig manure discharge load depends on (i) the generation of pollutants; (ii) the manure reuse rate; (iii) manure treatment; and (iv) manure discharge into the environment. An ABM was developed for simulating the interaction of all the above factors. The generation of pollutant load was not simulated as a behavior but calculated directly by multiplying the number of pigs with the emission coefficient as suggested by An et al. (2020). According to the information obtained from the survey in 2020, manure reuse activities were mainly composting (carried out at pig farms or sold to outside fertilizer manufacturers) while manure treatment was mainly by biogas digester operation. These two factors were directly related to the farm owners' behaviors and thus is modelling based on the TPB approach (Ajzen, 1991).

In terms of structure, the ABM model for predicting the pollutant load discharge from pig farming was constructed with 3 basic components, similar to the three basic structural components of the Netlogo software (Wilensky, 1999), which was designed specifically for agent-based modeling purposes (Macal and North, 2007, Railsback et al., 2006):

- Pig production household (HH-agent): is the agent where pollutants are discharged from, which is also a decision-making unit regarding manure recycling and treatment, thereby determining the pollutant load discharging to the environment. The pig production households are designed as agents, which are capable of making autonomous decisions based on their inherent characteristics and external environmental conditions. The hierarchical representation of household agents is by their own characteristics, manure management behaviors and group interactions or cooperation in the sub-basin.
- The physical environment (Land patches): is a grid of land parcels that consist of cells; each cell has x-y coordinates; neighbor relations between the cells; and biophysical properties of the patch. Land patches are represented in the ABM as the dynamics of the land properties (e.g., pollutant loads, pollutant density), which are updated after each model run.
- Global parameters or the global environment: is the set of external conditions that include important socio-economic and environmental policy-related parameters. The global environment is considered as the external driving force of environmental change. It varies from year to year but applies across the whole grid and influences all agents.

In the model, the HH-agents interact and change environmental properties by their pollutant discharge. The environment also changes the behavior of HH-agents by directly affecting the perception of each agent and indirectly through the social pressure created by neighboring agents. More details of the above components are provided in the sections that follow.

3.2.1 HH-agent

Each household is an autonomous agent that contains a background representing its household characteristics, and potential decision behaviors. The general structure of the household agent is provided in Fig. 3.

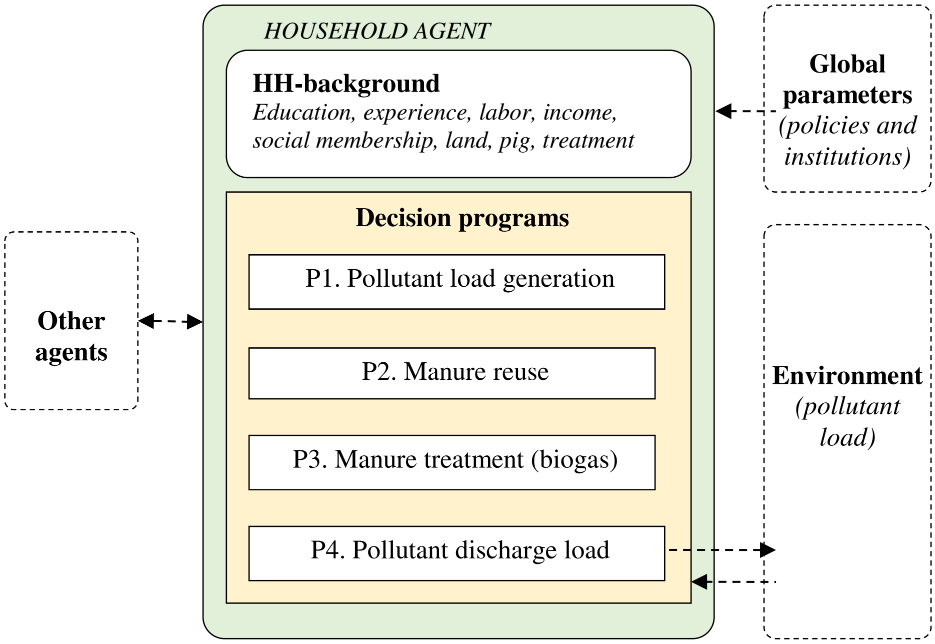

Fig. 3: The structure of the household agent

HH-background: is a set of household/farm characteristics that influence pig manure management such as education, experience of the household head, labor, income, social membership, land, number of pigs, and treatment facilities.

As shown in Fig. 3, there are four decision programs as follows:

P1. Pollutant load generation**:** is for calculating the pollutant load generated from pig farming. The load is the product of the number of pigs multiplied by emission coefficients. In this study, the pollutant coefficients (COD, BOD_5, TN and TP), which were officially recommended by the Vietnam Environmental Agency (VEA, 2019), were applied.

P2. Pig manure reuse**:** is for predicting manure reuse behavior with 2 routines as follows:

- *P2.1. Manure reuse behavioral intention (BI) routine*: The background factors, AT, SN and PBC have significant relationships with BI as confirmed by previous studies (Scalco et al., 2017b, Ajzen and

Fishbein, 2005, Edgerton et al., 2009, Mustafa-Msukwa et al., 2011, Chu and Chiu, 2003, Taylor and Todd, 1997). Therefore, this routine used coefficients associated with AT, SN and PBC from regression models to predict BI. First, AT, SN and PBC were predicted from background factors and global parameters. These three variables were then used to predict BI.

- *P2.2. Manure reuse behavior (B) routine*: According to the theoretical model, B is correlated with BI and PBC. This correlation has also been verified by previous studies (Ajzen and Fishbein, 2005, Chu and Chiu, 2003, Taylor and Todd, 1997). In this study, B is the actual manure reuse behavior (including on-farm and off-farm composting). The quantity of actual manure reuse by the HH-agents is predicted using the correlation coefficients from a statistical analysis. More details can be found in Ngo et al. (2021).

P3. Manure treatment (biogas)

According to the survey results, since composting is classified as reuse (i.e., not discharged into the environment), the only treatment method applied at the study site was by biogas digester. For households having a biogas digester, they usually discharge all the manure (except for the part collected for composting) through the biogas system before discharging it to the environment. For households without a biogas digester, the waste is discharged directly into the environment. Thus, the biogas treatment in the study site depends directly on the biogas digester related projects. Biogas digesters are usually only built by breeders who own a large number of pigs. Additionally, some households have built biogas digesters through support by government projects. Due to the specificity of the external interventions, the program predicting the biogas treatment behavior did not proceed sequentially as in P2 above. Instead, this routine used the correlation coefficients between the background factors and the probability that households have a biogas digester to make predictions for the entire study area.

P4. Pollutant discharge load

The pollutant discharge load is calculated as follows:

$$\text{Pollutant discharge load} = (\text{treated pollutant}) \times (1 - \text{treatment efficiency}) + (\text{untreated pollutant}) \quad [9]$$

Equation [9] calculates the amount of pollutant load discharged to the environment based on the integrated interactions between agents and the external environment, including the physical environment and the social context (environmental management policy) in the study area. The final result from each agent is determined through the model run.

3.2.2 Land patches

The physical environment must be designed in such a way that it accommodates household agents (Russell and Norvig, 1995). Accordingly, each land patch in this model is defined with both dynamic and deterministic properties. The main properties of the land patch include:

- Administrative unit (name)
- Land use type
- Sub-basin (basin ID)
- Pig density (pig/ha)
- Pollutant load (kg/year)
- Pollutant load density (kg/ha/year)

The discharge load of each agent is calculated and updated for the land patches using the basic algebra map calculation function. The discharge loads on patches are then aggregated by sub-basin, to calculate the average and maximum discharge load on each sub-basin. The loads of sub-basins are the values that create the pressures on agents in the form of a feedback/response index.

3.2.3 Global parameters and interactions between agents

The effects of social and environmental contexts are conceptualized in ABM terms as global parameters (Wilensky, 1999). The influence of global parameters in this model is shown through the indirect pressure of environmental policy from the top level. This pressure will be received by the local policy implementing agency who then translates this pressure into direct effects on pig production households. The environmental protection commitment of agents is closely related to SN and PBC (Ajzen and Fishbein, 2005, Edgerton et al., 2009).

According to the survey results, the global parameters in the model are defined as technical training policies, financial support and other socio-economic factors existing at the study site. In fact, positive beliefs about environmental behavior can be enhanced if local leaders are able to influence an individual's responsibility by the role of the organization (Taylor and Todd, 1995b). Financial and technical support will directly affect the perception and perceived behavioral control of an agent. Thus, the social pressure has a simultaneous impact on the expected benefits and the accountability of the household agent. Of course, this factor only takes effect when the policy is attached with pressure from superiors in its implementation. In the absence of enforcement pressure, it is clear that households/farm owners will only discharge manure according to their spontaneous behavior. In this case, there is only local customary law or culture that governs the waste management.

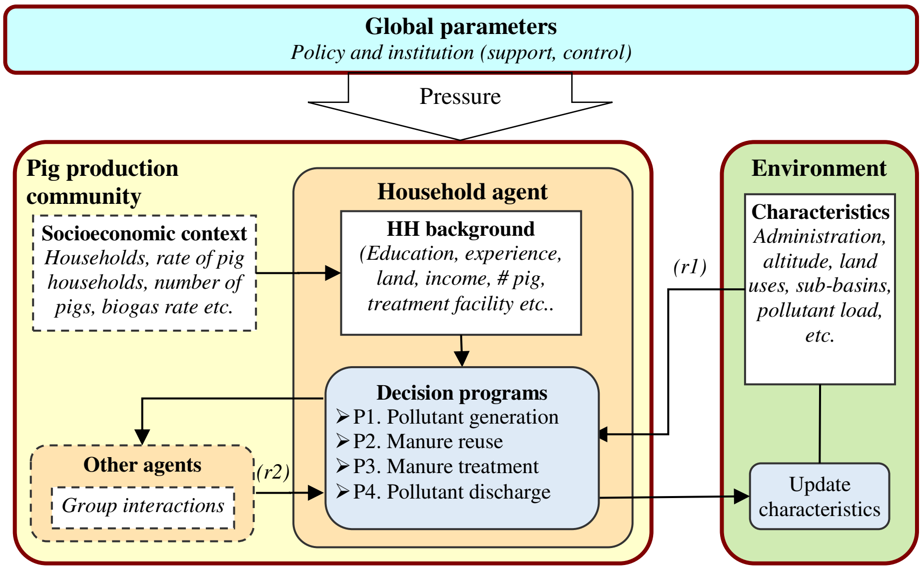

Fig. 4: The structure of the ABM for predicting pollutant load

The feedback mechanism of the ABM is designed to include the direct environmental impact due to pollutant discharges accumulated on patches (r_1), which in turn affects all agents. This feedback is expressed through the level of pollutant loads in the sub-basins which the agents in those sub-basins can receive. The second feedback (r_2) is expressed as an indirect influence through the social pressure that the surrounding community (other agents) put on the agent under a highly polluted situation. According to Ajzen and Fishbein (2005), these feedback flows reach the agents through the SN before they change their manure reuse behavior (B). The structure of the ABM is shown in Fig. 4.

3.3 Model simulation

The ABM predicting pollutant discharge operates iteratively on an annual cycle because the manure reuse and treatment (composting and biogas treatment) takes place over several months. Each simulation is for 1 year, consisting of 3 main phases as shown in Fig. 5: (1) initialization; (2) waste management decisions (generation, reuse, treatment and discharge); and (3) updating the state of the model.

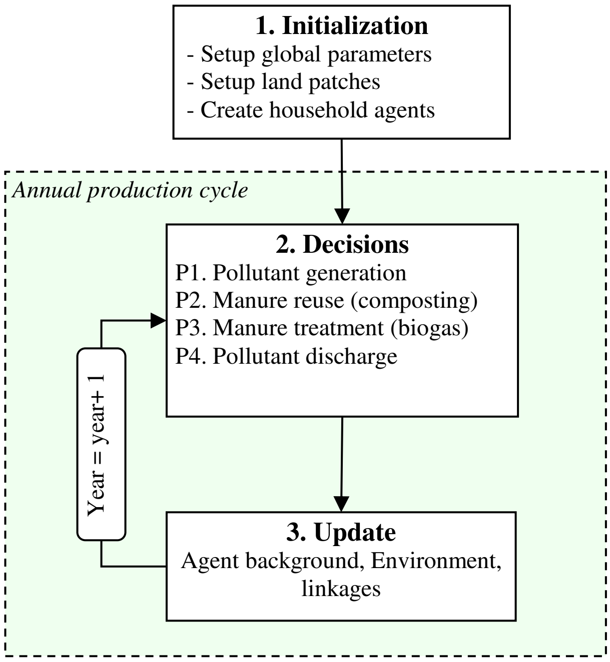

Fig. 5: A general simulation cycle of the ABM model for predicting pollutant load

The *initialization phase* is comprised of three main steps:

- Set the global parameters (policy, institution and socioeconomic context of the study area), which are assigned according to the values obtained through the survey in the study area, including basic parameters such as the percentage of households participating in the composting training program, the number of households having access loans to build a biogas digester, etc.
- Set up the physical environment by importing GIS data layers and attribute tables.
- Create the household agents, assign household characteristics, and locate households in settlement clusters.

The environment and its attributes are contained in administrative, land use and sub-basin maps (as shapefiles). The attributes from these three maps are assigned to the 'patches' properties using the GIS extension in the Netlogo software. The sub-basin map was delineated from DEM data (global SRTM 1 arc-second) by a separate procedure in the BASINS software (Fig. 6). First, the image was filtered for noise using the Fill-Sinks method (Wang and Liu, 2006) and to calculate the flow direction and flow accumulation of each cell using the Top-down Deterministic-8 method. The flow network with stream order plays an important role in determining the hierarchy of basins. In this study, we chose a limit of flow searching within 100 ha (corresponding to the average area of one village) to determine the sub-basins according to the method of Conrad (2015).

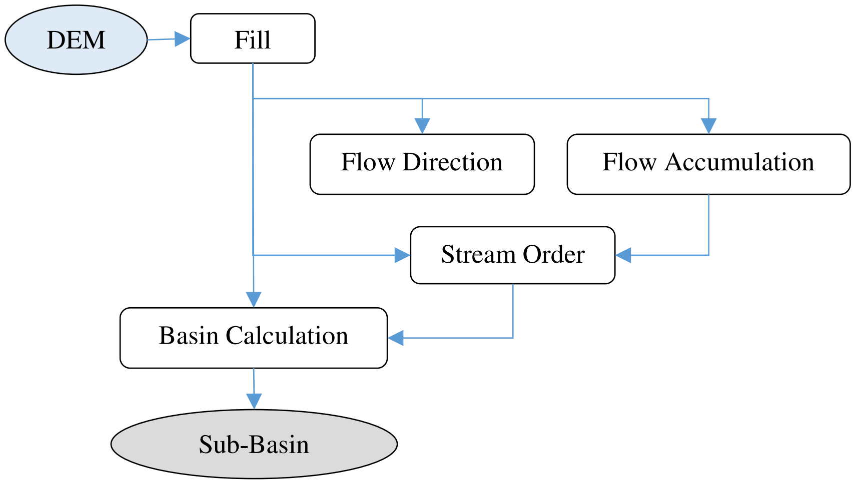

Fig. 6: Sub-basin delineation framework

Pig production household agents are created as shown in Fig. 7, where the number of agents is equal to the number of households in each administrative unit. Among the households created, only the surveyed households have exact geographical coordinates, which were imported directly into Netlogo while the remaining households were created and assigned attributes randomly based on descriptive statistics (i.e., mean, standard deviation) found in the study area.

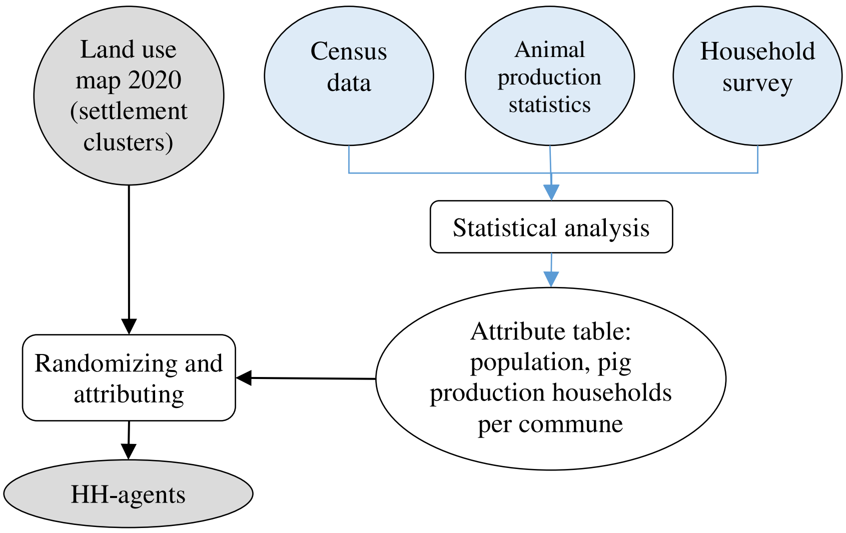

Fig. 7: The household agent creation framework

Waste management decision phase: This phase involves the operation of waste management programs. The steps in the simulation include the following:

- Calculate the pollutant load generated based on the number of pigs and the emission coefficients.

- Predict the manure reuse possibility using the correlation coefficients provided by the statistical analysis to predict AT, SN, PBC, BI and B from household background factors (Ngo et al., 2021).
- Calculate the treatment load based on the generated load, the manure reuse rate, and the status of biogas treatment.
- Calculate the pollutant discharge load based on the treatment load, treatment efficiency, reuse rate and non-treatment load.

Update phase: This phase includes the routines to update the agent properties such as manure reuse, the treatment and the discharge status during each run. Similarly, environmental properties are also regularly updated. The sequence of steps includes:

- Update household background factors.
- Update discharge load from HH-agent in each patch.
- Update pollutant load accumulated (mean, max) for the sub-basins.

A simulation cycle ends at the end of the update phase. However, the user of the ABM can specify for how long the model runs. Further model details are provided in the Supplementary Material, with an example of the model interface shown in Fig. S2.

3.4 Model validation

3.4.1 Sensitivity analysis

Sensitivity analysis is used to examine the impact of variables on the behavior or outputs of the model. As suggested in previous studies on ABM development (Ngo and See, 2012), this study applied the Sequential Bifurcation (SB) technique proposed by Bettonvil and Kleijnen (1997), which is essentially a method to determine the most important factors among those that affect the performance of the system. The relationship between the effects of factors on the model output is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_K x_K \qquad [10]$$

where y is the model output, $x_1,\ldots, x_j,\ldots, x_K$ are the standardized variables to be tested for sensitivity, β_j represents the main effects of corresponding variables and β_0 is the overall mean of the effects.

To conduct the sensitivity analysis, it is necessary to identify the dependent variable for observations (y), the factors (x_j) and the direction of the relationship, i.e., positive or negative. Then the x_{min} and x_{max} values must be determined. If the correlation between x_i and y is positive, x_{min} will be the lower bound of x_j (state 0); conversely, if the correlation is positive, then x_{max} will be the upper bound of x_j (state 1 or state K). The function y that operates with all factors x_{max} is y_K, and for all x_{min}, it is y_0. The contrast level between the two states is calculated by the following formula:

$$\left(y_{(K)} - y_{-(K)}\right) - \left(y_{(0)} - y_{-(0)}\right) \; or \; 2y_{(K)} - 2y_{(0)} \qquad [11]$$

where ($y_{-(j)}$) is the mirror observation of ($y_{(j)}$), which denotes the output y with the first j factors switched off (state 0) and the remaining factors switched on (state 1 or state K).

The SB is conducted in an iterative manner. First, the model is run with all factors set to low (state 0) and then all factors set to high to calculate y_K and y_0. If $y_0 < y_K$, then the sum of all the individual main effects is important and the second stage of SB is entered. Then, the *successive differences or contrasts* are calculated by splitting the factors into two subsets of equal size and the estimation process is continued for each subgroup, which is the same as that described in the first stage; the procedure continues in an iterative manner. SB terminates when the effect level (i.e., $y_j - y_0$) reaches the lower effect limit defined by the user. The lower effect limit was suggested as $\beta_j <$ 5% of $(2y_K - 2y_0)/N$ (Bettonvil and Kleijnen, 1997).

In this ABM, the model output is the pollutant discharge loads to the land patches. The sensitivity factors to be tested are the statistically significant background factors, which are used to predict B.

3.4.2 Relative Operating Characteristic (ROC) for validating the response of the agents

The ROC (Relative Operating Characteristic) is used to evaluate the performance of a classification or prediction scheme by identifying where instances fall in a certain class or group (Beck and Shultz, 1986). In this ABM, ROC is applied to validate the responses of the agents to environmental and social pressures (r_1 and r_2, presented in Section 3.2.3). It was hypothesized that the pressure from the surrounding environment (pollution loads in the sub-basin and waste recycling status of neighboring agents) determines the response of the agents. The response can change the load distribution pattern in the sub-basins when switching between scenarios (see section 3.6).

ROC validation is used to test the differences between the predicted and observed values related to the response mechanism. The comparison has four outcomes:

1. True positive (TP): both the prediction and the actual value are *P*.
2. True negative (TN): predicted value is *N* and the actual value is also *N*.
3. False positive (FP): prediction is *P* but the actual value is *N*.
4. False negative (FN): the predicted value is *N* while the actual value is *P*.

The four outcomes can be formulated in a two by two confusion matrix or contingency table as shown in Table 1 (Fawcett, 2004). Definitions of precision, accuracy and specificity are provided in this table.

Table 1: The confusion matrix to calculate the ROC

		True class	
		P	*N*
Predicted outcome	*P*	True Positives (*TP*)	False Positives (*FP*)
	N	False Negatives (*FN*)	True Negatives (*TN*)
Total		*P*	*N*

tp-rate = TP/N = sensitivity;
fp-rate = FP/N;
precision = TP/(TP + FP);
accuracy = TP+TN/(P+N);
specificity = TN/(FP+TN) = 1 – f-rate

The ROC evaluation is based on the ROC curve, which is a graphical representation of the relationship between the sensitivity or *tp-rate* and the specificity or *1 – fp-rate* of a test over all possible thresholds (Beck and Shultz, 1986). A ROC curve involves plotting the sensitivity on the y-axis and 1-specificity on the x-axis as shown in Figure S1 in the Supplementary Materials. This graphical ROC approach makes it relatively easy to grasp the interrelationships between the sensitivity and the specificity of a particular measurement. In addition, the area under the ROC curve provides a measure of the ability to correctly classify or predict those households having responses related to environmental and social pressures. The ROC area under the curve (AUC) would have a value of 1.0 for a perfect test, while the AUC would reduce to 0.5 if a test is no better than random (Fawcett, 2004).

3.4.3 Output (discharge) validation

Model output validation is performed on the basis of comparison between the predicted results on pollutant discharge load from the model and 125 wastewater samples collected at the end of effluents from households interviewed in the survey (Ngo et al., 2021). In addition, 30 wastewater samples from the front of the effluent, before discharge into the biogas digester, were analyzed. The sampling method was compliant with official standards (TCVN 5999:1995), and each site was sampled three times. The measured parameters consisted of Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD_5), Total Nitrogen (TN) and Total Phosphorous (TP). The results of the wastewater sample analysis were used to calculate the actual pollutant discharge load (i.e., the concentration multiplied by the volume of wastewater), for evaluating the efficiency of the waste treatment system, and to validate the results predicted by the ABM on the pollutant discharge load.

The validation was quantified using RMSE, which is one of the most commonly used error measures (Chatfield, 1992). This statistic measures the squared differences between the simulated or predicted values and the observed or reference values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i}-x_{2,i})^2}{n}} \quad [12]$$

where $x_{1,i} - x_{2,i}$ is the difference between variable i from data source 1 (i.e., the simulated result) and data source 2 (i.e., the reference or observed data) and n is the total number of variables, which in this study is 125.

3.5. Input data for the ABM

The input data to the ABM are listed in Table 2 along with the sources of information. Some of the variables on the household characteristics were taken from previous studies (Ajzen and Fishbein, 2005, Ittiravivongs, 2011, Scalco et al., 2017a, Supaporn et al., 2013, Scalco et al., 2018) while others were based on a household survey of 125 households located in five typical communes (25 households per commune) representing three typical pig farming densities (low, medium and high) (Ngo et al., 2021). The questionnaire included information on 16 background factors (Table S1 in the Supplementary Material) and 17 behavioral variables, where the mean value and standard deviation were used to assign properties to the agents during the random initialization phase.

Table 2: Input data to the ABM

Type	**Details**	**Source**
Pig production	Pig population and livestock farming household by commune, used to initialize the number of agents and pigs in each administrative unit	Yen Dung Department of Agriculture and Rural Development (2020)
Household characteristics	• Administrative unit • Age of farm owners • Education level • Pig farming experience (years) • Labor availability (person) • Income (income was classified into rich, medium, near-poor and poor level by the local classification scale) • Political and social membership (involvement of family member in social or political organizations) • Land area (ha) • Total number of pigs (sows, hogs) • Waste treatment infrastructure (biogas digester) (0,1) • Distance to the nearest neighboring farms (m) • The manure recycling status of the neighboring farms (1=yes) • The higher waste discharge rate compared to neighbors' (1=higher)) • Access to technical support (training) (0,1) • Access to financial support (0,1)	Informed by the literature (Ajzen and Fishbein, 2005, Ittiravivongs, 2011, Scalco et al., 2017a, Supaporn et al., 2013, Scalco et al., 2018) and from the household survey (Ngo et al., 2021)
Sub-basin map	Delineated from a global DEM at a 1 arc-second resolution	SRTM

Administrative map Land use map	Maps for 2020 to determine the location of pig farming households (in residential areas)	Bang Giang Department of Natural Resource and Environment (2020)
Pollutant emission coefficients	COD: 59.2; BOD_5: 32.9, TN: 7.3, TP: 2.3 (kg/head/year)	Decision No. 154/QD-TCMT (VEA, 2019)
Behavioral variables	17 variables to predict behavior derived from the household survey and statistical analysis	Ngo et al. (2021)

Regarding the behavioral variables, multiple regression and Structural Equation Modeling (SEM) were used to select the most significant variables and coefficient correlations for the ABM as suggested in previous studies (Chu and Chiu, 2003, Taylor and Todd, 1995b). In particular, the relation between the background factors and factors from the TPB were analyzed to select the determinants of PRB, SRB, PNB, SNB, SE and FC from the selected background factors that were statistically significant. The statistical analysis using SEM confirmed that the TPB approach was reliable for predicting the manure reuse behavior of the farm owners in Yen Dung district. Background factors, including farm characteristics (e.g., education, experience, income, number of pigs, land area) and the external context (e.g., financial and technical support, distance to the neighboring farms, pollutants of the sub-basin) were found to be significant determinants of the TPB factors, where the regression coefficients were then used as weights in the ABM. More details can be found in Ngo et al. (2021).

3.6 Scenario formulation

3.6.1 Potential developments in pig production in Bang Giang province

According to the Bac Giang department of Animal husbandry and Veterinary (2020), there were two development possibilities in pig production outlined in the report. The first is an optimistic scenario in which pig farms that are free from the current African cholera epidemic will actively repopulate. Thus, by the end of 2019, the number of pigs will increase, and by the end of 2021, the number of pigs in the district will reach that of two years before 2019. Moreover, by 2025, there will be a 10% increase in the number of pigs. The second is less optimistic and considers the situation in which the repopulation was implemented more cautiously, starting from the first quarter of 2020. If this was the case, then the number of pigs would be only two-thirds of that compared to 2018 by the end of 2020, and this number would only increase slightly until 2025. This latter situation would be considered a stable scenario in terms of the number of pigs.

3.6.2 Current environmental control policies related to pig manure

As discussed in the previous sections, existing environmental control policies related to pig production in the district have been mainly limited to controlling pig manure discharge to the environment. Only two policies were reported by respondents of the household survey, which were the financial and technical support (e.g., loans, training on biogas digester construction and operation, composting, awareness raising of environmental protection) provided by the district extension services. These activities have already been incorporated in the ABM as the accessibility of farmer agents to these support services, which were tested as determinants of manure reuse behavior. As reported by local government officials, future policy interventions would be realized as a change in the intensity of the support rather than on issuing new environmental control regulations. The officials suggested that the annual environmental support fund could fluctuate by 15%. Hence, in the scenarios formulated here, the assumption is that technical and financial support will increase by 15% compared to the current situation.

3.6.3 Scenarios for the ABM

Combining the information on future pig production (section 3.6.1) and environmental control policies (section 3.6.2), three scenarios have been formulated as outlined in Table 3. Scenario 1 is the baseline scenario using survey data as the main inputs, Scenario 2 is a positive scenario in which the number of pigs remains stable but there are increases in the supporting policies for environmental management, while Scenario 3 is a negative scenario in which the number of pigs increases but the current management measures are not increased.

Table 3: Scenarios for the ABM

Factors\Scenarios	**Scenario 1: Baseline**	**Scenario 2: Positive**	**Scenario 3: Negative**
Number of pigs	Assumes current number of pigs (on average 72.9 pigs, of which 11.4 sows per household)	Same as in scenario 1	Increased by 10%
Financial support	54.4% of households supported	Increased by 15%	Same as in scenario 1
Technical support	30.4% of households supported	Increased by 15%	Same as in scenario 1

4 Results and discussion

4.1 Sensitivity analysis of model input variables

4.1.1 Identification of experimental factors

An important output of the ABM model is the pollutant discharge load from pig farming. For simplification, BOD_5 was selected as representative of the four environmental parameters (COD, BOD_5, TN and TP) for the purpose of evaluation. The factors included in the sensitivity assessment were the global and background factors selected in the prediction model (Table 4). If the factors are sensitive to the model output, changes in pollutant loads will be observed when a range of factors are shifted, for example, from low (x_{min}) to high (x_{max}) values. Since the correlation between variables is well defined in the statistical analysis, the x_{min} (y_0 state) and x_{max} (y_1 state) are the lower and upper bounds of the variable, respectively.

Table 4: Parameters tested in the sensitivity analysis

No	Factor or Parameter	Low (y_0)	High (y_1)
1	Age of farm owners (years)	34	72
2	Education level (grades)	3	12
3	Labor availability (persons)	1	4
4	Social membership (0,1)	0	1
5	Household income (0 to 4)	1	4
6	Pig farming experiences (years)	2	47
7	Land area (m^2)	10800	150
8	Owning biogas digester (0,1)	1	0
9	Number of sows (heads)	0	30
10	Number of pigs (heads)	1	1520
11	Distance to the nearest farm (m)	500	10
12	The manure recycling status of the neighboring farms (0,1)	0	1
13	A higher waste discharge rate compared to neighbors (0,1)	0	1
14	Access to technical support (0,1)	0	1
15	Access to financial support (0,1)	0	1

Note: 3 variables have negative correlations: "land area", "biogas" and "distance to the nearest farm".

4.1.2 Estimation of the sensitivity coefficients or factor effects

The three stages in the estimation of the sensitivity coefficients are described below:

Stage 0: Equation [11] was used to calculate all x_{min} (y_0 state) and x_{max} values; an extremely low $y_{(0)}$ of 16.4 and an extremely high $y_{(k)}$ of 29738.3. The contrast between two states ($2y_{(K)} - 2y_{(0)}$) is (29738 - 16)*2 = 59444. This is the effect of all 15 factors ($\beta_{(1\text{-}15)}$) on the model output. According to Bettonvil and Kleijnen (1997), the contrast between variables must be greater than 5% of the average contrast; it means that the successive contrasts must be greater than 0.05*(59444/15) = 198.

Stage 1: As suggested by Bettonvil and Kleijnen (1997), this step starts with the separation of the 15 factors into groups for observation, namely $y_{(8)}$ and $y_{(15)}$. The $y_{(8)}$ group provides a $\beta_{(1\text{-}8)}$ of 10652 while the second group yields a $\beta_{(9\text{-}15)}$ of 48792. These results indicate that the factors in the second group have greater effects than the first group. However, the value of the first group is quite significant (18%) compared to the total effect. Thus, both groups require a further SB operation.

Stage 2: The SB process continues with all factors in the same manner as described above until the observed value of β matches the lower limit of 0.05 (198). This means that any factor with β < 198 is considered as having an insignificant effect on the model output. The overall SB process is illustrated in Fig. 8.

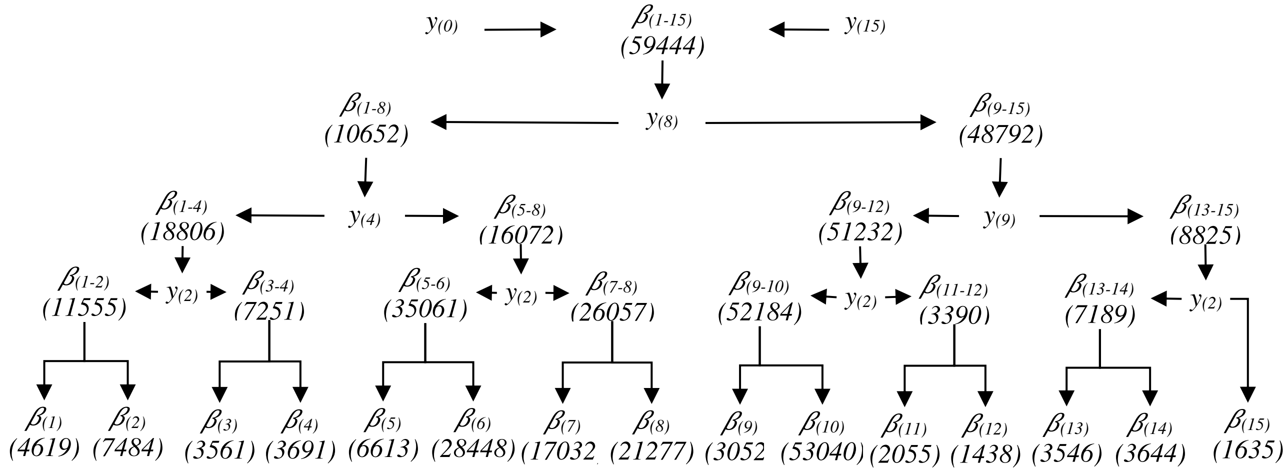

Fig. 8: An illustration of SB finding factors with significant effects in the SCM

Fig. 8 shows that at the end, the SB found all 15 factors as significant, where four factors: "number of pigs" (β_{10}); "pig farming experience" (β_6); "having biogas digester" (β_8) and "land area" (β_7) have quite strong effects compared to the others[1]. These results are technically reasonable in terms of model operation, which are embedded with deterministic regression functions. For example, the number of pigs is directly related to the pollutant load calculation. Therefore, even a small change in this factor value will lead to a large variation in the load of the study area.

4.2 Validation of model outputs

The result to be verified in the model is the amount of waste discharged from the breeding facilities (i.e., the agents) into the environment (after reuse and treatment). Waste load verification was performed by comparing the outputs of the model with the data obtained from the 125 interviewed households (Ngo et al., 2021). The model was run 30 times and the average RMSE value was calculated, which is shown in Table 5.

Table 5: Comparison between the outputs predicted from the ABM and the household survey data

No.	Parameters	Predicted results from the ABM (kg/year)		Root mean square error (RMSE) between predicted and household survey data		Ratio of RMSE and predicted results
		Mean$_{(load)}$	SD$_{(load)}$	Mean$_{(RMSE)}$	SD$_{(RMSE)}$	(%)
1	COD	207382	418(0.2%)	1304	28(2.1%)	0.6
2	BOD5	111860	282(0.3%)	747	13(1.7%)	0.7
3	TN	32809	98(0.3%)	109	6(5.5%)	0.3
4	TP	10591	40(0.4%)	33	2(6.1%)	0.3

[1] Please note that the sum of individual effects might not equal the group effect, which resulted from collective individual factor effects.

The results show that the operating model is relatively stable as the standard deviations of the RMSE error ($SD_{(RMSE)}$) and the standard deviation of the total forecast load ($SD_{(load)}$) of the agents are relatively low, ranging from 0.2-0.4% and 1.7-6.1%, respectively. In particular, the predictive ability of the model is very close to the actual survey data as the ratio between the RMSE and predicted results is only 0.3-0.7% (much smaller than the 5% threshold). Thus, the model can be used to evaluate different scenarios and assist in providing management solutions for the local district.

4.3 Results from the three scenarios

The ABM was run 30 times independently for each scenario and the average of the runs was calculated. The results for scenario 1 (baseline), which assumes the current situation for determining pollutant loads from pig farming, are provided in Fig. 9. The sub-basins having a high load (represented in dark colors) are concentrated in areas with high pig farm density such as Yen Lu, Tu Mai, Duc Giang, Dong Viet and Xuan Phu communes.

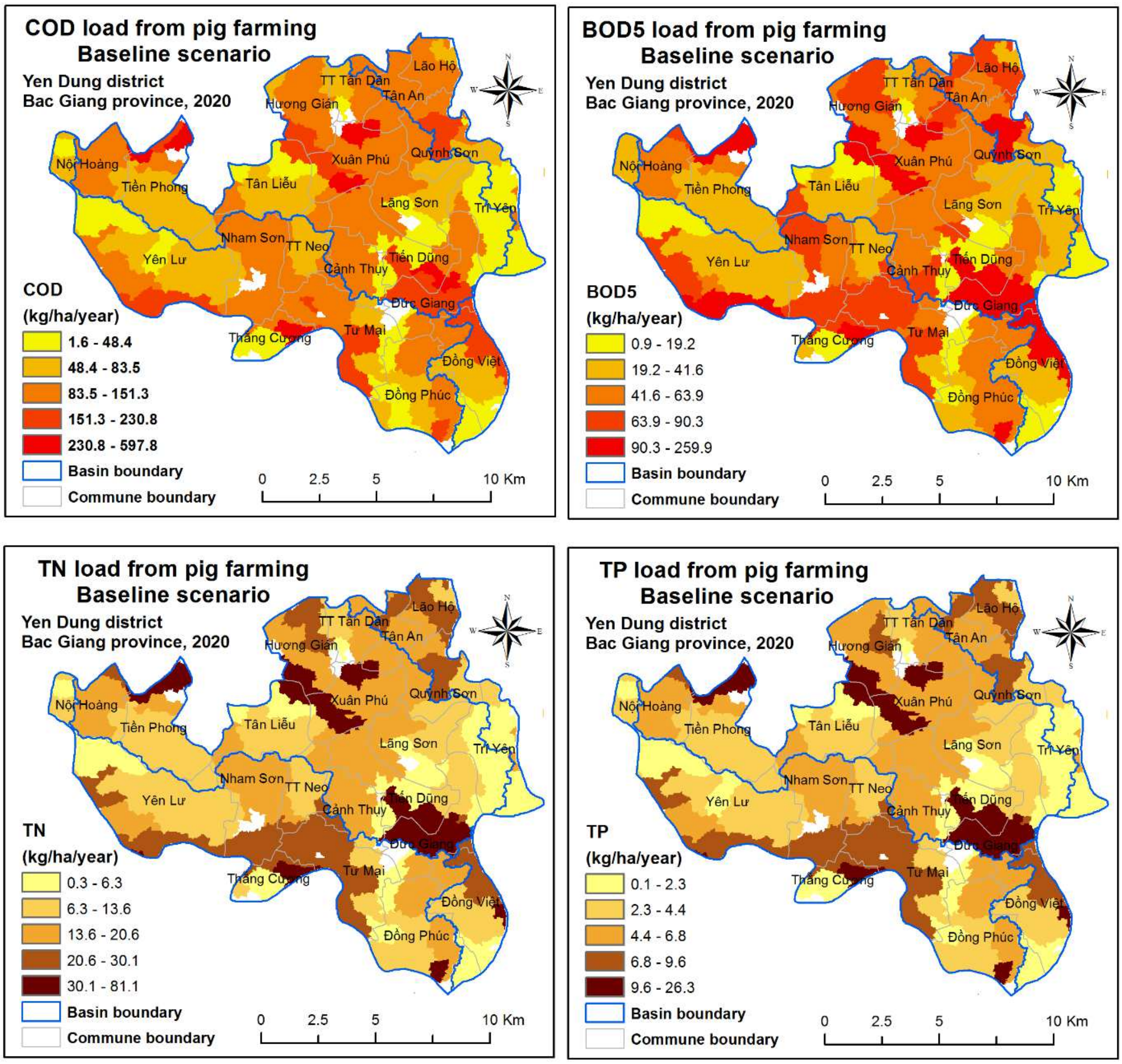

Fig. 9: Estimated pollutant load for the baseline scenario

The model outputs generated for the positive scenario (scenario 2) are presented in Fig. 10, where this scenario is based on the assumption that the determining factors in livestock waste management are improved. Regarding policies from the central government, this scenario assumes the highest pressure in fulfilling environmental protection commitments, which means that local governments are required to strictly implement central policies related to environmental protection. From Figure 10, one can see decreases in the pollutant load in some sub-basins as more attention is paid to waste control interventions such as environmental communications for better awareness and the implementation of environmental protection commitments.

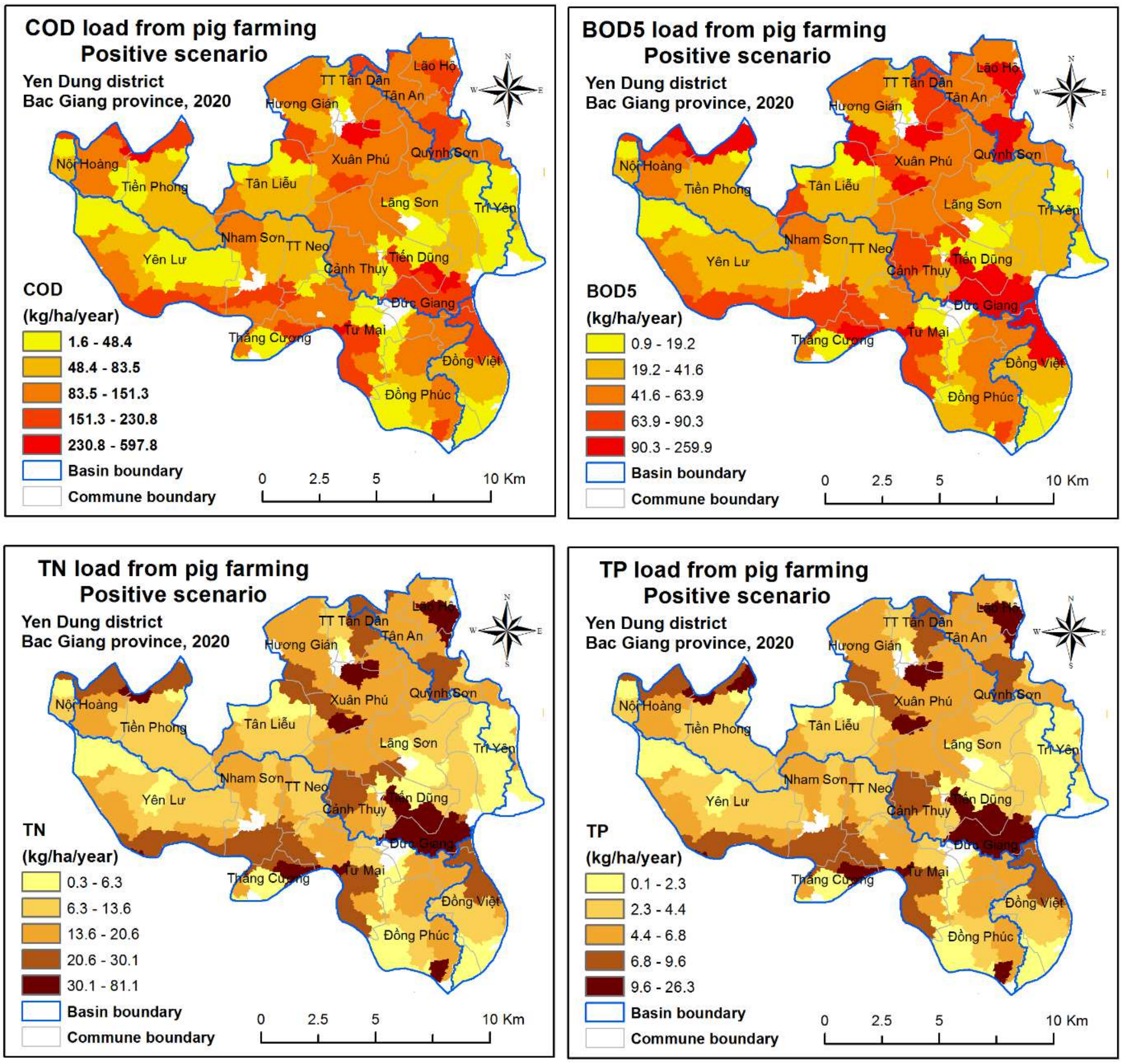

Fig. 10: Estimated pollutant load for the positive scenario

Scenario 3 (negative scenario) is the scenario with all negative impact factors, where the results are shown in Fig. 11. While the number of pigs is increasing, individual awareness and local government interventions are limited. Thus, this scenario generates a higher pollutant load than the previous scenarios.

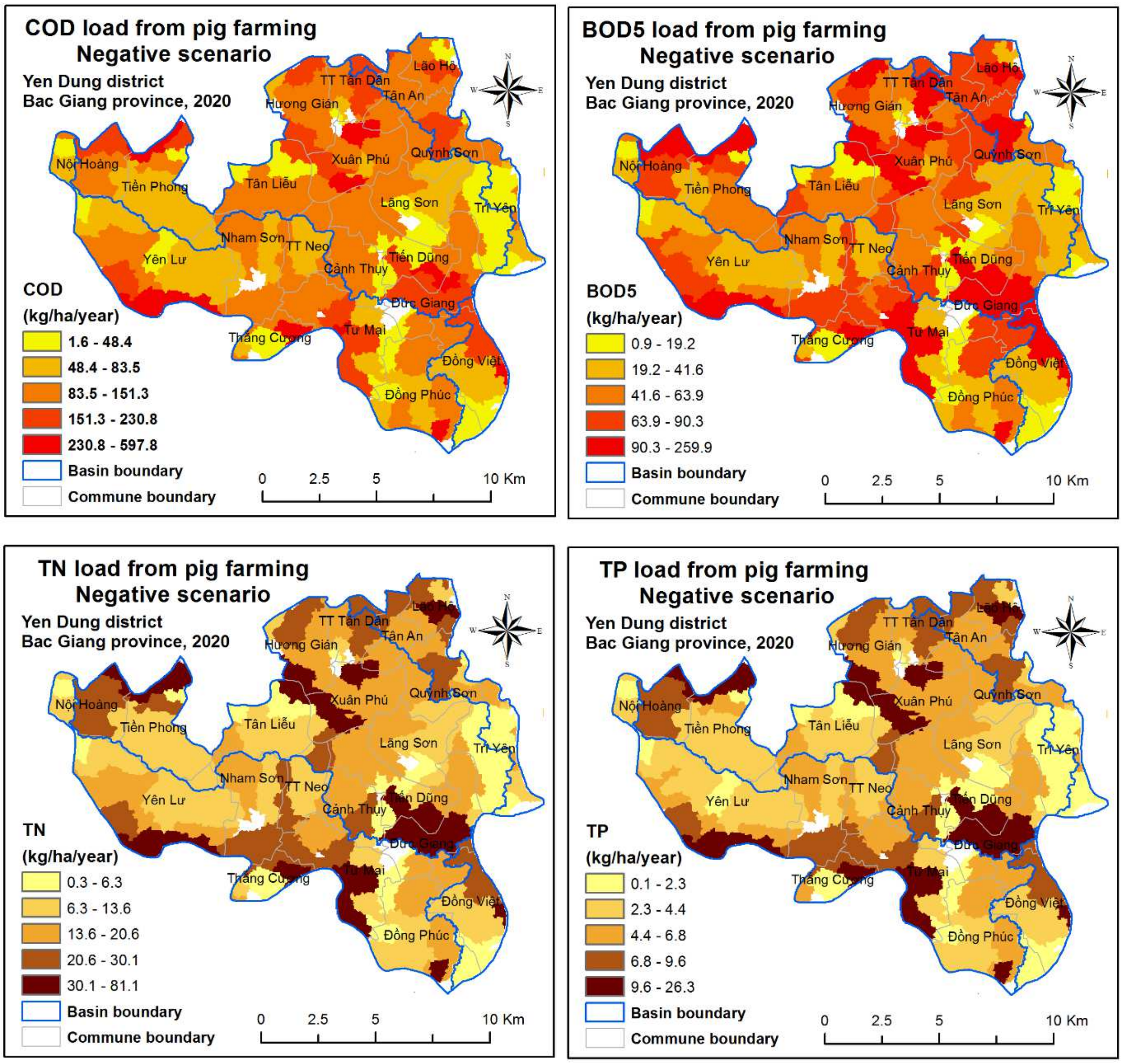

Fig. 11: Estimated pollutant load for the negative scenario

The pollutant loads of the four basic environmental parameters for each scenario are presented in Table 6. In the positive scenario, the environmental parameters (COD, BOD_5, TN and TP) are 95 to 98% of that in the baseline scenario while the negative scenario has a significantly higher pollutant load, i.e., 11-13% higher than that of the baseline scenario.

Table 6: Projected pollutant loads by scenarios

Pollutant loads by scenario (averaged over N=30 runs)	COD (tons/year)	BOD_5 (tons/year)	TN (tons/year)	TP (tons/year)
Baseline scenario	1626	872	263	85
Positive scenario	1591	850	251	83
Negative scenario	1831	985	291	94

Fig. 12 shows the spatial difference between the baseline scenario and the negative and positive scenarios, respectively, in terms of the most common pollutant (BOD_5). Sub-basins with a higher load than that of the baseline are represented in red, while those with a lower load than the baseline are shown in green. When comparing the negative scenario with the baseline (Fig. 12a), most sub-basins are red, which indicates that most have increased their loads of BOD_5 to the environment. However, there are a few sub-basins in Noi Hoang, Nham Son, Thang Cuong, Xuan Phu, Huong Giam and Tan An communes that have shown a decrease in pollutant loads, which are shaded in green. The sub-basins in Fig. 12b are almost entirely shaded green, which indicates that the positive scenario has contributed to the reduction of the discharged BOD_5 load in many farms. A small number of sub-basins are shaded red, indicating an increase in pollutant load.

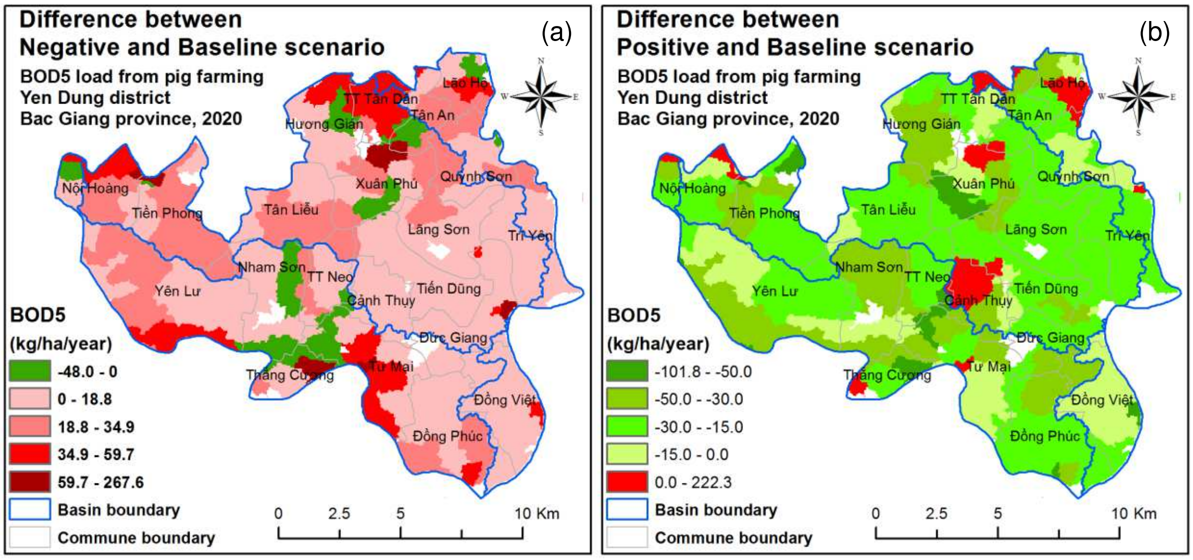

Fig. 12: The differences in BOD_5 load between (a) the baseline and the negative scenario and (b) the baseline and the positive scenario

In both scenarios, sub-basins with anomalous differences in the quantity of BOD_5 discharged can be explained by the different responses of the farm owners (i.e., agents) to the external environmental status. As explained in Section 3.2, the agents are designed with a set of rules that allows them to react autonomously in relation to other agents and to the external environment (i.e., the land patches). When neighboring agents reuse manure or there was an increase in the pollutant load in the environment, the agents will be under higher social pressure. Therefore, there is a possibility that an agent's reuse behavior could be higher in such circumstances. This is the reason why there is variability across sub-basins and illustrates the feedback mechanisms, r_1 and r_2, on the ABM outputs.

Further verification of the feedback r_1 and r_2 in the scenarios was performed based on the observations of two variables related to the number of household agents changing their manure recycling behavior due to social pressures (i) when the loads in the sub-basin (from neighboring agents) are higher or lower than the discharged level of the households; and (ii) when neighboring agents applied manure recycling measures. First, the model was run 30 times independently to examine the transitions between the baseline and the positive scenario (experiment-1), and then 30 times independently for transitions between the baseline and the negative scenario (experiment-2). During each run, the change in variable states was recorded. The variables (i) and (ii) are 0 or 1 (representing a no or yes state). A change in response occurs if the above variables change from 0 to 1 or from 1 to 0. Since variables (i) and (ii) are both sensitive to the model outputs (as discussed in section 4.1.2), the higher number of agents having switched states would result in larger differences in pollutant loads in the sub-basins between different scenarios.

Logically, the change in experiment-1 should result in changes from a higher load (in baseline) to a lower load (in positive scenario). However, an anomalous response by the agents will reverse this change, which means that the loads in the positive scenario will be higher than that in the baseline scenario. The change in experiment-2 is the opposite of experiment-1. This means that anomalous responses may cause the loads in the negative scenario to be lower than that of the baseline scenario. Through each run, the normal or anomalous variation (if any) of the loads in the sub-basins was recorded along with the ratio of the agents having switched their responses. An example of results recorded from these two experiments is presented in Table 7.

Table 7: Comparison between predicted and observed results regarding the behavioral response of agents by sub-basin

Sub-basin	From baseline scenario to negative scenario		From baseline scenario to positive scenario	
	Percentage of agents switching their responses (%)	Anomalous increase in pollutant load in the sub-basin (red color)	Percentage of agents switching their responses (%)	Anomalous decrease in pollutant load in the sub-basin (green color)
1	1	0	0	0
2	0	0	20	1
3	15	1	0	0
…	…	…	…	…
153	0	0	0	0

After running the model, the ROC area index was calculated and is shown in Table 8. The values range from 0.70 to 0.713, indicating that the model is able to predict much better than the random case (0.50). Therefore, the anomalous variations in the sub-basins can be explained and the overall scenario assessments are reliable.

Table 8: Area under the ROC curve (ROC area index) obtained from validating model results

Scenarios	ROC area index	Standard Error (SE)
From baseline scenario to negative scenario	0.700	0.020
From baseline scenario to negative scenario	0.713	0.021

4.4 Policy implications

The results of the scenario analysis show that to maintain the development of the pig herd while ensuring environmental protection, it is necessary to have financial and technical support for the pig producers. These investments directly affect the behavior of treating and discharging pollutants into the surrounding environment. The environmental effects related to the investment policies can be quantified based on the increase and decrease rate of the pollutant loads. For the positive scenario, this effect is the ratio of costs to achieve a reduction in the pollutants, which are the costs related to the financial and technical support, and the environmental benefits that the community gains due to the reduction in water treatment. For the negative scenario, the main effect is the offset of the profit obtained from increasing the number of pigs and the environmental costs incurred when handling a large additional pollutant load or the environmental damage caused by it. This result will be the basis for environmental managers to choose appropriate interventions according to the development goals of the locality.

In addition, the results related to the distribution of waste also strongly suggest that the waste tends to concentrate in low-lying areas and at the end of streams in the sub-basins of Duc Giang, Dong Viet and Tu Mai communes. The COD and BOD_5 loads in Duc Giang, Dong Viet and Tu Mai communes could reach to 598 and 260 kg/ha/year, respectively. Those areas with a low population density have many ponds, lakes and unused land. Therefore, a short-term solution that needs to be implemented specifically in the sub-basins in Duc Giang and Tu Mai communes is: (i) improving the treatment efficiency of the existing biogas system; (ii) reducing the waste load entering the biogas system; and (iii) building new biogas systems that are integrated with waste collected from biogas digesters.

The sensitivity analysis demonstrated that the number of pigs is the primary factor affecting the volume of pollutant discharge load. Other factors such as the experience and education level of the farm owner are the secondary factors that influence the manure reuse and treatment behavior. Therefore, in addition to technical measures, environmental communication to raise awareness is still an effective and necessary solution. The suggestion above is not new. However, with this model, we can quantify the effect and calculate the change in the pollutant discharge load. For example, suppose one goal for a particular phase is to control the load to a limited extent. By testing a few scenarios, it is possible to quantitatively optimize investment efficiency and the interventions needed to achieve this goal.

Realistically, the source of environmental pollution from livestock comes mainly from households and the load distribution is also in residential areas. Therefore, the solution in the long term is to establish concentrated livestock farms in the form of industrial farms, located in sub-basins with low pollutant emission potential but convenient for market access. The current model is fully scalable to solve this optimization problem in spatial planning as described above.

5. Conclusions

This paper presented an ABM to predict the pollutant load discharge from pig farming. Three basic components, namely, the household agent, the land patches and global parameters, were used to construct the model, where the behavior of the household agents was grounded in the TPB. The model was evaluated through a sensitivity analysis, validation using data collected from a household survey and ROC statistics. The sensitivity analysis showed that all background factors had a significant effect on the ABM outputs. The predictive ability of the model was shown to be good as the ratio between the RMSE and predicted results was less than 0.7%.

Based on the socioeconomic conditions at the study site, three integrated scenarios for pig production development and environmental control policy were then formulated. In addition to a baseline or business as usual scenario, a positive scenario was modeled, where the number of pigs was assumed to remain stable but supporting policies for environmental management were increased, and a more negative scenario was implemented, which assumed the number of pigs increases but management measures did not improve relative to the baseline. The results showed that the positive scenario contributes to a reduction in the discharged loads in many farms while in the negative scenario, most of the farms have increased loads discharged to the environment.

The results from the scenarios suggest that to maintain the development of pig production while ensuring environmental protection for the district, it is necessary to provide financial and technical support to the pig producers. These support mechanisms directly improve the manure treatment and reuse by farmers and thereby reduce the pollutants discharged into the surrounding environment. As the experience and education level of the farm owner were significant factors that influence the manure reuse and treatment behavior, environmental communication to raise awareness is still an effective and necessary solution in addition to the implementation of technical measures.

6. Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon request.

Competing interests

The authors declare that they have no competing interests.

Funding information

This research were funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.99-2018.318.

Authors contributions

Conceptualization, An The Ngo, Giang Huong Thi Nguyen, Linda See; Methodology, An The Ngo, Linda See; Software, An The Ngo; Validation, An The Ngo; Formal Analysis, An The Ngo, Giang Huong Thi Nguyen, Duong Huu Nong; Investigation, An The Ngo, Giang Huong Thi Nguyen, Duong Huu Nong; Resources, An The Ngo, Giang Huong Thi Nguyen, Duong Huu Nong; Data Curation, An The Ngo, Giang Huong Thi Nguyen, Duong Huu Nong; Writing-Original Draft Preparation, An The Ngo, Giang Huong Thi Nguyen, Duong Huu Nong, Linda See; Writing-Review & Editing, An The Ngo, Linda See; Visualization, An The Ngo, Linda See; Supervision, An The Ngo; Project Administration, An The Ngo; Funding Acquisition, An The Ngo. All authors have read and agreed to the published version of the manuscript.

7. References

AJZEN, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes,* 50, 179-211.

AJZEN, I. & FISHBEIN, M. 2005. The Influence of Attitudes on Behavior. *The handbook of attitudes.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

AN, N. T., LAN, N. P., CONG, V. H., DUONG, N. H. & HUONG GIANG, N. T. 2020. Environmental Pressure from Pig Farming to Surface Water Quality Management in Yen Dung District Bac Giang Province. *VNU Journal of Science: Earth and Environmental Sciences,* 36.

AN, N. T., GIANG, N.T.H., DUONG, N. H. & SEE, L. 2021. Understanding the behavour of pig farmers in waste management.

BAC GIANG DEPARTMENT OF ANIMAL HUSBANDRY AND VETERINARY 2020. Report on Animal production of the first 6 months of 2020 at Bac Giang province. Bac Giang province: Bac Giang department of Animal husbandry and Veterinary.

BECK, J. R. & SHULTZ, E. 1986. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med.,* 110, 13-20.

BERKES, F., COLDING, J. & FOLKE, C. 2000. Rediscovery of Traditional Ecological Knowledge as Adaptive Management. *Ecological Applications,* 10, 1251-1262.

BETTONVIL, B. & KLEIJNEN, J. P. C. 1997. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research,* 96, 180-194.

BONABEAU, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences,* 99, 7280.

CHATFIELD, C. 1992. A commentary on error measures. *International Journal of Forecasting,* 8, 100-102.

CHU, P.-Y. & CHIU, J.-F. Factors Influencing Household Waste Recycling Behavior: Test of an integrated Model. 2003.

CONRAD, O., BECHTEL, B., BOCK, M., DIETRICH, H., FISCHER, E., GERLITZ, L., WEHBERG, J., WICHMANN, V. & BÖHNER, J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development Discussions,* 8, 2271-2312.

EDGERTON, E., MCKECHNIE, J. & DUNLEAVY, K. 2009. Behavioral Determinants of Household Participation in a Home Composting Scheme. *Environment and Behavior,* 41, 151-169.

FAWCETT, T. 2004. ROC Graphs: Notes and Practical Considerations for Researchers. *HP Labs Tech Report.*

FENG, S. & HEERINK, N. 2008. Are farm households' land renting and migration decisions inter-related in rural China? *NJAS - Wageningen Journal of Life Sciences,* 55, 345-362.

HAPPE, K., HUTCHINGS, N. J., DALGAARD, T. & KELLERMAN, K. 2011. Modelling the interactions between regional farming structure, nitrogen losses and environmental regulation. *Agricultural Systems,* 104, 281-291.

ITTIRAVIVONGS, A. Factors Influence Household Solid WasteRecycling Behaviour In Thailand:An Integrated Perspective. 2011.

KARMAKAR, S., LAGUË, C., AGNEW, J. & LANDRY, H. 2007. Integrated decision support system (DSS) for manure management: A review and perspective. *Computers and Electronics in Agriculture,* 57, 190-201.

MACAL, C. M. & NORTH, M. J. Agent-based modelling and simulation: desktop ABMS. 2007 Winter Simulation Conference, 2007.

MACAL, C. M. & NORTH, M. J. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation,* 4, 151-162.

MENG, X., WEN, Z. & QIAN, Y. 2018. Multi-agent based simulation for household solid waste recycling behavior. *Resources, Conservation and Recycling,* 128, 535-545.

MONRE 2014. Rural environment: Country environment report 2014. Ministry of Natural resource and Environment (MONRE).

MUSTAFA-MSUKWA, A. K., MUTIMBA, J. K., MASANGANO, C. & EDRISS, A. K. 2011. An assessment of the adoption of compost manure by smallholder farmers in Balaka District, Malawi. *South African Journal of Agricultural Extension,* 39, 17-25.

NGO, T. A., DRAKE, F. & SEE, L. M. 2012. An agent-based modelling application of shifting cultivation. *In:* HEPPENSTALL, A. J., CROOKS, A. T., SEE, L. M. & BATTY, M. (eds.) *Agent-Based Models of Geographical Sytems.* Dordrecht, The Netherlands: Springer Verlag, .

NGO, T. A. & SEE, L. 2012. Calibration and Validation of Agent-Based Models of Land Cover Change. *In:* HEPPENSTALL, A. J., CROOKS, A. T., SEE, L. M. & BATTY, M. (eds.) *Agent-Based Models of Geographical Systems.* Dordrecht: Springer Netherlands.

NGUYEN, T. H. G., NGO T.A., LE T.T.H, YABE M., NGUYEN T.T., VU N.H. & T.S., C. 2021. Recycling Wastewater in Intensive Swine Farms: Selected Case Studies in Vietnam. *Fac. Agr., Kyushu Univ.*, 7.

NSIMBE, P., MENDOZA, H., WAFULA, S. T. & NDEJJO, R. 2018. Factors Associated with Composting of Solid Waste at Household Level in Masaka Municipality, Central Uganda. *Journal of Environmental and Public Health,* 2018, 1284234.

OSKAMP, S., HARRINGTON, M. J., EDWARDS, T. C., SHERWOOD, D. L., OKUDA, S. M. & SWANSON, D. C. 1991. Factors Influencing Household Recycling Behavior. *Environment and Behavior,* 23**,** 494-519.

PARKER, D. C., MANSON, S. M., JANSSEN, M. A., HOFFMANN, M. J. & DEADMAN, P. 2002. Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. *Association of American Geographers***,** 75pps.

RAILSBACK, S. F., LYTINEN, S. L. & JACKSON, S. K. 2006. Agent-based Simulation Platforms: Review and Development Recommendations. *SIMULATION,* 82**,** 609-623.

RUSSELL, S. & NORVIG, P. 1995. *Artificial Intelligence: A modern approach,* Englewood, New Jersey, Prentice Hall.

SCALCO, A., CESCHI, A., ITAD. SHIBOUB, RICCARDO. SARTORI, JEAN-MARC. FRAYRET & STEPHAN. DICKERT 2017a. The Implementation of the Theory of Planned Behavior in an Agent-Based Model for Waste Recycling: A Review and a Proposal. *In:* A. ALONSO-BETANZOS ET AL (ed.) *Agent-Based Modeling of Sustainable Behaviors. Understanding Complex Systems.* Springer International Publishing Switzerland.

SCALCO, A., CESCHI, A. & SARTORI, R. 2018. Application of Psychological Theories in Agent-Based Modeling: The Case of the Theory of Planned Behavior. *Nonlinear Dynamics Psychol Life Sci.,* 22**,** 15-23.

SCALCO, A., CESCHI, A., SHIBOUB, I., SARTORI, R., FRAYRET, J.-M. & DICKERT, S. 2017b. The Implementation of the Theory of Planned Behavior in an Agent-Based Model for Waste Recycling: A Review and a Proposal. *In:* ALONSO-BETANZOS, A., SÁNCHEZ-MAROÑO, N., FONTENLA-ROMERO, O., POLHILL, J. G., CRAIG, T., BAJO, J. & CORCHADO, J. M. (eds.) *Agent-Based Modeling of Sustainable Behaviors.* Cham: Springer International Publishing.

SI, H., SHI, J.-G., TANG, D., WEN, S., MIAO, W. & DUAN, K. 2019. Application of the Theory of Planned Behavior in Environmental Science: A Comprehensive Bibliometric Analysis. *International Journal of Environmental Research and Public Health,* 16**,** 2788.

SUPAPORN, P., KOBAYASHI, T. & SUPAWADEE, C. 2013. Factors affecting farmers' decisions on utilization of rice straw compost in Northeastern Thailand. *2013,* 114**,** 7.

TAYLOR, S. & TODD, P. 1995a. An integrated model of waste management behavior: A test of household recycling and composting intentions. *Environment and Behavior,* 27**,** 603-630.

TAYLOR, S. & TODD, P. 1995b. An Integrated Model of Waste Management Behavior:A Test of Household Recycling and Composting Intentions. *Environment and Behavior,* 27**,** 603-630.

TAYLOR, S. & TODD, P. 1997. Understanding the Determinants of Consumer Composting Behavior1. *Journal of Applied Social Psychology,* 27**,** 602-628.

TONGLET, M., PHILLIPS, P. S. & BATES, M. P. 2004. Determining the drivers for householder pro-environmental behaviour: waste minimisation compared to recycling. *Resources, Conservation and Recycling,* 42**,** 27-48.

VAN DER STRAETEN, B., BUYSSE, J., NOLTE, S., LAUWERS, L., CLAEYS, D. & VAN HUYLENBROECK, G. 2010. A multi-agent simulation model for spatial optimisation of manure allocation. *Journal of Environmental Planning and Management,* 53**,** 1011-1030.

VEA 2019. Decision number 154/QĐ-TCMT dated 15/2/2019 re. Issueing technical guideline on estimating Total maximal dialy load of rever water. Vietnam Environment Administration (VEA).

VU, T. K. V., TRAN, M. T. & DANG, T. T. S. 2007. A survey of manure management on pig farms in Northern Vietnam. *Livestock Science,* 112**,** 288-297.

WANG, L. & LIU, H. 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science,* 20**,** 193-213.

WILENSKY, U. 1999. *NetLogo* [Online]. Center for Connected Learning and Computer−Based Modeling. Available: http://ccl.northwestern.edu/netlogo/ [Accessed 2/11/2020.

WOOLDRIDGE, M., JENNINGS, N. R. & KINNY, D. 2000. The Gaia Methodology for Agent-Oriented Analysis and Design. *Autonomous Agents and Multi-Agent Systems,* 3**,** 285-312.

ZHENG, C., LIU, Y., BLUEMLING, B., CHEN, J. & MOL, A. P. J. 2013. Modeling the environmental behavior and performance of livestock farmers in China: An ABM approach. *Agricultural Systems,* 122**,** 60-72.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ABMmanuscriptSupplementaryMaterials.docx