

Accepted refereed manuscript of:

Falconer L, Telfer T & Ross L (2016) Investigation of a novel approach for aquaculture site selection, *Journal of Environmental Management*, 181, pp. 791-804.

DOI: [10.1016/j.jenvman.2016.07.018](https://doi.org/10.1016/j.jenvman.2016.07.018)

© 2016, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Investigation of a novel approach for aquaculture site selection

Lynne Falconer^{1*}, Trevor C. Telfer¹, Lindsay G. Ross¹

¹Institute of Aquaculture, University of Stirling, Stirling, FK9 4LA, United Kingdom

*Corresponding author: lynnefalconer1@gmail.com

Abstract

This study investigated the potential use of two “species distribution models” (SDMs), Mahalanobis Typicality and Maxent, for aquaculture site selection. SDMs are used in ecological studies to predict the spatial distribution of species based on analysis of conditions at locations of known presence or absence. Here the input points are aquaculture sites, rather than species occurrence, thus the models evaluate the parameters at the sites and identify similar areas across the rest of the study area. This is a novel approach that avoids the need for data reclassification and weighting which can be a source of conflict and uncertainty within the commonly used multi-criteria evaluation (MCE) technique. Using pangasius culture in the Mekong Delta, Vietnam, as a case study, Mahalanobis Typicality and Maxent SDMs were evaluated against two models developed using the MCE approach. Mahalanobis Typicality and Maxent assess suitability based on similarity to existing farms, while the MCE approach assesses suitability using optimal values for culture. Mahalanobis Typicality considers the variables to have equal importance whereas Maxent analyses the variables to determine those which influence the distribution of the input data. All of the models indicate there are suitable areas for culture along the two main channels of the Mekong River which are currently used to farm pangasius and also inland in the north and east of the study area. The results show the Mahalanobis Typicality model had more high scoring areas and greater overall similarity than Maxent to the MCE outputs, suggesting, for this case study, it was the most appropriate SDM for aquaculture site selection. With suitable input data, a combined SDM and MCE model would overcome limitations of the individual

approaches, allowing more robust planning and management decisions for aquaculture, other stakeholders and the environment.

Keywords: Aquaculture, GIS, Maxent, Mahalanobis Typicality, Multi-criteria evaluation, site selection

1.1. Introduction

One of the most important decisions for aquaculture is site selection as it provides the foundations not only for economic benefit, but also the sustainability, reputation and longevity of an individual farm and the industry as a whole. Site selection influences almost all aspects of aquaculture, including production and economic performance (Llorente and Luna, 2013), environmental impact (Wu, 1995), social acceptability (Katranidis et al., 2003) and the location may even have consequences for human health (Jang et al., 2006). As land and water are finite resources, space for aquaculture and competing industries is limited so it is vital that site selection is planned and managed appropriately. Unplanned development in the past has resulted in environmental, economic and social issues (Afroz and Alam, 2013; FAO, 2014; Suplicy et al., 2015) but even planned development can have negative consequences if insufficient information is available, plans are ill-defined and site allocation/selection is inappropriate for the species, system, community, other resource users and the environment. Decision support tools including spatial models are valuable sources of information when developing strategies and plans for development (Aguilar-Manjarrez et al., 2010; Ross et al., 2013). However, it is important to ensure the decision support tools and methodologies are both relevant and useful for the overall purpose and alternative approaches should also be considered alongside more established techniques.

Spatial modelling has been used to identify suitable sites for many different aquaculture species and systems throughout the world (Aguilar-Manjarrez et al., 2010; Ross et al., 2013) and one of the most common methods is the use of multi-criteria evaluation (MCE) (e.g.

Buitrago et al., 2005; Salam et al., 2005; Radiarta et al., 2008, Hossain et al., 2009; Ross et al., 2011). The MCE approach combines multiple variables in a structured model (e.g. temperature, depth, distance to markets etc) using a weighted overlay where the weights are proportional to importance (Nath et al., 2000). This is advantageous, as it allows assessment of the spatial variability of the biological, environmental, and socio-economic characteristics relevant to an aquaculture site, includes consideration of the different levels of importance amongst parameters and provides a qualitative and quantitative output which is useful and easy to understand for decision makers. However, development of such models requires knowledge not only of the species and systems but also their relationship with the relevant key parameters, so enabling reclassification to a common scale and assignment of weights. Model developers can employ their own experience and knowledge, values from literature and/or expert and stakeholder opinion within the process, but difficulties arise if there is disagreement or insufficient information to develop a robust model. Therefore in some cases an alternative approach which does not involve reclassification or weighting may be preferable.

“Species distribution models” (SDMs) are numerical tools which use observations of species occurrence or abundance together with environmental variables to predict probable species distribution across a study area (Elith and Leathwick, 2009). SDMs, habitat suitability models (HSMs), ecological niche models and bioclimatic envelope models all address similar issues and terminology can be confusing (Hirzel and Lay, 2008; Bradley et al., 2012). The term SDM is used in this study as it is one of the more popular terms. Generally, SDMs extrapolate species location data in space based on correlations of occurrence with selected environmental variables (Franklin, 2010). Their primary use is to explain or predict species distributions and the information provided can help with conservation planning, assist the understanding of evolution, predict climate change impacts and assess invasive species (Elith and Leathwick, 2009). However, recently, applications of SDMs have become more diverse and some studies have used them for other purposes, such as the estimation of the

monthly probability of wildfires (Peters et al. 2013) and to map landslide susceptibility (Felicísimo et al., 2013).

Furthermore, Evans *et al.* (2010) suggested that presence-only SDMs show potential as a method to assess the suitability of geographic regions for biofuel feedstock production. Using similar logic, SDMs could be used to assess the suitability of an area for aquaculture production. Replacing the species location data with farm location data would allow the model to assess conditions at those sites, extrapolating the analysis to the whole of the study area to identify further areas that have similar conditions to the input farms and are thus also suitable sites for aquaculture.

Unlike the MCE approach, the use of SDMs requires no prior information on how farmed species are influenced by the variables and there is no need to reclassify data or establish weightings. The data needed are the location of the existing farms and spatial layers of key variables which are thought to influence the location of those farms. Consequently, the use of an SDM could provide an alternative option for site selection assuming that the selected farms used within the SDM process are in suitable locations for aquaculture and that similar areas could also be made available for development. The aim of this study was to investigate the potential use of two SDMs, Mahalanobis Typicality and Maxent, for aquaculture site selection.

2.Methodology

2.1. Study area and farm locations

The study area (total land area approximately 46000km²) was located in the Mekong Delta in Southern Vietnam (Figure 1), which is the largest aquaculture production area in the country (Raux *et al.*, 2006). Rapid growth of the Vietnamese aquaculture industry has occurred in recent years and the culture of pangasius (*Pangasianodon hypophthalmus*) is the most

important sector in terms of production and value (FAO FishStat Plus, 2014). Pangasius farming occurs along two main branches of the Mekong River (Phan et al., 2009; De Silva and Phuong, 2011), however, the potential for expansion and development of new pangasius farms along the river is limited as there is competition with other users and land prices, which are expensive to begin with, continue to rise (Bosma et al. 2005; Phan et al. 2009). This area was selected as a suitable case study as alternative locations may be required for future aquaculture development.

Pangasius is an obligate air-breathing species that can tolerate high stocking densities (Phuong and Oanh, 2009) and environmental conditions that would otherwise be fatal to most aquaculture species (Belton et al., 2011). Consequently, it is farmed in highly intensive systems, with high production levels (De Silva and Phuong, 2011). Farms in the Mekong Delta typically consist of earthen ponds with an average water depth of 4 m and regular water exchange (Phan et al., 2009). As pangasius is a freshwater species, saline intrusion in the delta may impact production, particularly in the dry season. The dry season in the Mekong Delta lasts from December to April, with a rainy season from May to November (Sakamoto *et al.*, 2009).

The locations of 192 pangasius farms (Figure 1) were obtained from a survey conducted as part of the EU FP7 Sustainable Ethical Aquaculture Trade (SEAT) project (Little et al., 2009). SDMs require input points which normally represent the presence or absence of the studied species. In this study the input points represent the presence of an existing pangasius farm as the models will use this information to identify further similar areas based on the conditions at these locations. The surveyed farms were considered appropriate for use in the models as they had successfully produced at least one crop of pangasius and they were all located in the main area used for pangasius culture in Vietnam.

2.3. Models

In order to assess the potential use of SDMs as an alternative approach to MCE for aquaculture site selection, four models were compared; two SDMs and two MCE models. As there is almost always more than one way to construct an MCE model, two models were developed using different reclassification methods; a User-defined MCE model and a Fuzzy MCE model. Mahalanobis Typicality and Maxent were selected as two SDMs with a contrasting approach. The former considers the variables to be of equal importance, while the latter assesses the variables to identify the most important variables which explain the spatial distribution of the input points.

The main differences in required user/developer input for the MCE models and the two SDM models are highlighted in Figure 2. Both MCE and SDM approaches require user input for data identification, collection and processing into layers for use in the models. The MCE approach also requires the user to design the model structure, reclassify the data and weight layers and submodels. With regard to SDMs, Mahalanobis Typicality needs no user input beyond the development of data layers, while Maxent has settings which must be adjusted depending on the scope of the work.

Further details on each modelling approach are found in their respective sections below.

IDRISI Selva [Clarks Lab, Massachusetts, USA] was used as a modelling environment and each data layer was processed to have a spatial resolution of 30m and georeferenced using the UTM reference system (UTM-48N).

2.3.1. Variables

Variables were selected after visits to the study area and discussions with aquaculture experts with experience of Vietnamese pangasius production to identify key parameters which are important for a pangasius site (Table 1). The parameters were selected for initial

site identification and assessment of the physical features of the site, further analysis at a more local scale would be needed to assess other parameters and ecological and production carrying capacity (Ross et al., 2013). Although both categorical and continuous variables can be analysed by Maxent and can be integrated within an MCE, only continuous variables can be used within the Mahalanobis Typicality approach, therefore categorical variables, such as land use, were not included. Across the study area temperatures are constantly within optimal ranges for pangasius (22 to 30°C) throughout the year and there is little spatial variation, therefore temperature was also excluded from this study although it would be important in other locations. Rainfall was not considered as a potential water source as pangasius farming systems require a lot of water and generally use the main Mekong River or smaller rivers and canals as primary inputs (Phan *et al.* 2009).

2.3.2. Multi Criteria Evaluation (MCE) models

After discussion with aquaculture experts and stakeholders, the layers were arranged in four submodels; salinity (salinity dry season and salinity rainy season), land (slope and soil), water (waterbodies and population) and access (urban areas and roads) (Figure 3). The first stage of the MCE models involved data reclassification to ensure the parameters were on a common scale. There are numerous methods of reclassifying data and a fundamental part of model development is deciding the most appropriate reclassification scheme to use. For the purpose of this study, two methods which have been used in previous aquaculture site selection studies were selected; user-defined hard classification and fuzzy classification.

Both reclassification methods have advantages and disadvantages in their use. User-defined hard classification, where the user specifies the values within the classes (Eastman, 2012), is one of the most commonly used reclassification schemes in aquaculture site selection studies (Giap et al., 2005; Salam et al., 2005; Hossain et al. 2009; Ross et al., 2011). This method is beneficial as it provides a range of suitability classes which are easily interpreted

by decision makers. However, for some parameters (e.g. temperature) there can be uncertainty associated with distinct hard boundaries (Eastman, 2012), thus some studies have used fuzzy classification (Falconer et al., 2013) where there are no hard boundaries between the classes and the transition is gradual between membership and non membership (Zadeh, 1965; Eastman, 2012). Nevertheless, in the absence of distinct categories the results may be more difficult to interpret.

Suitable reclassification values for the two methods were developed from the literature; user-defined (Table 2) and fuzzy (Table 3). The user-defined reclassification adopted a five-point categorical scale (1 = "highly unsuitable", 2 = "unsuitable", 3 = "intermediate" (neither suitable nor unsuitable and requires further investigation), 4 = "suitable" and 5 = "highly suitable") while the fuzzy reclassification used a continuous scale from 0 to 1 (not suitable to highly suitable). All parameters within the fuzzy reclassification were considered to have a sigmoidal ("s-shaped") fuzzy function within the IDRISI fuzzy module (Eastman, 2012). Weights (Figure 3) were established after discussions with farmers and other stakeholders during visits to the region and consultation with a panel of six aquaculture experts with knowledge of pangasius farming in Vietnam. As the final site suitability model included the combination of four submodels, the Analytical Hierarchy Process (AHP) was used to calculate weights for the MCE (Saaty, 1977; Eastman, 2012).

2.3.3. Species distribution models (SDMs)

There are two categories of SDMs, presence/absence and presence-only. Presence/absence SDMs require absence data which infers the location is unsuitable for a species; however, in the case of aquaculture, absence does not necessarily mean the area is unsuitable; therefore presence/absence models were considered inappropriate here. Two presence-only approaches were used. The first was the Mahalanobis Typicality modelling approach, derived from Mahalanobis distance, where the results express the degree to

which the values of a set of variables at a location are typical of the conditions found in training data (Sangermano and Eastman, 2007; Eastman, 2012). The second approach involved the use of Maxent; a software which uses the principle of maximum entropy to estimate suitability (Phillips *et al.*, 2006). The principle of maximum entropy aims to find a marginal suitability function for each variable that matches the input data, is maximally uninformative elsewhere and has a mean equal to that of the input data (Warren and Seifert, 2011). Mahalanobis Typicality assumes that all variables are of equal importance (Hernandez *et al.*, 2006) whereas Maxent is a general-purpose machine learning method which analyses the variables to assess which contribute the most to the distribution of the input data (Phillips *et al.* 2006).

Both methods have their strengths and weaknesses as SDMs. The advantages of the Mahalanobis Typicality approach are that it can be executed rapidly and requires minimum computation (Sangermano and Eastman, 2007). On the other hand the variables must be continuous and they are all assumed to be of equal importance (Eastman, 2012) which may be incorrect as some variables may be more important in terms of distribution. By contrast, Maxent is a popular tool for ecologists (Phillips *et al.*, 2006; Phillips and Dudik, 2008; Elith *et al.*, 2011) as the software is easy to use (Merow *et al.*, 2013), it can include both continuous and categorical data (Phillips *et al.*, 2006) and often outperforms other predictive methods (Elith *et al.* 2006; Merow *et al.*, 2013). However, the software is flexible with multiple different settings and although this can be advantageous, many authors (including Phillips *et al.* 2006; Phillips and Dudik, 2008; Baldwin, 2009; Elith *et al.* 2011; Warren and Seifert, 2011; Merow *et al.* 2013) have highlighted the need for more research and the development of guidelines regarding the use and settings of Maxent. Although there are many different SDMs available which could have been used in this study, Mahalanobis Typicality and Maxent allowed a comparison between an SDM which considered equal importance of variables and another which assessed the importance of variables in explaining the spatial distribution of farms.

The Mahalanobis Typicality modelling approach within IDRISI Selva's Land Change Modeler (see Sangermano and Eastman, 2007; Eastman, 2012 for details) was used. The model was run ten times and then averaged to produce the final output. For each of the ten runs a random selection of 75% of the input data (144 pangasius farms) was used to train the model and the remaining 25% (48 pangasius farms) was used to test the model. Maxent 3.3.3K [www.cs.princeton.edu/~schapire/maxent/] was used (see Phillips et al., 2006; Phillips and Dudik, 2008; Elith *et al.* 2011 for details). Again, 75% of the input data points (pangasius farms) were used for training and the remaining 25% were used for testing through 10 subsampled replicate runs with a limit of 5000 iterations. The final output was averaged after the ten runs.

Unlike Mahalanobis Typicality, Maxent does not assume equal importance of the variables and throughout the analysis it records which variables are contributing to fitting the model (Phillips, 2006). One of the methods which assesses variable importance is percent contribution; an estimate of the relative contribution of the variable to the model (Phillips, 2006). The Jackknife test provides an alternative estimate of variable importance. The Jackknife approach excludes one variable at a time as the model runs, which provides information on variable performance, particularly, how important each variable is at explaining the distribution of points (farms) and how much unique information each variable provides (Baldwin, 2009).

Both Maxent and Mahalanobis Typicality were tested using the area under the receiver operating characteristic curve (AUC) (Fielding and Bell, 1997). AUC is on a linear scale where values closer to 1 indicate better results and values of 0.5 or less represent a model with random predictions (Baldwin, 2009). Swets (1988), cited in Araújo et al. 2005, recommended that AUC values are interpreted as: $AUC > 0.90$ is excellent, $0.80 > AUC < 0.90$ is good, $0.70 > AUC < 0.80$ is fair, $0.60 > AUC < 0.70$ is poor and $0.50 > AUC < 0.60$ is a fail.

262

263 **2.4. Comparison of MCE and SDM approaches**

264 To allow comparison between the approaches, the results were reclassified to a suitability
265 scale. A five-point categorical scale was considered appropriate, thus there was no need to
266 reclassify the results of the User-defined MCE model. The output from the Fuzzy MCE
267 model used a continuous range of suitability scores from 0 to 1 and was reclassified to the
268 five-point categorical scale. Scores of 0 to 0.2 were considered “highly unsuitable”, 0.2 to 0.4
269 were “unsuitable”, 0.4 to 0.6 were “intermediate” (neither suitable nor unsuitable and
270 requires further investigation), 0.6 to 0.8 were “suitable” and 0.8 to 1.0 were “highly suitable”.

271 The outputs of the Mahalanobis Typicality and the Maxent models are represented by
272 typicality and Maxent scores, respectively. As noted by Eastman (2012), caution must be
273 applied when interpreting low typicalities (and scores) as although a low typicality may be
274 unusual it could still be within a species range. This makes establishing thresholds difficult,
275 something which is further complicated here as the typicalities and Maxent scores represent
276 similarity to existing aquaculture sites, and thus suitability for culture, rather than species
277 range. To account for this, the “unsuitable” and “highly unsuitable” categories were not used
278 for the SDMs, so scores of 0.001 to 0.6 were considered “intermediate”. The “intermediate”
279 category is neither suitable nor unsuitable for aquaculture and such areas would require
280 further investigation to determine suitability. The “intermediate” category is considered
281 appropriate for use here rather than the “unsuitable” and “highly unsuitable” categories as,
282 unlike the MCE models, low scores for SDMs do not necessarily indicate the area is
283 unsuitable for pangasius culture, it only suggests the conditions (of the selected criteria) are
284 less similar than at input farms. Scores between 0.6 to 0.8 “suitable” and scores greater than
285 0.8 were “highly suitable”. Scores below 0.001 were considered “not suitable” as these areas
286 are not similar to the input farms.

This allowed a visual and qualitative comparison of the results from the User-defined MCE model and the reclassified Fuzzy MCE model, reclassified Maxent and reclassified Mahalanobis Typicality models. A quantitative comparison was also carried out by calculating the area covered by each suitability category and evaluating the locations of the farms (Figure 1) against the four outputs.

3. Results

3.1. MCE models

3.1.1. User-defined MCE

The results of the User-defined MCE model are shown in Figure 4A. "Unsuitable" and "highly unsuitable" areas (approximately 13500km² and 100km² respectively) are mainly found in the south and east of the study area. Almost 25% (~11300km²) of the area is considered "suitable", while a further 4% (~1900km²) is "highly suitable" for pangasius culture. The "suitable" and "highly suitable" areas are both found inland along the main Mekong river and other channels in the north and west of the study area.

3.1.2. Fuzzy MCE

The results of the Fuzzy MCE model are shown in Figure 4B. Over 46% of the study area (approximately 21600km²) has a score above 0.5 and 9% of the study area (approximately 4100km²) has a score greater than 0.8. Similar to the results of the User-defined MCE model (Figure 4A), most of the high scoring areas are found along the Mekong River and other rivers and channels in the north of the study area.

3.2. SDMs

3.2.1. Mahalanobis Typicality

The results of the Mahalanobis Typicality model are shown in Figure 4C. Each run of the 10 Mahalanobis Typicality model runs had an AUC above 0.9. The output of Mahalanobis Typicality is in the form of typicality probabilities and how typical each pixel is to the training data. The highest achievable value is 1, which would indicate the pixel is identical to the mean variable conditions of the input training data (Eastman, 2012). The model results show approximately 11500km² of the study area has a typicality of 0.001 or above.

The areas with the highest typicality, and therefore most similar to the mean conditions of the input farms, are located along the two main branches of the Mekong, where most pangasius culture already occurs, and in the north of the study area. Over 2200km² has typicality greater than 0.5 and approximately 900km² has a typicality of more than 0.8. The Mahalanobis Typicality model also indicates approximately 75% (35000km²) of the study area has a typicality below 0.001 suggesting these areas are not similar to the mean conditions of the input farms. These areas are predominately located away from the main channels of the Mekong, extending across most of the south of the study area.

3.2.2. Maxent

The Maxent output has a probabilistic interpretation where there is a smooth gradient from least to most suitable conditions (Phillips, 2006) and in this study represents the suitability of the area for further pangasius farms based on the conditions defined by the input datasets (Figure 4D). Each run of the 10 Maxent model runs had an AUC above 0.9. Over half the study area (almost 25000km²) has a score of 0.001 or above, most of these areas are found inland along the Mekong and in the north of the study area. However, low scores dominate as only 700km² (<2% of the study area) has a score greater than 0.5 and less than 40km²

($<0.1\%$) has a score more than 0.8. These areas are found along the main Mekong river and adjacent channels, primarily near where the input farms (Figure 1) were located.

The two most important variables with regard to percent contribution were the clay content of soil and distance to waterbodies (Table 4). With regard to the Jackknife test (Figure 5), the variable with the highest gain when used in isolation was the clay content of soil which means it was the variable with the most useful information when used alone. The variable which decreased the gain the most when omitted was distance to waterbodies suggesting it had the most information that was not present in other variables. As in the percentage contribution analysis (Table 4), this again indicated that the two most important variables to model development according to the Jackknife test were the clay content of soil and distance to waterbodies.

Response curves generated from Maxent (Figure 6) can be used to evaluate how the model responds to the variables, which in turn, can be used to help determine model performance. Figures 6F and 6H show how the model responded to the two most important variables in model development; clay content of soil and distance to waterbodies. As the distance to waterbodies increases there is also a sharp decrease in probability of farm presence (Figure 6H) which would be expected as pangasius farming practices often require a lot of water. Likewise, as salinity increases (in both seasons) the predicted probability of presence (*ergo* suitability for farms) decreases (Figures 6C, 6D) as pangasius is usually grown in freshwater conditions.

3.3. Comparison of the MCE and SDM approaches

The visual comparison of the reclassified model outputs is shown in Figure 7. Tables 5 and 6 contain information on the area covered by each suitability category and the number of farms located within each suitability category respectively. For both MCE models, "suitable"

and "highly suitable" areas are mainly found inland, along the Mekong River and in the north east of the study area (Figure 7). Over 40% of the study area has a "intermediate" suitability score from the User-defined MCE model, which is greater than the 27% classed as "intermediate" by the reclassified Fuzzy MCE model (Table 5). The latter is a consequence of more areas being classed as "highly unsuitable" or "highly suitable".

Most of the study area (75%) is considered "not suitable" by the reclassified Mahalanobis Typicality model, while just less than half (46%) is considered "not suitable" by the reclassified Maxent model (Table 5). However, although the output of the reclassified Maxent model covers a greater extent of the study area, the reclassified Mahalanobis Typicality has more "suitable" and "highly suitable" scoring areas with 900km² for each category compared to 400km² ("suitable") and less than 100km² ("highly suitable") for the reclassified Maxent model. This is also highlighted in Figures 7C and 7D, which show the reclassified Mahalanobis Typicality model (Figure 7C) has more "suitable" and "highly suitable" areas along the Mekong river than the reclassified Maxent model (Figure 7D); these areas would be expected to have high scores as the input farms were located here (Figure 1). Some areas in the north of the study area, near the large waterbody and close to the Cambodian border are classed as "suitable"/"highly suitable" by the reclassified Mahalanobis Typicality model.

When farm location is evaluated against the model results almost all farms are found in "intermediate", "suitable" or "highly suitable" areas (Table 6). Only one farm (<1% of all farms) is found in an "unsuitable" area with regard to the reclassified User-defined MCE model, while 4 farms (<3%) are found in "unsuitable" or "highly unsuitable" areas for the reclassified Fuzzy MCE model. Likewise, only one farm (<1% of all farms) is found in a "not suitable" area when compared to the reclassified Mahalanobis Typicality model results and no farms are found in "not suitable" areas for the reclassified Maxent model. This suggests that all four models are fit for purpose as site selection tools. Nonetheless, it must be noted that the SDMs used the farms as part of the modelling process whereas the MCE models did

not. Furthermore, the thresholds and suitability categories are more appropriate for the MCE models as 75% of farms are found in "suitable" or "highly suitable" locations for both models, while only 47% and 57% of farms are located in "suitable" or "highly suitable" areas for the reclassified Mahalanobis Typicality and reclassified Maxent model respectively.

4. Discussion

This study investigated the potential use of two SDMs, Mahalanobis Typicality and Maxent, for aquaculture site selection. The SDMs were compared to two models which were developed using the popular MCE approach. It is clear there are advantages and disadvantages to each method. The key advantages of the MCE approach are that it is a well established technique for aquaculture site selection, can combine multiple biological, environmental and social parameters within one framework and models can be developed to consider different priorities and scenarios, providing extra support for decision makers and stakeholders. However, the disadvantages include the requirement for a high amount of user input during development of the MCE models and the uncertainty that may arise over the many different ways to construct a logical model. Conversely, a strength of the SDM approach is that it requires minimal user input compared to the MCE approach (Figure 2), and there is no need for the user to reclassify or weight the data. This removes some of the uncertainty in the modelling process, however one of the main disadvantages is the difficulty in interpreting the output of SDMs compared to the output of the MCE models, which makes establishing thresholds for decision makers a challenge. The main difference between the approaches is that MCE models are developed specifically to determine suitability, whereas SDMs assess the similarity of an area to input farms and suitability is implied from this, based on the assumption that input farms are located in suitable areas.

For all four models, as with any modelling approach, selection of variables is a key step, and use of alternative variables may result in different model outcomes. Within the MCE

approach there are further decisions which will affect the results; data reclassification, weighting and overall model structure (Figure 2). Although based on scientific knowledge and experience, most reclassification and weighting values will still rely on an individual or group judgement as universally accepted scores and weights rarely exist. Expert opinion can vary substantially (Nath et al., 2000) and alternative decisions may generate different results with no single definitive “correct” answer (Carver, 2008). Although this can be an advantage as it allows users to analyse different scenarios and evaluate alternative stakeholder opinions (Carver, 2008) it can also add to uncertainty and confusion for decision makers. As shown here, even changing the reclassification method can impact the results. Initially, the outputs of the MCE models appear to have similar results (Figures 4A, 4B), however, once the fuzzy results have been reclassified further to a categorical suitability scale it is clear there are differences (Figure 7, Table 5). The reclassified Fuzzy MCE model has more “highly suitable” and “highly unsuitable” areas than the User-defined MCE model, whereas the User-defined MCE model has more areas that are classed as “intermediate” (Table 5). This suggests the User-defined MCE model averaged out scores more than the Fuzzy MCE model and thus lost some sensitivity in the final results.

Unlike the MCE approach, the use of SDMs is dependent on the GPS locations of farms. Such information may be difficult to collect, particularly across large areas and if no, or insufficient, information is available then SDMs cannot be used; potentially excluding areas with inadequate data/records as well as undeveloped locations where aquaculture has never been established. Here, the locations of pangasius farms were obtained from the large-scale farmer survey (Little et al., 2009); however some studies may be data-poor which could impact model performance. Likewise, the quality of the input data must also be considered, as any farm which is used as an input point is assumed *prima facie* to be a suitable location for aquaculture. The results of this study (Figures 4, 7) show both SDMs, are heavily influenced by farm location (Figure 1) as most of the higher scoring areas are found near where the farms were located. Some clustering of high scoring areas along the two main

branches of the Mekong would be expected due to the distribution of the input farm data and may also be a consequence of the variables. The distance layers in particular may have constrained the higher scores in localised areas, notably the distance to waterbodies layer which was one of the two most important variables in model development (Table 4, Figure 5). Another reason for high scores localised around the input points in the Maxent output could be to do with model performance. Maxent is known to overfit results and predicted distributions are often clustered around points (Baldwin, 2009). Maxent settings can be adjusted to account for this (Phillips and Dudik, 2008; Baldwin, 2009; Warren and Seifert, 2011; Merrow et al., 2013), however, even within standard applications of SDMs this is an area that still needs investigation (Anderson and Gonzales Jr, 2011; Warren and Seifert, 2011) and further studies related to aquaculture site selection are necessary.

One of the key strengths of Maxent is the ability to analyse variable contribution (Figure 5, Table 4) and response curves (Figure 6). This allows the user to understand the contribution each variable makes to the final model and is essential when interpreting the results. The response curves can also be used to identify trends in farm locations, however they must be interpreted with caution. Although most of the response curves (Figure 6) show patterns that would be expected (e.g. probability of farm presence decreases as distance to waterbodies increases (Figure 6H)) some require further analysis. Slopes greater than 5% would generally be considered too steep for pond construction (Hajek and Boyd, 1994), however the response curve (Figure 6E) suggests only a gradually declining probability of presence beyond 5%, whereas in reality a sharper decline after 5% would be expected as very few farms would be developed in such areas. All of the input farms are found in low slope areas, yet because the slope layer had a very low contribution to the model (0.2%, Table 4) this was not reflected in the final model output. This highlights the importance of analysing both variable contribution and the response curves. The use of a constraints layer(s) which excludes areas where aquaculture must not and should not occur would also be useful for future studies.

With regard to accuracy assessment, AUC values above 0.9 were obtained for both Maxent and Mahalanobis Typicality, suggesting they are excellent predictors of site locations. However, these values are largely due to the clustered nature of the input data and are not necessarily informative of the overall results. Although useful for species distribution studies, AUC and similar tests may be misleading for site suitability assessment which is more abstract. For this reason, the model results were compared and analysed in a more qualitative approach.

One of the challenges in using Mahalanobis Typicality and Maxent for aquaculture site selection is interpreting the outputs in terms of site suitability. Decision makers rely on thresholds which allow them to make an informed choice but establishing thresholds is difficult and often relies on judgements. Ultimately, threshold selection will depend on the aim and scope of the study and there is no definitive method that can be applied universally. To allow a comparison between the four different approaches, a five-point suitability classification was used in this study (Figure 7), however an alternative approach for SDMs may be more appropriate and research is needed into interpreting SDM model outputs for aquaculture site selection. Caution must be applied when reclassifying SDM outputs in terms of suitability as SDMs estimate similarity to the input variables rather than optimal values. It is recommended that any suitability reclassification is transparent so users understand the context of the model outputs and results.

Assuming all scored areas in the SDM outputs indicate a suitable location for aquaculture (ignoring degree of suitability), the results of the Mahalanobis Typicality model and Maxent model show a similar spatial distribution to the “suitable” and “highly suitable” areas of the MCE models (Figure 7). The south of the study area is considered “not suitable” by the SDMs and generally “unsuitable”/“highly unsuitable” by the MCE models. Along the Mekong River is identified suitable for development by all models, as are areas found inland in the north of the study area. This suggests that, despite their different methodologies, the MCE and SDMs have similar outputs thus could be used in a complementary approach. A

combination of both MCE and SDM would provide extra support to decision makers, alleviating some concerns over the subjective nature of MCEs, while providing a way to interpret the SDM results in terms of site suitability. The complementary approach could involve qualitative assessment and evaluation of the individual models, as in this study, or a more formalised hybrid model where both SDM and MCE model outputs are organised within a new model structure to produce the final result.

Based on this study, Mahalanobis Typicality appears to show more promise for aquaculture site selection than Maxent. The Maxent model has a large spatial distribution compared to Mahalanobis Typicality, however some areas, albeit with low scores, are located nearer the eastern coast where it is thought that salinity would be a limiting factor, particularly in the dry season due to saltwater intrusion. Mahalanobis Typicality appears to be more conservative and has a spatial distribution that would be expected, with scores found in low sloping, inland areas that are close to waterbodies. The differences in the results are due to the different modelling approaches and algorithms used. Mahalanobis Typicality considers the mean variable conditions of the input farms whilst the MCE approach uses optimal values and ranges. Thus, assuming suitable farms are used as input points, Mahalanobis Typicality could not only be used on its own for site selection it could also complement the MCE approach by providing information on areas that are similar to those already in use and it could even be used to help validate the MCE models. However, Maxent should not be ruled out completely as it may be useful for other case studies and the ability to analyse variable contribution is informative for decision makers.

This study has shown that SDMs, particularly Mahalanobis Typicality, could be used, together with MCE models in a complementary approach, providing a more robust method of aquaculture site selection. The study focused on several parameters that are important for the initial site identification and site selection process at the large scale of the Mekong delta, but there will be other biological and environmental parameters of importance and more detailed assessment of carrying capacity would be required at the local scale, as discussed

in Ross et al., (2013). Further work is necessary, particularly with regard to interpreting the outputs in terms of site suitability if SDMs are to be used on their own as an alternative to MCE. Multiple models could also be combined in an ensemble model rather than focusing on a single model for a result (Araújo and New, 2007). Additional case studies would be advantageous and there is also the potential to explore other SDMs in addition to Mahalanobis Typicality and Maxent. SDMs provide additional information to support the MCE approach that would otherwise be difficult to acquire. The extra confidence obtained from a combined MCE SDM approach would be of vital importance for the aquaculture sector as sustainable development is dependent on the planning and management decisions associated with site selection.

Acknowledgements

This work was funded by the EU-FP7 Sustaining Ethical Aquaculture Trade (SEAT) project (project number 222889) and the Institute of Aquaculture, University of Stirling.

References

- Afroz, T. and Alam, S., 2013. Sustainable shrimp farming in Bangladesh: a quest for an Integrated Coastal Zone Management. *Ocean & Coastal Management*, 71: 275-283.
- Aguilar-Manjarrez, J., Kapetskey, J.M. and Soto, D. 2010. The potential of spatial planning tools to support the ecosystem approach to aquaculture. FAO/Rome. Expert workshop. 19-21 November 2008, Rome, Italy. *FAO Fisheries and Aquaculture Proceedings*. No. 17. FAO, Rome. 176pp.
- Anderson, R.P. and Gonzales Jr, I. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, 222(15): 2796-2811.
- Araújo, M.B., Pearson, R.G., Thuiller, W. and Erhard, M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, 11(9):1504-1513.
- Baldwin, R.A. 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, 11: 854 - 866.

550 Belton, B., Little, D.C. and Sinh, L.X. 2011. The social relations of catfish production in
551 Vietnam. *Geoforum*, 42(5): 567-577.

552 Bosma, R.H., Udo, H.M.J., Verreth, J.A.J., Visser, L.E. and Nam, C.Q. 2005. Agriculture
553 diversification in the Mekong Delta: farmers' motives and contributions to livelihoods. *Asian*
554 *Journal of Agriculture and Development*, 2 (1 & 2): 49 - 66.

555 Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E. and
556 Zachmann, L.J. 2012. Species detection vs habitat suitability: are we biasing habitat
557 suitability models with remotely sensed data? *Ecological Modelling*, 244: 57-64.

558 Buitrago, J., Rada, M., Hernández, H. And Buitrago, E. 2005. A single-use site selection
559 technique, using GIS, for aquaculture planning: choosing locations for mangrove oyster raft
560 culture in Margarita Island, Venezuela. *Environmental Management*, 35(5): 544-556.

561 Carver, S.J. 2008. Multicriteria Evaluation. In: Kemp, K. ed. *Encyclopedia of Geographic*
562 *Information Science*. Sage Publications Inc, California, USA. pp. 290-297.

563 De Silva, S.S. and Phuong, N.T. 2011. Striped catfish farming in the Mekong Delta, Vietnam:
564 a tumultuous path to a global success. *Reviews in Aquaculture*, 3(2): 45-73.

565 Eastman, J.R. 2012. *IDRISI Selva Manual*. Clark Labs, Worcester 324pp.

566 Elith, J. and Leathwick, J.R. 2009. Species distribution models: ecological explanation and
567 prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, 40:
568 677-697.

569 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J.,
570 Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion,
571 G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC.M., Peterson, A.T., Phillips,
572 S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz,
573 M.S. and Zimmermann, N.E. 2006. Novel methods improve prediction of species'
574 distributions from occurrence data. *Ecography*, 29(2): 129-151.

575 Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. and Yates, C.J. 2011. A statistical
576 explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17: 43-57.

577 Evans, J.M., Fletcher, R.J. and Alavalapati, J. 2010. Using species distribution models to
578 identify areas for biofuel feedstock production. *GCB Bioenergy*, 2: 63-78.

579 Falconer, L., Hunter, D.C., Scott, P.C., Telfer, T.C. and Ross, L.G. 2013. Using physical
580 environmental parameters and cage engineering design within GIS-based site suitability
581 models for marine aquaculture. *Aquaculture Environment Interactions*, 4: 223-237.

582 FAO. 2014. *The State of World Fisheries and Aquaculture*. Rome. 223pp.

583 FAO FishStat Plus. 2014. FishStatJ - software for fishery statistical time series. [Database].
584 FAO, Rome. Available< <http://www.fao.org/fishery/statistics/software/fishstatj/en>> [Accessed:
585 4/08/2014].

586 FAO/IIASA/ISRIC/ISS-CAS/JRC. 2012. *Harmonized World Soil Database (version 1.2)*.
587 FAO, Rome, Italy and IIASA, Laxenburg, Austria.

588 Felicísimo, A.M., Cuartero, A., Remondo, J. and Quirós, E. 2013. Mapping landslide
589 susceptibility with logistic regression, multiple adaptive regression splines, classification
590 and regression trees, and maximum entropy methods: a comparative study. *Landslides*,
591 10(2): 175-189.

592 Fielding, A.H. and Bell, J.F. 1997. A review of methods for the assessment of prediction
593 errors in conversation presence/absence models. *Environmental Conservation*, 24(1): 38-49.

594 Franklin, J. 2010. Moving beyond static species distribution models in support of
595 conservation biogeography. *Diversity and Distributions*, 16(3): 321 - 330.

596 Giap, D.H., Yi, Y. and Yakupitiyage, A. 2005. GIS for land evaluation for shrimp farming in
597 Haiphong of Vietnam. *Ocean & Coastal Management*, 48: 51-63.

598 Hajek, B.F. and Boyd, C.E. 1994. Rating soil and water information for aquaculture.
599 *Aquacultural Engineering*, 13: 115-128.

600 Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. 2006. The effect of sample
601 size and species characteristics on performance of different species distribution modelling
602 methods. *Ecography*, 29(5): 773-785.

603 Hirzel, A.H. and Lay, G.L. 2008. Habitat suitability modelling and niche theory. *Journal of*
604 *Applied Ecology*, 45: 1372-1381.

605 Hossain, M.S., Chowdhury, S.R., Das, N.G., Sharifuzzaman, SM. And Sultana, A. 2009.
606 Integration of GIS and multicriteria decision analysis for urban aquaculture development in
607 Bangladesh. *Landscape and Urban Planning*, 90(3-4): 119-133.

608 Jang, C.S., Liu, C.W., Lin, K.H., Huang, F.M. and Wang, S.W. 2006. Spatial analysis of
609 potential carcinogenic risks associated with ingesting Arsenic in aquacultural tilapia
610 (*Oreochromis mossambicus*) in Blackfoot Disease hyperendemic areas. *Environmental*
611 *Science & Technology*, 40(5): 1707-1713.

612 Little, D.C, Murray F, Telfer T, Young J, Ross L, Hill B, Dalsgaard A, van den Brink PJ,
613 Guinee J, Kleijn R, Mungkung R, Yi Y, Min J, Liping L and Huanan L (2009) Sustaining
614 ethical trade in farmed aquatic products between Asia and the EU. In: Millar K, Hobson
615 West P, Nerlich B (ed.). *Ethical Futures: Bioscience and Food Horizons* , Wageningen,
616 The Netherlands: Wageningen Academic Publishers, pp. 226-231.

617 Llorente, I. and Luna, L. 2013. The competitive advantages arising from different
618 environmental conditions in Seabream, *Sparus aurata*, production in the Mediterranean
619 Sea. *Journal of the World Aquaculture Society*, 44(5): 611-627.

620 Katranidis, S., Nitsi, E. and Vakrou, A. 2003. Social acceptability of aquaculture development
621 in coastal areas: the case of two Greek islands. *Coastal Management*, 31(1): 37-53.

622 McLeod, I., Pantus, F. and Preston, N. 2002. The use of a geographical information system
623 for land-based aquaculture planning. *Aquaculture Research*, 33; 241-250.

624 Merow, C., Smith, M.J. and Silander, J.A. 2013. A practical guide to MaxEnt for modeling
625 species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36: 1-12.

626 NASA. 2009. *Shuttle Radar Topography Mission: The mission to map the world*. [Internet].
 627 NASA, Washington D.C. Available from <<http://www2.jpl.nasa.gov/srtm/>>

628 Nath, S.S., Bolte, J.P., Ross, L.G. and Aguilar-Manjarrez, J.A. 2000. Applications of
 629 geographical information systems (GIS) for spatial decision support in aquaculture.
 630 *Aquacultural Engineering*, 23: 233-278.

631 Oak Ridge National Laboratory. 2008. Landsat population database 2008. [Internet] Oak
 632 Ridge National Laboratory, Tennessee, USA. Available from :
 633 <<http://web.ornl.gov/sci/landsat/>>

634 Peters, M.P., Iverson, L.R., Matthews, S.N., Prasad, A.N. 2013. Wildfire hazard mapping:
 635 exploring site conditions in eastern US wildfire-urban interfaces. *International Journal of*
 636 *Wildland Fire*, 22(5): 567-578.

637 Phan, L.T., Bui, T.M., Nguyen, T.T.T., Gooley, G.J., Ingram, B.A., Nguyen, P.T. and De
 638 Silva, S.S. 2009. Current status of farming practices of striped catfish, *Pangasianodon*
 639 *hypophthalmus* in the Mekong Delta, Vietnam. *Aquaculture*, 296(3-4): 227-236.

640 Phillips, S.J. 2006. A brief tutorial on Maxent. AT & T Research. 38pp. Available
 641 <<http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>>

642 Phillips, S.J. and Dudik, M. 2008. Modeling of species distributions with Maxent: new
 643 extensions and comprehensive evaluation. *Ecography*, 31(2): 161-175.

644 Phillips, S.J., Anderson, R.P. and Schapire, R.E. 2006. Maximum entropy modeling of
 645 species geographic distributions. *Ecological Modelling*, 190: 231-259.

646 Phuong, N.T. and Oanh, D.T.H. 2009. Striped catfish aquaculture in Vietnam: a decade of
 647 unprecedented development. In: De Silva, S.S. and Davy, F.B. eds. *Success stories in Asian*
 648 *aquaculture*. Springer Science+Business Media, New York, USA. pp133- 150.

649 Radiarta, I.N., Saitoh, S.I. and Miyazono, A. 2008. GIS-based multi-criteria evaluation
 650 models for identifying suitable sites for Japanese scallop (*Mizuhopecten yessoensis*)
 651 aquaculture in Funka Bay, southwestern Hokkaido, Japan. *Aquaculture*, 284: 127-135.

652 Raux, P., Bailly, D. and Nhung, T.V. 2006. Vietnamese shrimp farming at a key point in its
 653 development: a review of issues examining whether development is being carried out in a
 654 sustainable way. 107-130pp. In : Leung, P.S. & Engle, C.R. eds. *Shrimp culture: economics,*
 655 *market and trade*. Blackwell Publishing Ltd, Oxford, UK, pp. 107-130.

656 Ross, L.G., Falconer, L., Campos Mendoza, A. and Martinez Palacios, C.A. 2011. Spatial
 657 modelling for freshwater cage location in the Presa Adolfo Mateos Lopez (El Infiernillo),
 658 Michoacán, México. *Aquaculture Research*, 42(6): 797 - 807.

659 Ross, L.G., Telfer, T.C., Falconer, L., Soto, D. & Aguilar-Manjarrez, J., eds. 2013. Site
 660 selection and carrying capacities for inland and coastal aquaculture. FAO/Institute of
 661 Aquaculture, University of Stirling, Expert Workshop, 6–8 December 2010. Stirling, the
 662 United Kingdom of Great Britain and Northern Ireland. *FAO Fisheries and Aquaculture*
 663 *Proceedings No. 21*. Rome, FAO. 46 pp. Includes a CD-ROM containing the full document
 664 (282 pp.).

- Saaty, T.L. 1977. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15: 234-281.
- Salam, M.A., Khatun, N.A. and Ali, M.M. 2005. Carp farming potential in Barhatta Upazilla, Bangladesh: a GIS methodological perspective. *Aquaculture*, 245(1-4): 75-87.
- Sakamoto, T., Phung, C.V., Kotera, A., Nguyen, K.D. and Yokozawa, M. 2009. Analysis of rapid expansion of inland aquaculture and triple rice-cropping areas in a coastal area of the Vietnamese Mekong Delta using MODIS time-series imagery. *Landscape and Urban Planning*, 921: 34-46.
- Sangermano, F. and Eastman, J.R. 2007. Linking GIS and ecology. The use of Mahalanobis Typicalities to model species distribution. Proceedings of XI CONFIBSIG (Iberoamerican Conference of GIS). Buenos Aires, Argentina, May 2007.
- Suplicy, F.M., de Novaes Vianna, L.F., Rupp, G.S., Novaes, A.L.T., Garbossa, L.H.P., de Souza, R.V., Guzenski, J., da Costa, S.W., Silva, F.M. and dos Santos, A.A. 2015. Planning and management for sustainable coastal aquaculture development in Santa Catarina State, south Brazil. *Reviews in Aquaculture*. DOI: 10.1111/raq.12107
- Trieu, T.T.N. and Phong, N.T. 2015. The impact of climate change on salinity intrusion and Pangasius (*Pangasianodon Hypophthalmus*) farming in the Mekong Delta, Vietnam. *Aquaculture International*, 23(2): 523-534.
- Tucker, C.S. and Hargreaves, J.A. 2008. Environmental best management practices for aquaculture. Blackwell Publishing Ltd, Oxford, UK. 592pp.
- USGS. 2013. Landsat missions. [Internet]. USGS, Reston, Virginia, USA. Available from: <<http://landsat.usgs.gov/>> [Accessed: 12/09/2013].
- Warren, D.L. and Seifert, S.N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological applications*, 21(2): 335 - 342.
- Wu, R.S.S. 1995. The environmental impact of marine fish culture: towards a sustainable future. *Marine Pollution Bulletin*, 31: 4-12.
- Zadeh, L.A. 1965. Fuzzy sets. *Information and Control*, 8: 338-353.

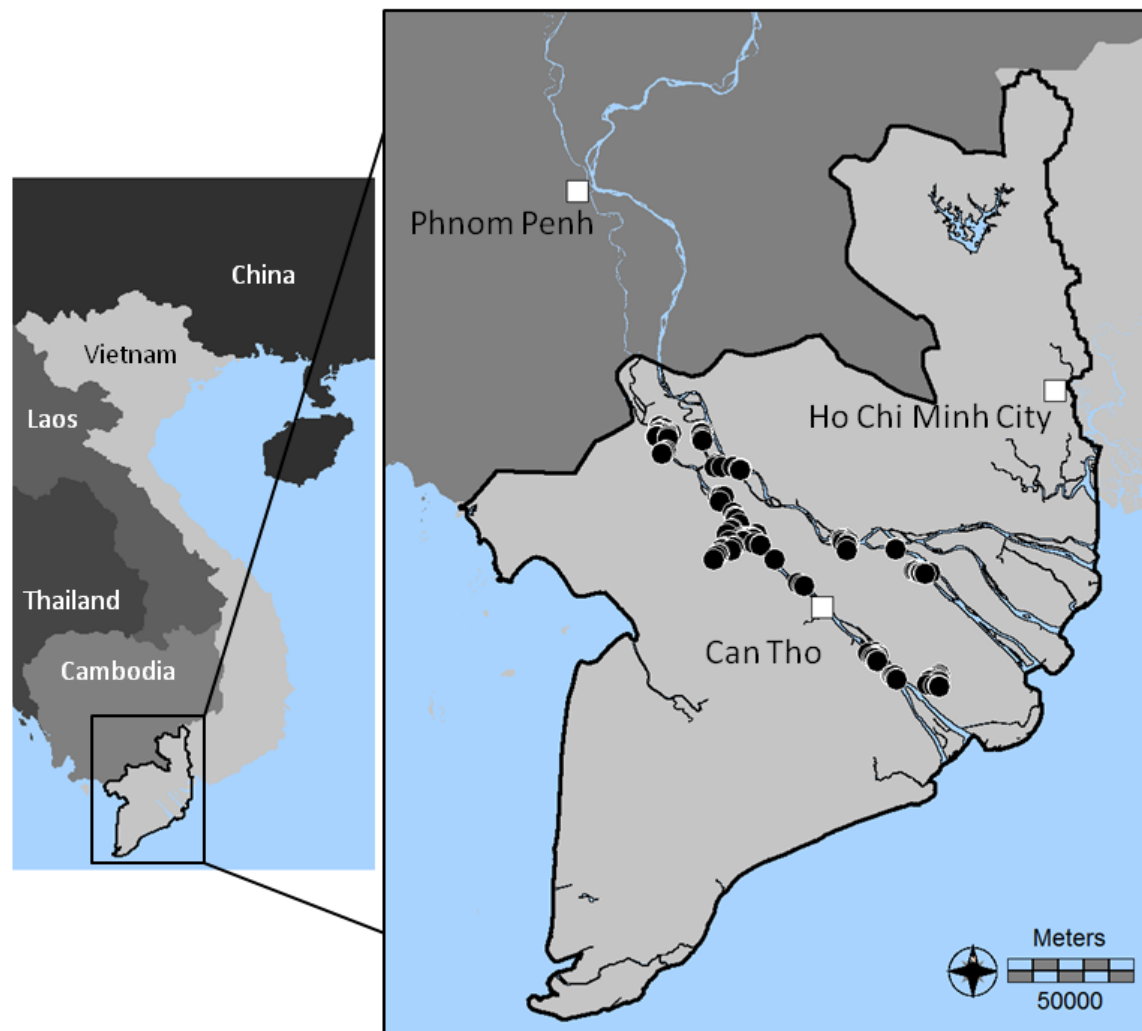


Figure 1: Study area (outlined in black) in the Mekong Delta, Vietnam. The black points represent pangasius farms.

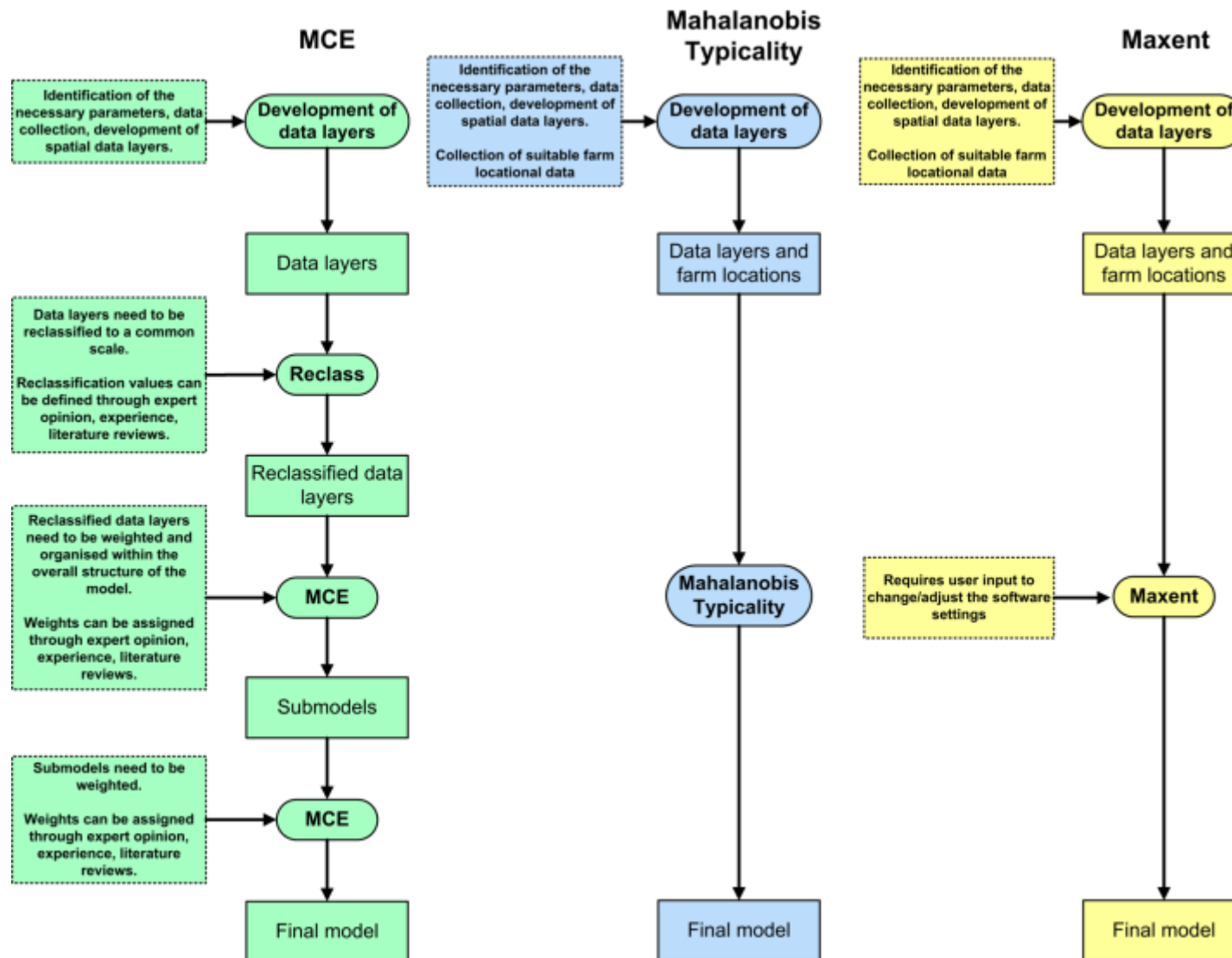


Figure 2: Comparison of the required input from the user/developer for the MCE, Mahalanobis Typicality and Maxent models.

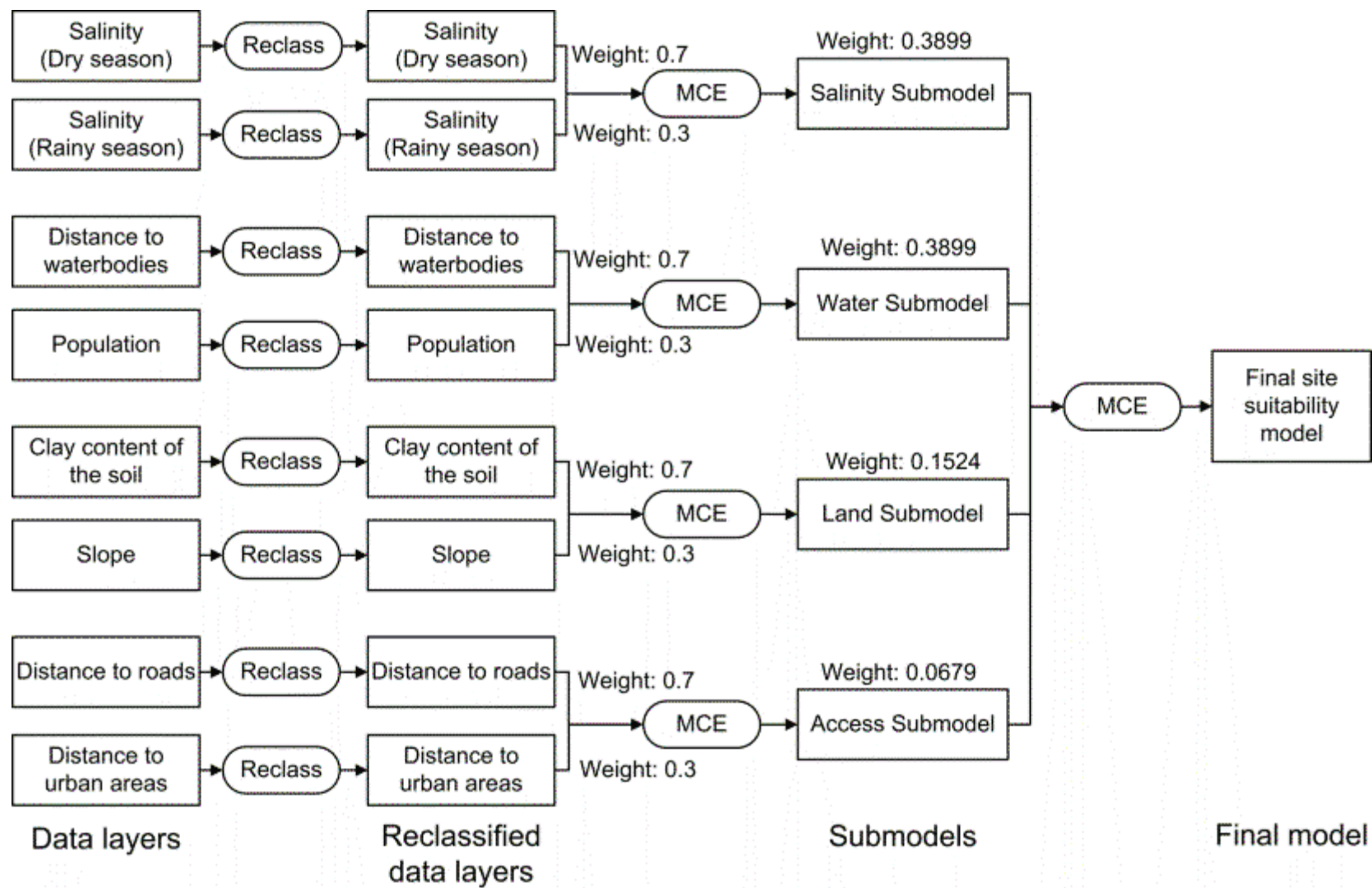


Figure 3: Structure of the MCE models

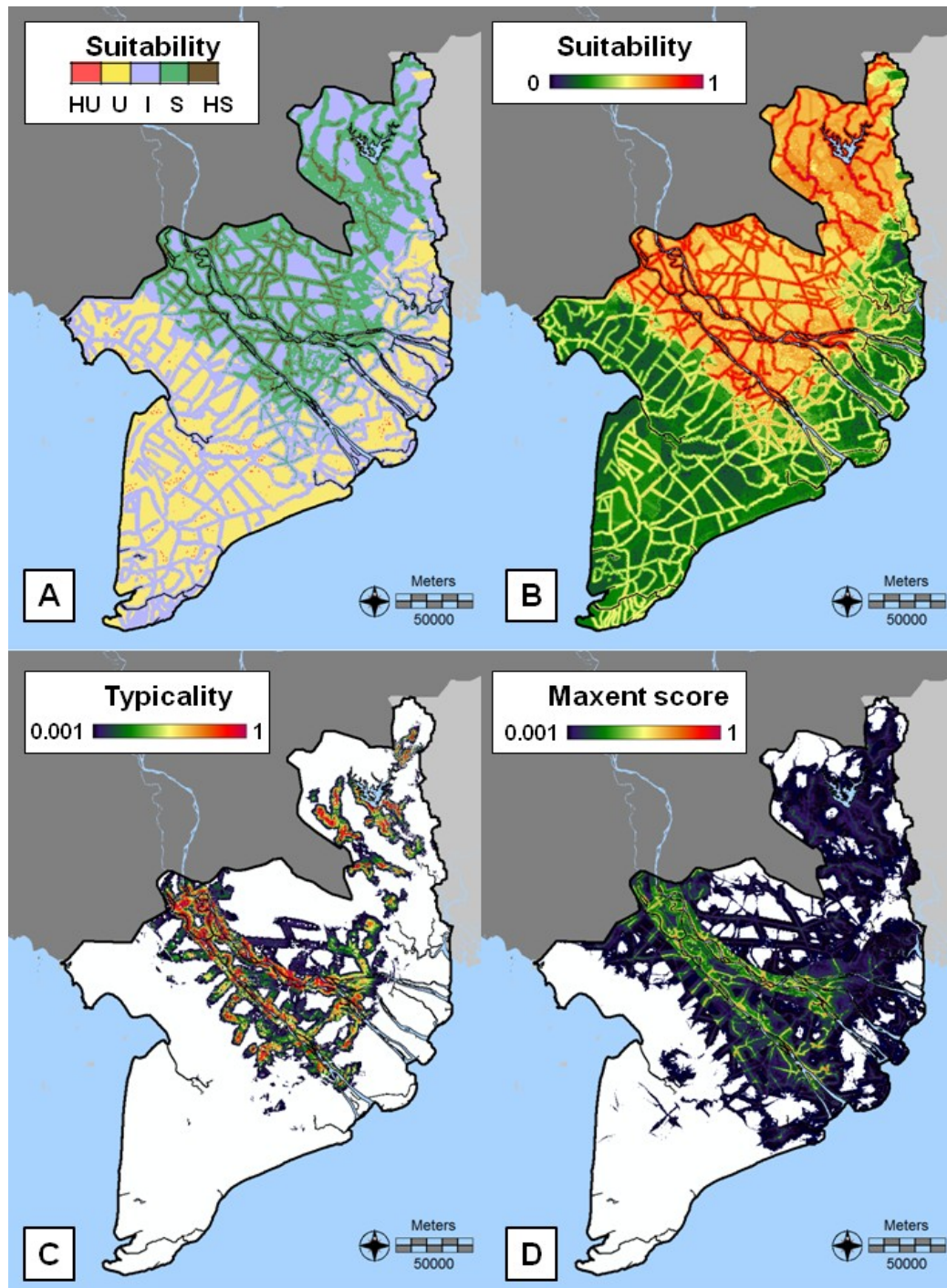


Figure 4: Model results for A) the user-defined MCE model, the categories of suitability show how suitable the area is for pangasius culture (HU = “Highly Unsuitable”, U = “Unsuitable”, I = “Intermediate”, S = “Suitable”, HS = “Highly Suitable”. B) Fuzzy MCE model, the continuous scale from 0.001 to 1 represents suitability where the most suitable areas would have a score of 1, C) Mahalanobis Typicality model, the higher the typicality, the more similar that location is to the mean variable conditions of the input farms. D) Maxent model, higher scores represent greater probability of the conditions needed for pangasius culture.

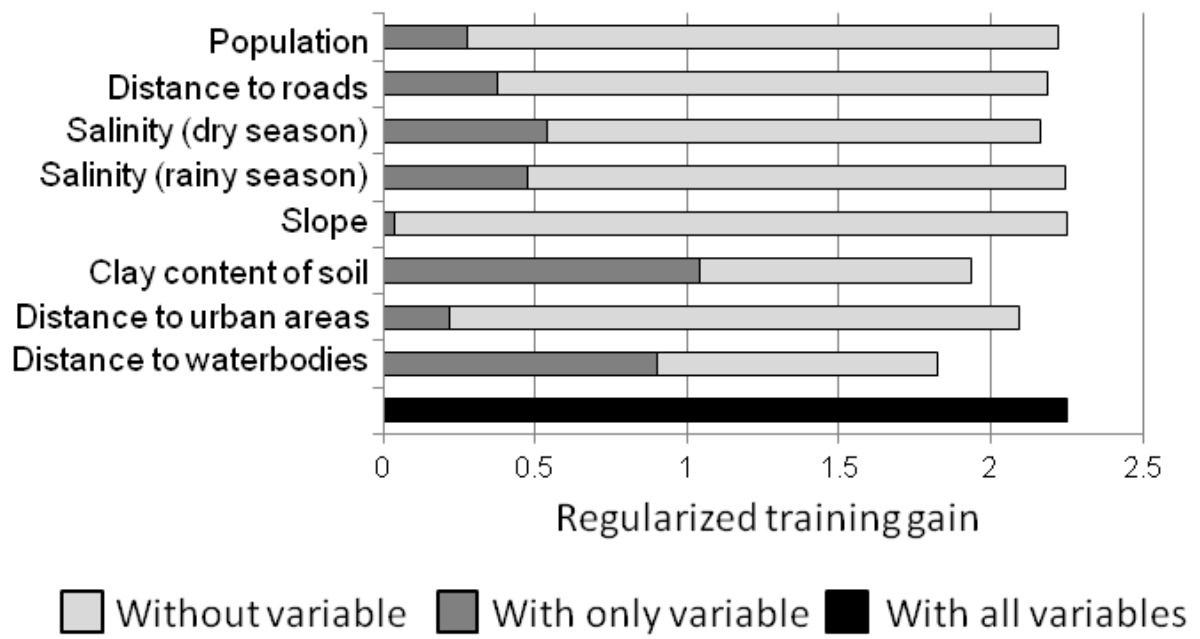


Figure 5: Jackknife test of variable importance for the Maxent model.

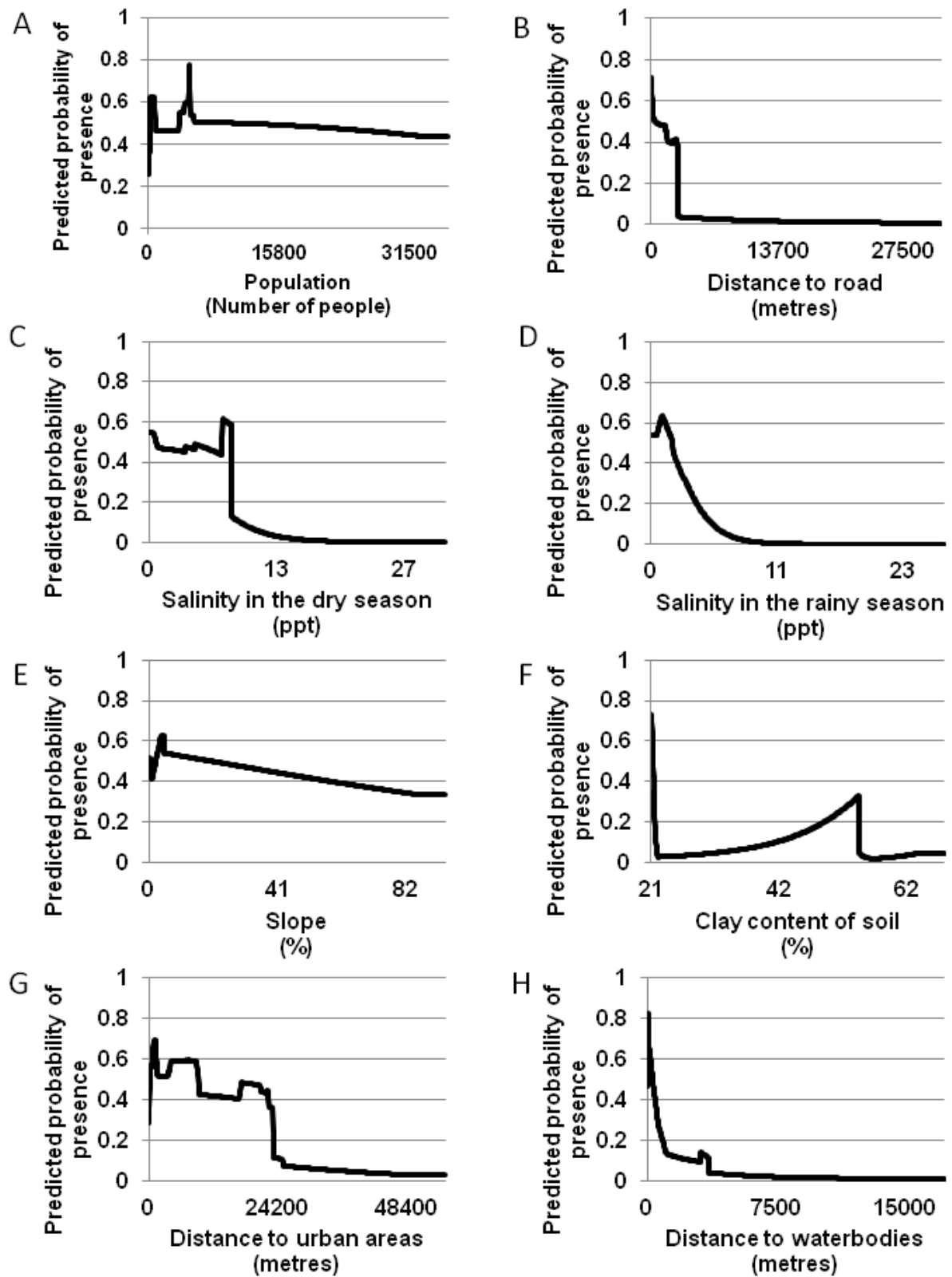


Figure 6: Response curves for the Maxent model

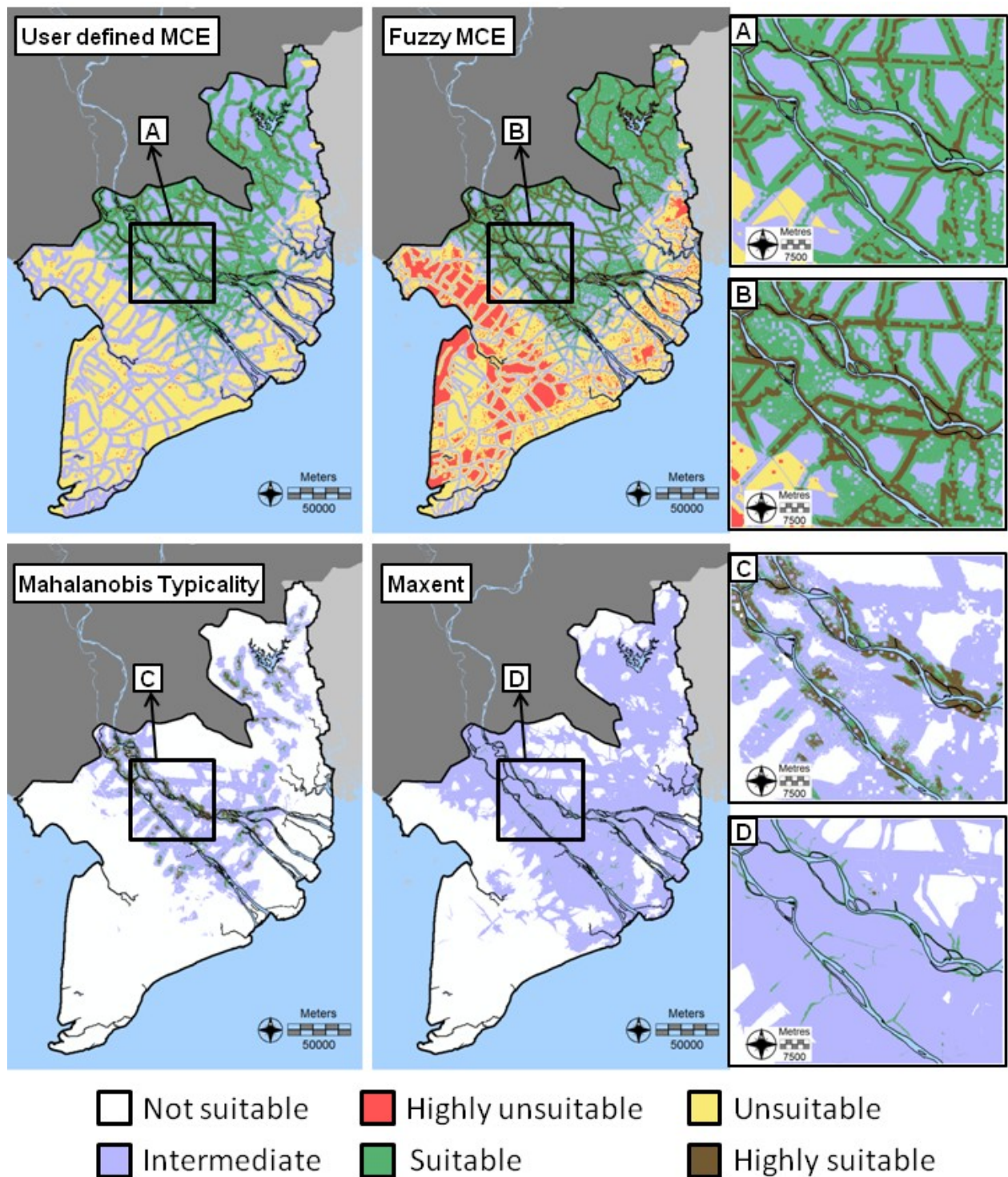


Figure 7: Comparison of the User-defined MCE model, reclassified Fuzzy MCE model, reclassified Mahalanobis Typicality model and reclassified Maxent model. Zoomed in boxes (A, B, C, D) highlight an area already used for pangasius culture in the Mekong Delta, Vietnam.

Table 1: The eight variables used in the SDMs and MCE models.

Variable	Description	Source
Population	People represent a labour force, market source and opportunity to sell products. However, they can also put more pressure on water resources, lead to conflict and increase potential water quality issues.	Landscan 2008 population dataset (Oak Ridge National Laboratory, 2008)
Distance to Roads	Farm location could be influenced by distance to road in terms of access and transport networks.	Roads were digitised using Google Earth and LandsatETM+ data (USGS, 2013) in IDRISI Selva. Distance was then calculated using IDRISI Selva.
Salinity (dry season)	Salinity values outside the tolerance range of an aquaculture species can affect the health and welfare of the animal and/or prevent culture.	Interpolated using data from Vietnam Research Institute for Aquaculture (RIA2).
Salinity (rainy season)	Salinity values outside the tolerance range of an aquaculture species affect the health and welfare of the animal and/or prevent culture.	Interpolated using data from Vietnam Research Institute for Aquaculture (RIA2).
Slope	The slope of an area can influence farm construction and pond drainage.	Calculated from SRTM DEM (Nasa, 2009).
Clay content of soil	The clay content of a soil determines the porosity and the potential for water to leak from the pond.	Clay content of the dominant subsoil in the Harmonised World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISS-CAS/JRC. 2012)
Distance to urban areas	Distance to urban areas is important as they are a source of supplies, labour and could also be valuable markets for trade.	Urban areas with a population were digitised using Google Earth and LandsatETM+ data in IDRISI Selva. Distance was then calculated using IDRISI Selva.
Distance to waterbodies	Distance to rivers, channels, lakes and reservoirs which can all be a source of water for the system.	Larger waterbodies were obtained from the SRTM Water Body Data (SWDB) (http://www2.jpl.nasa.gov/srtm/). Smaller rivers and channels were digitised using Google Earth and LandsatETM+ data in IDRISI Selva. Distance was then calculated using IDRISI Selva.

Table 2: Reclassification values for the variables within the User-defined MCE model

User-defined classification	Highly unsuitable	Unsuitable	Intermediate	Suitable	Highly suitable	References
Population (people per km ²)	>2000	1500-2000	1000-1500	500-1000	<500	Adapted from: Giap et al., 2005
Distance to roads (m)	>2000	1500-2000	1000-1500	500 - 1000	0 - 500	Adapted from: Giap et al., 2005
Salinity (ppt)	>4	-	-	2 - 4	0 - 2	Adapted from: Trieu and Phong, 2015
Slope (%)	>10	5 - 10	0 - 0.5	2 - 5	0.5 - 2	Adapted from: Hajek and Boyd, 1994;
Clay content of soil (%)	<15 >60	15 - 25	25 - 30	40 - 60	30 - 40	Giap et al., 2005 Adapted from: Hajek and Boyd, 1994;
Distance to urban areas (km)	<1 > 15	10 - 15	4 -10	2 - 4	1 - 2	Tucker and Hargreaves, 2008 Adapted from: Giap et al., 2005
Distance to waterbodies (m)	>2000	1500-2000	1000-1500	500 -1000	0 - 500	Adapted from: McLeod <i>et al.</i> , 2002; Salam <i>et al.</i> , 2005

Table 3: Reclassification values for the variables within the Fuzzy MCE model

Fuzzy classification	Membership function	a	b	c	d	References
Population (people per km ²)	Sigmoidal decreasing	-	-	500	2000	Adapted from: Giap et al. 2005
Distance to roads (m)	Sigmoidal decreasing	-	-	0	2000	Adapted from: Giap et al. 2005
Salinity (ppt)	Sigmoidal decreasing	-	-	2	4	Adapted from: Trieu and Phong, 2015
Slope (%)	Sigmoidal symmetric	0	0.5	2	10	Adapted from: Hajek and Boyd, 1994; Giap et al., 2005
Clay content of soil (%)	Sigmoidal symmetric	15	30	40	60	Adapted from: Hajek and Boyd, 1994; Tucker and Hargreaves, 2008
Distance to urban areas (km)	Sigmoidal symmetric	0	1	2	15	Adapted from: Giap et al. 2005
Distance to waterbodies (m)	Sigmoidal decreasing	-	-	0	2000	Adapted from: McLeod <i>et al.</i> , 2002; Salam <i>et al.</i> , 2005

Table 4: Percent contribution of the variables to the Maxent model.

Variable	Percent contribution (%)
Clay content of soil	33.6
Distance to waterbodies	30.9
Salinity (dry season)	17.1
Distance to roads	8.2
Distance to urban areas	5.7
Salinity (rainy season)	2.6
Population	1.7
Slope	0.2

Table 5: Areas (% and km²) covered by each suitability category for the User-defined MCE model and the three reclassified models; Fuzzy MCE model, Mahalanobis Typicality SDM and Maxent SDM

	Model			
	User-defined MCE model	Reclassified Fuzzy MCE model	Reclassified Mahalanobis Typicality model	Reclassified Maxent model
Not suitable	- -	- -	34900km ² 75%	21400km ² 46%
Highly unsuitable	100km ² <1%	5300km ² 11%	- -	- -
Unsuitable	13500km ² 29%	13500km ² 29%	- -	- -
Intermediate	19500km ² 42%	12700km ² 27%	9700 km ² 21%	24500km ² 53%
Suitable	11300km ² 24%	10800km ² 23%	900km ² 2%	400km ² 1%
Highly suitable	1900km ² 4%	4100km ² 9%	900km ² 2%	<100km ² <1%

Table 6: Number of SEAT pangasius farms located in each suitability category for the User-defined MCE model and the three reclassified models; Fuzzy MCE model, Mahalanobis Typicality SDM and Maxent SDM

	Model			
	User-defined MCE model	Reclassified Fuzzy MCE model	Reclassified Mahalanobis Typicality model	Reclassified Maxent model
Not suitable	-	-	1	0
	-	-	<1%	0%
Highly unsuitable	0	1	-	-
	0%	<1%	-	-
Unsuitable	1	3	-	-
	<1%	<2%	-	-
Intermediate	46	44	101	83
	24%	23%	53%	43%
Suitable	85	41	31	83
	44%	21%	16%	43%
Highly suitable	60	103	59	26
	31%	54%	31%	14%