Published in final edited form as:
Ann Assoc Am Geogr. 2013 January 1; 103(4): . doi:10.1080/00045608.2012.702479.

Land Suitability Modeling using a Geographic Socio-Environmental Niche-Based Approach: A Case Study from Northeastern Thailand

Benjamin W. Heumann[a,b], **Stephen J. Walsh**[a,b], **Ashton M. Verdery**[b,c], **Phillip M. McDaniel**[b], and **Ronald R. Rindfuss**[b,c,d]

[a]Department of Geography, University of North Carolina at Chapel Hill

[b]Carolina Population Center, University of North Carolina at Chapel Hill

[c]Department of Sociology, University of North Carolina at Chapel Hill

[d]East-West Center

Abstract

Understanding the pattern-process relations of land use/land cover change is an important area of research that provides key insights into human-environment interactions. The suitability or likelihood of occurrence of land use such as agricultural crop types across a human-managed landscape is a central consideration. Recent advances in niche-based, geographic species distribution modeling (SDM) offer a novel approach to understanding land suitability and land use decisions. SDM links species presence-location data with geospatial information and uses machine learning algorithms to develop non-linear and discontinuous species-environment relationships. Here, we apply the MaxEnt (Maximum Entropy) model for land suitability modeling by adapting niche theory to a human-managed landscape. In this article, we use data from an agricultural district in Northeastern Thailand as a case study for examining the relationships between the natural, built, and social environments and the likelihood of crop choice for the commonly grown crops that occur in the Nang Rong District – cassava, heavy rice, and jasmine rice, as well as an emerging crop, fruit trees. Our results indicate that while the natural environment (e.g., elevation and soils) is often the dominant factor in crop likelihood, the likelihood is also influenced by household characteristics, such as household assets and conditions of the neighborhood or built environment. Furthermore, the shape of the land use-environment curves illustrates the non-continuous and non-linear nature of these relationships. This approach demonstrates a novel method of understanding non-linear relationships between land and people. The article concludes with a proposed method for integrating the niche-based rules of land use allocation into a dynamic land use model that can address both allocation and quantity of agricultural crops.

Keywords

Land Suitability; *Machine-Learning*; *Modeling*; *Niche*; *Thailand*

Correspondence: Department of Geography and Global Environmental and Climate Change Centre, McGill University, Montreal, QC H3A 2K6, Canada, benjamin.heumann@mail.mcgill.ca (Heumann); Department of Geography and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, swalsh@unc.edu (Walsh); Department of Geography and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, verdery@email.unc.edu (Verdery); Department of Geography and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, pmmcdani@email.unc.edu (McDaniel); Department of Geography and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, ron_rindfuss@unc.edu (Rindfuss).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Introduction

Understanding the pattern-process relationships of land use/land cover change is important to a whole host of scientific questions and policy applications that revolve around human-environment interactions. Land use decision making is often a focal point of human-environmental interaction. This decision-making is influenced by the natural, built, and social environment. For example, land use decisions can be based on soils, climate, drainage characteristics, access to markets, transportation costs, and household assets and labor availability. Furthermore, off-farm activities such as out-migration and remittances can affect land use decision making.

Land use modeling consists of allocating land use and quantifying land use activities (Chen and Pontius 2010). For example, when modeling an agricultural landscape there is the question of which crop is grown and the question of how much of that crop is grown. In this article, we concern ourselves primarily with the former. The question of where and why certain land uses occur in geographic space is similar to understanding other geographic patterns that occur from a complex set of conditions such as species distributions. One common approach to modeling species distributions is to model the niche in environmental space and project the niche across geographic space (Austin 2002). The niche is often described in terms of population dynamics where the niche is the set of environmental conditions in which the population is stable or growing without immigration (Pulliam 2000). An important theoretical advance of this article is our adaptation of niche space modeling to human-managed landscapes by replacing population dynamics with changes in agricultural yield and household income. In other words, the niche of a crop in a human-managed landscape is the set of environmental, both natural and built, and social conditions that provide a yield and income that can sustain and enhance the household. Our goal is to understand the land suitability for crops, i.e., where crops are likely to occur, using available data on the natural abiotic conditions, the built environment, and household social factors that either directly address or serve as proxies to the human modification of the niche described above. Based on a socio-environmental niche where the likelihood of crop occurrence is a function of natural, built, and social environmental conditions that affect the processes determining land use choices in a human managed landscape, crop occurrences can be modeled by determining the relationships between crop presences and the total environmental space and projecting those relationships across geographic space.

Background

The socio-economic activities of households involve the formation of livelihood strategies designed to meet present and future needs, according to household perceptions, values, and goals (Rajbhandari 2006). A livelihood is a dynamic set of activities that may change over time according to available opportunities, environmental conditions, economic trajectories, and exogenous shocks. Further, livelihood alternatives are mediated by the complex interactions of social, cultural, demographic, economic, institutional, and ecological processes (Robbins 2004; Zimmerer 2006). Individuals and households adapt to shifts in socio-economic opportunities and constraints (Netting 1979; Chambers and Conway 1992),which are influenced by environmental settings, including land cover types and conditions and land use practices (Rindfuss et al. 2004a). While assets, capabilities, and activities shape livelihoods (Ellis 2000), the relationship between a household and its social and ecological contexts influences the decision-making processes. This often reflects an alternate household livelihood strategy that is affected by real and/or perceived risks to household wealth, a combination of natural, social, physical, human, and financial resources (Ellis 1998; de Sherbinin et al. 2007).

Risk abatement strategies reflect household preferences that involve, for instance, labor diversification through kinship relations and broader social networks, off-farm employment and remittances to the household, and changes in planting and/or harvesting practices, timing, and crop type selection (e.g., cash vs. subsistence crops) to reflect constraints in resources, such as water deficits in times of drought (Goland 1993; Winterhalder, Lu, and Tucker 1999). Coping behaviors of households often involve the adoption of a broad and dynamic portfolio of assets and activities (Little and Leslie 1999) to minimize the threat to a single livelihood activity (Cashdan 1990). In addition to livelihood diversification, population migration to cities or other countries is also a strategy adopted to reduce economic risks by moving all or part of the household to areas of higher wages and more employment opportunities (Massey 1990; Stark 1991).

Both risk-mitigation through alternate household livelihood strategies and migration occur in the Nang Rong District study area. This has implications for the selection of crop types planted in optimum and less than optimum ecological settings, household diversification through farm and off-farm employment, and cultivation of upland and lowland crops relative to social and ecological factors, including geographical, sociological, and demographic conditions. Assessing the social and ecological drivers of land use change at multiple and interacting space-time scales is embodied in Land Change Science, an integrated set of theories and practices that are used to assess land use/land cover change in diverse environments, including Nang Rong District, northeastern Thailand (Walsh et al. 2001; Rindfuss et al. 2004a, b; Entwisle et al. 2005).

In the past decade, increasingly sophisticated modeling techniques have been developed to investigate the links between patterns and processes of land use/land cover change. These new techniques include the use of spatially explicit models that incorporate aspatial and spatial data. Two common approaches are cellular automata (CA) and agent-based models (ABM). The relatively simple CA model uses the cell, with fixed geographic coordinates in a regular grid lattice as the unit of analysis, with decisions of land use/land cover changed derived based on the attributes of the cell as well as neighboring cells (e.g., Clark et al. 2007; Walsh et al. 2006). The more sophisticated ABM focuses on modeling the agents of change and how they interact with the spatial and aspatial environment as they adapt to dynamic social and ecological situations (e.g., Entwisle et al. 2008; Manson 2005). For example, the agent of change is commonly the household, so household-level decisions are modeled based on a wide variety of social, economic, geographic, and economic factors. By modeling the agent of change, an ABM can demonstrate large-scale patterns that emerge from an interconnected set of small-scale processes. This ability makes them well suited for theory testing. There are two fundamental approaches commonly taken with ABMs. In the first, testing is often achieved using hypothesized rules and synthetic data (e.g., Bruch and Mare 2006; Bearman, Moody, and Stovel 2004). The other approach uses empirical data in combination with hypothesized rules to focus on the implications of theoretical mechanisms given what has been observed.

One of the greatest challenges to creating these types of models is creating the rule sets that drive them. As with any model, assumptions and simplifications have to be made, but these assumptions and simplifications can introduce unexpected error. For example, while assuming linear relationships can greatly simplify a model, the assumption is that the linear relationship is a justified approximation of reality. However, in many cases, the data and relationships are messy (e.g., the data are not normally distributed or uni-modal, the relationships are interactive and non-linear and may exhibit different patterns at different areas of the parameter space). All of this makes finding and setting rules based on real world data complicated.

One solution to this problem is the use of machine-learning algorithms. Machine-learning algorithms have been applied to a wide range of scientific problems revolving around pattern recognition. One area in the natural sciences where machine-learning has emerged is species distribution models (SDMs). SDMs are a computational model that applies the niche concept to link species observations to environmental conditions and projects a likelihood of occurrence across the landscape (Guisan and Zimmerman 2000; Austin 2002; Peterson 2003; Phillips, Anderson, and Schapire. 2006; Thuiller, Lavorel, and Araujo 2005; Peterson, Papes and Eaton, 2007; Hirzel and Le Lay 2008). There are a variety of models and techniques that have been used to assess and predict habitat patterns and species dynamics including climatic-envelope models, statistical models such as generalized linear models and generalized additive models, and machine-learning algorithms such as GARP and MaxEnt. While numerous models exist (see Elith et al. 2006 for a review of 16 different models), there is a consensus that the methods used in machine-learning algorithms are consistently among the best performing according to validation using the receiver operating characteristic. Such models optimize their performance based on iterative calculations that use defined geographic positions of target species and environment conditions to calculate habitat suitability or likelihood of species presence (Elith et al. 2006; Hernandez et al. 2006; Phillips, Anderson, and Schapire 2006; Hirzel and Le Lay 2008; Ortega-Huerta and Peterson 2008; Phillips et al. 2009).

The development of novel machine-learning SDMs requiring presence-only data (rather than definite data on absences) provides a great opportunity for land use/land cover change modeling. SDMs can find complex species-environmental relationships based on geographic species locations and environmental data. This approach can be further modified for land use by adapting the niche concept to a human-managed system. For example, Heumann, Walsh, and McDaniel (2011) found that, using only natural environment variables, MaxEnt could reasonably predict land suitability for common agricultural land uses.

Adapting the Niche to a Human-Managed Landscape

Although the niche concept is nearly a century old, considerable debate remains about definitions of the niche and how the niche concept is applied to geography and space (Pulliam 2000; Soberon 2007; Hirzel and Le Lay 2008; Araujo and Guisan 2006; Peterson, Papes, and Eaton 2007; Jimenez-Valverde, Lobo, and Hortal 2008). This is particularly important given the explosion of species data and the methods for deriving species distributions and habitat models. At the very outset, niche-based modeling requires a clear statement of how the niche concept is defined and applied (Austin 2007). We propose a theoretical framework that carefully considers the niche concept as it applies to agriculture. This is an important contribution to the literature, particularly given the context of a human-managed ecosystem and the substantial human impact across the landscape, and our modeling approach and data source provides a unique perspective on this topic.

We accept the concept of the niche and associated distributions to be a function of dispersal capability, abiotic environmental conditions or geographic setting, and the biotic environment including competition and the dynamic nature of resource availability (Soberon 2007). In an agricultural setting, human management modifies environmental conditions and ecological processes in the following ways. First, natural dispersal capability and limitations are replaced with managed dispersal based on human selection of crops as a function of the environmental needs of the crop (both real and as perceived by farmers), resource constraints in the crop year (i.e., droughts or floods), and socio-economic and demographic factors that influence household decision-making including commodity prices, wealth and assets, government quotas, and labor availability. It is important to note that household decision-making depends not only on socio-economic and demographic factors, but also on the abiotic environmental conditions, the biotic processes, and the extent to which these can be improved for crop productivity. Second, abiotic environmental conditions can be modified to improve crop suitability, though the scale of modification depends on the conditions. For example, climatic conditions such as temperature and precipitation are part of a larger earth-system, but water availability can be modified using irrigation and the construction of bunds that are used to retain water in the rice paddy well beyond the monsoonal rains; the use of fertilizer to improve growing conditions of otherwise less suitable soils. These modifications can change the potential distribution of the crops in the study area and, hence, need to be integrated in the interpretation of model results. Third, agricultural practices fundamentally change the biotic environment, at least at the local scale. Weeds or competitors can be removed and pests/pathogens can be manually and/or chemically controlled.

Research Questions and Objectives

The aim of this research is to understand what factors of the natural, built, and social environment correspond to crop occurrence based on land use choices made by households, the relationship between those factors and crop occurrence, and the relative importance of each factor. Specifically, we ask the following questions: 1) What are the conditions of the natural, built, and social environment that correspond to the likelihood of crop presence? 2) What factors of the built and social environment are related to crop choice? 3) What is the relationship between important social and ecological factors and crop suitability? 4) What differences exist between the relationship of crop suitability and the environmental factors for low yield and high yield locations?

To address these objectives, the niche-based geographic species distribution model, MaxEnt (Phillips, Anderson, and Schapire 2006), is used for several common crops: cassava, heavy rice, and jasmine rice as well as an emerging crop, fruit trees. Additionally, environmental and social data have been transformed into geospatial data using geographic information systems to calculate surfaces and link household characteristics with the landscape through use of the land.

Study Area

Nang Rong District is located in the Buriram province in northeast Thailand (Figure 1 --note the location of actual study villages has been removed for confidentiality related to Human Subject research). The District is approximately 1300 sq. km and is positioned in the southwest portion of a wide shallow basin underlain by Cretaceous sandstones, shales, and siltstones called the Khorat Plateau. Nang Rong District has amongst the lowest per capita income in Thailand. Farming is the dominant livelihood, with generally low levels of agricultural productivity and high levels of inter-annual uncertainty due to poor soil fertility related to acidity or salinity and inconsistent seasonal rains (Ghassemi, Jakeman, and Nix 1995; Parnwell 1992). Environmental conditions vary across the district from steep slopes susceptible to erosion or infertile soils from laterization to alluvial soils in the lowlands that support high yield rice production (Ghassemi, Jakeman, and Nix 1995). Agriculture is largely rain-fed by the seasonal monsoons that occur between April and November and provide around 80 percent of the annual precipitation (Kaida and Surareks 1984). The strong seasonality of the rainfall in the region can lead to soil moisture deficit and drought during the dry season (Fukui 1993). Broadly, agricultural land use corresponds to water availability and soil fertility with rice cultivation in the lowlands and cassava or other cash crops like sugarcane grown in the uplands. Recently, agroforestry has been expanding, likely in relationship to global demands for wood and rubber.

In Nang Rong District, dwelling units are clustered together in villages with agricultural land distributed separate from households in the areas surrounding the village. The land holdings of individual villages are interspersed with those of neighbouring villages. In the lowlands, where rice cultivation is the rule, land parcels tend to be long and thin such that the short side of many parcels is adjacent to streams or rivers. Within each parcel, there are many bunds for rice cultivation with water managed between bunds. In the highlands, land parcels tend to be larger and have a lower perimeter to area ratio, which is more suitable for plantation type agriculture.

Nang Rong District's agricultural history is relatively recent. In the early 20th century and until the 1950s, Nang Rong District was an agricultural frontier (Rindfuss et al. 2007). Forested and relatively unattended, it drew new arrivals from other areas of the country. The recent migrants transformed the landscape from one dominated by upland and lowland forest to one dominated by rain-fed, lowland paddy rice. This land transformation primarily occurred through processes associated with human migration and settlement, deforestation, and agricultural extensification (Walsh et al. 2005). The initial transformation from forest to lowland rice, cultivated for subsistence, was followed by a shift to the cultivation of jasmine and heavy rice for the commercial market starting in the 1970s. During this same period, extensive changes were occurring in the forested uplands, as large tracts of land were converted to field crops, primarily cassava and sugar cane, for the international market.

Until quite recently, Nang Rong District experienced substantial amounts of population growth. During the early 20th century, immigration was the source of much of this growth, as death rates were high. Death rates dropped in the 1950s, though fertility remained high until the 1970s (Knodel 1987). This classic demographic transition scenario led to substantial population growth. Indeed, though fertility fell rapidly through the 1970s and 1980s, population sizes in the villages of the district continued to grow until at least 2000. Emigration from the district increased roughly in line with population growth, with the region becoming a substantial exporter of migrants to Bangkok and other industrializing areas of Thailand in the 1990s (Entwisle et al. *In Review*). Migration is now a fundamental aspect of villagers' lives as almost all households have members who are currently emigrants or returned migrants. Still, there is substantial variation from village to village in all of these patterns (Entwisle et al. 1996). These demographic features of the district can be expected to have important consequences for the management of land.

Data and Methods

Data

In the year 2000, a social survey was administered to all households in 92 study villages in the Nang Rong District as part of a longitudinal demographic study (Rindfuss et al. 2003). This survey reached nearly 10,000 households. Part of the survey, “Form 6,” asked questions about agricultural practices and land use such as the amount of land used, types of crops grown, yields of rice and cassava, and other agricultural practices.

In addition, households were asked to list the users of neighboring parcels. In Nang Rong District, farmers's dwelling units are separate from the *plangs* that are used for the cultivation of crops: generally, farmers living in nuclear village settlements do not live on the land that they farm, rather, their land parcels are geographically arrayed in a discontinuous fashion around the village, with some parcels 5 km or more away from the village centroid. This makes the linking of parcels to the people who farm them a challenging and complex matching process (cf. Rindfuss et al. 2004b). *Plangs* used by households were geo-located using existing cadastral maps and integrated with other data assembled within a geographic information system. In addition, this project used community participatory mapping that involved large-format, contemporary, panchromatic, aerial photography and GIS overlays of roads and other landmark features for recognition by discussion groups. Discussion groups were composed of village headman, hunters, long-term farmers, and other key informants who had an extensive knowledge about the land and the use of *plangs* by households in his/her village and nearby villages.

The list of variables is shown in Table 1. Crop presence locations were analyzed based on two types of variables – household variables, and plang (i.e. parcel) variables. Variables derived from the Form 6 questionnaire required linking households with plangs, which was performed through a land use matching process during the year 2000 fieldwork (Rindfuss et al. 2004b). For plang-level variables, the plang centroid and the associated environmental conditions at that location were used to represent the entire plang. Any households with plangs that could not be located were excluded from analysis. Most of the data came from the year 2000 social survey. However, some variables relied on earlier data. For example, digitized polygons from air photos from 1994 were used to map residential area and roads and paper maps from as early as the 1970s were digitized to obtain soil and topographic characteristics. It was assumed that these data accurately represented the built and natural environment in the year 2000.

A series of household-level variables were derived from Form 6 for the analysis. The variables [hh_size] and [hh_size1355] are measures of household size. While [hh_size] is a count of the total number of persons of any age within each household, excluding migrants, [hh_size1355] serves as a proxy of agricultural labour availability as this is the age group most likely to be farming. The presence of remitting household members is reflected in [num_hh_remit], which is a household tally of the number of persons that have migrated to Bangkok or other areas of the country and remit wages or commercial products back to the household.

Within Form 6, households reported a variety of information about the land that they use. There are two variables that relate to the landholdings of each household. The first, [plang_sum], is the total area, in square meters, of agricultural land in production by the household, derived by calculating the sum of all plangs used by the household, rented or owned. The second, [plang_sum], is the total number of separate parcels of land either used or rented by households. The variable [all_asset] measures, in baht, the assets of each household in the year 2000, through the combination of productive assets, consumptive assets, and mixed assets (Tong and Entwisle 2005). Productive assets refer to items possessed that contribute to the generation of marketable goods, such as sewing machines, tractors, and *itans* (a multipurpose vehicle with a variety of on-farm uses). Consumptive assets, or consumer goods, relate to items such as televisions, washing machines, cellular phones, computers, and refrigerators. Finally, mixed assets, are made up of those items that may fall into either or both of the previous asset groups, such as motorcycles and mopeds, cars, and pickup trucks.

In addition to the Form 6 data described above, several geographic and environmental attributes were used. Geographic attributes relating to the site characteristics of land surrounding a given plang, such as roads, and village centroids, were derived from digitized government maps, remotely sensed imagery, and field collected GPS points gathered during the social survey. The variable [pct_resid_pl] is a plang-level variable that was derived from photointerpreted aerial photographs that measures the development within the neighborhood of a plang. Mosaicked aerial photographs from 1994 were used to interpret and map urban clusters, or areas surrounding villages's centroids comprised of dwelling units, household-level land uses, and other residential structures. Next, the household-linked plangs were buffered to a distance of 5 km. These residential cluster polygons were then overlaid on the 5 km plang buffers, and the percentage of each buffer within the residential buffer was calculated. A related measure is [cnt_resid_pl], the count of residential clusters within the 5 km buffer area. Whereas the previous measure tabulates the area within residential land uses, this variable is simply the count of residential clusters intersected by each plang's buffer, which serves as a measure of village density. A third plang-level variable, [cent_dist], provides a measure of the household-accessibility of each plang by calculating the distance in kilometers to its associated village centroid. Similarly, [road_dist] is a plang-level variable that measures the accessibility of each plang to markets by calculating the distance in kilometers to the nearest major road.

Two environmental variables, [soils] and [dem], were generated from digitized government maps. [dem] is elevation based on a digital elevation model derived from the digitization of all 10 m topographic contour lines on the 1:50,000 scale Thai military base maps. As a supplement to the contours, a total of 1,373 spot elevation points, collected from the topographic maps, were used to obtain a finer level of detail. ESRI's ArcGIS was used for the preparation and processing of the DEM. All topographic sinks were removed prior to the construction of the DEM. The Topogrid tool was then used to integrate the contour lines with spot elevations, and to ensure hydrographic enforcement during the core processing of the DEM. [soils] is the UN soils classification derived from two 1:100,000 scale Thai military base maps. A 1974 soil survey map, created using aerial photography and field verification, contains detailed soil characteristics information relating to the soil Mapping Units, soil series, physical and chemical characteristics, crop suitabilities and capabilities, and crop limitations. The 1974 soil map contained a gap in the extreme southwest of the study district, outside the extent of the survey villages. A companion dataset, derived from a set of 1992 1:100,000 scale topographic/geomorphic maps, covers the entire study area, though it contains less descriptive information and has coarser spatial detail than the 1974 soil survey. Consequently, a combined dataset was created primarily using the 1974 soils map, with the 1992 topographic maps integrated through edge-matching and attribute comparison to extend the areal extent of soils data. Adjacent soil polygons within the two soils datasets that could be reliably matched based on aggregated mapping units were manually edited to create a surface without discontinuities in boundaries. In the event that compatible adjacent soil groups were not present, soil polygons from each source dataset remained unedited. In the areas where the 1992 topographic/geomorphic maps were used, the spatial accuracy is less than that of the 1974 map, but suitably comparable for this application.

The crops examined represent the current and emerging dominant crops in the district – cassava, fruit trees, heavy rice, and jasmine rice. As a subset analysis, jasmine rice was examined in terms of high, medium, and low yield productivity as determined by yield (kg/ha) to understand any correlations between productivity and household or land characteristics. High and low yields indicate the highest and lowest quartiles, respectively.

Based on the survey data, the number of plangs with cassava, jasmine rice, and heavy rice were 415, 5167, and 3950, respectively. The number of plangs with high yields and low yields of jasmine rice were 247 and 260, respectively. Fruit trees were uncommon in the study area in 2000 and found only only 16 plangs. It is important to note that the total area of each crop could not be calculated even though we have a good estimate of each plang size because multiple crops could be reported for each plang.

Methods

To test the relationships of the household and land characteristics with crop distributions, we used the geographic species distribution model, MaxEnt. MaxEnt is a geographic species distribution model that uses a deterministic machine learning algorithm to optimize environment-species relationships based on maximum entropy (Phillips et al. 2004; Phillips, Anderson, and Schapire 2006; Phillips et al. 2009). The inputs to MaxEnt are geographic presence locations of the species of interest and geographic environmental variables, both continuous and categorical. Data are fit using linear and non-linear functions and different function types can be hinged together. MaxEnt has been widely used in ecology to understand a wide range of species distribution applications and has been found to consistently perform among the best of available species distribution models based on validation using the receiver operating characteristic (Elith et al. 2006) and the spatial distribution of predictions from small training sets with full species occurrence data (Hernandez et al. 2006). MaxEnt provides several forms of output for analysis and assessment. Environmental variables can be examined using jack-knife parameter contribution to the model and species-environment relationship curves, both raw and controlling for other environmental variables. Different models and model runs can be compared using the “area under the curve” of the receiver operating characteristic (AUC-ROC), a common classification statistic adapted to the presence-only methodology.

The AUC-ROC statistic was developed for binary presence/absence classifications. In a presence-only framework, the AUC-ROC compares the results not to absence data but to the random background. Thus the interpretation of the AUC-ROC is not true “accuracy” but the ability of the model to define a pattern different from the random background (Lobo, Jimenez-Valverde, and Real.2008; Heumann, Walsh, and McDaniel 2011). Finally, MaxEnt produces maps of species likelihood occurrence. As several scientists have noted (Austin 2007; Jimenez-Valverde, Lobo, and Hortal 2008; among others), it is important to have a good ecological understanding of the species of interest and to analyze each component of the model to ensure accurate results. On-going development of the MaxEnt model based on feedback from the ecological modeling community has led to a number of improvements such as adjustments for spatial and environmental sampling bias (Phillips et al. 2009). For a more comprehensive explanation of MaxEnt, see Elith et al. (2011).

This analysis consisted of two phases. The first phase was an exploratory analysis to narrow down the relevant variables. For each crop, we conducted 10 runs using a 90/10 split for training and testing with replacement. A map of all plangs was used as the bias surface to account for environmental sampling bias. Over a series of 10-run iterations, variables were removed based on one of the following conditions: 1) the variable decreased training or testing gain; 2) the variable contributed less than 1 percent to the model; and 3) the species-variable response curve was static when all other variables were held at their mean.

As Pontius and Pacheco (2004) note, the strength of model prediction can be overestimated based on the calibration goodness-of-fit. Thus once the list of variables had been narrowed down for each crop, we conducted a second more rigorous set of 100 runs with cross-validation using a 60/40 split with replacement between training and test datasets. The results presented are the summary of these 100 runs for each crop type. For all model runs, linear, quadratic, and product functions were used and hinging of these functions was allowed. Threshold functions were removed because we observed that the combination of thresholds and hinging produced unrealistic over-fitting of environment-species relationship curves.

Assumptions, Simplifications, and Limitations

Our model has several assumptions and simplifications. First, human management of the agricultural landscape supersedes natural dispersal and biotic processes, resulting in environmental, both natural and built, and social conditions as the set of factors influencing crop distribution that can thus be used to define the niche. Second, human management is captured through a series of variables that either directly affects crop decision making or serve as a proxy for other factors. Third, since our household social data is limited to the agricultural plots of our study households, we can only assess the areas with social data and cannot extrapolate these results for the entire district like typical species distribution models. Forth, since we are working with presence-only data and analyzing a single crop at a time, our results indicate a potential rather than realized niche and competition between crops is not explicitly considered. Fifth, just as ecological niches have regional variation (Murphy and Lovett-Doust 2007), the socio-economic context has regional variation and thus the model is not directly transferable beyond our study villages, although the methodology is.

Results

Major Crops – Cassava, Heavy Rice, and Jasmine Rice

The contribution of each variable to the model of each crop is shown in Table 2. Not surprisingly, overall elevation [dem] and soil type are the dominant factors. Only two household level variables, total household assets [all_assets] and number of plangs used by the household [plang_num], contributed to all three crop models. (i.e. cassava, heavy rice, and jasmine rice). While household assets were consistently a very minor factor, the number of plangs used contributed 7 percent to heavy rice. The total area of plangs used [plang_sum] was also minor factor only for heavy rice.

The contribution of the built environment variables varied. While soils and elevation contributed less to heavy rice, the built environment variables were larger contributors. For example, the number of nearby residential areas [cnt_res_pl] was a major contributor (24.5 percent) for heavy rice, but a relative minor factor for cassava (3 percent) and jasmine rice (8.7 percent). Similarly, distance to the nearest major road [road_dist] contributed 7.1 percent to heavy rice, but just 2.9 percent for jasmine rice, and it was not a factor for cassava. The percent of nearby area that was residential [pct_resid_pl] was also a minor contributor to all three crop models.

The area under the curve of the receiver operating characteristic (AUC-ROC) can serve as a useful indicator of model fit. Figure 2 illustrates ROC curves of model sensitivity as a function of 1 – specificity, calculated as fraction predicted area in the Maxent model. The median and standard deviation of AUC values for training and testing for each crop are listed with the corresponding ROC curve. The differences between the testing and training datasets as well as the standard deviations are small, indicating robustness in the results. For the three dominant crop types, cassava, heavy rice, and jasmine rice, the AUC values are indicative of each crops relative distribution. Specifically, cassava, which is generally constrained to higher elevations where rice cannot grow, has a very high AUC, but the heavy rice and jasmine rice have lower AUC because of their widespread distribution throughout Nang Rong District. This point is discussed in more detail later in the article.

Figures 3, 4, and 5 illustrate the response curves of the likelihood of presence based on each input variable for cassava, heavy rice, and jasmine rice, respectively. For the two major contributors to the cassava model, elevation and soils, a clear response is shown. Cassava was mostly likely to grow in the highlands on specific types of soils, namely Korat and Phon Phisai, as well as Surin and Chok Chai to a lesser degree (Figure 3B). Plangs with cassava also tended to be in areas with a moderate number of smaller nearby villages (Figure 3 C, D). Household assets slightly increased the likelihood of cassava from 0 to 1000 baht, however, above 1000 baht, the uncertainty increased with assets such that no trend is detectable. Similarly, the number of plangs a household used slightly increased the likelihood of cassava.

Heavy rice was more likely to occur in locations with alluvial complex, Ratchaburi, Phen, Nam Phong, or Slope Complex soils and at higher elevations of the lowlands, between 190 and 200 m (Figure 4A). Locations more likely to have heavy rice also had a moderate to high number of residential areas nearby (Figure 4B), although the areal extent of this development did not seem to have a strong effect except at very low levels, and were within 2 km of a major road (Figure 4D). The household characteristics that increased the likelihood of heavy rice were a small number of plangs used but medium to large plang size and generally fewer household assets.

The response curves of jasmine rice fittingly contrast heavy rice. The natural environmental conditions where jasmine rice was more likely to occur had a single soil type, Slope Complex, and were located in the lower part of the lowlands (Figure 5A). The built environment conditions differed from the heavy rice in that jasmine rice was more likely to occur in areas with either high or low numbers of residential areas and within 4.5 km of a major road (Figure 5F). In terms of household characteristics, heavy rice and jasmine rice were similar in that fewer plangs used and fewer household assets were associated with a high likelihood of occurrence, although for jasmine rice variation in the response curves was less than that of the heavy rice.

Crop Yield – Jasmine Rice High and Low Yield

High yield jasmine rice was most strongly associated with plang area, [plang_sum], the majority contributing factor (See Table 2). The natural environmental variables, soils and elevation, contributed 20 percent and 14 percent, respectively. Other built environmental variables included the number of residential areas [cnt_res_pl] and distance to road. None of the household social variables contributed to the model.

The natural environmental factors, primarily soils, contributed the majority to low yield jasmine rice. The same built environmental variables as high yield jasmine rice also contributed to low yield jasmine rice, although [plang_sum] contributed just 9.7 percent to low yield. Additionally, one household social variable, the number of people in the household of agricultural working age [hh_size1355] contributed 9.1 percent.

The AUC of the high yield and low yield jasmine rice, considered separately, were higher than all jasmine rice yields (Figure 2). High yield jasmine rice had a higher AUC than low yield jasmine rice, indicating a more distinct pattern for the location of high yield jasmine rice. The test dataset AUC was lower than the training dataset for the high yield and low yield jasmine rice than the other crop datasets. Similarly, the standard deviation of AUC values was higher, indicating higher variability within a smaller dataset.

The response curves illustrate some distinct patterns for high yield jasmine rice locations (Figures 6 and 7). High yield jasmine rice was more likely to be found on very small plots and very unlikely to be found on larger plots. High yield jasmine rice was more likely to occur on several soil types, namely Slope Complex, Buri Ran, Roi Et / Renu Assoc., Rock Outcrop, and Nam Phong from 175 to 200 m elevation, with a moderate decrease in likelihood with increasing elevation but a sharp decrease at the lowest elevations. The built environmental conditions associated with increased likelihood were moderately high number of residential areas and locations within 2 km of a major road.

The response curves for low yield jasmine rice differed in several ways (Figure 7). The natural environmental conditions associated with increased likelihood of low yield jasmine rice included many soil types, namely Korat, Roi Et / Korat Assoc., Roi Et / Renu Assoc., and Satuk, and the upper parts of the lowlands with maximum likelihood at 185 m and a strong decrease in likelihood with either an increase or decrease in elevation especially above 205 m and below 185 m. The built environmental conditions associated with increased likelihood were moderately high number of residential areas, though with less severe decrease in likelihood towards moderate and moderately low numbers, and a maximum likelihood between 1 km and 1.5 km from a major road with a decrease in likelihood both closer and further away.

Emerging Crop – Fruit Trees

The last crop type is an emerging crop that is relatively uncommon but observed to be increasing in presence based on field observations in 2000 and 2010. In contrast to the other crops, social household variables were the dominant contributing factors to fruit trees (Table 2). Household assets contribute almost half to the model (49 percent). The number of household remitters [hh_remit] was a contributor only for fruit trees. Additionally, the number of plangs used by the household [plang_sum] contributed more for fruit trees than other crops. Soil type was the only natural environmental contributing factor, contributing 21 percent to the model, similar to cassava and high yield jasmine rice. The only built environmental factor was the number of residential plangs [cnt_res_pl] though it was the smallest contributing factor. The AUC for the fruit tree models was reasonably high, although the difference between the training and test AUC was much larger than other crop datasets, as was the standard deviation (Figure 2).

The household social characteristics associated with increased likelihood of fruit trees were household assets above 1000 baht, few remitters, and small plang size (Figure 8). Fruit tree were more likely to be grown on a few types of soil, namely Roi Et, Tha Uthen, Satuk, and Korat / Satuk Assoc., although the amount of likelihood for each soils type varied greatly between model runs. Fruit trees were also more likely to be grown in locations with a relatively moderate amount of development, although this factor was also highly variable such that no distinct pattern is evident.

Discussion

The role of the environmental setting for crop choice was consistently dominant among the major crop types. Although Nang Rong District consists largely of marginal land, how marginal the land is varies by location. The low soil fertility and water availability in the highlands generally excludes rice cultivation, hence the dominance of cassava, a crop promoted by the Thai government in these areas. In the lowlands, the choice between jasmine rice and heavy rice was related to topographic positioning and soil type, but the environmental conditions are not exclusive. Part of the co-presence of jasmine and heavy rice across environmental space can be explained by other factors such as distance to road or the density of residential areas nearby. However, these geographic factors may be a proxies for an unmeasured social phenomenon. During a focus group interview in 2010 with a village that grew both types of rice in similar environmental conditions, the question was asked about how and why certain types of rice are grown (Entwisle et al. unpublished data). The focus group responded that heavy rice was less valuable and generally planted in locations where jasmine rice grew poorly but that, at the household level, the decision to grow jasmine or heavy rice was largely traditional. In other words, households tended to grow heavy or jasmine rice based on a decision made years or decades ago and have continued the practice without consideration of change – not surprising in an area where much of the rice cultivation is for domestic consumption.

Overall, household-level factors about land or characteristics of the household itself played a relatively minor role in crop suitability with two notable exceptions – high yield jasmine rice and fruit trees. For high yield jasmine rice, the area that the household used was the strongest factor. The likelihood response curve for high yield jasmine rice shows that small areas have a stronger likelihood than large areas. While the likelihood of low yield jasmine rice also decreases with area, the strength of the contribution and effect are much lower. There are a few possible non-exclusive interpretations: 1) the extent of environmental conditions in which jasmine rice thrives is limited and this area is divided into small parcels or 2) households that have high yield parcels need less land for the equivalent yield compared to less productive land.

The other crop type in which a social factor was a strong contributor was fruit trees. In Nang Rong District, fruit trees are an emerging crop seen as a form of income diversification away from cassava or rice. Fruit trees are a relatively large, long-term investment as the cost of establishing a small orchard is expensive and the returns are delayed until the trees mature. This is exemplified by our results. The number one factor for the likelihood of fruit trees is household assets in which higher assets increased the likelihood of fruit trees, up to about 1500 baht. Above 1500 baht, fruit trees remained very likely to occur. Interestingly, the number of remitters in a household was also a substantial contributor but had a negative effect. This is unexpected as we expected that as the number of remitters increased, off-farm income increases and agricultural labour decreases. There may be interactive effects between assets and remitters that obscure the relationships and it may be that the loss of migrants (all remitters are, by definition, migrants) constrain household labor in such a way that they cannot afford to diversify (Tong and Entwisle 2005). However, remittances and assets are poorly correlated, likely due to confounding effects such as time lags.

At first glance, some of the relationships observed seem dubious, but they can be explained through a deeper understanding of the data and place. For example, the soil associated with high likelihood of jasmine rice was Slope Complex (i.e. soils too steep for agriculture) compared to the expected alluvial soils. This result occurred due to inconsistencies between the soils map and land use and the relative abundance of each soil type. Specifically, there were a substantial number of jasmine rice locations in an area of the relatively uncommon Slope Complex soils. Thus while Jasmine rice is grown on many types of soil and Slope Complex is not best suited to rice cultivation, the model identified jasmine rice on Slope Complex as a strong predictor of presence. The overlap of jasmine rice and Slope Complex soils is likely the factor of the relatively course soils mapping based largely on historical air photo interpretation and human-modification of the land. In other words, while the landscape looked like Slope Complex at the time of soils mapping, clearing and terracing of the land transformed it to suitable for jasmine rice. This result does not mean that the data or model are wrong, but that one is likely to find jasmine rice located on the lower side of a hillside with Slope Complex soils as these areas have been transformed into suitable locations. This instance strongly demonstrates the need to have in-depth understanding of the data and place of modeling.

Application to Land Use Modeling

One of the main rationales for this line of research is to help create the rules and relationships required for allocation in land use modeling. This section outlines one potential method of applying the relationships identified into a land use model. The methodology outlined creates rules based on the niche-based model that can be used to dynamically calculate land use likelihood for every parcel of land in a model as environmental, geographic, and social factors change over time. This approach aims to estimate the likelihood of all land use type occurring for a particular parcel from which a land use type can be selected stochastically. The steps of this methodology include the following: 1) creation of look-up tables (LUTs) of land use likelihood based on environmental factors, 2) weighted calculation of land use likelihood based on all factors according to proportion of contribution for a single land use, 3) conversion of single land use likelihoods to land use outcomes from multiple land use options.

The first step is to create LUTs of the functions illustrated in Figures 3 – 8. An LUT provides a simple and quick solution to solving these non-continuous functions. The LUTs are a static dataset that can be quickly referenced for future calculations. The intervals of the LUT are created at the discretion of the modeller based on the precision required for the application.

The second step is to calculate the likelihood of each land use type individually for each parcel of land. This relatively simple calculation is a weighted sum of the likelihoods calculated from the LUT for each factor. For example, to calculate the likelihood of fruit trees for a given parcel, the values of assets [all_assets], soil type [soils], number of parcels [plang_num], number of remitters [num_remit], and percent of residential area [pct_res_pl] are used to look-up the associated likelihood for each factor. These likelihoods are added together using the contribution values from Table 2. Equation 1 illustrates the weighted sums using fruit trees as an example.

$$\begin{aligned}\mathrm{TLH}=&0.489^{*}\ [\mathrm{LH\_\ all\_\ assets}]\\&+0.21^{*}\ [\mathrm{LH\_\ soils}]\\&+0.162^{*}\ [\mathrm{LH\_\ plang\_\ num}]\\&+0.098^{*}\ [\mathrm{LJ\_\ hh\_\ remit}]\\&+0.041\,[\mathrm{LH\_\ pct\_\ resid\_\ pl}]\end{aligned} \quad \text{(eq. 1)}$$

where TLH is total likelihood and LH is likelihood.

The third step is to convert the likelihood of a single land use type into a value that can be used to estimate the likelihood among all land use types. Each land use type is modeled individually with an outcome of relatively likelihood compared to the whole landscape. But to select a land use type compared to other land uses, the relative likelihood is required. For example, a given parcel maybe is very likely to have either heavy rice or jasmine rice compared to overall landscape, but the choice of land use is relative. To calculate the relative likelihood, the likelihood of each land use type is divided by the sum of likelihoods for all land use types. For example, if a given parcel had a likelihood of 0.7 for jasmine rice and 0.4 for heavy rice, the relative likelihoods would be 0.636 and 0.363, respectively.

It is important to note that the approach described above does not account for several aspects of land use often included in a dynamic land use model. Agent learning or adaptation, cultural inertia (growing a crop because it is traditional), or changes from external factors such as market prices are not explicitly considered. However, these factors can be included after the fact as coefficients that adjust the relative likelihoods. For example, the relative likelihoods of jasmine and heavy rice from the example above can be adjusted to reflect any other process that may affect land use decision making. This analysis also does not address the issue of crop quantity or yield. Crop yield would require another set of analysis based on socio-economic factors for farming inputs such as labour and biological processes of agricultural crops.

Conclusions

The patterns and processes of land use are intertwined with environmental, geographic, and socio-economic factors that affect the household decision making process. The relationship between these factors is often complicated and non-linear in nature. Tools based on machine learning algorithms such as MaxEnt can be useful for examining complicated human-environment interaction by adapting the niche theory to economic theory. Nang Rong District has transitioned through several stages of land use over the past decades, starting with forest conversion to rice cultivation in the 1950s, followed by land conversation for calorie crops in the 1970s. Since 2000, there has been increasing diversification of livelihoods and land uses particularly agro-forestry in terms of fruit trees, eucalyptus for wood, or rubber plantations. This research demonstrates how different factors affect the likelihood of presence for different types of crops and that these factors vary between crops. Specifically, our research shows that while the distribution of the dominant crops, rice and cassava, are largely environmental, emerging crops that require investment such as fruit trees are primarily related to household attributes such as household assets. Furthermore, the land use-environment relationships are non-linear and often discontinuous functions, which machine-learning algorithms are well suited to identify. We have also outlined a proposed methodology for incorporating our findings into a dynamic land use model such as an agent-based model. This research further contributes to our understanding of how the decision making process of land use allocation and the resulting land use patterns are often complex involving spatial and aspatial factors from both the landscape and household.

Acknowledgments

This research was supported by grants from the National Science Foundation-HSD: “Marginality in a Marginal Environment: An Agent- Based Approach to Population-Environment Relationships” (B. Entwisle, PI), NSF 0728822; and the National Institutes of Health-NICHD “Modeling Household Dynamics and Land Use” (B. Entwisle, PI), 1 R21 HD051176. Thanks to Peter Mucha, George Malanson, Megan Rua, Josh Gray, Carla Jensen and the anonymous reviewers for comments that greatly improved this manuscript.

References

ArcGIS 9.2 ESRI (Environmental Systems Resource Institute). Redlands; California, USA:

Araujo MB, Guisan A. Five (or so) challenges for species distribution modelling. Journal of Biogeography. 2006; 33(10):1677–1688.

Austin MP. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling. 2002; 157(2–3) PII S0304-3800(02)00205-3.

Austin MP. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling. 2007; 200(1–2):1–19.

Bearman PS, Moody J, Stovel K. Chains of affection: The structure of adolescent romantic and sexual networks. American Journal of Sociology. 2004; 110(1):44–91.

Bruch EE, Mare RD. Neighborhood choice and neighborhood change. American Journal of Sociology. 2006; 112(3):667–709.

Cashdan, E. Risk and uncertainty in tribal and peasant economies. Westview; Boulder: 1990.

Chambers, R.; Conway, G. IDS Discussion Paper. 1992. Sustainable rural livelihood: Practical concepts for the twenty-first century.

Chen H, Pontius RG. Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. Landscape Ecology. 2010; 25:1319–1331.

Clarke, KC.; Gazulis, N.; Dietzel, CK.; Goldstein, NC. A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. In: Fisher, PE., editor. Classics from IJGIS. Twenty Years of the International Journal of Geographical Information Systems and Science. Taylor and Francis, CRC; Boca Raton, FL: 2007. p. 413-425.

de Sherbinin A, VanWey L, McSweeney K, Aggarwal R, Barbieri A, Henry S, Hunter L, Twine W, Walker R. Household demographics, livelihoods and the environment. Global Environmental Change. 2007; 18:38–53. [PubMed: 19190718]

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE. Novel methods improve prediction of species's distributions from occurrence data. Ecography. 2006; 29(2):129–151.

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions. 2011; 17(1):43–57.

Ellis F. Household strategies and rural livelihood diversification. Journal of Development Studies. 1998; 35(1):1–38.

Ellis, F. Rural livelihoods and diversity in developing countries. Oxford University Press; New York, NY: 2000.

Entwisle B, Rindfuss RR, Guilkey DK, Chamratrithirong A, Curran SR, Sawangdee Y. Community and contraceptive choice in rural Thailand: A case study of Nang Rong. Demography. 1996; 33(1): 1–11. [PubMed: 8690134]

Entwisle B, Rindfuss RR, Walsh SJ, Page PH. Population growth and its spatial distribution as factors in the deforestation of Nang Rong, Thailand. GeoForum. 2008; 39(2):879–897.

Entwisle B, Walsh SJ, Rindfuss RR, VanWey LK. Population and upland crop production in Nang Rong, Thailand. Population and Environment. 2005; 26(6):449–470.

Entwisle B, Verdery AM, Rindfuss RR, F. K. Social networks and migration: A view from origin. American Sociological Review. In Review.

Fukui, H. Food and population in a northeast Thai village. University of Hawaii Press; Honolulu: 1993.

Ghassemi, F.; Jakeman, AJ.; Nix, HA. Salinisation of land and water resources: Human causes, extent, management and case studies. University of New South Wales Press Ltd; Englewood Cliffs, NJ: 1995.

Goland C. Field scattering as agricultural risk management: a case study from Cuyo, Department of Puno, Peru. Mountain Research and Development. 1993; 13:317–338.

Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecological Modelling. 2000; 135(2–3):147–186.

Hernandez PA, Graham CH, Master LL, Albert DL. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography. 2006; 29(5):773–785.

Heumann B, Walsh SJ, McDaniel P. An assessment of a presence-only model for crop suitability mapping, Nang Rong, Thailand. Ecological Informatics. 2011; 6(5):257–269. [PubMed: 21860606]

Hirzel AH, Le Lay G. Habitat suitability modelling and niche theory. Journal of Applied Ecology. 2008; 45(5):1372–1381.

Jimenez-Valverde A, Lobo JM, Hortal J. Not as good as they seem: The importance of concepts in species distribution modelling. Diversity and Distributions. 2008; 14(6):885–890.

Kaida, Y.; Surarerk, V. Climate and agricultural land use in Thailand. In: Yoshino, MM., editor. Climate and agricultural land use in Monsoon Asia. University of Tokyo Press; Tokyo: 1984. p. 231-254.

Knodel J. Thailand's reproductive revolution. Social science. 1987; 72(1):52–6. [PubMed: 12341219]

Little, MA.; Leslie, PW. Turkana herders of the dry savanna: Ecology and biobehavioral response of nomads to an uncertain environment. Oxford University Press; Oxford: 1999.

Lobo JM, Jimenez-Valverde A, Real R. AUC: A misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography. 2008; 17(2):145–151.

Massey DS. Social structure, household strategies, and the cumulative causation of migration. Population Index. 1990; 56:3–26. [PubMed: 12316385]

Manson SM. Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. Agriculture, Ecosystems and Environment. 2005; 111(1):47–62.

Murphy HT, Lovett-Doust J. Accounting for regional niche variation in habitat suitability models. Oikos. 2007; 116(1):99–110.

Netting RM. Household dynamics in a nineteenth-century Swiss village. Journal of Family History. 1979; 4:39–58.

Ortega-Huerta MA, Peterson AT. Modeling ecological niches and predicting geographic distributions: A test of six presence-only methods. Revista Mexicana De Biodiversidad. 2008; 79(1):205–216.

Parnwell MJG. Southeat-Asia - A region in transition - A thematic human-geography of the Asean region. Progress in Human Geography. 1992; 16(4):651–652.

Peterson AT. Predicting the geography of species's invasions via ecological niche modeling. Quarterly Review of Biology. 2003; 78(4):419. [PubMed: 14737826]

Peterson AT, Papes M, Eaton M. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography. 2007; 30(4):550–560.

Phillips, SJ.; Dudík, M.; Schapire, RE. A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-First International Conference on Machine Learning; 2004. p. 655-662.

Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecological Modelling. 2006; 190(3–4):231–259.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications. 2009; 19(1):181–197. [PubMed: 19323182]

Pontius RG, Pacheco P. Calibration and validation of a model of forest disturbance in the Western Ghats, India. GeoJournal. 2004; 61(4):325–334.

Pulliam HR. On the relationship between niche and distribution. Ecology Letters. 2000; 3(4):349–361.

Rajbhandari, B. Sustainable livelihood and rural development in South Asia. In: Behera, MC., editor. Globalising rural development: Competing paradigms and emerging realities. Sage Publications; New Delhi: 2006. p. 211-241.

Rindfuss RR, Entwisle B, Walsh SJ, Mena CF, Erlien CM, Gray CL. Frontier land use change: Synthesis, challenges, and next steps. Annals of the Association of American Geographers. 2007; 97(4):739–754.

Rindfuss, RR.; Turner, BL., II; Entwisle, B.; Walsh, SJ. Land Cover/Use and Population. In: Gutman, G., editor. Land Change Science: Observing, Monitoring, and Understanding Trajectories of Change on the Earth's Surface. Academic Publishers; Kluwer: 2004a. p. 351-366.

Rindfuss RR, Walsh SJ, Turner BL II, Fox J, Mishra V. Developing a science of land change: challenges and methodological issues. Proceedings of the National Academy of Science. 2004b; 101(939):13976–13981.

Rindfuss, RR.; Walsh, SJ.; Entwisle, B.; Sawangdee, Y.; Vogler, JB. Household-parcel linkages in Nang Rong, Thailand: Challenges of large samples. In: Rindfuss, RR.; Walsh, SJ.; Jefferson Fox, VM., editors. People and the environment: Approaches for linking household and community surveys to remote sensing and GIS. Kluwer Academic Publishers; Boston, MA: 2003. p. 131-172.

Robbins P. Comparing invasive networks: Cultural and political biographies of invasive species. The Geographical Review. 2004; 94(2):139–156.

Soberon J. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters. 2007; 10(12):1115–1123. [PubMed: 17850335]

Stark, O. The Migration of Labor. Basil Blackwell; Cambridge, MA: 1991.

Thuiller W, Lavorel S, Araujo MB. Niche properties and geographical extent as predictors of species sensitivity to climate change. Global Ecology and Biogeography. 2005; 14(4):347–357.

Tong, Y.; Entwisle, B. Impact of migration and remittances in households in rural Thailand. Annual Meeting of the Population Association of America; Philadelphia. 2005.

Walsh, SJ.; Entwisle, B.; Rindfuss, RR.; Page, PH. A Cellular automata approach to land use change modeling in northeastern Thailand. Paper read at Proceedings, GIS Planet 2005, International Conference and Exhibition on Geographic Information, at Estoril Congress Center; Portugal. 2005.

Walsh SJ, Entwisle B, Rindfuss RR, Page PH. Spatial simulation modeling of land use/land cover change scenarios in northeastern Thailand: a cellular automata approach. Journal of Land Use Science. 2006; 1(1):5–28.

Walsh SJ, Crawford TW, Welsh WF, Crews-Meyer KA. A multiscale analysis of LULC and NDVI variation in Nang Rong District, northeast Thailand. Agriculture Ecosystems and Environment. 2001; 85(1–3):47–64.

Winterhalder B, Lu F, Tucker B. Risk sensitive adaptive tactics: Models and evidence from subsistence studies in biology and anthropology. Journal of Archaeological Research. 1999; 7(4): 301–347.

Zimmerer, KE. Globalization and new geographies of conservation. The University of Chicago Press; Chicago: 2006.

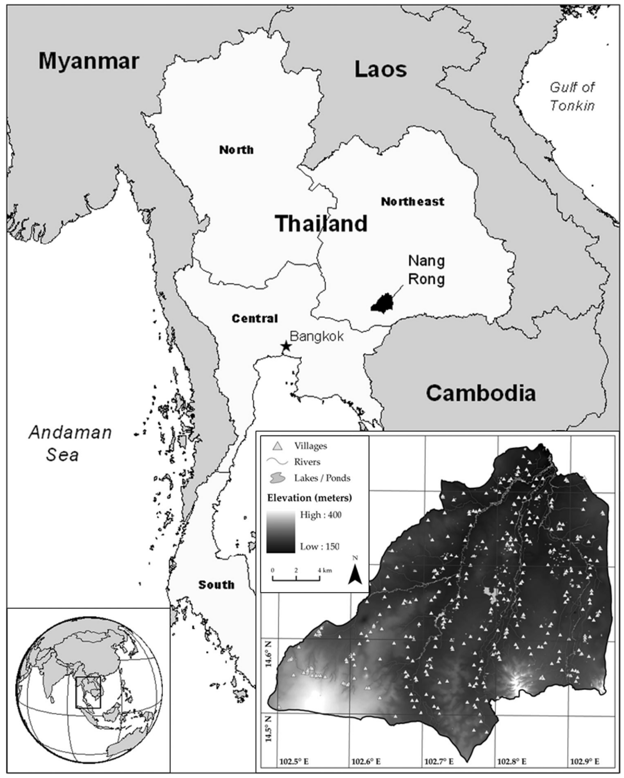

Figure 1.
Study area of Nang Rong District in northeastern Thailand

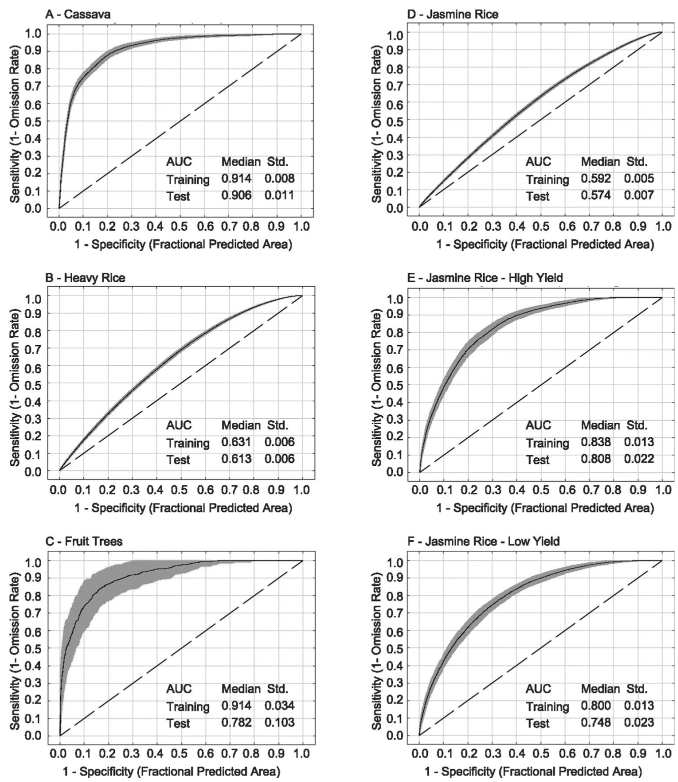

Figure 2.
Receiver operating characteristic (ROC) curves for cassava (A), heavy rice (B), fruit trees (C), jasmine rice (D), jasmine rice – high yield (E), and jasmine rice – low yield (F). The median and standard deviation (std) of the area under the curve (AUC) for the training and testing data are shown in each subplot.

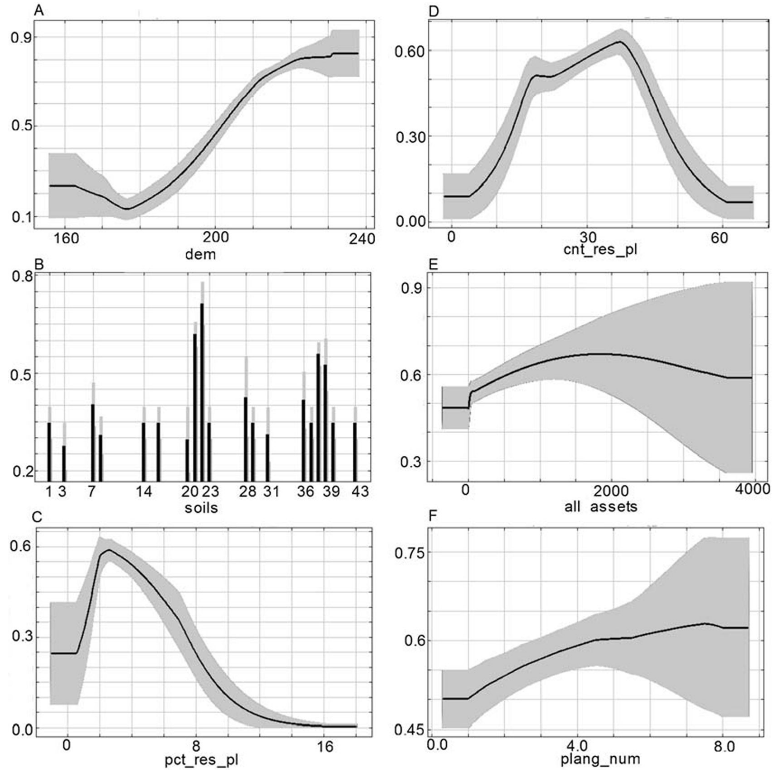

Figure 3.

Cassava response curves for elevation (A), soils (B), percent residential area (C), count of residential areas (D), household assets (E), and number of plangs (F).

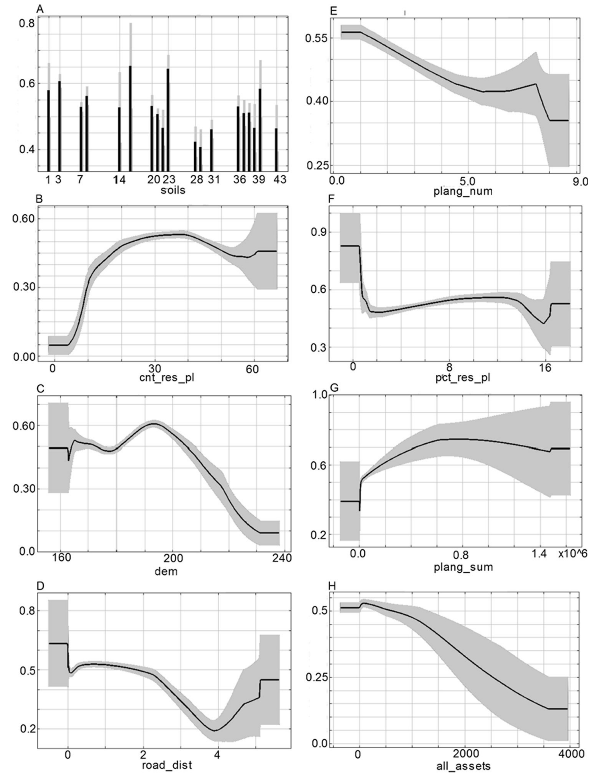

Figure 4.
Heavy rice response curves for soils (A), count of residential areas (B), elevation (C), distance to road (D), number of plangs (E), percent residential area (F), sum of plang area (G), and household assets (H).

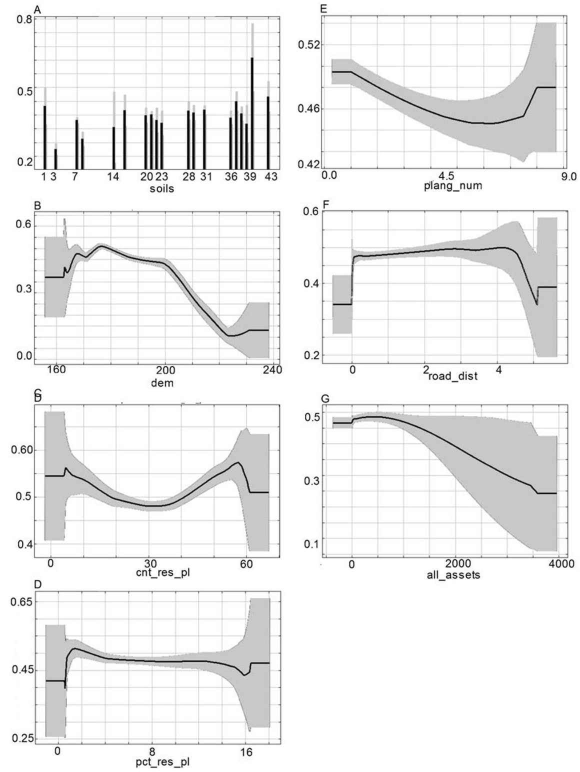

Figure 5.
Jasmine rice response curves for soils (A), elevation (B), count of residential areas (C), percent residential area (D), number of plangs (E), distance to road (F), and household assets (G).

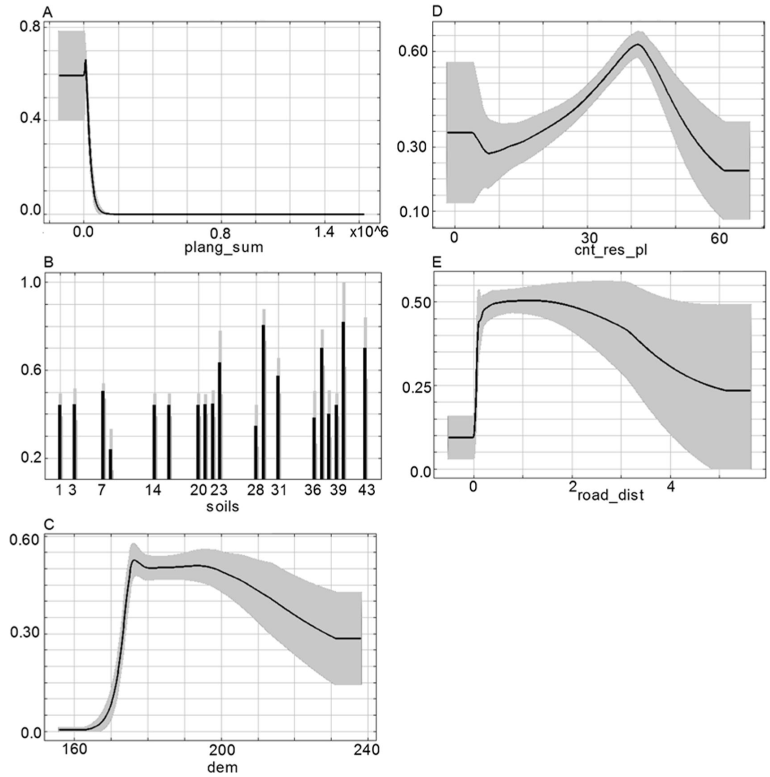

Figure 6.

High yield jasmine rice response curves for sum of plang area (A), soils (B), elevation (C), count of residential areas (D), and distance to road (E).

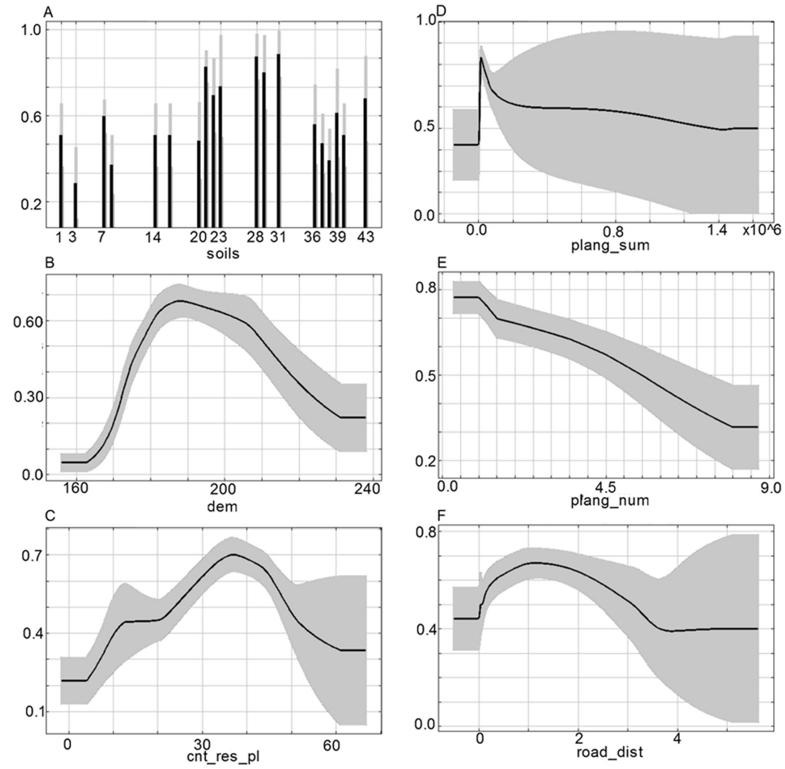

Figure 7.

Low yield jasmine rice response curves for soils (A), elevation (B), count of residential areas (C), sum of plang area (D), number of plangs (E), and distance to road (F).

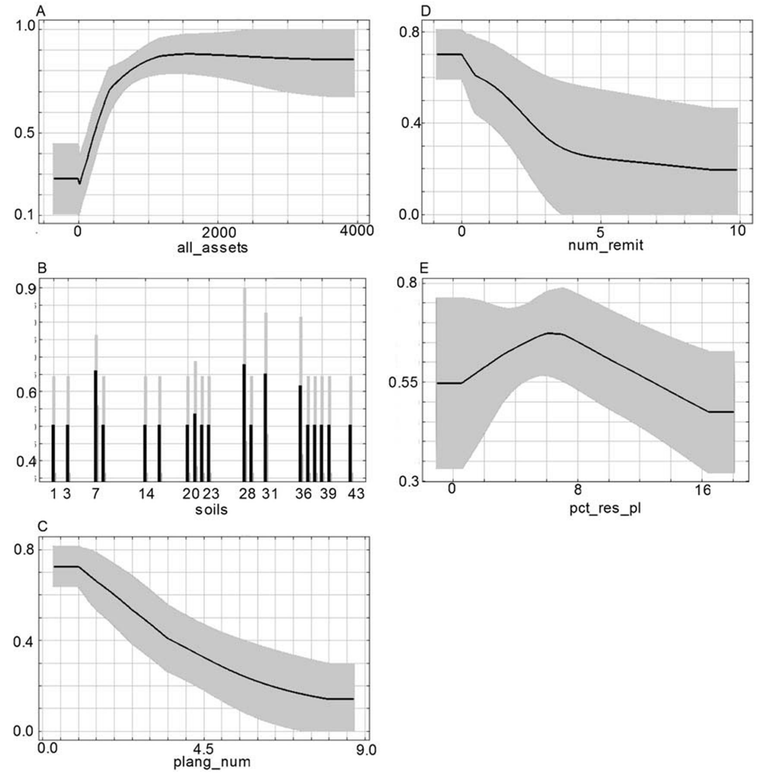

Figure 8.

Fruit tree response curves for household assets (A), soils (B), number of plangs (C), number of remitters (D), and percent residential area (E).

Table 1

List of social, built, and natural environmental variables and descriptions.

Variable	Type	Range	Description	Year of Data
HH_SIZE	HH	0 – 15	Number of persons in a household	2000
HH_SIZE1355	HH	0 – 8	Number of persons age 13–55 in a household - proxy of available agricultural labor	2000
ALL_ASSET	HH	0 – 4000	The value of all household assets in year 2000 baht (Entwisle and Tong, 2005)	2000
NUM_HH_REMIT	HH	0 – 10	Number of persons in the household that have migrated to Bangkok and remit wages back to the household	2000
PLANG_SUM	HH	0 – 1,600,000	Total area (units) of agricultural land used by the houshold, both rented and owned	2000
PLANG_NUM	HH	0 – 10	Total number of separate parcels the houshold used, both rented and owned	2000
PCT_RESID_PL	P	0 – 18	Percentage of area within a 5km buffer of the plang that is residential (i.e. non-agriculture)	1994
CNT_RES_PL	P	0 – 65	Number of residential polygons within a 5km buffer of the plang - an indicator of number of villages or village density	1994
CENT_DIST	P	0 – 5	Distance (km) from the plang to the household's village center - an indicator of accessiblity of plang by the household	2000
ROAD_DIST	P	0 – 5	Distance (km) from the plang to the nearest major road - an indicator of accessibility ofplang to markets	1994
SOIL	P	0 – 43	Categorical soil classes based on UN soil classification scheme	1974
DEM	P	155 – 240	Elevation (m) of the plang based on interpolated contour intervals from military topographic maps	1992

Table 2

Percent contribution of each variable for each crop model.

	Cassava	Heavy Rice	Jasmine Rice	JR - High	JR - Low	Fruit Trees
HH_SIZE	*	*	*	*	*	*
HH_SIZE1355	*	*	*	*	9.1	*
ALL_ASSET	1.4	1.8	1.6	*	*	48.9
HH_REMIT	*	*	*	*	*	9.8
PLANG_SUM	*	2.8	*	53.1	9.7	*
PLANG_NUM	1.1	7	3.2	*	*	16.2
PCT_RESID_PL	4.4	5.6	4.4	*	*	4.1
CNT_RES_PL	3	24.5	8.7	8.3	13.6	*
CENT_DIST	*	*	*	*	*	*
ROAD_DIST	*	7.1	2.9	4.4	7.6	*
SOIL	17.5	27.2	40.5	19.9	40.7	21
DEM	72.2	24.2	38.6	14.3	19.2	*

* indicates variables excluded due to negative or negligible contribution.