The Hawai`i Island Crop Probability Map: An Update of the Crop Growth Parameters for the Hawai`i County Crop Model

Final Report prepared for the County of Hawai`i Research and Development Department

Karen K. Kemp

GIS research laboratory

Technical Report No. 13

Abstract

The Hawai`i Island Crop Probability Map is an interactive desktop computer application that provides easy access to the results of a mathematical agricultural land suitability model. This project used a Maximum Entropy modeling technique to analyze information about environmental conditions at the current locations of agricultural crops on Hawai`i Island in order to predict the probability of suitable conditions existing for the same crop at other locations on the Island. This document describes the model and data used in this analysis and explains how to interpret and use the results.

Contents

Introduction

The Hawai`i Island Crop Probability Map is an interactive desktop computer application that provides easy access to the results of a mathematical agricultural land suitability model. This project analyzed information about environmental conditions at the current locations of agricultural crops on Hawai`i Island to predict the probability of suitable conditions existing for the same crop at other locations on the Island.

This project was initially designed to address a shortfall in the outcome of a 2008 Hawai`i County R&D funded project called the Hawai`i County Crop Model. The original one-year funding in July 2008 to the University of Redlands and The Kohala Center focused on the development of an interactive, web-based agricultural land planning tool which would assist country agricultural planners and others to assess environmental conditions at given locations on the Island and advise on the kinds of crops that might be grown there.

The original system was delivered and deployed on a web server behind the County of Hawai`i firewall in May 2009. It was not possible at that time, due to County data system restrictions and lack of funding, to make the system available to the general public. Eventually, as a courtesy to the County, the system was replicated and updated, through an unfunded partnership between the Redlands Institute of the University of Redlands and the Spatial Data Analysis and Visualization Lab (SDAL) at the University of Hawai`i at Hilo. It thus became available to the public for one year while surplus system resources at UHH could be used for this purpose. Unfortunately, that site is no longer active. This project, therefore, not only builds upon this earlier project as outlined in my original proposal, but it now provides updated, easy access to a related product.

The development plan for the original Hawai`i County Crop Model project intended to make use of existing published materials about crop growth parameters for a wide range of Hawai`i Island crops. Unfortunately, publicly available information about growth parameters did not exist in a readily computable format; information from CTAHR was in qualitative textual format, not easily converted to numeric ranges. Therefore, the developers relied on a limited set of information from a 1997 report on Hamakua agriculture potential. As a result, the "crop model" in the internal and public websites was purely for functionality demonstration purposes.

This current project has developed a more scientifically-sound crop model. It was initially intended to "reverse engineer" a preliminary 2008 dataset created by the NRCS called "CroplandsRangelandsForestlands". Fortunately, in the same year as this current project began, the CoH R&D Department also funded the SDAL at UHH to undertake the *Hawai'i County Food Self-Sufficiency Baseline Study* (Melrose and Delparte, 2012). In early 2012, the lab produced a new crop dataset that provided the basis for the statistical modeling undertaken in this project.

The modeling framework used here is the maximum-entropy approach to species habitat modeling (Phillips et al., 2004). By analyzing georeferenced occurrence locations, it is possible to predict the probability of other locations having suitable growth conditions based on the existence of similar environmental characteristics. The outcome is a series of Crop Probability Maps, one for each of eight main crop types included in the UHH data set.

Materials and Methods

This project relied heavily upon the use of GIS (Geographic Information Systems), in particular ArcGIS from Esri. It was required to create all of the data used in the modeling effort and thanks to the ability to directly import the output files from the modeling process, all results are available as map layers. These and all environmental data used are included in the CropProbabilityMap map document provided as the main product of this project.

The Modeling Technique

The basis for this project is the general notion that knowledge about environmental conditions at locations where particular plant species are successfully grown should provide a basis for summarizing crop growth parameters throughout the region. This notion is borne out by the popularity of the general purpose machine learning technique called Maximum Entropy Modeling (MaxEnt). A maximum-entropy estimate, as originally described by Jaynes (1957: 620), is "the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information". It has found strong support in the ecology domain as a means for predicting the spatial distribution of species from a limited set of occurrence or presence-only records.

The MaxEnt technique estimates an unknown probability distribution that "satisfies any constraints on the unknown distribution that we are aware of, and that subject to those constraints, the distribution should have maximum entropy" (Phillips et al., 2004: 233). In information theory, entropy is randomness or unpredictability, meaning that the portion that is not explained by the probability distribution has no remaining information with respect to the distribution of the prior data. Thus the result of a maximum entropy model is the best possible "description" of the distribution of the prior data. The benefit of MaxEnt is that you do not need to specify the determining conditions completely. In the case of this project, the result is a probability distribution of a crop type that reflects the environmental constraints that have been observed to be associated with the locations of existing crops.

The modeling process uses an iterative heuristic technique. Thus a first "guess" is made of the solution and its result is tested. The solution is repeatedly permuted slightly and retested. Solutions that improve the model are kept as the preferred solution for further permutation and ones which do not are discarded. The process is iterated hundreds of times, a process that is called "training the model", gradually moving toward a stable solution that is far better than the original random solution.

The MaxEnt modeling tool used here is available as a free software download from http://www.cs.princeton.edu/~schapire/maxent/. Several articles describe its use in ecological modeling and explain the various parameters and measures involved (see Elith et al., 2011; Phillips et al., 2006; Phillips et al., 2004). Importantly for this project, this technique can take the environmental conditions at occurrence locations and produce a probability distribution that can then be used to assess every other location for its likely occurrence. The result is a map of the probability of conditions being favorable to occurrence.

Since this report is intended for an audience with only basic statistical background, those who wish more depth on this technique are directed to those excellent sources. More about the choices made in the application of this technique in this study are provided in the Results section below. This tool requires the input of a set of raster data layers which provide the spatial distribution of the environmental constraints in the landscape and a set of point data to represent the occurrence data. Output from the model comes in the form of excel tables and raster layers. All of these are easily produced and managed within GIS.

Existing Crop Data

For the MaxEnt model, we need occurrence data. Fortunately, the crop dataset developed recently at UHH (Melrose and Delparte, 2012) provides a reliable and sound basis for this analysis. This data was developed through an examination of various data sources and was confirmed by analysis of satellite and photographic imagery and updated by local knowledge. No better representation of the current distribution of Hawai`i Island agriculture is available.

The UHH dataset includes only crop areas greater than 3 acres in area. Existing croplands were classified into 11 categories. In this study only eight of those were used. Aquaculture and Dairy were not included since their locations are determined by factors beyond the basic land-based environmental conditions considered for the remaining crops. There are so few Taro plots recorded in that dataset that prediction results are very unstable. The crop classes covered by this study and the resulting Crop Probability Maps are described in Table 1.

Table 1 - Crop categories used in this study (source Melrose and Delparte, 2012)

Crop Class	Crop Types
Flowers and Foliage	Potted Plants, Ornamental, Dracena, Orchids, Antherium, Nursery plants, Cut Flowers, Landscape plants, lei flowers
Tropical Fruits	Lichee, Rambutan, Longon, Mangosteen, Mango, Dragon Fruit, Avocado, Oranges, other
Papaya	Active production and fallow/cleared
Banana	
Coffee	All varieties and locations
Macadamia Nuts	
Specialty Crops	Mushrooms, Vanila, Cacao, Tea, Noni, Awa, Heart of Palm
Truck Crops	Commercially grown Vegetables, Melons, Squash, leaf and root crops

To produce the set of sample points required by the MaxEnt tool, the polygons of the crop data were intersected with the 100m grid framework used for the environmental data (see below) and with the soil map unit polygons. This produced a large number of smaller polygons identified with a single crop type and soil map unit nesting within the cell boundaries of the other environmental data. To reduce the size of the dataset, all intersected polygons with an area smaller than 1000 sq m were removed. Given the number of polygons, this loss of information was considered irrelevant. Finally a point was calculated in the center of each of these remaining polygons (Figure 1). Note that in the legend there is a Macadamia Nuts/Coffee category. Points falling in this category were duplicated in the dataset and retyped once as “Coffee” and once as “Macadamia Nuts”.

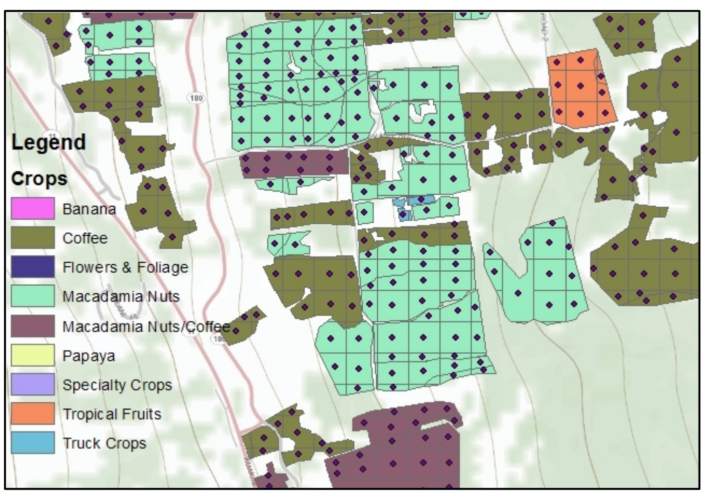

Figure 1- Crop point samples at center of each intersected polygon.

These points then became the “samples” dataset used in the MaxEnt tool. After intersecting these points with all of the data layers described below, associated with each point were all of the environmental conditions found at that location. The final samples dataset has 31,807 points with the count breakdown shown in Table 2.

Table 2 - Count of crop points used in the MaxEnt samples dataset

Crop Types	Count
Banana	446
Coffee	6014
FlowersFoliage	1547
Macadamia Nuts	15173
Papaya	1973
Specialty Crops	284
Tropical Fruits	2938
Truck Crops	3432

Environmental Data Layers

In order to describe the environmental constraints determining the distribution of the existing crops and to provide a basis for predicting where similar conditions exist, it was necessary to compile a set of GIS data layers representing a large number of environmental conditions. Fortunately, these data are easily available on-line. In addition to soil characteristics, rainfall, solar radiation, elevation, slope and temperature provided variables for the modeling effort. The source and development of each these data layers is described briefly in this section. Soils data are supplied and extracted quite differently from the other kinds of data, so the soils characteristics are handled separately later in this section.

It was necessary that all of the raster environmental layers used in the modeling process be converted into the same coordinate system, projection and cell resolution. After some initial data exploration and consideration of the scale and precision of the available data, the raster modeling framework was set to UTM NAD 83 Zone 5N coordinate system with a 100mx100m cell size. All environmental data layers were transformed and extracted into this geographic framework. Final rasters have a size of 1315 by 1507 cells which of course includes a large number of bounding "no data" cells over the ocean.

Table 3 lists the set of non-soil environmental layers included in this study. It provides their sources and outlines the transformations that were used to create the final raster layers.

Table 3 – Non-soil environmental layers included in this study.

Theme **- layer name**	**Source**	**Comments**
Elevation - elevation	10m elevation grid available from gis.ess.washington.edu/data/raster/tenmeter/hawaii/index.html	This grid was transformed to the correct projection and aggregated to the larger cell size by calculating the mean value of all included cells.
Slope - slope	Calculated from 10m elevation grid available from gis.ess.washington.edu/data/raster/tenmeter/hawaii/index.html	Calculated from the original elevation data using the Slope command in ArcGIS.

Theme - layer name	Source	Comments
Temperature - tempmaxann - tempminann	PRISM Climate Group at Oregon State University, http://prism.oregonstate.edu/index.phtml	Interpolated temperature surfaces for average temperatures over the period 1971-2000. Original data in 15 arc second cells (in Hawai ` i th approximately 500m). Data transformed and downsampled to 100m resolution.
Rainfall -rain[month] -rainmin -rainmax -rainann	2011 Rainfall Atlas of Hawai ` i, http://rainfall.geography.hawaii.edu/.	Data averaged over the 30 year base period 1978-2007. Original data source has 250m resolution. Produced 15 separate raster layers for average rainfall for each month, average total annual rainfall and two additional rasters that provide the maximum and minimum monthly rain recorded in each cell.
Solar Radiation - solrad	Two sources considered: • Global Horizontal Irradiance from the National Renewable Energy Laboratory www.nrel.gov/gis/data_solar.html). • "Contour" polygons of estimated daily solar insulation from Hawai ` i S GIS Program, www.state.hi.us/dbedt/gis/.	GHI data available as gridded data at 10km resolution and very smoothed contours at an interval of 50 calories/sq.cm/day provided only very coarse data in either case. An attempt to integrate the two sources by converting units and overlaying was not successful and the large jumps in value between cell or between-contour-area boundaries caused problems in the model. *This data was not included in the final model.*

Several different soil conditions were included in the initial modeling effort. Patrick Niemeyer, USDA Soil Scientist (now retired), provided a list of characteristics to Caroline Neary (Neary, 2011) that could be considered the key determining soil factors for crop growth in Hawai`i (Table 4 – Soil layers included in this study ). Using a very useful tool available from the USDA called the *Soil Data Viewer* that works within ArcGIS and the most recent soil survey data for Hawai`i Island available from the USDA, I extracted separate data layers for each of these components and produced 100mx100m cell rasters for each of them.

Table 4 – Soil layers included in this study

Characteristic	Layer name	Description	Depth Range Choice	Value Range
pH	pHwater	Measurement of acidity or alkalinity.	0-30 inches	0, 3.9-8.4, mean 6.0
Bulk density	Db3rdbar	Indicates pore space available at roots.	0-30 inches	0,.15-1.81, mean .59
Available Water Capacity	AWS100	Quantity of water that soil is capable of storing for use by plants	0-30 inches	0, 1-32.16, mean 10.0
Organic matter	OrgMatter	Plant and animal residue. Source of nitrogen and nutrients.	0-30 inches	0, .05-52, mean 9.2
Surface texture	TextCode	Percentages of sand, silt and clay.	All Layers	Many codes, no 0
Depth to any soil restrictive layer	Dep2ResLyr	Depth to layer that impedes movement of water and air.	All Layers	0, 6-500, mean 131

Characteristic	Layer name	Description	Depth Range Choice	Value Range
Drainage Class	DrainCode	Frequency and duration of wet periods under conditions similar to those in which soil is formed	All Layers	0, 1-7 populated
Flood Class	FloodCode		All Layers	2 and 3 only
Map unit	NameCode	Descriptive name	Map Unit	
Representative slope	SlopeSoil	Gradient in difference in elevation between two points	All Layers	0, 1-85, mean 16.5 AND 0-85, mean 16.4
Effective Cation-Exchange Capacity	ECEC	Ability to retain cations reduces hazard of ground water pollution. Lower CEC may require more fertilizer.	0 - 30 inches	0, 2-60, mean 12.5
Crop Productivity Index		Index developed to indicate crop productivity	Did not populate, insufficient data.	

Finally, the crop points were intersected with all of these environmental raster layers to extract environmental values for each point. This produced a samples table in which each of the 31,000+ rows includes entries for Crop_Type, X (longitude), Y (latitude), and a value for each included environmental layer.

Results

With all the data in hand, it was possible to begin making test runs with the MaxEnt tool to explore the impact of various tool settings and different combinations of environmental variables. Output from the tool provides extensive graphics and tables that make it easy to assess the results. An example of the full output for a single replication run is included as an "exhibit" in the Appendix.

Reasoning that the distribution of current crops does reflect in general most of the areas where they would do well, assessing how the resulting distribution maps reflected the known distribution of crops seemed to be the most reasonable way to determine the somewhat arbitrary settings. In most cases, tweaking of these parameters changed the final results only slightly with the same good places for a crop consistently shown in any variation. According to the maximum entropy technique foundations, any logically relevant set of environmental conditions should generally predict the same distribution. This appears to be borne out in the various attempts made here. The final settings used to produce the probability maps in the map document product are shown in

Table 5.

Table 5 – MaxEnt Parameters used in final model

Parameter	Value	Description
Model output type	Logistic	The default output is logistic, which is the easiest to conceptualize: it gives an estimate between 0 and 1 of probability of presence. (This number was multiplied by 100 for the final Probability maps.)
Replicates	100	Since the model is heuristic, a sufficiently high number of replications is needed to converge in the average upon the most suitable result.
Default Prevalence	0.8	While the tool tutorial states that this value is "fairly arbitrary", it is used to indicate the probability of presence at typical presence locations. A value of .5 is the default, but I found that a setting of .8 produced maps that showed a spread of higher probabilities that looked more like the expected distribution of crops. Reasoning that these maps should reproduce the existing situation and provide information between the sampled points, a distribution that reflected the known distribution of crops seems reasonable.
Regularization multiplier	1	The regularization multiplier parameter determines how closely-fitted the probability distribution is. The default is 1. A smaller value can result in overfitting while a larger value will give a more spread out prediction. After some testing, it was determined that the default value produced the best fit.
Random seed	yes	Given the high number of replicates used, a random seed will cause the most variation in the heuristic results, thus give the best range of possible results.
Maximum iterations	500	Default value, seemed to be sufficient
Convergence threshold	.00001	Default value, seemed to be sufficient

Fortunately, as the tool documentation suggests, the MaxEnt technique can produce good results even when given a large number of correlated variables. Unlike traditional statistical techniques, such correlations do not invalidate the modeling process. However, after many runs and permutations, I determined that the most realistic looking results were produced with just the eight environmental layers shown in Table 6.

Table 6 – Final set of environmental layers included in all MaxEnt models

Variable	Description
dep2reslyr	Depth to restrictive layer
elevation	Elevation
rainannual	Total annual rainfall
rainmax	Maximum monthly rainfall
rainmin	Minimum monthly rainfall
slope	Slope
tempmaxann	Maximum annual temperature
tempminann	Minimum annual temperature

Interestingly, the soil data provided very little informational contribution. I believe this is probably a result of its very coarse nature and the homogeneity of the values within and between soil map units. Only one soil property, depth to restrictive layer, produced important contributions to the final MaxEnt models. This is not to say that soil properties are not important determinants in crop success, but within the maximum-entropy model framework, and given the scale of the data available, they did not provide sufficiently unique, crop-type-correlated information to weigh into the final results.

Model runs using the full set of 12 monthly rainfall layers were tested based on the premise that *when* the rainfall occurs might have an important impact on crop distribution. However, this did not result in models any more successful than ones using the three rainfall variables (rainannual, rainmax and rainmin) included in the final models. It seems, in Hawai`i, just the magnitude of maximum and minimum rainfall are at least as important as the timing.

It is important to note that in the interpretation of the final results, given the environmental data included, my analysis assumes natural rainfall. Since there is no data available about which of the crops in the original crop dataset are irrigated, I made the assumption that, irrespective of the application of additional water, the general rainfall regime in any location is an important determinant.

The extensive output for any run of the tool includes a few useful and easily interpreted tables. A particularly interesting one, though it must be interpreted with some care, shows the percent contribution and permutation importance of each variable. Table 7 shows the average of these values of 100 replicates in the final run of the model used to generate the Probability Maps included in the map document. This table has been color coded to make interpretation easier. Given the heuristic nature of this technique, another set of 100 replicates would produce slightly different values, though they are likely to (and did in my various runs) hover around these numbers and rankings would be similar.

Table 7 – Variable percent contribution and permutation importance (in brackets) for final model run using 100 replicates. Dark green indicates the highest stable contribution values, light green shows high contribution and permutation, and pink highlights the lowest values.

Variable	Species							
	Banana	Coffee	Flowers	MacNuts	Papaya	Specialty	TropFruits	TruckCrops
dep2reslyr	7 (3)	5 (1)	2 (1)	9 (7)	5 (1)	6 (3)	6 (7)	21 (10)
elevation	31 (9)	3 (3)	16 (27)	4 (5)	19 (16)	20 (22)	9 (35)	32 (65)
rainannual	26 (34)	11 (20)	32 (15)	14 (16)	7 (8)	29 (23)	12 (14)	13 (5)
rainmax	14 (9)	20 (65)	12 (16)	5 (39)	11 (32)	3 (12)	6 (5)	5 (3)
rainmin	20 (45)	32 (7)	27 (40)	15 (14)	34 (42)	23 (38)	27 (31)	10 (3)
slope	1 (0)	6 (0)	0 (1)	1 (2)	1 (0)	1 (0)	1 (1)	2 (1)
tempmaxann	1 (0)	10 (1)	0 (0)	18 (16)	4 (1)	1 (0)	9 (5)	12 (3)
tempminann	0 (0)	14 (2)	10 (1)	34 (2)	19 (0)	18 (0)	30 (2)	6 (10)

The meaning of the values shown in Table 7 is explained in the MaxEnt tutorial included with the software. Percent contribution is explained as follows:

> "While the Maxent model is being trained, it keeps track of which environmental variables are contributing to fitting the model. Each step of the Maxent algorithm increases the gain of the model by modifying the coefficient for a single feature; the program assigns the increase in the gain to the environmental variable(s) that the feature depends on… In addition, when there are highly correlated environmental variables, the percent contributions should be interpreted with caution." (Phillips, no date)

The permutation importance

> "depends only on the final Maxent model, not the path used to obtain it. The contribution for each variable is determined by randomly permuting the values of that variable among the training points (both presence and background) and measuring the resulting decrease in training AUC [the entropy test]. A large decrease indicates that the model depends heavily on that variable." (Phillips, no date)

My summary analysis of Table 7 is shown in Table 8. Here I have separated the values of high contribution with low permutation (stable) from those with high contribution and high permutation (unstable). Parentheses indicate values that are unstable but have lower contributions in this particular summary of the model run. Multiple runs of the tool have shown me that those with high permutation values tend to vary in contribution more substantially between model runs than others. I am not sure if it is possible to make any general conclusions from this, particularly since the direction of importance is not included (i.e. whether any variable is a positive or negative influence), but it is interesting to examine.

Table 8 – Analysis of variable importance

Species	High Stable	Moderately High Stable	High Unstable	Low
Banana	Elevation		Rainann Rainmin	Slope Tempmax Tempmin
Coffee	Rainmin		Rainmax Rainann	Elevation
Flowers and Foliage			Elevation Rainann Rainmin	Slope Tempmax
Macadamia Nuts	Tempmin		Rainmin (Rainmax)	Slope
Papaya		Tempmin	Rainmin Rainmax Elevation	Slope
Specialty Crops		Tempmin	Rainann Rainmin Elevation	Slope Tempmax
Tropical Fruits	Tempmin		Rainmin (Elevation)	Slope
Truck Crops			Elevation Dep2reslyr	Slope

GIS data of the probability surfaces for each crop are automatically generated as raster layers by the MaxEnt tool. These are the most interesting and useful results. Each run of the model with replicates summarizes the probability surfaces over the full set of replicates into a set of raster images that show the maximum, minimum, median, average and standard deviation of probability values for each cell. With the high number of replicates used in the final model, the average and median results are quite similar and both represent a good summary of the overall results. I, thus, decided to provide the set of average results from the final model run as the core content of the CropProbabilityMap document. They can be viewed and explored in ArcReader as explained in the accompanying Tutorial document. Figure 2 on the next page is a snapshot of the final MaxEnt output for Coffee.

Close examination of the final MaxEnt raster outputs and a recognition of the coarseness and level of imprecision in the input data suggested that the final version of these maps would benefit by a simple smoothing filter. In this process, the value of each cell is replaced with the average of the nine cells in a 3x3 window centered on the cell. As shown in Figure 3, this reduces the appearance of raggedness and spottiness of the distribution of values when zoomed in and produces a probability surface that is more likely to approach reality. The highest values are reduced, but the output of this entire process is sufficiently imprecise that slightly lower maximum values will make no difference in their interpretability.

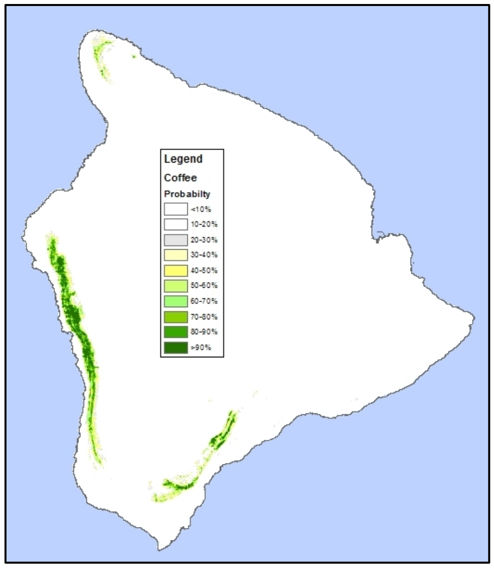

Figure 2 – Coffee Probability Map showing the predicted probability that conditions are suitable for coffee crops. Full scale, interactive versions of these maps for all crops included in this study are included in the associated CropProbabilityMap document.

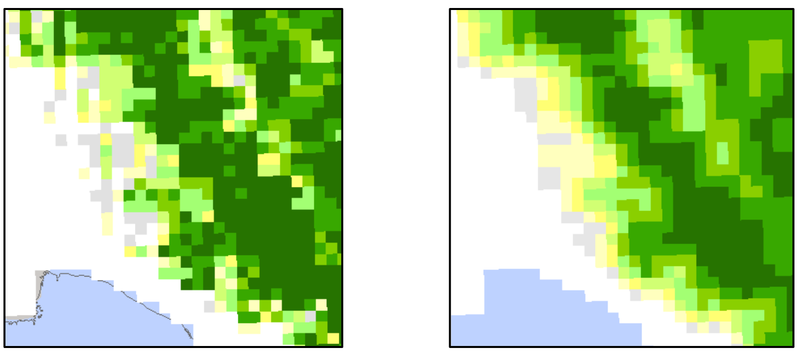

Figure 3 – The final MaxEnt output (left) and the smoothed final probability surface (right).

Discussion

As pointed out above, the results of this MaxEnt modeling effort are not definitive; slightly different results can be generated with each model run. However, they are sufficiently stable that I believe the final Crop Probability Maps can become an important contribution to land use planning efforts on Hawai`i.

Since all of these results are only as good as the input data, the data do NOT have sufficient spatial resolution to give an exact value for a single property. However, while the precise values should not be considered significant, relative values certainly are. It is useful to zoom in on any parcel and to explore the values of cells which fall within or around that parcel and the adjacent parcels. It is quite appropriate to consider drawing generalized lines around the areas where particular crop growth conditions are strongly favorable as a means of delineating areas that are potentially "good" for that purpose.

The maps distributed as the core product of this project stand on their own as a useful planning resource. However, the nature of the heuristic modeling process begs further exploration for those who wish to consider other variations. The role of the soil parameters could be further explored. Certainly the role of depth to restrictive layer in the final model output is worth reconsidering. The Truck Crops model, in particular, gives some importance to this value. Examination of the data used implies this likely arises from the fact that the crops dataset includes a few polygons that occur in areas that are designated in the soils data with surface texture of bedrock. I have included the dep2reslyr layer as a transparent polygon overlay in the final map document so that it can be used for this kind of further exploration.

Thus, an additional, unexpected output from this project is the full set of environmental layers and crop sample points that can be made available for such exploratory purposes. If others wish to test options, all the data now exists so it is quite easy to explore variations. I can imagine a small group session in which planners and agricultural specialists examine the available data and consider alternate sets of variables to test in a model run.

In conclusion, by reviewing the original goals for this project, I highlight what has been accomplished and improved. The original proposal for this project laid out the following steps:

1. *Further refinement of the original NRCS data to correct for existing data entry errors* – This task, fortunately, became unnecessary due to the valuable effort completed by Melrose and Delparte under a related R&D grant.
2. *Checking for updated data for the various GIS layers to be used in the analysis* – This was an key part of the model preparation. Some important updated data was uncovered and incorporated into the model. Sources are listed in Table 3.
3. Completion of a series of GIS overlay and analytical manipulations to extract the various landscape conditions found in all locations where each individual current crop is found – A much better, more defensible solution was found in the maximum entropy modeling technique.
4. Aggregation of data into a large table of max/min/average growth conditions – Superseded by superior results from the MaxEnt models.
5. *Application of these ranges to the landscape conditions found in each island TMK* - Summarizing values by TMK was one of the more questionable outcomes of the original web service. I decided in this project to keep the output of the model in its final format and to leave the assessment of values to a more quantitative style of analysis. This can be performed by visual analysis of the Crop Probability Maps in the ArcReader map document.
6. *Production of a revised base dataset that underlies the Hawai`i County Crop Model* – Since this system is no longer functional, this task is not relevant. As a replacement, I have included in the ArcReader document the relevant environmental data. Much more is possible, of course, but rather than expecting anyone who wishes to access this project's outcomes to have access to complex, expensive technology, I felt it was most expedient to prepare something anyone could use and access easily by sharing a simple (though large) zipped file.

The final product of this project is available as of October 2012 as a zipped file approximately 40MB in size. The user of the map document will also need to download and install the ArcReader software, available free on-line. Instructions for accessing the software and using the map document are included in the companion document "Crop Probability Map Tutorial".

Acknowledgments

I must begin by acknowledging Dayday Hopkins for her enthusiasm for the work that I and others have been able to complete with her support. Although I am sure she is enjoying her retirement, I am sorry that this new product will not be in her hands to be used for the purposes she foresaw. I would also like to acknowledge two of my Masters students at the University of Southern California, Eric Peña and

Caroline Neary (now Caroline Carl), who made initial attempts with this project, both before and after it was formally funded. While I have not used any of their results, the research they undertook provided an important foundation for this effort. Patrick Niemeyer from the USDA contributed several times with information about critical soil characteristics for crop growth and updated soils data. Gilbert Bailado's helpful contribution of updated County GIS data is always appreciated. Travis Longcore, a colleague at USC, unknowingly contributed the invaluable lead of using the MaxEnt model in this project when he mentioned it during a student's thesis defense. Finally, I want to acknowledge and thank Keola Childs for his always insightful comments and thought-provoking contributions to many of my scientific and academic efforts.

Contact Information

Karen K. Kemp, PhD GISP
Professor of the Practice of Spatial Science
Spatial Sciences Institute, University of Southern California
PO Box 821, Holualoa HI 96725
kakemp@usc.edu or karenkemp@geokemp.net

Literature Cited

Elith J, Phillips SJ, Hastie T, et al. (2011) A statistical explanation for MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.

Jaynes ET. (1957) Information Theory and Statistical Mechanics. *The Physical Review* 106: 620-630.

Melrose J and Delparte D. (2012) Hawaii County Food Self-Sufficiency Baseline 2012. Geography and Environmental Studies Department, University of Hawaii at Hilo.

Neary C. (2011) Evaluation of Land Suitability for the Expansion of Agriculture on the Island of Hawai'i. Unpublished final report for Geog 583: Spatial Sciences Institute, University of Southern California.

Phillips SJ. (no date) A brief tutorial on MaxEnt. Unpublished tutorial distributed with software at www.cs.princeton.edu/~schapire/maxent.

Phillips SJ, Anderson RP and Schapire RE. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

Phillips SJ, Dudik M and Schapire RE. (2004) A Maximum Entropy Approach to Species Distribution Modeling. *Proceedings of the Twenty-First International Conference on Machine Learning.* Banff, Canada, 655-662.

Appendix – Example MaxEnt output

On the following pages is a copy of the HTML summary output from a model run of 5 replicates for Coffee.

USC Dornsife

Spatial Sciences Institute