A Description of BAMM, the Bio-Environmental Appraisal and Mapping Model

TABLE OF CONTENTS

A Dsecription of BAMM - the Bio-Environmental Appraisal and Mapping Model

A new species distribution model, the Bio-environmental Appraisal and Mapping Model (BAMM) was developed primarily to serve as a tool for regulatory risk analysis of exotic plants pests. The modeling needs of ecologists and phytosanitary risk analysts can differ substantially in terms of model complexity, ease of use, scale and model performance. Consequently, the design of BAMM incorporates considerations for reproducibility, ease of use, a variable-rich global database and quality control statistics. The versatility of BAMM enables it to be used to determine habitat zones of species in other kingdoms. BAMM simulations can be made through a web interface or in off-line mode if large batches of species distribution simulations are desired.

Overview

The purpose of BAMM is to provide life sciences researchers with a method that could be applied in a consistent fashion to determine potential global habitat zones of multi-kingdom life forms. This goal is accomplished by finding characterization relationships of species native environments and applying those characterization relationships globally to identify non-native habitats.

Species presence data is used to derive statistical relationships from weather and soil data to define their habitats. These statistical relationships are applied to global gridded data sets to determine other areas in the world where those species may thrive. In additional to weather data, databases such as elevation, slope, aspect land, cover/land use, and soil properties may be used to further refine global habitat zones. BAMM uses a simple percent match method derived from the weather and soil data to define the suitability of species environments. These percent match results are displayed on geo-referenced maps as a "zone map" with the results classified into five-percent intervals.

Like other types of models that use large volumes of data, the BAMM model process consists of three parts (Figure 1).

First, a pre-processor prepares the various raster environmental data created from CFSR data for input into the BAMM model. This pre-processing converts the input data into a form that minimizes data ingestion time. Pre-processing includes summarizing the gridded daily data to monthly, seasonal, and annual values. The second part of the process is the BAMM engine that uses the presence data to generate the statistical relationships derived from the weather and soil data and applies those statistics to all land areas in the world. The final process consists of post-processing that is used to create the maps, tables, and model performance evaluations. BAMM often is run on a single processor but large batches of species may be run on multiple processors to speed up computation.

Weather Data Sets

The source of the weather data is the hourly Climate Forecast System Reanalysis (CFSR) data. These data are generated from a global 38-km coupled ocean-atmosphere model. The data extend from 1979 to the near present. Uncompressed, the size of the CFSR data would be well over 100TB. A downscaling technique was used to transform the 38-km data to a 10-km resolution and gain value-added information for thermal variables such as temperature, dew point, etc. For these variables, the downscaling algorithm employed a 3-dimensional technique using CFSR upper air data and surface data to calculate lapse rates. The lapse rates were used with anelevation model to adjust the thermal variables to the 10-km grid point elevations. Other variables, such as precipitation, were simply interpolated to the 10km resolution. The hourly data for all variables were summarized to daily values. Companion datasets at a resolution of about 32-km (one quarter degree) were extracted from the 10-km data. The advantage of using the 32-km data is that processing time is much less, owing to the much smaller amount of data that BAMM needs to ingest to create products. The CFSR data was used to obtain the basic weather variables such temperature, relative humidity, and precipitation. Other datasets were derived from the basic ones, such as leaf wetness, temperature during the leaf wetness events, soil temperatures at 2 and 4 inch depths, snow depth, etc. Currently, there are 64 weather variables available to BAMM. A list and description of these variables is shown in Table 1.

Description of Environmental Layers (filters)

BAMM uses various environmental data sets to identify habitable zones and to derive statistics such as the human population that co-exists with species. Among the environmental data are elevation, slope, aspect, human population data to estimate the number of people living in areas where species may exist, land cover/land use data, and soil property data. Elevation data are available at various spatial resolutions from the United States Geological Survey. Gridded human population data are available from the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University. The land use/land cover data used by BAMM were obtained from the Global Land Cover Facility of the University of Maryland. Soil property data was obtained from the Harmonized World Soil Database (HWSD) developed the Food and Agricultural Organization. The last two databases are available at a 30 second resolution (about 1-km) and have been reduced to the quarter degree and 5-minute resolutions used by BAMM. This reduction was done using only the most dominant features or properties found in the larger grid cells.

Pest Observations

BAMM can use presence data from many sources. The most common source of presence data currently used in the BAMM model is found at the Global Biodiversity Information Facility (GBIF). The GBIF contains detailed multi-kingdom presence data. Among the information contained in the presence data are geophysical coordinates, date of observation, provider of the information, specie scientific name, country, etc. Data from GBIF, or any other source, are organized into a common format and stored as comma separated variable (csv) format. GBIF data is checked manually to ensure its integrity and by the BAMM model, which upon ingestion of presence data, checks the geophysical coordinates to eliminate erroneous entries.

Presence data can also be generated on the web using tools on the NAPPFAST site. The web software automatically converts the presence data information into a format used by BAMM. Data can be uploaded into NAPPFAST as xls or csv formats using standard headings and data formats (Table 1.1). Data uploaded into NAPPFAST must also pass through the Data Manager tool to remove data errors. This is not covered in this manual.

BAMM Statistical Tests

BAMM uses simple statistical algorithms to determine how suitable the habitat of a target grid cell is for given specie. Its strength lies in having available very diverse sets of global climate and environmental inputs and its ability to apply derived statistical relationships and rules consistently anywhere in the world across multiple kingdoms. BAMM employs both continuous and discontinuous variables in its analyses. Continuous variables are those which are reported everywhere, such as temperature. Discontinuous variables, such as precipitation, may or may not occur at a given location. Those variables are better resolved statistically using gamma distributions.

Much of the statistical work done by BAMM consists of finding tolerance intervals for weather variables using Anderson-Darling normality tests. Before the statistical analyses begin, the user can choose a confidence value that a given proportion of the population is represented by the tolerance interval for each variable. BAMM uses the same user defined confidence and proportion entries for all variables. Default values of 90 percent are set for both the confidence and proportion. The tolerance intervals are determined as follows. Using the locations of the presence data, values from the gridded weather data are gathered for each variable. Data for each variable are ordered from minimum to maximum values. The data in each ordered variable are then ranked. Next, the uniform order statistic median value for each observation is calculated for the variables. The percent point function (inverse standard normal distribution) is evaluated for the uniform order statistic median value. Linear regressions are derived by evaluating the percent point function against the sorted data. Next the k-factor for a two-sided limit and chi-square values are calculated. These values are then used to determine the lower and upper bounds for each variable.

Because BAMM uses a large number of weather variables to define species habitats, some variables are more representative than others. It therefore is desirable to eliminate those variables that poorly define species habitats. This process, which was known as the “smart mode” in earlier versions of BAMM is automatically performed in each BAMM simulation.

The process of evaluating how well each weather variable helps to define species habitats is done as follows. Using the tolerance intervals calculated earlier, the weather data for all training data locations are evaluated to determine the percentage match for each weather variable. Next, a standard deviation is calculated from the results of (100-percent match) for all weather variables. Any weather variable whose 100-percent match value that is greater than one standard deviation is discarded. Below is an example of how this process works. Suppose we are using ten weather variables and their percent matches calculated from the training points' weather data are 72.6, 12.1, 92.4, 8.5, 89.3, 4.2, 93.5, 72.7, 93.1, and 88.6. Subtracting the percent matches from 100 yields 27.4, 87.9, 7.6, 91.5, 10.7, 95.8, 6.5, 27.3, 6.9, and 11.4. The standard deviation of those ten values is 38.4. Any variable whose 100-percent match value, (e.g. variables 2, 4, and 6) exceeds 38.4 is discarded.

BAMM Limiting Value Checks

Limiting value checks are applied to BAMM output to ensure results are within the likely physical limits of a species habitat. The limiting values are applied as follows:

Extreme daily maximum temperature: 130 degrees F
Extreme daily minimum temperature: -90 degrees F
Upper limit of three-month average temperature: 105 degrees F
Lower limit of three-month average temperature: 30 degrees F
Minimum annual precipitation: 1.00 inch
Minimum annual precipitation/evaporation ratio: 1.00 inch

If any datum for a given pixel fails one or more of these checks, it is assumed the environment in that pixel is not suitable for the specie.

Use of Environmental Characteristics as Filters in BAMM Habitat Zone Maps

Environmental characteristics such as elevation, slope, aspect land, cover/land use, and soil properties may be used to further refine global habitat zones. Land cover/land use (LCLU) data having a resolution of 1/360 degree (about 300m) was obtained from the European Space Agency ESA) and processed to a quarter degree and 10km resolution. During this processing of the ESA data, the twenty most abundant LCLU classifications were identified for each quarter degree and 10km pixel. The LCLU characteristics for more than 99 percent of both the quarter degree and 10km pixels could be adequately described by this method. Elevation data obtained from the USGS. The Harmonized World Soil Database (HWSD), having a resolution of 30-seconds (about 1km) was obtained from the FAO for use in this project. The HWSD database contains information for both top soil and subsoil and they also provide databases on slope and aspect. They also were used in this project. The HWSD databases were process into a quarter degree and 10km resolution using the same process that was used on the LCLU database. Additional environmental characteristics now available for use in BAMM are population density and Koppen climate biomes.

The training data derived from the presence observations are used to define the various environmental characteristics that describe the species habitat. Each environmental characteristic for each quarter degree or

10km pixel is compared to the range of values found in the training data. For some environmental variables, such as elevation and slope, the presence data are used to find a range of values for a characteristic. For other variables, such as land use /landcover, the training data are used to define categories a species occpies. For each grid cell, each environmental characteristic for that cell is compared to the ranges or categories defined by the presence data. If a match is found for a particular characteristic, the grid cell passes that particular check and the total number of passes is tallied and compared to the total possible number of environmental characteristics activated in the configuration file. If a pixel passes a set percentage of the filters (usually 80 percent), the habitat zone defined by the climate data is kept active. If the grid cell fails the filter test, the habitat zone is turned off for that cell. The result of the filtering is written to a tiff file having “filtered_zone_map” as part of its file name. Currently there is a maximum of 55 environmental characteristics available for filter use in BAMM. A description of these characteristics is given in Table 2 below.

Products Created by BAMM

The results of BAMM simulations include maps, graphs, and tables files to help researchers define species habitat zones and to provide estimates of model quality and performance. These results can either be displayed on the web or researchers can retrieve files via ftp or email. The file names from maps and graphs are created from the specie names and date of creation. Map products depicting habitat zones word “zone” as part of their file name. Maps showing location of the input (presence data) species have the word “location” as part of their file name. TFW files are available for geo-referencing.

BAMM also generates several kinds of scatter plots made from selected weather variables and presence data. These include moisture plots, temperature plots, temperature-precipitation plots, and temperature-precipitation/evaporation ratio plots. These scatter plots are helpful in providing an indication of how widely dispersed species habitable ranges may be.

Map products include percent match classes from the tests described above and maps showing the locations of the input GBIF or other data used in the climate matching process (Figure 2). The percent map classes seen on the maps define how (un)likely a specie is likely to be able to thrive in a particular location. The percent matches are shown in five-percent class intervals starting with 1 percent. The blue-red color scheme starts at dark blue (1 -5 percent) to dark red (95-100 percent). Lower percent matches indicate it is less likely a species could thrive in such a location. Map areas having lower percent classes may be conducive for a species to thrive only part of a year or perhaps only in some years. On the other hand, the higher percentage matches (70 percent or more) indicate a species may thrive most or all of the year and most or all of the years used in the simulation. A habitat zone maps if shown in Figure 3.

A second zone map is made for each species that reflects the results of the environmental filtering process. The class values in the filtered zone map are similar to the unfiltered zone map. The major difference between the two maps is all pixels that fail the filtering test are turned off. A filtered habitat zone maps if shown in Figure 4.

A third map is a cumulative distance map that shows the difference between each pixel and the training data locations. The resulting values are expressed as (1/km) X 1000. The purpose of this map is to provide a measure of distance constraints of the presence data. An example of this map is given in Figure 5.

BAMM output includes several kinds of tables that provide information on data and model performance. These tables are the All Tests Summary, the Individual Test Summary, the Population Summary, and a statistical data summary file.

An example of the All Tests Summary is given in Table 3. This summary provides an overall estimation of model performance and data quality. In the upper portion of the table, are the simulation date and the user-specified proportion of the population represented and the confidence interval. Below that information is basic information on the presence data quality used in the simulation. An estimate of the overall quality of the data follows the individual checks. The model quality assessment always will contain values for the first four lines on the table. If a user makes a simulation in “withheld” mode (i.e. presence data for the United States is NOT used for training), information on the Test Accuracy and the Accuracy Ratio is provided in the table. The Model Quality Assessment concludes with an overall rating. Below the body of the table is information about each data or model quality check as well as the scale used to make the overall rating.

BAMM also provides a report on the results of the individual tests used in the simulation. In the upper portion of the table, are the simulation date and the user-specified proportion of the population represented and the confidence interval. Below that, for each of the tests used, information given is the average percent match for all presence data used as input (column: Pct Match), the extremes of the data (maximum and minimum values) found using the training data, and the low and high values for the tolerances for each test. An example of the Individual Tests Summary is given in Table 4.

Another table BAMM generates is a summary of the species distribution and human population exposure to a species. There are two parts to this table. Below the table description is a listing of the species distribution in terms of the percentage of class interval ranges and also in terms of the total area that particular class interval occupies. These distributions are for the entire world, the fifty United States and Puerto Rico, the fifty United States, the contiguous United States, and Hawaii only. At the bottom of the table is a listing of the estimated human population exposed to selected each percent match intervals. The Gridded Population of the World, Version 3 was used to make these estimates. Table 5 is an example of the species distribution and human population summary.

In order to compare BAMM results for statistical tests, BAMM provides csv output for this purpose. This information consists of the coordinates of each location, whether a location is a training or withheld location, and the percent match results for all tests (test-yrs passed). A partial example of this output is shown in Table 6.

Selected weather variables are used to create scatter plots. The weather data from grids where training locations are located are used in these plots. These include moisture plots, temperature plots, temperature-precipitation plots, and temperature-precipitation/evaporation ratio plots. These scatter plots are helpful in providing an indication of how widely dispersed species habitable ranges may be. Some examples of these plots are shown in Figure 5.

Figure 5. Example of scatter plots. (A) Extreme minimum and maximum temperatures, (B) extreme two inch soil temperatures, (C), growing season length and growing season growing degree day totals, and (D) annual precipitation vs the annual precipitation-evaporation ratio (percent).

In addition to the tabular format, the results of the individual tests are presented in image format. These images provide a quick overview of that tests that were used (tests not used are shown in white) in the simulation as well as their relative performance. An example of given in Figure 6.

Future Development of BAMM

Suggestions on changes to the BAMM model are welcome. These suggestions may be directed to:

Joe Russo
President
ZedX, Inc.
369 Rolling Ridge Drive
Bellefonte PA 16823
814-357-8490 x 238
russo@zedxinc.com

Jay Schlegel
Senior Research Scientist
ZedX, Inc
369 Rolling Ridge Drive
Bellefonte PA 16823
814-357-8490 x 221
schlegel@zedxinc.com

References

Gridded Population of the World, Version 3 (GPWv3) Data Collection,
URL: http://sedac.ciesin.columbia.edu/gpw/

Nachtergaele, Freddy, van Velthuizen, Harrij, and Verelst, Luc, Harmonized World Soil Database, March, 2009.
URL: http://www.iiasa.ac.at/Research/LUC/luc07/External-World-soildatabase/HTML/HWSD_Data.html?sb=4

Suranjana Saha, et. al., The NCEP Climate Forecast System Reanalysis**,** Bulletin of the American Meteorological Society, Volume 91, Issue 8, pp 1015-1058.

USGS Digital Elevation Models, URL: http://eros.usgs.gov/

The GlobCover Land Cover version V2.2, Source data: © ESA / ESA GlobCover Project, led by MEDIAS France, URL: http://ionia1.esrin.esa.int/index.asp

Figure 1 . Process Overview of the Bio-environmental Appraisal and Mapping Model (BAMM)

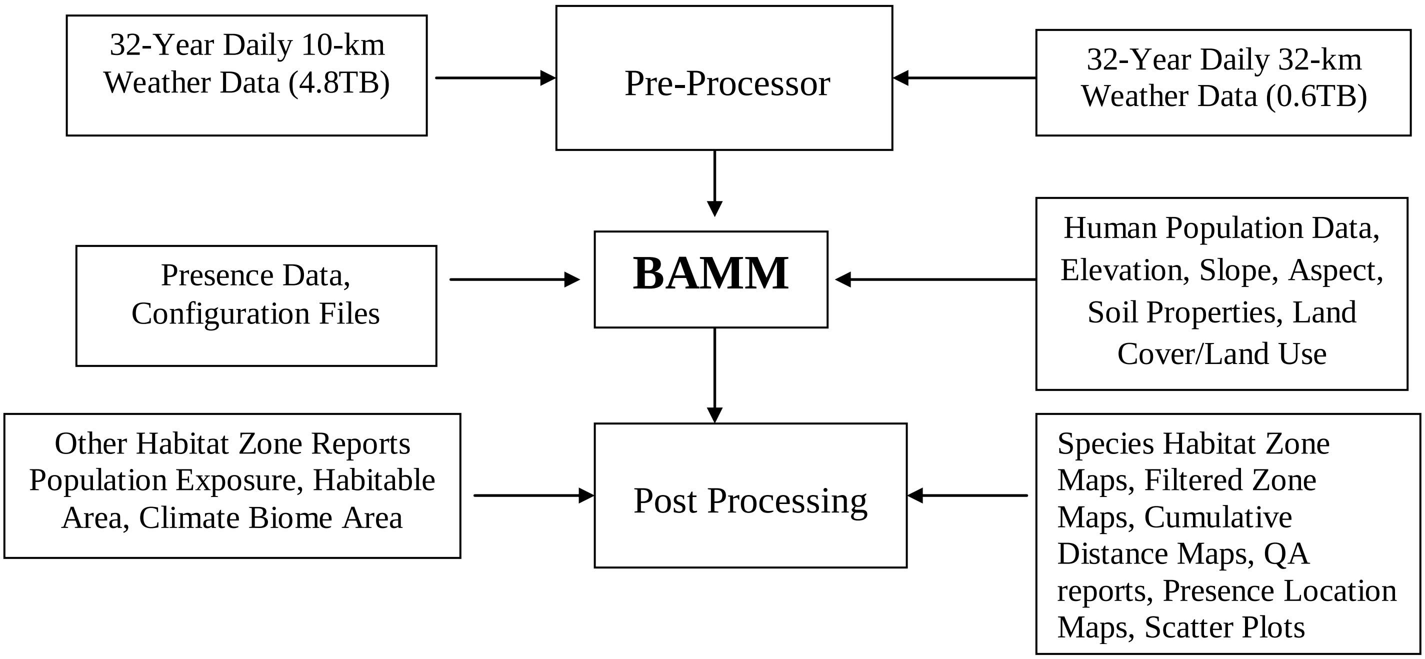

Figure 2. An example of a location map showing the training data locations.

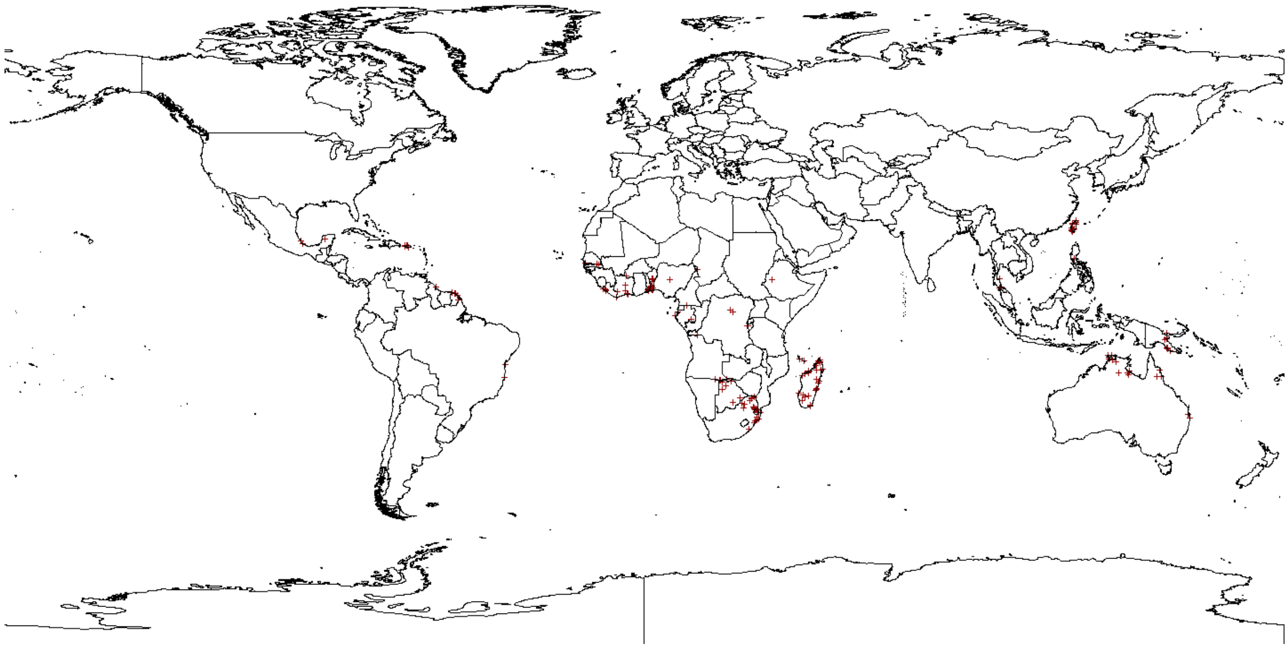

Figure 3. An example of a habitat zone map.

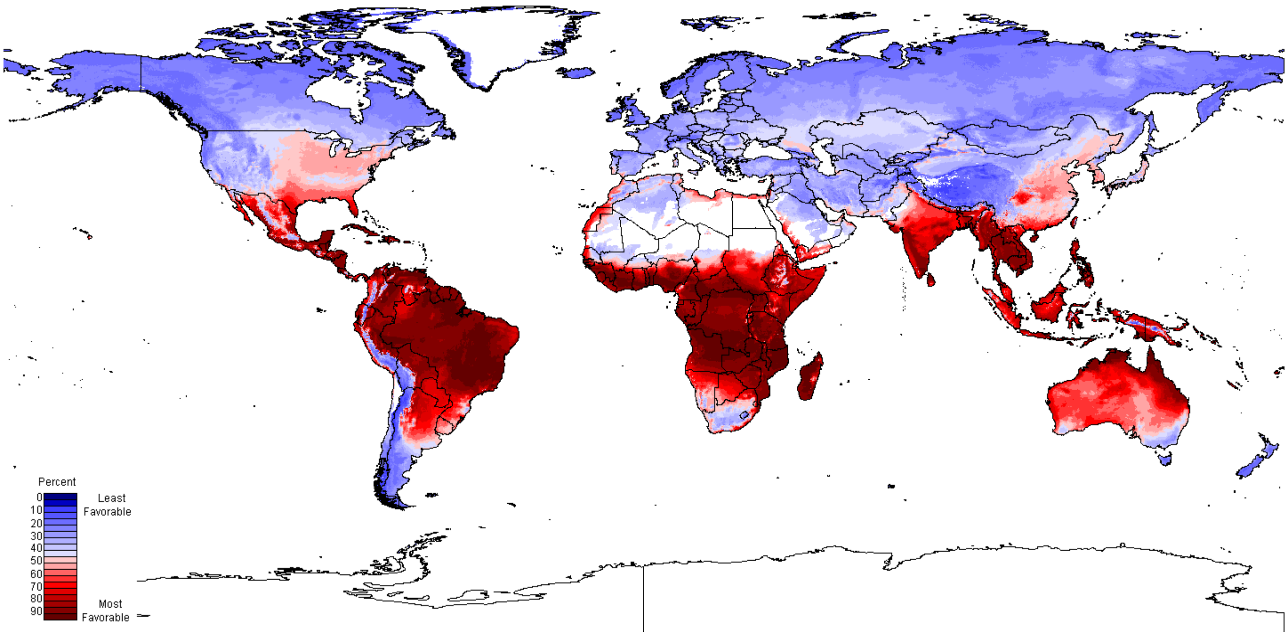

Figure 4. An example of a filtered habitat zone map.

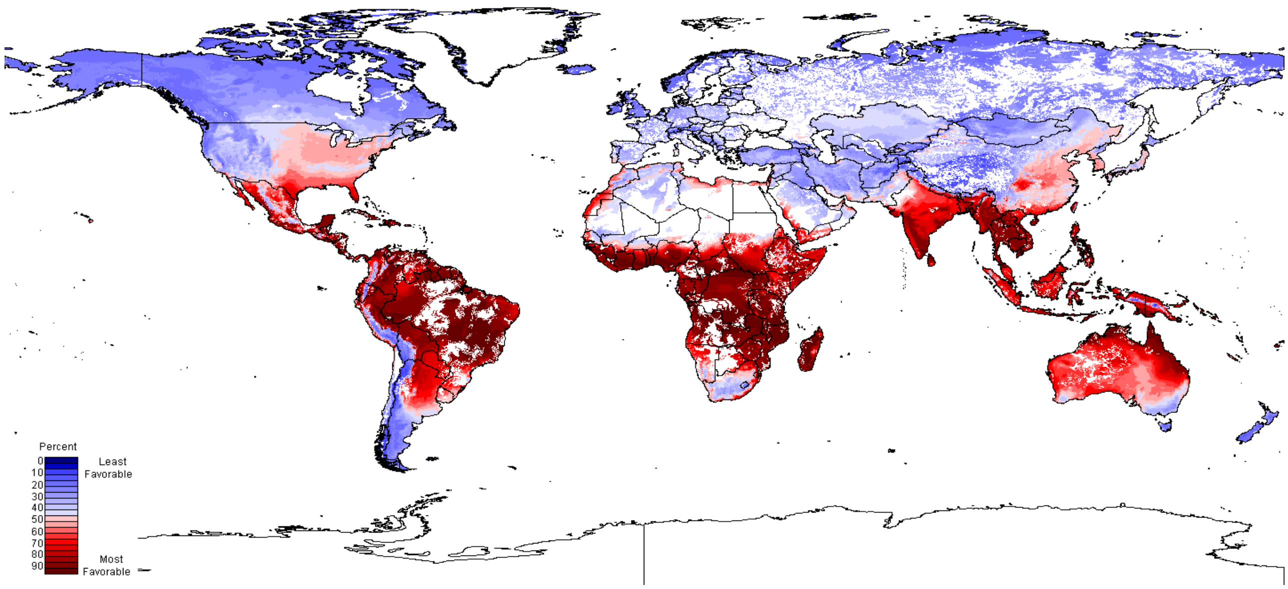

Figure 5. An example of a cumulative distance map.

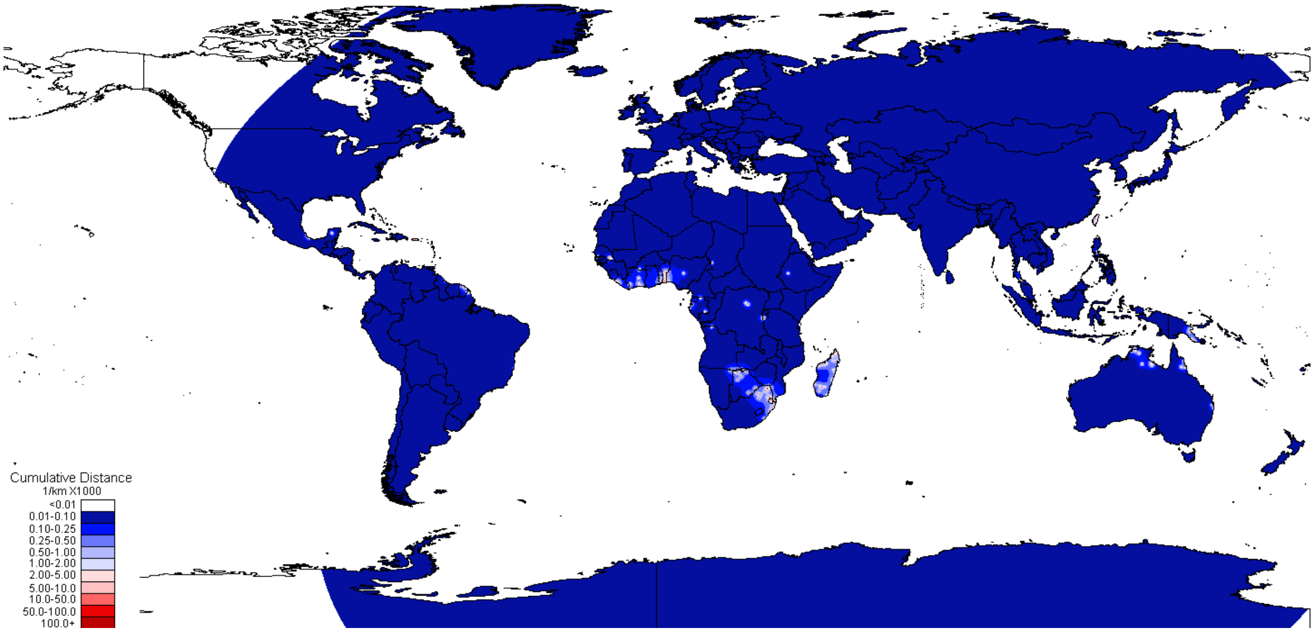

Figure 6A Climate space diagram

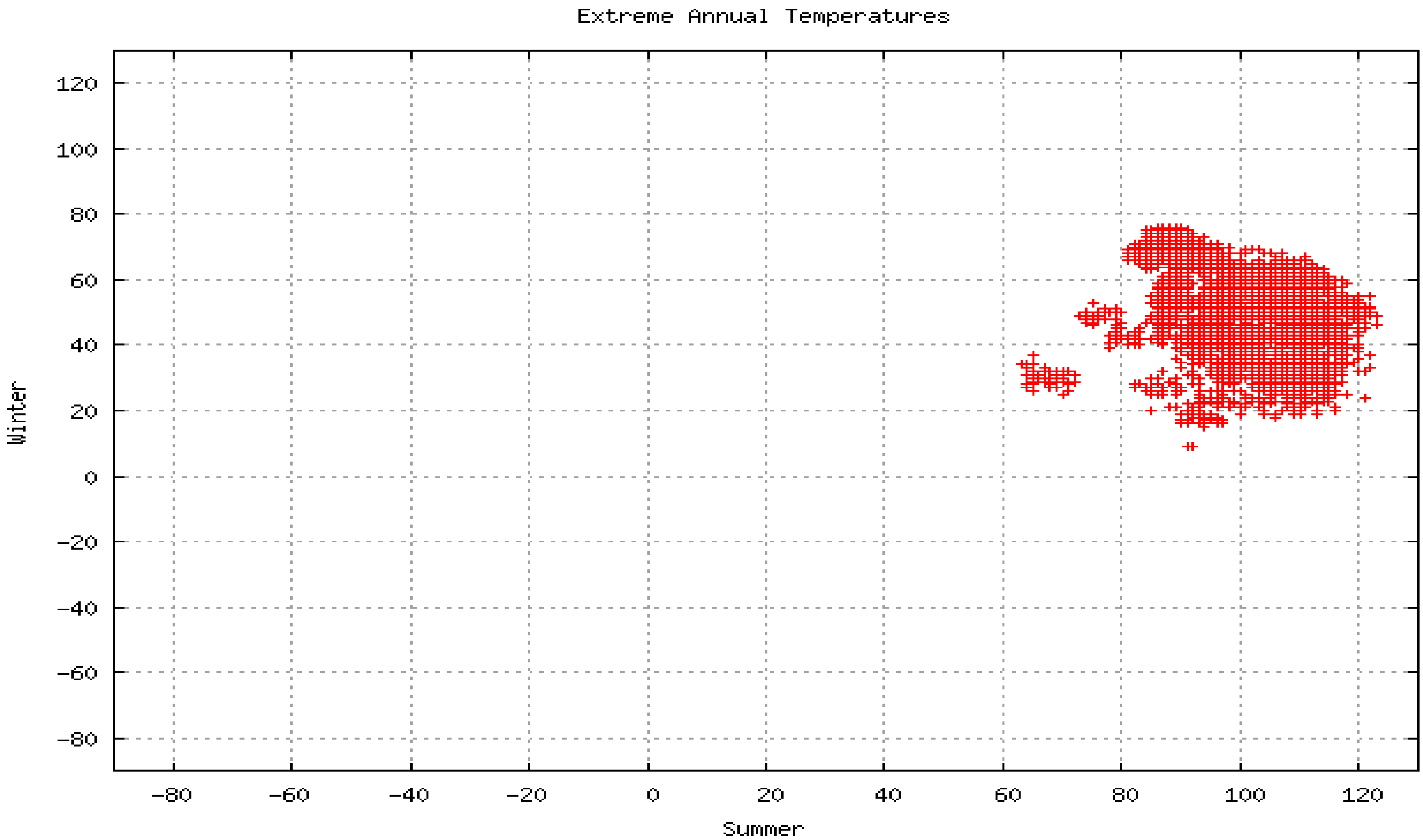

Figure 6B Climate space diagram

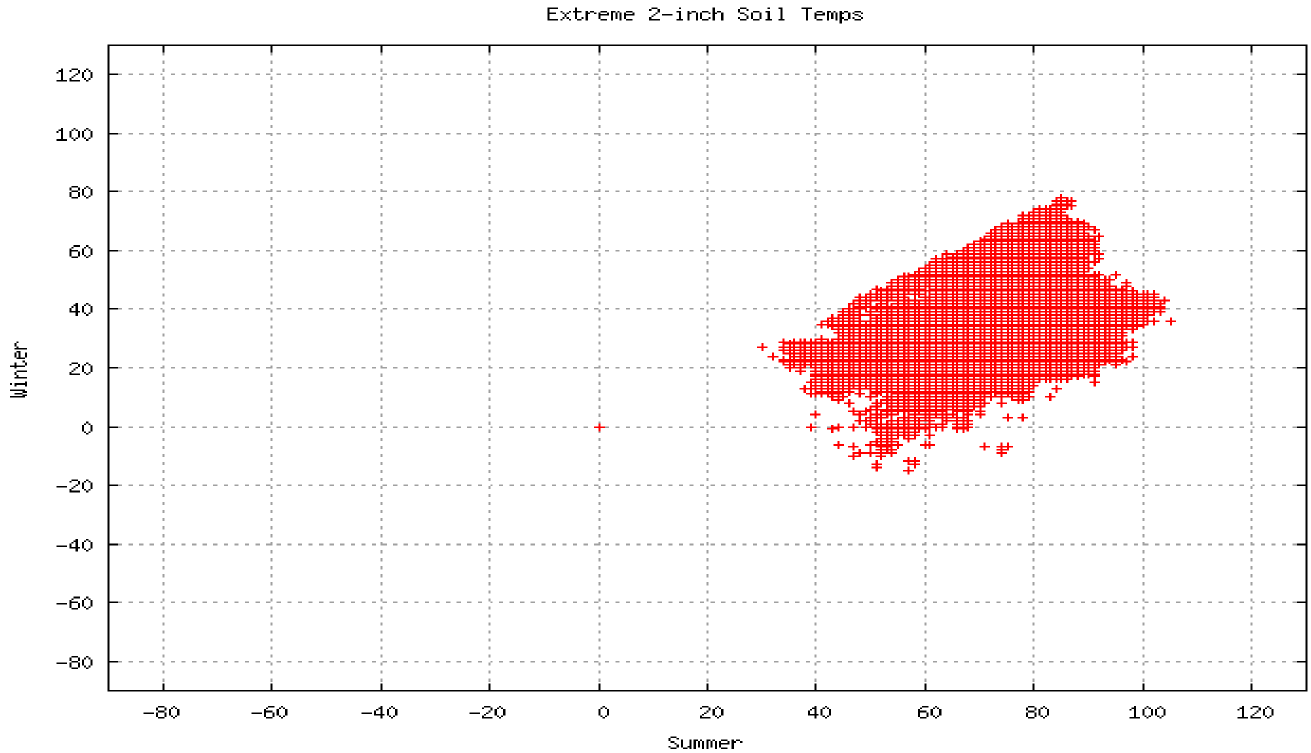

Figure 6C

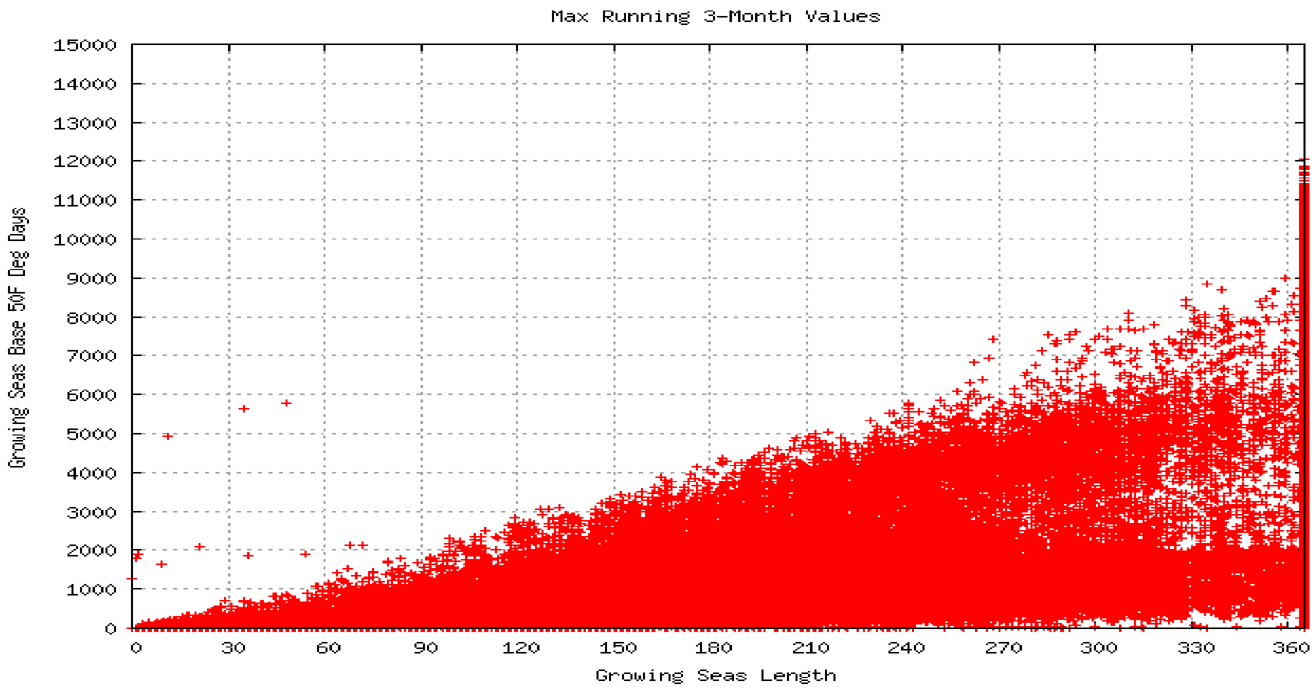

Figure 6D

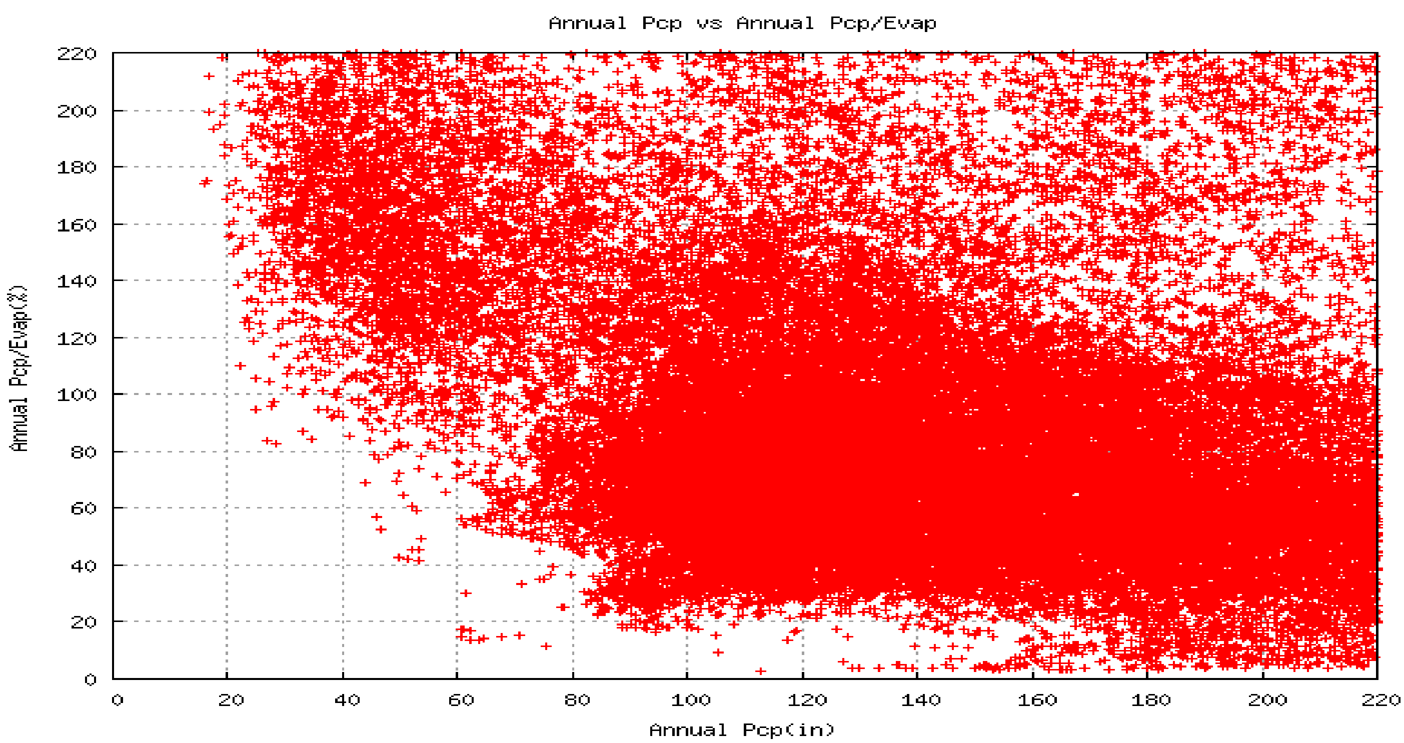

Figure 7. Example of individual test results in graphical format.

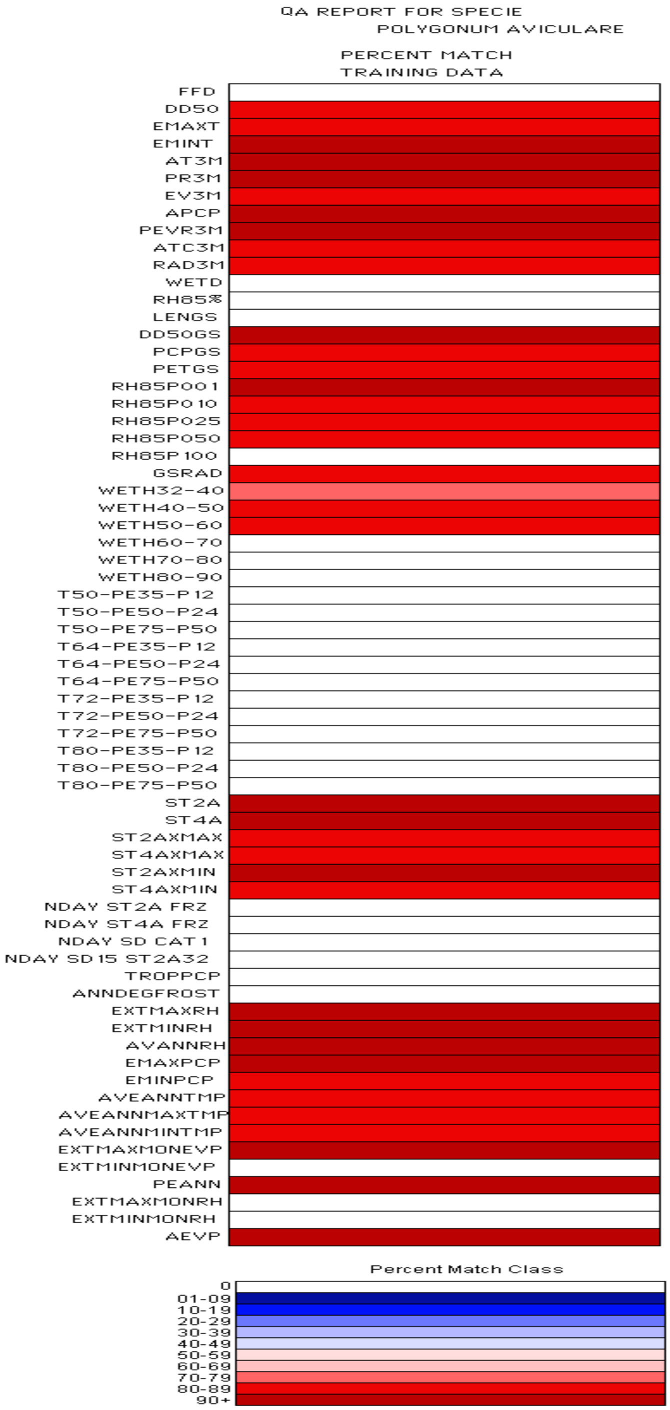

Table 1. Weather Variables Currently Used by BAMM

Varable	Description	Period
AEVP	Total evaporation	Annual
ANNDEGFROST	Total degrees of frost calculated by (SUM OF DAYS OF 32-DLYAVETMP)	Annual
APCP	Total precipitation	Annual
AT3M	Three-month average temperature	Running 3-month
ATC3M	Average temperature for the coldest 3 months	3-month
AVEANNMAXTMP	Average maximum temperature	Annual
AVEANNMINTMP	Average minimum temperature	Annual
AVEANNTMP	Average annual temperature	Annual
AVANNRH	Average annual relative humidity	Annual
DD50	Growing degree day total, base 50, modified	Annual
DD50GS	Growing degree days, base 50 for LENGWRSEAS	Season
EMAXPCP	Highest monthly precipitation	Annual
EMAXT	Highest daily maximum temperature	Annual
EMINPCP	Lowest monthly precipitation	Annual
EMINT	Lowest daily minimum temperature	Annual
EV3M	Three-month evaporation total	Running 3-month
EXTMAXMONEVP	Highest monthly evaporation	Annual
EXTMAXMONRH	Highest monthly relative humidity	Annual
EXTMAXRH	Highest daily average humidity	Annual
EXTMINMONEVP	Lowest monthly evaporation	Annual
EXTMINMONRH	Lowest monthly relative humidity	Annual
EXTMINRH	Lowest daily average humidity	Annual
FFD	Number of frost free days	Annual
GSRAD	Average daily radiation for LENGWRSEAS	Season
LENGS	Length of the growing season	Season
NDAY-SD-CAT1	Number of days with snow depth 15+”	Annual
NDAY-SD15-ST2A32	Number of days with snow depth 15+” and 2-inch soil temps of 30-34F	Annual
NDAY-ST2A-FRZ	Number of days with 2-inch soil temps <32F	Annual
NDAY-ST4A-FRZ	Number of days with 4-inch soil temps <32F	Annual
PCPGS	Precipitation for LENGWRSEAS	Season
PEANN	Precipitation/evaporation ratio	Annual
PETGS	Precipitation/evapotranspiration ratio for LENGWRSEAS	Season
PEVR3M	Three-month precipitation/evaporation ratio	Running 3-month
PR3M	Three-month precipitation total	Running 3-month
RAD3M	Three-month average monthly radiation	Running 3-month
RH85%	Days with RH of 85% or more during wettest 3-months	Running 3-months
RH85P001	Number of days with RH 85+ and 0.01+ rain	Annual
RH85P010	Number of days with RH 85+ and 0.10+ rain	Annual
RH85P025	Number of days with RH 85+ and 0.25+ rain	Annual
RH85P050	Number of days with RH 85+ and 0.50+ rain	Annual
RH85P100	Number of days with RH 85+ and 1.00+ rain	Annual
ST2A	Three-month average 2-inch soil temperature	Running 3-month
ST4A	Three-month average4-inch soil temperature	Running 3-month
ST2AXTMAX	Highest daily maximum 2-inch soil temperature	Annual
ST4AXTMAX	Highest daily maximum 4-inch soil temperature	Annual
ST2AEXTMIN	Lowest daily minimum 2-inch soil temperature	Annual
ST4AEXTMIN	Lowest daily minimum 4-inch soil temperature	Annual
T50-PE35-P12	Months with average temp 50F,+, Prec/ET ratio of 35%+ and 1.2+” Prec	Annual
T50-PE50-P24	Months with average temp 50F,+, Prec/ET ratio of 50%+ and 2.4+” Prec	Annual
T50-PE75-P50	Months with average temp 50F,+, Prec/ET ratio of 75%+ and 5.0+” Prec	Annual
T64-PE35-P12	Months with average temp 64F,+, Prec/ET ratio of 35%+ and 1.2+” Prec	Annual
T64-PE50-P24	Months with average temp 64F,+, Prec/ET ratio of 50%+ and 2.4+” Prec	Annual
T64-PE75-P50	Months with average temp 64F,+, Prec/ET ratio of 75%+ and 5.0+” Prec	Annual

Varable	Description	Period
T72-PE35-P12	Months with average temp 72F,+, Prec/ET ratio of 35%+ and 1.2+" Prec	Annual
T72-PE50-P24	Months with average temp 72F,+, Prec/ET ratio of 50%+ and 2.4+" Prec	Annual
T72-PE75-P50	Months with average temp 72F,+, Prec/ET ratio of 75%+ and 5.0+" Prec	Annual
T80-PE35-P12	Months with average temp 80F,+, Prec/ET ratio of 35%+ and 1.2+" Prec	Annual
T80-PE50-P24	Months with average temp 80F,+, Prec/ET ratio of 50%+ and 2.4+" Prec	Annual
T80-PE75-P50	Months with average temp 80F,+, Prec/ET ratio of 75%+ and 5.0+" Prec	Annual
TROPPCP	Total Prec for all months with average temps of 64F+ and 2.4+ inches Prec	Annual
WETD	Number of wet days during the wettest 3-months	Running 3-month
WETH32-40	Number of wetness hours with temp 32-40F	Annual
WETH40-50	Number of wetness hours with temp 40-50F	Annual
WETH50-60	Number of wetness hours with temp 50-60F	Annual
WETH60-70	Number of wetness hours with temp 60-70F	Annual
WETH70-80	Number of wetness hours with temp 70-80F	Annual
WETH80-90	Number of wetness hours with temp 80-90F	Annual

Table 2. Example of Pest observation data

Data publisher	OZCAM (Online Zoological Collections of Australian Museums) Provider
Dataset	Queen Victoria Museum Art Gallery provider for OZCAM
Dataset Rights	2002-12-18
Date collected	
Scientific name (interpreted)	Cochlicella barbara
Country (interpreted)	Australia
Latitude	-43.0333
Longitude	147.417
GBIF portal url	http://data.gbif.org/occurrences/243043291

Table 3. Environmental Characteristics used by BAMM

Type Filter	Pixel Test Passes if
Elevation	Elevation is within range of found in training data
Population	Population is within range of found in training data
LCLU	LCLU class found in training data
Koppen Climate	Koppen biome found in training data
Slope Aspect-North	North aspect of pixel is similar to that found in training data
Slope Aspect-East	East aspect of pixel is similar to that found in training data
Slope Aspect-South	South aspect of pixel is similar to that found in training data
Slope Aspect-West	West aspect of pixel is similar to that found in training data
Slope Aspect-Undefined	Undefined aspect of pixel is similar to that found in training data
Slope 0.0-0.5 Percent	Slope of pixel is similar to that found in training data
Slope 0.5-2.0 Percent	Slope of pixel is similar to that found in training data
Slope 2.0-5.0 Percent	Slope of pixel is similar to that found in training data
Slope 5.0-8.0 Percent	Slope of pixel is similar to that found in training data
Slope 8.0-16.0 Percent	Slope of pixel is similar to that found in training data
Slope 6.0-30.0 Percent	Slope of pixel is similar to that found in training data
Slope 30.0-45.0 Percent	Slope of pixel is similar to that found in training data
Slope 45 Percent or more	Slope of pixel is similar to that found in training data
Top Soil Texture	Texture class found in training data
AWC Class	Available Water Content class found in training data
Rooting Depth	Rooting Depth class found in training data
Top Soil Drainage	Top Soil Drainage class found in training data
Soil Water Regime	Soil Water Regime class found in training data
Impermeable Layer	Impermeable Layer class found in training data
Top Soil Gravel	Top Soil Gravel percentage class found in training data
Top Soil Sand	Top Soil Sand percentage class found in training data
Top Soil Silt	Top Soil Silt percentage class found in training data
Top Soil Clay	Top Soil Clay percentage class found in training data
Top Soil USDA Texture Class	Top Soil USDA Texture class found in training data
Top Soil Bulk Density	Top Soil Bulk Density percentage class found in training data
Top Soil Organic Carbon	Top Soil Organic Carbon class found in training data
Top Soil Ph	Top Soil Ph class found in training data
Top Soil CEC Clay	Top Soil Cation Exchange Capacity class for clay found in training data
Top Soil CEC Soil	Top Soil Cation Exchange Capacity class for soil found in training data
Top Soil Base Saturation	Top Soil Base Saturation class class found in training data
Top Soil Total Exchangeable Bases	Top Soil Total Exchangeable Bases class found in training data
Top Soil Calcium Carbonate	Top Soil Calcium Carbonate class found in training data
Top Soil Gypsum	Top Soil Gypsum class found in training data
Top Soil Sodicity	Top Soil Sodicity class found in training data
Top Soil Salinity	Top Soil Salinity class found in training data
Subsoil Soil Gravel	Subsoil Soil Gravel percentage class found in training data
Subsoil Sand	Subsoil Sand percentage class found in training data
Subsoil Silt	Subsoil Silt percentage class found in training data
Subsoil Clay	Subsoil Clay percentage class found in training data
Subsoil USDA Texture Class	Subsoil USDA Texture class found in training data
Subsoil Bulk Density	Subsoil Bulk Density percentage class found in training data
Subsoil Organic Carbon	Subsoil Organic Carbon class found in training data
Subsoil Ph	Subsoil Ph class found in training data
Subsoil CEC Clay	Subsoil Cation Exchange Capacity class for clay found in training data
Subsoil CEC Soil	Subsoil Cation Exchange Capacity class for soil found in training data
Subsoil Base Saturation	Subsoil Base Saturation class class found in training data
Subsoil Total Exchangeable Bases	Subsoil Total Exchangeable Bases class found in training data
Subsoil Calcium Carbonate	Subsoil Calcium Carbonate class found in training data
Subsoil Gypsum	Subsoil Gypsum class found in training data

Type Filter	Pixel Test Passes if
Subsoil Sodicity	Subsoil Sodicity class found in training data
Subsoil Salinity	Subsoil Salinity class found in training data

Table 4. An example of the All Tests Summary made by BAMM

BAMM QA All Tests Summary
Simulation Date: 2011_0903
Specie Name: Abrus_precatorius
Proportion of the population represented: 0.90
Confidence level: 0.90

Data Quality Tests

QA Measurement	Pass/Fail Threshold	Value	Result (Pass/Fail)
Number of Tests Used		34	
Number of Training OBS	50	15	P
Ave Year of OBS	1970	99	P
Geodiversity-Num of Continents	3	5	P
Geodiversity-Conts w/30+ Samples in 1-Deg Areas			F
Percent of 1-Deg Squares		0.46	

Data Quality QA Rating: VG

Model Quality Assessment

Pct of Training OBS Outside Tolerance Limit	1	10.2	F
Proportional Predicted Area	75.0	65.9	F
Cum Distance-Pct Match Co-located Areas	75.0	89.2	P
Global Raw Accuracy	75.0	86.9	P
Test Accuracy	75.0	85.0	P
Accuracy Ratio	0.90	0.98	P

Model Quality QA Rating: G
Bamm Version: BAMM_22.0

Information about the QA Assessment

Number of Tests - The number of tests used in this evaluation.

Number of Training OBS - The number of presence observations used to train the model

Ave Year of OBS - The average year of the observations of the presence OBS. The value 9999 signifies missing data.

Geodiversity-Num of Continents - The number of continents on which the presence data were observed.

Geodiversity-Conts w/30+ Samples in 1-Deg Areas - The number of continents having training points in at least 30 1-degree latitude by 1-degree longitude areas.

Percent of 1-Deg Squares - The percentage of the total possible land 1-degree latitude-longitude squares on which the presence data were observed.

Proportional Predicted Area - A measure of over prediction. The Proportional Predicted Area is:100 - the total percentage of all areas in the contiguous US having percentage match values of 70 or greater.

Cumulative Distance-Percent Match Co-Located Areas - A measure of the habitat suitability near presence data. This value represents the percent of the area where cumulative distance values were 0.50 or greater and percent match values were 70 or greater.

Pct of Training OBS Outside Tolerance Limits - The percent of all training OBS that are less than the lower tolerance limit or higher than the upper tolerance limit.

Global Raw Accuracy - The average percent match values for all training data locations using the gridded data depicted in the habitat zone maps.

Rating Scale for Data and Model Tests

	Percent of Tests Passed/Total Tests	
	Lower Threshold (>=)	Upper Threshold (<)
E=Excellent	84	100
VG=Very Good	67	84
G=Good	50	67
F=Fair	33	50
P=Poor	16	33
VP=Very Poor	0	16

Table 5. An example of the Individual Tests Summary made by BAMM

BAMM QA Individual Tests Summary
Simulation Date: 2011_1115
Specie Name: Aegilops_cylindrica_formatted
Proportion of the population represented: 0.90
Confidence level: 0.90

Training Data Individual Statistics-Used Variables

		Data Extremes		Tolerance				
Test Code	Pct Match	Minimum	Maximum	Low	High	Type of Dist	Units	Status
AEVP	89.0	14.3	133.9	6.6	74.7	G	INCHES	Y
ANNDEGFROST	94.8	0.0	4457.0	0.0	3117.1	G	DEG F	Y
APCP	88.1	0.2	158.2	-1.5	1.6	N	INCHES	Y
AT3M	88.0	36.0	96.0	42.1	84.5	G	DEG F	Y
ATC3M	88.3	-3.0	79.0	1.1	57.3	G	DEG F	Y
AVEANNMAXTMP	89.1	29.8	98.6	28.7	84.5	G	DEG F	Y
AVEANNMINTMP	89.6	15.1	77.9	18.5	60.6	G	DEG F	Y
AVEANNTMP	89.2	24.1	84.3	23.9	72.2	G	DEG F	Y
AVANNRH	92.5	21.1	91.5	41.6	95.9	G	PCT	Y
DD50	88.1	500.0	16169.0	0.0	11340.0	G	DEG F	Y
DD50GS	88.3	0.0	12519.0	0.0	7900.0	G	DEG F	Y
EMAXPCP	89.6	0.1	27.8	-1.6	1.6	N	INCHES	Y
EMAXT	92.6	57.0	125.0	71.9	113.2	G	DEG F	Y
EMINPCP	42.1	0.0	6.2	-0.3	1.0	N	INCHES	N
EMINT	90.0	-60.0	64.0	-40.7	40.4	G	DEG F	Y
EV3M	93.3	0.0	9.4	0.0	1.6	G	INCHES	Y
EXTMAXMONEVP	88.2	3.7	12.1	-1.6	1.6	N	INCHES	Y
EXTMAXMONRH	0.0	33.8	99.5	0.0	0.0	G	PCT	N
EXTMAXRH	91.2	51.0	100.0	83.8	100.0	G	PCT	Y
EXTMINMONEVP	22.7	0.0	10.0	0.2	0.9	N	INCHES	N
EXTMINMONRH	0.0	8.4	82.8	0.0	0.0	G	PCT	N
EXTMINRH	86.9	5.0	64.0	8.2	52.8	G	PCT	Y
FFD	0.0	56.0	366.0	0.0	0.0	G	DEG F	N
GSRAD	93.9	0.0	17627.1	0.0	9572.7	G	MJ/M2	Y
LENGS	0.4	1.0	365.0	0.0	2.9	G	DAYS	N
NDAY_SD_CAT1	20.1	0.0	227.0	0.1	1.0	N	DAYS	N
NDAY_SD15_ST2A	20.7	0.0	103.0	0.3	1.0	N	DAYS	N
NDAY_ST2A_FRZ	45.8	0.0	263.0	-0.3	0.9	N	DAYS	N
NDAY_ST4A_FRZ	49.9	0.0	266.0	-0.3	1.0	N	DAYS	N
PCPGS	90.2	0.0	158.2	-1.6	1.6	N	INCHES	Y
PEANN	94.8	0.2	628.3	0.0	199.8	G	INCHES	Y
PETGS	86.1	0.0	84.4	3.9	68.5	G	INCHES	Y
PEVR3M	84.5	39.7	1700.0	211.0	862.0	G	INCHES	Y
PR3M	87.8	0.0	18.6	-1.6	1.7	N	INCHES	Y
RAD3M	91.9	5.1	30.8	9.5	28.5	G	MJ/M2	Y
RH85%	1.3	0.0	30.7	1.2	2.0	G	DAYS	N
RH85P001	56.9	0.0	291.0	-0.7	1.1	N	DAYS	N
RH85P010	61.7	0.0	180.0	-0.6	1.0	N	DAYS	Y
RH85P025	65.7	0.0	139.0	-0.6	1.0	N	DAYS	Y
RH85P050	52.6	0.0	80.0	-0.6	0.9	N	DAYS	N
RH85P100	5.2	0.0	32.0	0.5	1.1	N	DAYS	N
ST2A	88.4	34.5	97.4	39.0	83.0	G	DEG F	Y

	Data Extremes	Tolerance						
Test Code	Pct Match	Minimum	Maximum	Low	High	Type of Dist	Units	Status
ST4A	88.4	32.8	92.7	37.1	79.0	G	DEG F	Y
ST2AXMAX	93.4	48.0	109.0	58.6	97.7	G	DEG F	Y
ST4AXMAX	93.1	46.0	104.0	55.7	92.9	G	DEG F	Y
ST2AXMIN	88.2	-12.0	74.0	12.2	48.7	G	DEG F	Y
ST4AXMIN	88.6	-11.0	70.0	11.9	46.3	G	DEG F	Y
T50-PE35-P12	14.9	0.0	12.0	-0.4	0.6	N	MONTHS	N
T50-PE50-P24	8.8	0.0	12.0	0.0	0.8	N	MONTHS	N
T50-PE75-P50	0.8	0.0	12.0	1.0	1.6	N	MONTHS	N
T64-PE35-P12	3.4	0.0	12.0	0.8	1.2	N	MONTHS	N
T64-PE50-P24	0.7	0.0	12.0	1.1	1.6	N	MONTHS	N
T64-PE75-P50	0.2	0.0	12.0	1.7	2.2	N	MONTHS	N
T72-PE35-P12	0.6	0.0	12.0	1.5	2.0	N	MONTHS	N
T72-PE50-P24	0.2	0.0	12.0	1.8	2.3	N	MONTHS	N
T72-PE75-P50	0.1	0.0	12.0	2.1	2.4	N	MONTHS	N
T80-PE35-P12	0.2	0.0	7.0	2.0	2.3	N	MONTHS	N
T80-PE50-P24	0.1	0.0	6.0	2.1	2.4	N	MONTHS	N
T80-PE75-P50	0.0	0.0	5.0	2.3	2.5	N	MONTHS	N
TROPPCP	4.7	0.0	158.2	1.5	1.8	N	INCHES	N
WETD	5.8	0.7	30.7	0.0	7.8	G	DAYS	N
WETH32-40	71.1	0.0	1785.0	-0.8	1.2	N	HOURS	Y
WETH40-50	77.7	0.0	2074.0	-0.9	1.2	N	HOURS	Y
WETH50-60	76.6	0.0	1927.0	-1.0	1.3	N	HOURS	Y
WETH60-70	70.9	0.0	2772.0	-1.0	1.2	N	HOURS	Y
WETH70-80	38.6	0.0	5896.0	-0.1	0.9	N	HOURS	N
WETH80-90	14.3	0.0	2356.0	0.8	1.2	N	HOURS	N

Note: The "Type of Dist" column indicates the kind of distribution assumed. (G) is Gaussian and (N) is Non gaussian. A gamma distribution was assumed for non gaussian variables. Variables having values of -999 were not used in the simulation.

The "Status" column indicates whether or not a variable was used [Yes/No] in the simulation.

Table 6. An example of the Species Distribution and Human Population Exposure to a Species Table

BAMM Specie Distribution and Population Exposure
Simulation Date: 2011_0903
Specie Name: Abrus_precatorius

Percent and Total Area in Each Match Class Interval

Percent Class	Region	Percent of Area/Legend Class	Area (sqkm) Per Legend Class
0%	Global	14.8	22914874.0
01-09%	Global	0.0	0.0
10-19%	Global	0.0	6839.0
20-29%	Global	0.3	398432.4
30-39%	Global	1.5	2251887.5
40-49%	Global	6.8	10543565.0
50-59%	Global	9.2	14272423.0
60-69%	Global	8.1	12471291.0
70-79%	Global	6.2	9644333.0
80-89%	Global	5.8	8963106.0
90-100%	Global	5.2	8011315.0
0%	US48	0.0	0.0
01-09%	US48	0.0	0.0
10-19%	US48	0.0	0.0
20-29%	US48	0.0	0.0
30-39%	US48	0.1	4285.4
40-49%	US48	1.3	96319.8
50-59%	US48	3.8	292358.5
60-69%	US48	7.3	575900.8
70-79%	US48	7.5	611577.0
80-89%	US48	10.8	877788.6
90-100%	US48	15.8	1258387.4
0%	US48-AK-HI	0.0	3362.5
01-09%	US48-AK-HI	0.0	0.0
10-19%	US48-AK-HI	0.0	0.0
20-29%	US48-AK-HI	0.0	0.0
30-39%	US48-AK-HI	0.5	26448.9
40-49%	US48-AK-HI	14.4	934145.4
50-59%	US48-AK-HI	15.3	1130970.4
60-69%	US48-AK-H	5.4	578531.6
70-79%	US48-AK-HI	5.5	612307.6
80-89%	US48-AK-HI	7.9	878516.9
90-100%	US48-AK-HI	11.5	1258387.4
0%	US48-AK-HI-PR	0.0	3362.5
01-09%	US48-AK-HI-PR	0.0	0.0
10-19%	US48-AK-HI-PR	0.0	0.0
20-29%	US48-AK-HI-PR	0.0	0.0
30-39%	US48-AK-HI-PR	0.5	26448.9
40-49%	US48-AK-HI-PR	14.4	934145.4
50-59%	US48-AK-HI-PR	15.3	1130970.4

Table 6, continued

Percent Class	Region	Percent of Area/Legend Class	Area (sqkm) Per Legend Class
60-69%	US48-AK-HI-PR	5.4	578531.6
70-79%	US48-AK-HI-PR	5.5	612307.6
80-89%	US48-AK-HI-PR	7.9	878516.9
90-100%	US48-AK-HI-PR	11.5	1258387.4
0%	HI	0.0	0.0
01-09%	HI	0.0	0.0
10-19%	HI	0.0	0.0
20-29%	HI	0.0	0.0
30-39%	HI	0.0	0.0
40-49%	HI	0.0	0.0
50-59%	HI	0.0	0.0
60-69%	HI	0.0	0.0
70-79%	HI	1.8	730.6
80-89%	HI	1.8	728.3
90-100%	HI	0.0	0.0

Population Exposure to Each Match Class Interval

Percent Class	Population
0%	96099960
01-09%	0
10-19%	35016
20-29%	1686708
30-39%	6407624
40-49%	60999008
50-59%	162221408
60-69%	339923392
70-79%	332595488
80-89%	281484000
90-100%	409669920

Table 7. An example of a file that can be used to calculate statistics for model validation and comparison.

x	y	train or test	percent test-yrs passed
29.6250	-31.3750	train	89.4
31.1250	-29.6250	train	81.7
31.3750	-29.1250	train	73.3
31.8750	-28.8750	train	94.0
32.3750	-28.6250	train	94.7
153.5000	-28.4000	train	83.0
31.8750	-28.1250	train	91.2
153.0000	-27.4000	train	88.9
-82.7300	27.9200	test	59.3
50.4667	-15.2167	train	92.7
48.2500	-14.0667	train	97.6
-38.9520	-13.9880	train	97.2
48.5000	-13.6167	train	96.9
49.6553	-13.2678	train	99.5
132.8000	-13.2000	train	99.3
49.5833	-13.0667	train	99.6
45.2003	-12.9764	train	96.6
49.1000	-12.9167	train	96.5
49.5167	-12.6500	train	98.0
131.8000	-12.5000	train	98.4
130.8000	-12.4000	train	94.1
43.8522	-12.3758	train	97.1
130.5000	-11.3000	train	93.1
131.5000	-11.3000	train	96.8
148.1750	-10.0917	train	98.7
147.3420	-9.5917	train	98.8
147.4170	-9.4167	train	99.1
147.0080	-9.2583	train	97.8
146.6580	-7.1750	train	84.4
147.0080	-6.7583	train	78.4
147.0080	-6.6250	train	52.3
15.0667	-5.8500	train	99.3
147.0920	-5.3417	train	88.3
29.3500	-3.3833	train	93.1
29.2100	-3.3200	train	56.6
13.2833	-1.6667	train	91.1
8.7833	-0.7167	train	93.8
25.1100	0.2900	train	92.8
9.5333	0.3500	train	92.5
24.4500	0.7667	train	93.2
12.0167	2.1833	train	93.4
-52.0333	3.5667	train	96.1
-7.5500	4.3667	train	90.4
-52.2708	4.9073	train	94.5