A Predictive Model of Rock Art Sites in Wisconsin's Driftless Area: Preliminary Results

Abstract

This poster presents the preliminary results of independent research undertaken to develop a predictive model for rock art site sensitivity within Wisconsin's Driftless area. Forward stepwise logistic regression was utilized to determine the independent variables that were included in the model and raster values were extracted to the dataset using bilinear interpolation. The model run resulted in a Geographic Information Systems-based mapped predictive surface. The model was then analyzed in conjunction with SYSTAT output statistics to determine model quality. The resultant Kvamme's Gain Statistic indicates a strong predictive model. The results placed 97% of reported rock art sites in high/medium probability areas constituting 18% of the study area, while 85% of the random point dataset was placed in low probability areas constituting 82% of the study area. Field testing of the model is planned for 2013.

Dataset

The WiDARAS model was developed with a binary classified dataset composed of a control group and a random point group.The control group is comprised of 104 rock art sites recorded in the Wisconsin Historical Society's Archaeological Site Inventory Database. The control group includes sites located in the counties of Dane (11), Grant (1), Iowa (36), Jackson (8), Juneau (18), La Crosse (7), Monroe (6), Pepin (1), Richland (2), Sauk (2), and Vernon (12). Reported cultural affiliations represented in the control group include Early and Late Paleo-Indian, all stages of the Archaic period, all stages of the Woodland period, Oneota, Historic Native American, Historic Euro-American, as well as unknown prehistoric, unknown historic, and simply unknown. Reported site types represented in the control group include rock art, cave/rockshelter, quarry, cemetery/burial, isolated finds, workshop site, and campsite/village. The random point group was generated using ESRI's random point generator tool and includes a total of 196 random points.

Environmental Variables

A total of 61 independent variables were tested during this first phase of modelling. These variables include:

National Elevation Dataset (NED) Derivatives:

Elevation
Slope
Aspect

Distance Variables:

Bedrock Depth (split based on percentage of area within a certain distance to surface)
Bedrock Geology (based on unit age and rock types)
Hydrology Junctions (confluences, shoreline intersection, main drain, headwater)
Intermittent Hydrology (Backwater, Lake/Pond and River/Stream)
Perennial Hydrology (Backwater, Lake/Pond and River/Stream)
River/Stream Orders
Original Vegetation types
Ecotones

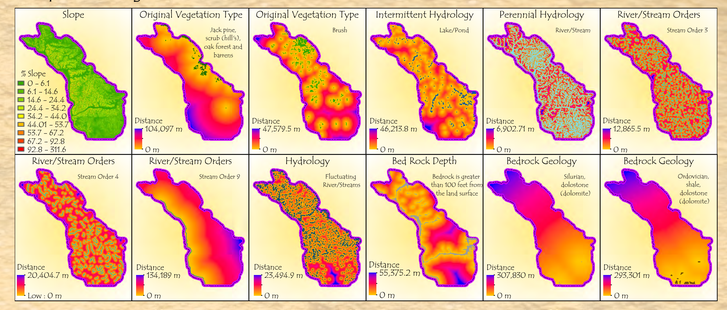
One Kilometer Densities:

Confluence
Stream/River
Shoreline
Main Drain
Headwater

Other Raster Datasets:

Ecological Diversity (1 kilometer)

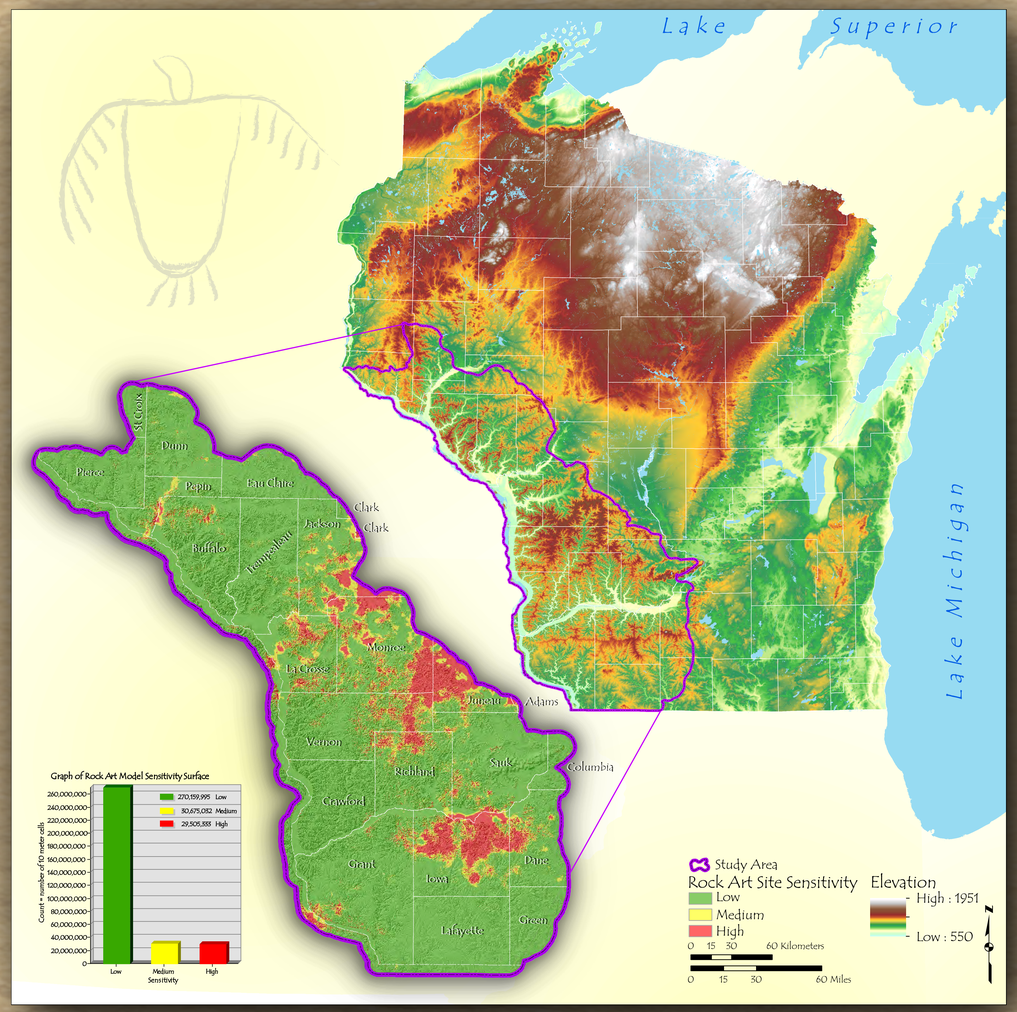
Forward stepwise logistic regression was utilized to determine the subset of independent variables used to create the predictive surface. Twelve of the 61 tested variables were deemed significant and used to create the predictive surface, and are illustrated below. It is interesting to note that logistic regression did not identify Aspect as a significant variable.



Allison M. Lange Mueller and Kevin J. Mueller

Wisconsin Driftless Area Rock Art Site (WiDARAS) Predictive Model

Wisconsin's Driftless area contains approximately 41 percent of all reported rock art sites in the state (Lange 2006). This is likely due to the environmental aspects of the Driftless area; prehistorically the region was unaffected by glacial advance and therefore harbours well-preserved Upper Cambrian sandstone and Ordovician limestone bedrock outcrops. These outcrops contain caves and shelters situated alongside sloping valleys. Thus, the region presents an ideal setting for rock art production. The WiDARAS study area includes the Driftless area with a 5 kilometer surrounding buffer. The model was developed to guide the authors in targeting discrete high probability areas for rock art within the region, in order to locate new rock art sites. To that end, the WiDARAS predictive model is non-parametric, as it is the intention of the authors to refine and develop the model to incorporate new rock art site data as it is identified and recorded over time.



WiDARAS Predictive Model: Preliminary Results

Statistical results placed 97% of reported rock art sites in high/medium probability areas constituting 18% of the study area, while 85% of the random point dataset was placed in low probability areas constituting 82% of the study area. While the Kvamme's Gain Statistic, Mann Whitney U test, and the ROC and AUC results indicate that the WiDARAS model is highly accurate, it is clear that the predictive surface is not precise. This is likely due to the high density of reported rock art sites in Iowa and Juneau counties, as sites within these two counties account for approximately 49 percent of all reported rock art sites in the Driftless area (Lange 2006). Thus it appears that the Iowa and Juneau county sites are skewing the predictive surface, although it is important to note that the model identified many smaller, discrete areas of high/medium probability in counties which have little or no reported rock art sites.

Future Steps

The next step in the modelling process is field testing of the smaller, discrete high/medium probability areas, primarily located outside of Iowa and Juneau counties. The results of the field testing will augment the binary classified data set and periodic refinement of the WiDARAS model with new data should result in the creation of a more precise predictive surface. Field testing of the high/meduim probability areas identified in the WiDARAS predictive model will begin in 2013.

Kvamee's Gain Statistic

The Kvamme's Gain Statistic is one of the most prominent model evaluation methods. It is based on the assumption that if the high potential area is small relative to the overall study area and the number of sites found within it is large in relation to the total for the entire study area, then it is a fairly accurate model. Numerically, the Kvamme's Gain Statistic ranges from 0 to 1, with 1 indicating a perfect model that predicts all possible instances with little known information and 0 indicating a model that predicts no better than random chance. The WiDARAS model produced a 0.97 Kvamme's Gain Statistic, indicating a very strong predictive model.

Mann-Whitney U

The Mann-Whitney U test is a non-parametric hypothesis test used to determine whether one of two sample datasets has larger significance than the other, and it is one of the most well-known non-parametric significance tests utilized to determine the strength of a predictive model. The results of the Mann-Whitney U test indicate that the distribution of variable values is significantly different (at the 0.05 level) from the random point dataset compared to the control dataset for all except two variables, Distance to Fluctuating Rivers/Streams and Distance to Stream Order 4. Thus the model effectively identifies a set of variables which discriminates between the two populations. This signifies that rock art sites are not randomly distributed across the landscape.

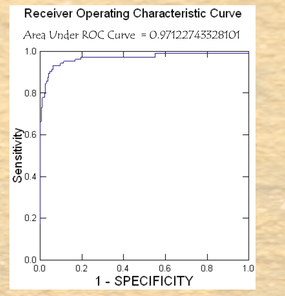
Mean Comparison

Variables	Modeled Sites	Low Prob. Sites	Random Points	Mann-Whitney
BG_WIOM0_DIST	64,333m	109,684m	98,137m	0.0001723938840
BG_WISU0_DIST	66,132m	116,439m	107,044m	0.0000014717210
BRD_100FT_DIST	16,951 m	16,305m	11,961m	0.0000334327450
HYDRO_FLUCTUATING_RIVERSTREAMS_DIST	2,834m	3,225m	3,271m	0.8876735814753
HYDRO_INTERMITTENT_LAKEPOND_DIST	342 m	5,605m	6,700m	0.0268407153230
HYDRO_PERENNIAL_RIVERSTREAM_DIST	568 m	514m	807m	0.0004453590173
HYDRO_ST_ORDER_3_DIST	2133m	1,031m	1,488m	0.0000001983185
HYDRO_ST_ORDER_4_DIST	3200m	904m	3,264m	0.9253478333344
HYDRO_ST_ORDER_9_DIST	66,884 m	22,870m	44,625m	0.0000000000035
OV_BRUSH_DIST	15,843 m	2,747m	8,843m	0.0000009528512
OV_JPSOFB_DIST	16,713 m	31,808m	35,092m	0.0000000000489
SLOPE	36	12	14	0.0000000000000

Receiver Operating Characteristic (ROC) Curve

The relationship between a binary classified system of sensitivity and specificity is analyzed and illustrated using a ROC curve, a graphical plot which illustrates the performance of a binary classified system. The ROC analysis is related to the cost/benefit analysis of diagnostic decision making and allows for the independent choice of optimal models from suboptimal models based on cost/benefit context.

The diagnostic accuracy of the ROC curve is determined by the Area under the Curve (AUC). The AUC is typically represented by a traditional academic point system, with the score of 1 representing perfect accuracy, a score of 0.5 representing random chance, and a score of less than 0.5 representing worse than random chance. The result of the AUC on the WiDARAS model is 0.97, indicating a near perfect score of diagnostic accuracy.



Thanks to

COMMONWEALTH CRG CULTURAL RESOURCES GROUP, INC.