Quantification of Water-Level Variability Effect on Plant Species Populations Using Paleoecological and Hydrological Time Series Data

Edwin A. Roehl Jr[1], Paul A. Conrads[2], and Christopher Bernhardt[3]

AUTHORS: [1]Advanced Data Mining Services, LLC, Greer, SC, [2] U.S. Geological Survey, Columbia, SC, [3]U.S. Geological Survey, Reston, VA
REFERENCE Proceedings of the 2012 South Carolina Water Resources Conference, held October 10-11, 2012 at the Columbia Metropolitan Convention Center

ABSTRACT. Soil cores provide valuable data on historical changes in vegetation and hydrologic conditions. Empirical models were developed to quantify the effect of meteorological and hydrologic forcing on plant species distributions over a 110-year period in Water Conservation Area 1 (WCA1) in the Florida Everglades, also known as the Arthur R. Marshall Loxahatchee National Wildlife Refuge. Empirical models that predict plant species distributions at sites within WCA1 were developed by linking temporally sparse seed bank data from soil cores with continuous multi-decadal daily meteorological and hydrologic time series data. The meteorological data included rainfall and maximum daily temperatures that spanned the entire study period of 110 years. The hydrologic data included stage data from two gages in WCA1 established in 1954. These stage data were hindcasted to be concurrent with the meteorological data by using correlation models that fit measured stages as a function of the meteorological parameters. The historical plant species data came from seven peat cores from WCA1. Different depths from each core were carbon-dated and assayed for relative percentages of 83 plant species using pollen counts. The oldest dates were more than 1,000 years old; however, only core data that overlapped the study period were used, for a total of 67 assays among the seven cores. Twenty-three of the species had ratios of at least 5 percent for one or more of the 67 assays, hereafter referred to as the "top23".

Using the assays as input vectors, the top23 were grouped using the k-means clustering into four plant classes that represented the extent to which the various species have historically appeared together. This reduced the modeling problem to one of predicting the relative ratios of the four plant classes from the hindcasted stage time-series data. A separate empirical model was developed for each class using a multi-layer perceptron artificial neural network, which provides multivariate, nonlinear curve fitting. The models predicted the relative ratios of the classes, and the sums of the predictions are near 1. The coefficient of determination (R^2) of the models varied from 0.87 to 0.96, indicating that the relative ratios of the plant classes are predictable, and therefore controllable, from stage forcing. Similar soil cores are available for the Coastal Plain of North Carolina and are planned for the Congaree National Park in South Carolina.

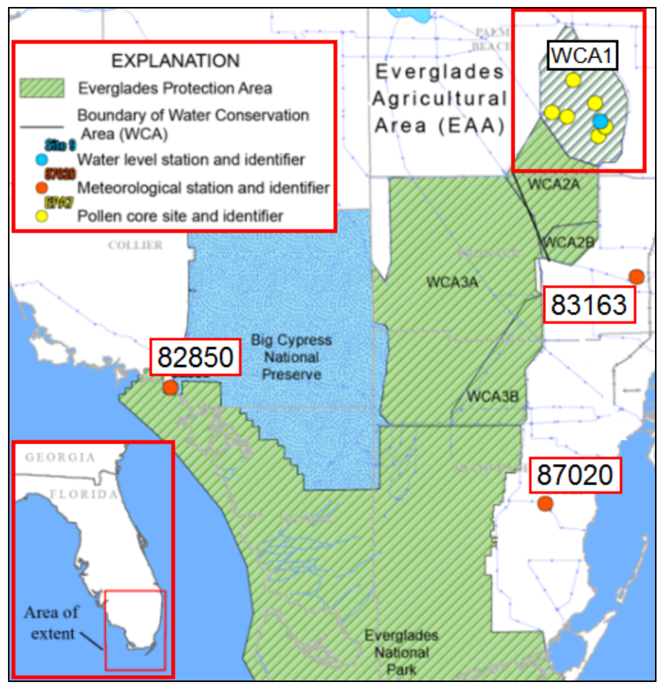

Figure 1. Data collection sites. Core and water level sites are in Water Conservation Area 1 (WCA1) at upper right.

INTRODUCTION

Long-term changes in meteorology and hydrology cause variations in plant species distributions and dependent wildlife. For more than a decade work in the Florida Everglades has focused on understanding this relationship for the purpose of restoring and managing this national treasure. The development of similar knowledge is now of interest in the Coastal Plain of North Carolina and the Congaree National Park in South Carolina. This paper describes the development of a process model that predicts plant species distributions for given water level scenarios for a portion of the Everglades.

PROJECT OBJECTIVE

The objective of this project was to develop a method for predicting plant species distributions as a function of water level to be used as a planning tool for habitat restoration and management.

PROJECT DESCRIPTION

Intact soil cores from the marshes provide valuable data on historical changes in vegetation and hydrologic conditions. Pollen and surface water-level data from the WCA1 and data from long-term meteorological monitoring stations (fig. 1) were used to develop empirical predictive models of plant distributions from a specified water-level history. The data were:

- *Meteorological Data* – three precipitation and air temperature datasets were downloaded from the National Oceanic and Atmospheric Administration's Global Historical Climatology Network (http://www.ncdc.noaa.gov/ghcnm/). Period of record: 1895 to 2011.
- *Hydrologic Data* – water-level data from Site in WCA19 (fig. 1) were downloaded from the South Florida Water Management District DBHYDRO database (http://www.sfwmd.gov). Period of record: 1954 to 2010.
- *Plant Species Assays* – U.S. Geological Survey data (unpublished) from seven cores were used for this study (fig. 1). The data included the relative abundance of 83 plant species using pollen counts and age models for each core. The age models for the cores varied from 380 to 1,470 calibrated years before the present (Traverse, 2007).

Transforming large numbers of parameters, such as the 83 plant species' relative abundance ratios, into a smaller set that accurately represents observed process behaviors is a means to reduce the dimensionality and complexity of analysis and modeling problems. The method for clustering the time series into a small set of classes is described by Roehl and others (2006). Only data overlapping the meteorological data were used in the study, leaving 67 (of the 83) assays from the seven coring sites. Twenty-three species with relative abundance of at least 0.05 (5 percent) for one or more of the 67 assays were used for the cluster analysis. Table 1 lists the resulting four class assignments of the "top 23" plant taxa time series.

Table 1. "Top 23" taxa and their class assignments.

Taxa	Class	Taxa	Class
Blechnum	1	Amaranthaceae[2]	3
Casuarina	1	*Ambrosia*	3
Nymphaea	1	*Cladium*	3
Quercus	1	Cyperaceae	3
Thyelypteris	1	*Pinus*	3
Ambrosia -like	2	*Sagittaria*	3
Asteraceae[1]	2	Triporate pollen	3
Chenopodiaceae/Amaranthaceae[2]	2	Asteraceae indet.[1]	4
Morella	2	*Cephalanthus*	4
Osmunda regalis[3]	2	*Ilex*	4
		Monolete fern spores	4
		Osmunda spp.[3]	4
		Trilete fern spores	4

[1, 2, 3] These taxa were not combined for this analysis

To obtain concurrent data between the three datasets, the surface water-level data from Site 9 were appended with hindcasted data back to 1923 using an artificial neural network (ANN) model as described by Jensen (1994). The inputs to the model were created by decorrelating and decomposing the raw rainfall time series into different spectral ranges from one month to 10 years. Figure 2 shows the measured water level with the model predictions. Note that the long-term trend shows an increase of approximately 2 feet, which has likely been a principal cause of habitat change in the WCA1.

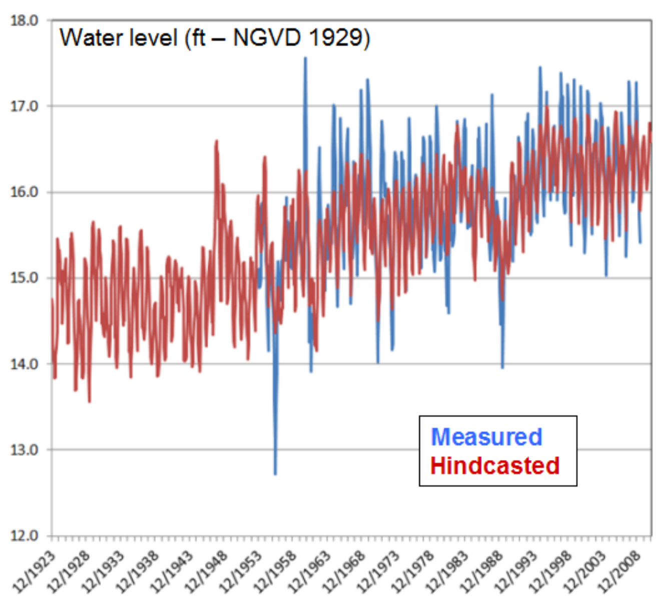

Figure 2. Measured and hindcasted monthly water levels for Site 9 for the period 1923 to 2010.

The modeling goal was to develop numerical models that predict the relative abundance of the four vegetation classes (table 1) as functions of water level. The inputs to the models are derived monthly water levels for Site 9 in

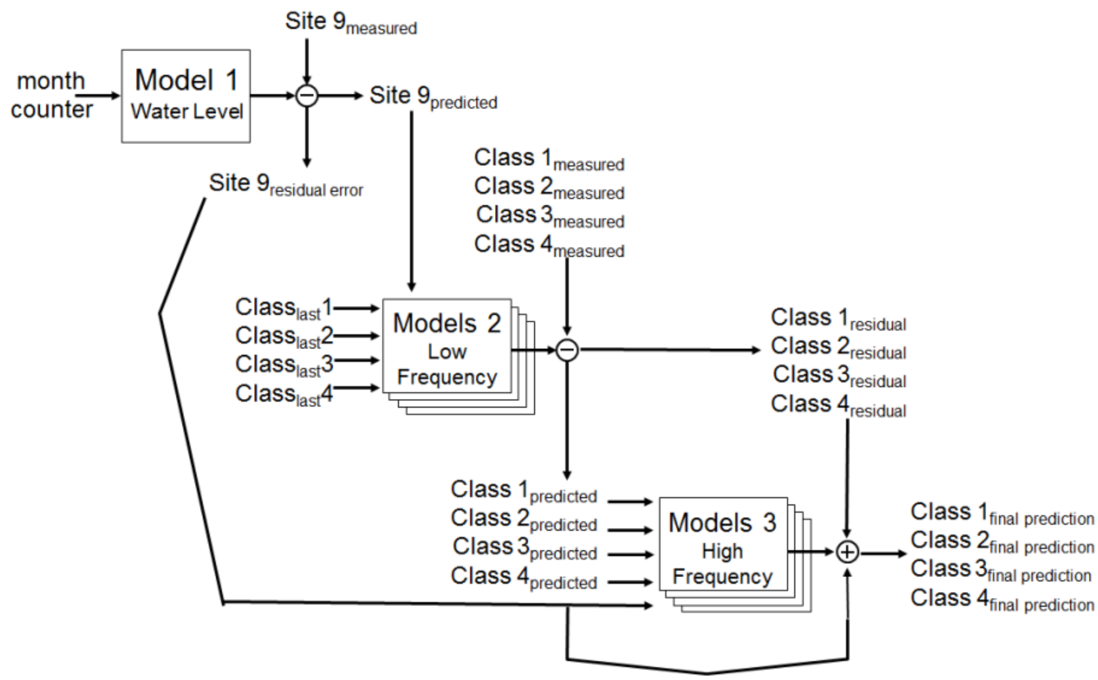

Figure 3. Super-model architecture showing connections of sub-models.

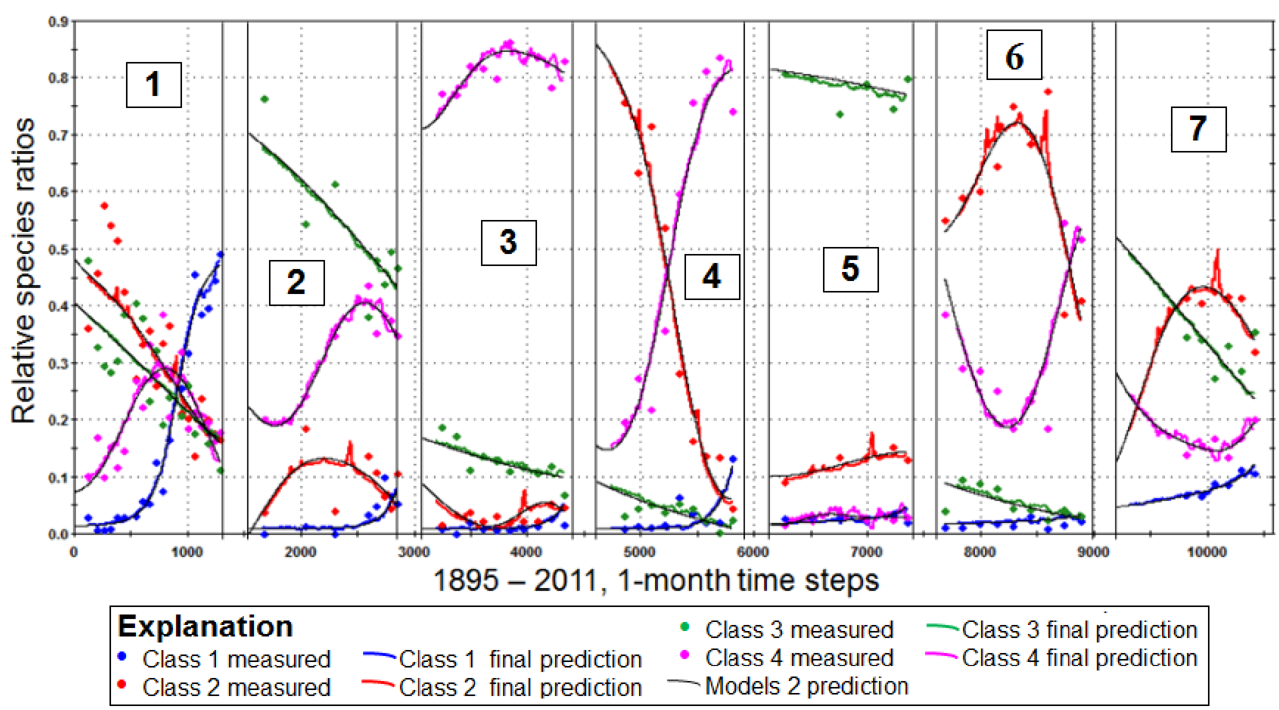

Figure 4. Measured and predicted class assignments from the Models 2 and Models 3 for seven cores.

addition to the most recent class abundance, which represent an “end condition”. Note that water depths vary throughout WCA1, and consequently, plant distributions also vary site-to-site. Over time, channels have been cut along the perimeter of WCA1 to move water from agricultural areas to the north to other water conservation areas to the south. In addition, smaller channels have been cut into the interior for access to the marsh. The consequence of the channelization has been the transport of non-point source nutrient loads from agriculture and land development to parts of WCA1. The end condition inputs enable the model to predict what the current plant distribution will change into for a given water level scenario.

The vegetation models are “sub-models” that collectively comprise a “super-model” (fig. 3). The steps taken to develop the super-model were as follows.

1. Develop Model 1 to generate a low-frequency component of Site 9 water levels using monthly counter input by fitting the hindcasted data (fig. 2) with a least-squares regression straight line.
2. Configure a stacked dataset that combines static (categorical) and dynamic (time series) data. The complete datasets for the cores are stacked one on top of the other. This provides for training ANN models to learn input-output relations that are common to all of the cores. The dynamic data included the hindcasted hydrology and class relative abundance ratios. The static data included the locations of cores and end-condition ratios.
3. Develop Models 2 to predict the low-frequency variability of each class ratio using the stacked dataset. A separate ANN was trained for each ratio.
4. Develop Models 3 to predict the high-frequency variability of each class ratio. A separate ANN was trained for each ratio. The inputs were the Model 1 residuals (prediction error = measured – predicted values), and the outputs were the Models 2 residuals.
5. The final predicted class ratios are the summation of the predictions from the Models 2 and Models 3 (fig. 4).

RESULTS AND DISCUSSION

From the prediction plots (fig. 4) and the model performance statistics (coefficient of determination and percent model error of the ANN training and testing datasets listed on table 2), it appears that long-term rather than short-term water-level change is the primary driver of the plant population distribution. The high frequency variability in the final model predictions (fig. 4) is not much different than the Models 2 predictions and the coefficients of determination for the Models 3 indicate that the models capture less than 10 percent of the high frequency variability of the data. While there are potentially several sources of error, such as the hindcasted Site 9 water-level, unaccounted nutrient loading, and ambient temperature change, it is perhaps most likely that the assay dates are insufficiently accurate to be correctly synchronized with the stage and meteorological data. Errors of plus or minus a year or two for each assay would prevent the ANNs from learning cause-effect relationships on a seasonal time scale.

Table 2. Performance statistics for the ANN sub-models [N, count; R^2, coefficient of determination; PME[1], percent model error = root mean square error of the model predictions divided by the range of the observed data.]

		N		R^2		PME	
Model	**Output**	**training**	**testing**	**training**	**testing**	**training**	**testing**
Models 2	C1	54	13	0.979	0.905	3.5	6.6
	C2	50	13	0.956	0.875	6.5	6.9
	C3	54	13	0.972	0.919	4.9	6.7
	C4	46	13	0.980	0.955	4.6	4.9
Models 3	C1 Residuals	52	12	0.040	0.046	17.5	33.6
	C2 Residuals	52	12	0.102	0.099	18.2	14.9
	C3 Residuals	51	12	0.014	0.075	16.9	20.5
	C4 Residuals	51	12	0.031	0.069	18.1	16.7

[1] The percent model error (PME) is the root mean square error of the model predictions divided by the range of the observed data.

Fig. 4 shows significant consistency in the model predictions for all of the cores. Note that for all cores the Class 1 ratios increase monotonically from low initial values; and that as a likely consequence, the Class 3 ratios decrease monotonically from values that are initially higher than those of Class 1. The Class 4 ratios vary non-monotonically in five cores (1, 2, 3, 6, and 7), and clearly vary oppositely to the Class 2 ratios for significant portions of the time in all but core 5. These results appear to confirm that the long-term increasing trend in water level (fig. 2) is a primary driver of plant distributions, and that the Models 2 are adequate predictors of the outcomes of alternative water level management scenarios.

LITERATURE CITED

Jensen, B.A., 1994, Expert systems - neural networks, instrument engineers’ handbook third edition: Chilton, Radnor PA, p. 48-54.

Roehl E., Risley J., Stewart J. and Mitro M., “Numerically optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems – Part 1”, Proceedings for the Environmental Modeling and

Software Society Conference, Burlington, Vermont, USA, (2006), pp 1-6.
Traverse, A., 2007, Paleopalynology (Second edition): The Netherlands, Springer, 813 p.