USGS
science for a changing world

http://www.evergladesplan.org

Quantifying How Water Level Variability Affects Plant Species Populations Using Paleoecological and Hydrological Time Series Data

Paul Conrads[1], Edwin A. Roehl Jr[2] and Christopher Bernhardt[3] [1]U.S. Geological Survey, Columbia, SC [2]Advanced Data Mining Services, LLC, Greer, SC [3]U.S. Geological Survey, Reston, VA

Changes in meteorology and hydrology have historically led to variations in the populations of different plant species in the Florida Everglades (fig. 1). Intact soil cores from the Everglades marshes provide valuable data on historical changes in vegetation and hydrologic conditions. Pollen and surface water-level data from the Arthur R. Marshall Loxahatchee National Wildlife Refuge and data from three long-term meteorological monitoring stations were used to develop empirical predictive models of plant distributions from a specified water-level history.

Data Used in the Study

Meterological 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/, fig 1). Period of record: 1895 to 2011.

Hydrologic Data

Water-level data from Site 9 (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U.S. Geological Survey data (unpublished) from seven cores were used for this study (fig. 1).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).

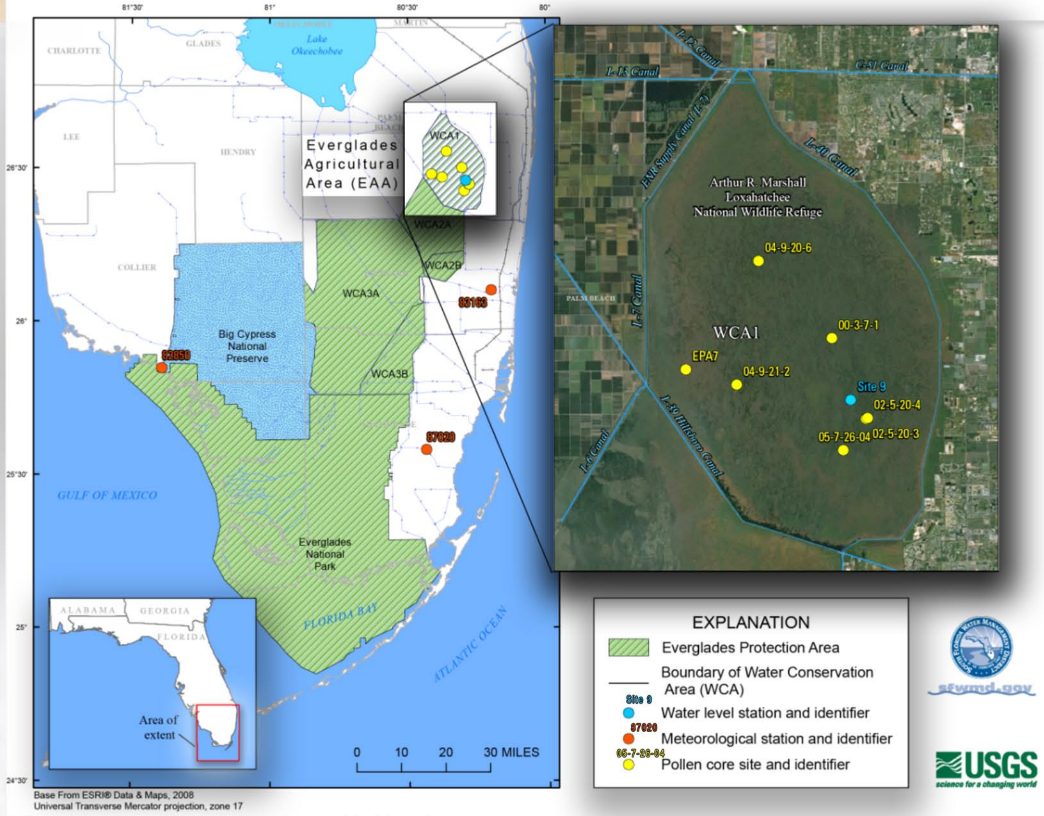


Figure 1. Location of data collection sites used in this study.

Cluster Analysis

Transforming large numbers of parameters, such as the 83 plant species' relative abundance ratios, into a small 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.

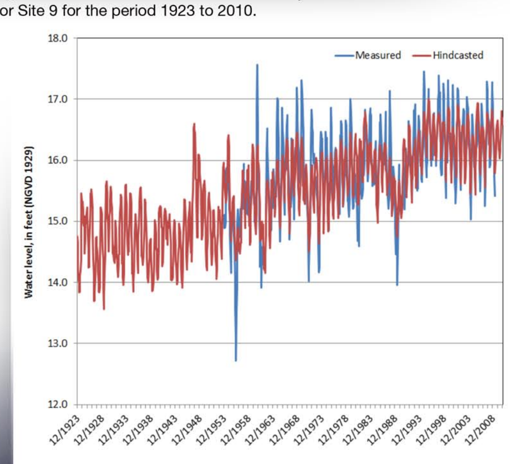
Taxa	Class	Taxa	Class
Blechnum	1	*Amaranthaceae* [2]	3
Casuarina	1	*Ambrosia*	3
Nymphaea	1	*Cladium*	3
Quercus	1	*Cyperaceae*	3
Thyelypteris	1	*Pinus*	3
Ambrosia -like	2	*Sagittaria*	3
Asteraceae [1]	2	*Triporate pollen*	3
Chenopodiaceae/Amarantaceae [2]	2	*Asteraceae indet* [1]	4
Morella	2	*Cephalanthus*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

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

Hindcasted Hydrology

To obtain concurrent data between the three datasets, the surface water-level data from Site 9 (fig. 1) 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 temporal ranges from one month to 10 years. Figure 2 shows the measured water level with the model predictions.

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



Vegetation Modeling Approach

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 addition to the most recent class abundance, which represent an "end condition". 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.

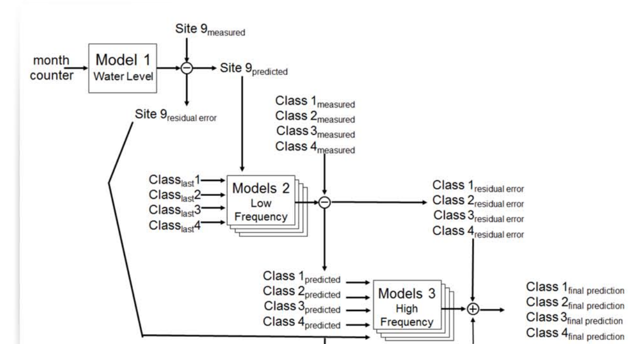
2. Configure a stacked dataset that combines static (categorical) and dynamic (time series) data. The complete datasets for the cores are stacked one on top of the other. This provides for training ANN models to learn input-output relations that are common to all of the cores. The dynamic data included 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 model was trained for each ratio.

4. Develop Models 3 to predict the high-frequency variability of each class ratio. A separate ANN model 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).

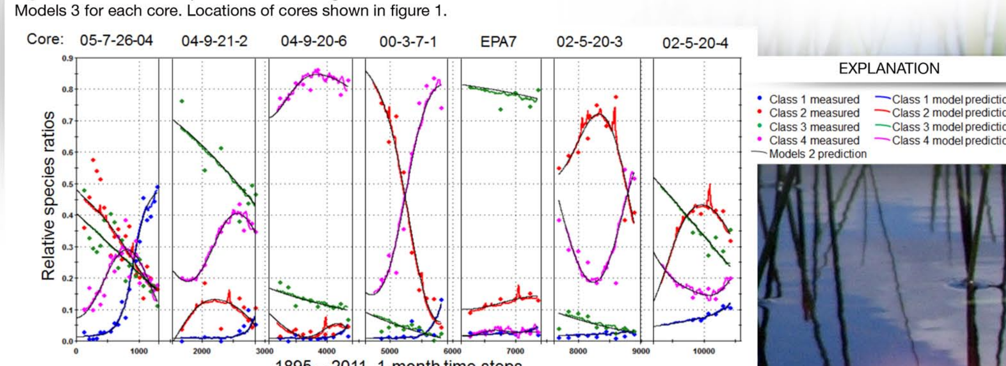
Figure 3. Super-model architecture showing connections of sub-models.



Model Results and Discussion

From the prediction plots (fig. 4) and the model performance statistics (coefficient of determination and percent model error of the ANN Model 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 coefficient of determination for the Models 3 indicate that the models capture less than 10 percent of the high frequency variability of the data. The Models 2 are probably adequate for estimating long-term plant distribution; Models 3 predict only a little of the high frequency variability in the class assignments. While there are potentially several sources of error, such as hindcasted Site 9 water-level data and unaccounted 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 ANNs from learning cause-effect relationships on a seasonal time scale.

Figure 4. Measured and predicted class assignments from the Models 2 and Models 3 for each core. Locations of cores shown in figure 1.



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

Table 2. Performance statistics for the artificial neural network sub-models.
[N, count; R2, coefficient of determination; PME[1], percent model error]

Model	Output	N training	N testing	R^2 training	R^2 testing	PME training	PME 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-Residual	52	12	0.040	0.046	17.5	33.6
	C1-Residual	52	12	0.102	0.099	18.2	14.9
	C1-Residual	51	12	0.014	0.075	16.9	20.5
	C1-Residual	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.

Water Conservation Area 3A of the Florida Everglades
Photo by Christa Zweig, Research Assistant, Florida Cooperative Fish and Wildlife Research Unit, University of Florida).

References

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.

For more information, please visit the EDEN web at:
http://sofia.usgs.gov/eden

EVERGLADES DEPTH ESTIMATION NETWORK EDEN

U.S. Department of the Interior
U.S. Geological Survey

June 2012