Establishing ecological criteria and thresholds for monitoring hydrologic impacts on Old Woman Creek wetlands

Project Final Report – Task 2: modeling hydrologic conditions

Introduction

In conjunction with task one, we also carried out two modeling tasks, with an objective of quantifying the impact of projected climate changes on hydrological conditions. Because most ecological indicators will be affected by wetland hydrology, understanding the changes in hydrological condition due to climate change can help us anticipate potential adverse impact of climate change on the Old Woman Creek (OWC) ecosystem. These tasks were proposed to achieve the following objectives.

```
1. Exploring the relationship between OWC water levels and
hydrologic and climate conditions to develop a predictive model of
water level,
2. Exploring the feasibility of mapping the changes in surface
water area using remote sensing techniques via un-manned ariel
vehicles (drones), and
3. Establishing the relationship between surface water area and
OWC water level.
```

In the rest of the report, we document the progress made to achieve these goals. As stated in the project proposal, we will develop a working model for predicting OWC water level using relevant predictors widely available from USGS and regional climate data services. The remote sensing component will be a feasibility assessment, largely because of the limited windows of drone operation in the project duration. The remote-sensing component of the projrct will continue after the close of the project, funded through other mechanisms.

Hydrological Modeling

Data Sources and Processing

Using OWC's long-term water quality monitoring program data collected continuously from 1983 to present, as well as daily discharge record from a USGS gauging station and weather data from a nearby weather station, we developed an empirical model to predict the changes of water depth at ??? site. The automatic data logger collects water quality and other environmental data. Water level data were calculated based on measured puresure near the bottom of the data logger. Using the silmulteneously measured berometric presure, the presure measured at the bottom of the data logger is converted to the depth of water, after adding the known distance from the data logger to the bottom of the wetland. The water depth is measured once every 15 minutes. These high frequency water level data were aggregated to daily average water depth, in accordance with the temporal scales of most potential predictors.

Daily flow rates were downloaded from USGS gauging station 04199155 (Old Woman Creek at Berlin Road, near Huron, OH, Latitude 41°20′5″N, longitude 82°30′50″W). Daily weather (precipitation and temperature) records from the on-site weather data logger were used. The flow and precipitation represent input of water to the reserve while temperature was used as a surrogate for evapotranspiration. Because a temporal lag may exist between the changes in water level and the changes in input and output, various lengths (10-, 20-, and 30-day) moving averages of these predictor variables were also included as potential predictors.

Data were grouped based on two factors, season and Lake Erie connection barrier status. Seasonal changes in the water level response to changes in input and output are expected because of the seasonal changes in relative importance of each predictor. OWC connects to or separates from Lake Erie via a sand barrier, shifting from open to close depending on wheather and wind conditions. When the barrie is open, OWC water level is also affected by Lake Erie water level, which may offset the effects of input and output variables. Two categorical predictors are included to represent seasonal and Lake Erie connectivity effects.

Data Analysis

Exploratary Data Analysis

A typical exploratory data analysis for developing a predictove model consists of plots of the response variable (water depth) against potential predictors. These plots enable us to make proper variable selection and transformation decisions. When categorical predictors are involved, we also use conditional plots to explore potential variations due to these categorical predictors.

Regression Analysis

Based on EDA, we list as many plausible predictors as possible, largely decided by data plots and our understanding of the relationship. In this case, our understanding of the factors affecting water level changes. For developing a useful predictive model using regression, we follwo a number of general rules developed by Qian (2016). These rules were intended for developing predictive models, not for testing hypotheses of whether a predictor is causally linked to the response. As a result, statistically insignificant predictors can be included as long as the slope is of the correct sign. Including an insignificant predictor will not improve or reduce a model's predictive capability. But if the included predictor is scientifically relevant, the inclusion of a statistically insignificant predictor would allows us to keep the option of updating the model later when more data are available. The general guidelines are 1. Include all predictors that are substantively relevant in the model 2. Using known mechanisms to guide the selection of variables 3. When little or no knowledge about the relationship is available, use tree-based model to find relevant predictors 4. For those predictors that have strong effects, consider including interaction terms 5. When a predictor has a slope that is statistically not different from 0, include the predictor if the slope is of the right sign, or expected sign based on our understanding of the subject.

Furthermore, we have considered various transformations. Through a systematic manual search over a number of combinations of transformations and model forms, we examined over 25 model forms and determined the best predictive model based on regression model diagnostic statistics based on residuals, as well as predictive accuracy based on Akaike Information Criterion.

Results

Exploratory Data Analysis

All relevant variables were considered. These variables include daily discharge (flow), daily precipitation (prec), daily mean temperature (temp), daily solar radiation (Ra), and their 10-, 20-, and 30-day moving averages. As water level is an aggregated summary of all input and output of water to the system, we expect that the role of each variable to vary by season. Furthermore, water level can be influenced by Lake Erie water level when the barrier mouth is open. Numerous plots were explored to generate a list of potential predictors (See on-line supplement materials).

The final model

Although each predictor cannot adequately predict the changes in water depth, combining four predictors in a log-log linear model (with one interaction term) is shown to be adequate. The best model, among the over 25 models examined, is of the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 \times x_2$$

where y is the water depth measured at site OL, x_1 is the 10-day moving average of daily discharge (measured at the USGS gauging station 04199155), x_2 is the 30-day moving average of daily precipetation, $x_3 is the 10-day moving average of solor radiation, and x_4 is the 30-day moving average of daily mean temperature. All variables are (natural) log-transformed. Regression coefficients

$(\beta_0, \cdots, \beta_5)$ are allowed to vary by season and by Lake Erie connection status. That is, $\beta_k (k = 0, 1, \cdots, 5)$ represents the sum of three terms:

$$\beta_k^{ij} = \theta_k + \delta_{skj} + \delta_{oki}$$

where ij represent ith opening status ($i = 1, 2$, for open and closed, respectively) and jth season, $\sum_i \delta_{oki} = 0, \sum_j \delta_{skj} = 0$. Because of the inclusion of an interaction term, all predictors were centered at their respective geometric means. That is,

$$\begin{aligned}
x_1 &= \log(flow_{ma_{10}}) - \overline{\log(flow_{ma_{10}})} \\
x_2 &= \log(prec_{ma_{30}}) - \overline{\log(prec_{ma_{30}})} \\
x_3 &= \log(Ra_{ma_{10}}) - \overline{\log(Ra_{ma_{10}})} \\
x_4 &= \log(temp_{ma_{30}}) - \overline{\log(temp_{ma_{30}})}.
\end{aligned}$$

As a result of centering, θ_0 represents the mean of water level (y) when all four predictors are at their respective medians.

Model Assessment

Diagnostic statistics and plots based on model residuals were used.

Overall, the residual plots suggest that the model provides an unbiased fit to the depth data and there is no obvious heteroscedasticity in the residuals. These plots and a modest R^2 value (0.33) are indicative that the model is an adequate summary of water level changes due to these main factors. Water level fluctuation due to other unmeasured variables is summarized in model residual variance, which will be reflected in the model's predictive uncertainty.

Model Documentation

Detailed documentation of the model, including computer code and model diagnostic plots are stored online at GitHub (https://github.com/songsqian/owc).

Discussion

Potention Uses of the Model

Because a statistical model can only include measured variables as predictors, not all relevant factors affecting water level changes can be accounted for. Our model includes main sources of input (discharge from Old Woman Creek and precipitation), but not variables of output (e.g., evapotranspiration and discharge to Lake Erie). We do not have variables to quantify water exchange between groundwater and the wetland (nor between Lake Erie and the wetland). The loss of water due to evaportranspiration may be partially represented by temperature and solar radiation. Although the model has an R^2 value of ??, the model should be a useful tool for evaluating future changes in water level due to the projected climate changes, represented in the increased temperature and the increased varaibility in precipitation.

The main use of the model is to predict the changes in water level due to future climate conditions for the region. Once the prejected climate/weather scenarios are available, we can use the model to make predictive simulation and summarize potential fluctuations due to the change. Results of the simulation can be used for evaluate the anticipated impact of climate change to the ecosystem as most wetland ecological indicators are functions of water depth.

Model Updating

Continued monitoring will enable model updating as more data are made available. Updating the model will enhance the relevance of the model for future climate conditions.

Connecting to Remote Sensing Data

As the third task of the project continue to accumulate data, we will link water level and surface water area through empirical modeling. Because water level and surface water area affect the ecological indicators we proposed, such a model can help forecast the potential impacts of climate change on the ecosystem, and inform the development of management strategies that are effective under future climate conditions.