Calcasieu Statistical Methods Document

Created by RTI International

December 2015

Data Source

Louisiana Coastwide Reference Monitoring System (http://www.lacoast.gov/crms2/Home.aspx)

- Stations: 0644, 0685, 0687, 1743
- All available data was downloaded for each station: Salinity, temperature, water level, water elevation, and specific conductance
- Hourly data averaged to daily values

Streamflow data from Calcasieu River modeled using the Watershed Flow and Allocation model (WaterFALL®). See documentation with in the HydroFlows App for more information.

- Daily streamflow values
- Two locations: downstream (river confluence with Calcasieu Lake) and at Lake Charles
- Current conditions modeled (2011 land use and current water uses)

Climate data from PRISM as reported for the location corresponding to the Calcasieu River entry into Calcasieu Lake (http://www.prism.oregonstate.edu/)

- Available data: precipitation (PRCPCM) and temperature (TAVG)
- Daily

Initial variable filtering

Variable "specific conductance" is almost in perfect linear association ($r^2 > 0.999$) with the dependent variable "salinity", thus was excluded from the predictor variable list in the analysis. Downstream water flow is strongly correlated with that of Lake Charles ($r^2 = 0.976$), thus provide redundant predictive value. Downstream flow data was selected per the clients' choice. Although measures, such as water temperature, water level, and water elevation, from the four stations are also correlated, because the geological distribution of the four stations, it is unclear which one of those provides the most predictive value to the dependent variable. Therefore, measurements from all four stations were kept for further selection.

Variable selection and model fitting

After initial variable filtering, 19 variables were included in the initial model. Variable selection was conducted using Least absolute shrinkage and selection operator (LASSO) regression with absolute penalty, as implemented in R package "glmnet" (Friedman et al., 2010). The most predictive independent variables for the four stations were selected separately in four LASSO regression models. The four models are listed below:

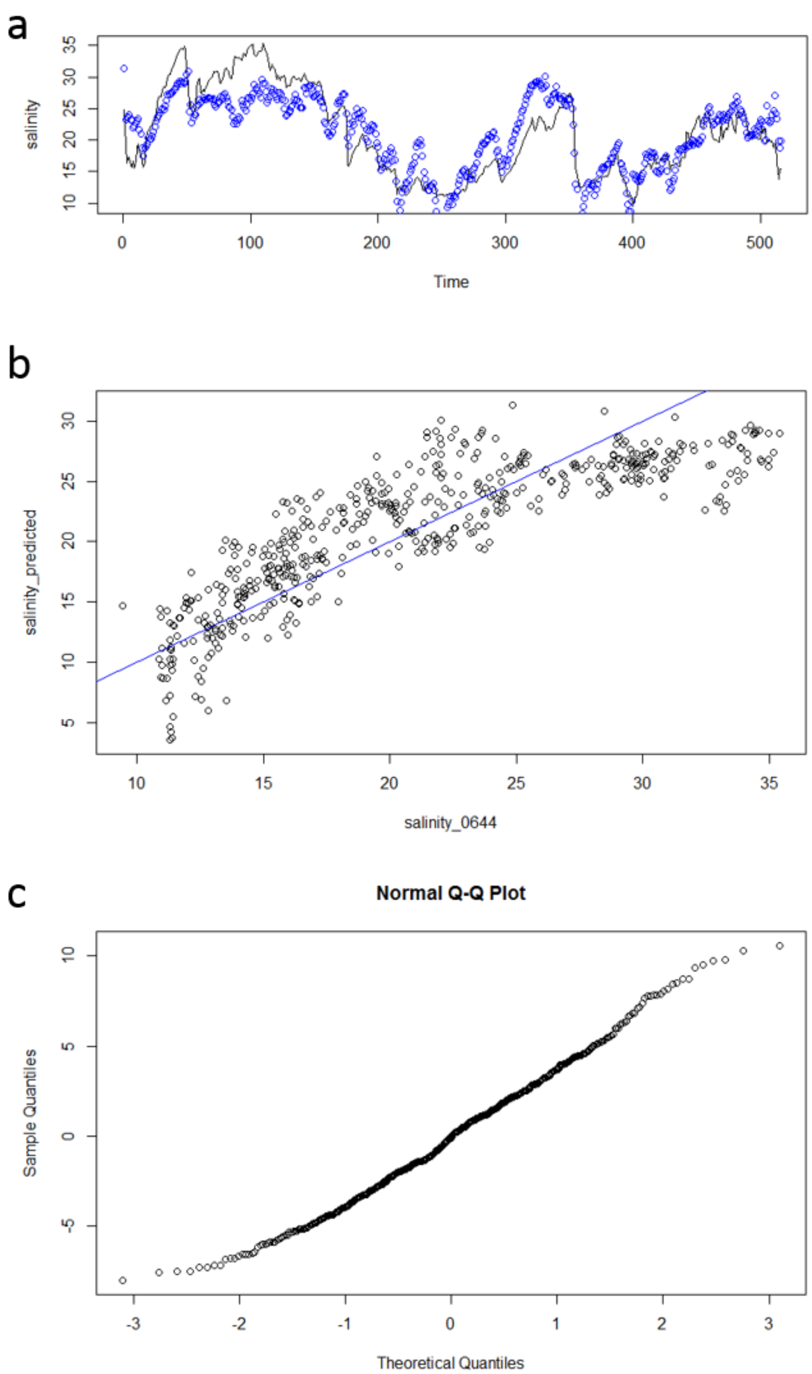
salinity_0644 = β_0 + β_1 * dCRMS1743_Wlevel + β_2 * Flow_Downstream + β_3 * dCRMS0685_Wlevel

salinity_0685 = β_0 + β_1 * Flow_Downstream + β_2 * dCRMS1743_WElevMarsh + β_3 * dCRMS0644_Wlevel + β_4 * dCRMS0685_WTempC + β_5 * GageHght_ft + β_6 * PrecipCM + β_7 * dCRMS0687_WElevMarsh

salinity_0687 = β_0 + β_1 * Flow_Downstream + β_2 * dCRMS1743_WElevMarsh + β_3 * dCRMS0644_Wlevel + β_4 * dCRMS0685_WTempC + β_5 * GageHght_ft

salinity_1743 = β_0 + β_1 * Flow_Downstream + β_2 * dCRMS1743_WElevMarsh + β_3 * dCRMS0644_Wlevel + β_4 * dCRMS0685_WTempC + β_5 * GageHght_ft

Figure 1. Diagnostic plots for station 0644. 1a) Plot of salinity over time. Measured salinity levels are shown in black lines, predicted salinity values are shown in blue dots. 1b) Plot of measured salinity vs. predicted salinity. The blue line indicates the 45 degree line where measured salinity is equal to predicted salinity. 1c) Residual Q-Q plot. Diagnostic plots for the other three stations are similar to this one (data not shown).



Simulation steps

To predict salinity value for the withdrawal scenario, we made the assumption that the measured variables could be well represented by historical data for the corresponding month, e.g. water level for January of the scenario year would be within similar range and distribution comparing to historical January water level data. For each of the twelve months, 1000 samples were sampled from historical data. Predictions of salinity were made correspondingly with the above specified models.

Future directions

1. Auto-correlation: to account for auto correlation, additional modeling variables is needed to reduce standard error of predicted salinity.
2. Dependency of independent variables: Water level may be dependent on water flow, temperature and precipitation. Explicitly modeling these relationship may further improve model.
3. Non-linear relationship: e.g. water level may affect salinity only if it is below/above some threshold; Temperature may affect salinity non-linearly. These factors are not considered in the current statistical model.

Reference

Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22.