Machine Learning Models to Predict Water-Level Anomalies in Annapolis, MD

Midshipman 1/C Forest Wan, Class of 2021; *Advisor(s): Instructor Alexander R. Davies and Dr. Joseph P. Smith*

Background

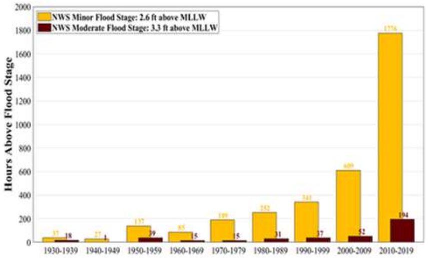


Figure 1. Hours above flood stage in Annapolis, MD by decade. The number of hourly water level observations in Annapolis, MD above minor (yellow; 2.6 ft above MLLW) and moderate (red; 3.3 ft above MLLW) flood stages by decade, as defined by the National Weather Service. Annapolis has experiencing an exponential increase in nuisance flooding since the end of the century. Relative sea level (RSL) is increasing in Annapolis, MD at a rate of +3.51 mm/year.

Annapolis MD is experiencing an acceleration in coastal nuisance flooding, or high-tide flooding, instances and duration (*Sweet et al., 2014; Dahl et al., 2019; Davies et al., 2021*) as a result of relative sea level rise, among other factors (**Fig. 1**). The drivers of coastal nuisance flooding, in general, are a superposition of global, regional, and local influences that occur across spatial and temporal scales to determine water levels (WLs) relative to a coastal datum (*Sweet et al., 2018*).

Most research to date related to coastal flooding has been focused on high-impact episodic events, decomposing global and regional drivers of sea-level rise (SLR), or assessing seasonal-to-multi-decadal trends in flooding. On short (hourly) time scales, WLs are primarily a function of the astronomical tides, regional and seasonal scale factors, and local-to-mesoscale meteorological conditions and forcing. This study investigates only the meteorological influences on WLs. In this study the difference between the observed WLs and predicted tidal WLs is defined as the water level anomaly (WLA), which is assumed to be primarily driven by meteorological influences. Meteorological factors that affect WLAs include: sustained wind forcing (direction and speed) relative to the orientation of the coast (*Davies et al, 2021*), atmospheric patterns that affect mean sea level pressure (*Sheridan et. al, 2017*), precipitation events (*Wolf, 2009*), and other factors. The relative impact of these influences on WLA is complicated by complex interactions with local morphology and the lag or inertial response time of WLs to changes in local forcing.

This study utilizes machine learning methods (Linear and Gaussian Process Regression analysis) to develop a model to better predict water level anomalies (WLAs) in Annapolis, MD using locally-available water level and meteorological oceanographic data.

Study Area and Data

Water Level Data

Annapolis, MD is home of the U.S. Naval Academy (USNA) and located at the mouth of the Severn River, a tidal tributary of the mesohaline Chesapeake Bay (**Fig. 2**). Quality controlled hourly water level observations were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Annapolis, MD tidal gaging Station (https://tidesandcurrents.noaa.gov/; NOAA- 8575512). The water level record in Annapolis spans 1929 through 2020, but was cropped to match the meteorological record (see below) Hourly tidal predictions for the Annapolis tide gauge were also downloaded, and matched with coincident water level observations. WLAs were calculated by subtracting hourly tidal predictions from coincident hourly water level observations.

Figure 2 (left). Map of study area including Annapolis, MD, USNA, and the lower Severn River. Red triangle marks the location of the KNAK ASOS weather station and the blue circle marks the location off the NOAA Annapolis Tide Gauge.

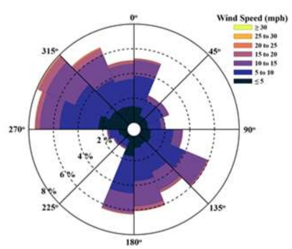


Figure. 3 (top). Wind observations between 2004-2020 at the Annapolis KNAK ASOS weather station directionally binned by 22.5 degrees (**Fig 2**). Wind speeds < 3.5 mph were removed from analysis.

Meteorological Data

Meteorological data (wind speed, wind direction, mean sea level pressure (MSLP), air temperature, and precipitation) were collected from 2004-2020 at the Automated Surface Observing System (ASOS) Annapolis (KNAK) weather station located on USNA Hospital Point about 1.2 km upriver from the NOAA Annapolis, MD Tidal Gauge (**Fig. 2**). ASOS-KNAK data were downloaded from the Iowa State University IEM archive (https://mesonet .agron.iastate.edu/). The winds were predominately out of the northwest, south, and southeast in Annapolis (**Fig. 3**), which aligns with the orientation of the Severn River (Davies et al., 2021; **Fig. 2**). Hourly meteorological data was matched with coincident hourly water level and tidal observations.

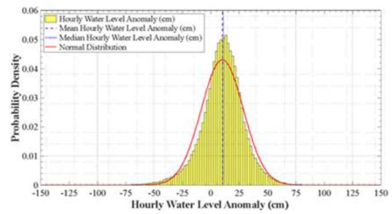
Variable Selection

The hourly WLA was selected as the response variable with 29 predictor variables. The predictor variables are: 3-, 6-, 12- and, 24-hour average MSLP; 3-, 6-, 12-, and 24-hour change in MSLP; 24-hour average air temperature; 24-hour change in air temperature; 6-, 12-, and 24-hour precipitation accumulation; 3-, 6-, 12-, 24-hour average wind direction; 3-, 6-, 12-, and 24- hour change in the wind direction; 3-, 6-, 12-, and 24-hour average wind speed; and 3-, 6-, 12-, and 24-hour change in the wind speed.

References:

Dahl et al., 2017: Sea level rise drives increased tidal flooding frequency at tide gauges along the U.S. East and Gulf Coasts: Projections for 2030 and 2045. *PLoS ONE*, 12(2), e0170949, https://doi.org/10.1371/journal.pone.0170949; Davies et al., 2021: Sustained wind forcing and water level anomalies in Annapolis, MD. *Earth Interactions* (in preparation); Grbić et al., 2013: Stream water temperature prediction based on Gaussian process regression. Expert Systems with Applications, 40, 7407–7414, http://dx.doi.org/10.1016/j.eswa.2013.06.077; Sheridan et al., 2017: Atmospheric drivers of sea-level fluctuations and nuisance floods along the mid-Atlantic coast of the USA. *Reg. Environ. Change*, 17, 1853–1861, https://doi.org/10.1007/s10113-017-1156-y.; Sweet et al., 2014: Sea level rise and nuisance flood frequency changes around the United States. NOAA Tech. Rep. NOS Center for Operational Oceanographic Products and Services 73, 58 pp., https://doi.org/10.13140/2.1.3900.2887.

Modeling Approach and Initial Results



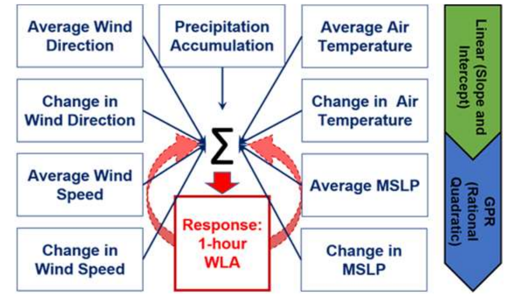


Figure 4. (*above*) Probability density distribution of WLAs between 2004-2020. (*right*) Conceptual modeling approach. The MATLAB 2020b regression learner applet was used to conduct all supervised modeling in this study.

The modeling approach used is outlined in **Figure 4**. The 2004-2020 coincident meteorological and WLA data was randomly subsect into training (70% of the data) and testing (30% of the data) datasets. A linear regression model was initially selected based on basic meteorological and oceanographic principles assuming water levels directly respond to chosen predictor variables. Models were trained to minimize the mean absolute error (MAE) and maximize the R^2 fit. The trained linear regression model had an R^2 fit of 0.49 with a MAE of 13.35 cm (**Fig. 5a, b**). The linear regression model failed to capture some of the complexities between the predictor variables and WLA response (i.e. nonlinear directional wind forcing, lag time, inertia, etc.), especially at higher amplitudes. A Rational Quadratic Gaussian Process Regression (GPR) model was next implemented similar to the approach used by *Grbić et al. (2013)*. Rather than modeling the direct response of an output variable to multiple predictors, the GPR fits a gaussian distribution to the predicted dataset in search of patterns; therefore GPR models and predictor variables need to be chosen carefully based on general principles. In this case, inputs to the GPR model were taken from the previous linear model based on the previous assumption of direct response. The trained GPR had an R^2 fit of 0.73 with a MAE of 9.65 cm, showing an improvement relative to the linear model, particularly in dealing with high amplitude observations (**Fig. 5c, d**).

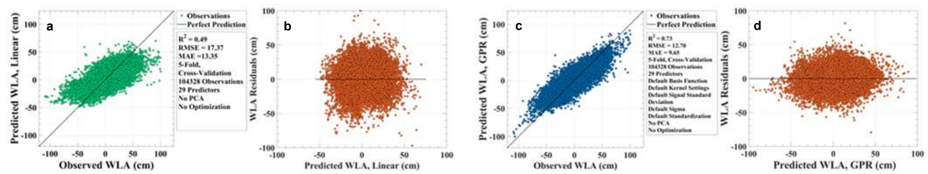


Figure 5. (**a**) Trained linear regression model results showing predicted vs. observed WLA values (cm). The solid black line indicates a perfect 1:1 prediction; (**b**) Distribution of residuals for observed vs. predicted WLA (cm) for the linear regression model; (**c**) Trained GPR model results showing predicted vs. observed WLA values (cm), and; (**d**) Distribution of residuals in the observed vs. predicted WLA (cm) for the GPR model. The MATLAB 2020b regression learner applet was used to conduct all supervised modeling in this study.

Discussion, Conclusions, and Future Work

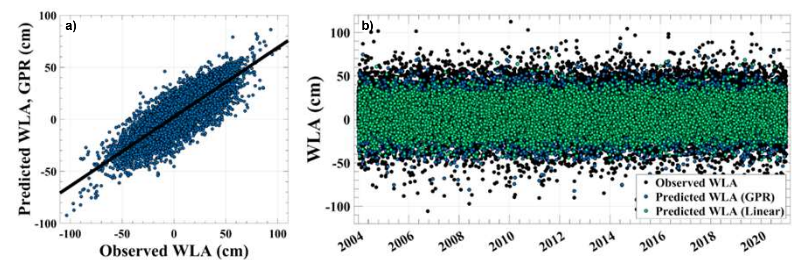


Figure 6. (**a**) Testing GPR model results showing predicted vs. observed WLA values (cm) and (**b**) Observed WLAs vs. predicted WLAs by the linear (green) and GPR (blue) models for the entire time record from 2004-2020.

The trained GPR model run using the testing dataset resulted in a fit with an R^2 = 0.69 and a MAE = 10.43 cm (**Fig. 6a**). **Figure 6b** shows the full time series of WLAs with linear (green) and GPR (blue) model predictions. The GPR model used predictor variables informed by the linear modeling approach and adequately captured the observed WLA variability and amplitude. However this modeling effort still had shortcomings that were likely due to the selection of predictor variables. In future interactions of the model, the predictor variables need to be selected using a more rigorous statistical approach. One example of this was found in *Grbić et al.* (2013) whereby predictor variables were selected based on mutual information (MI), a nonparametric, nonlinear measure of relevance between variables that can identify relationships. In addition to refining the statistical techniques for predictor variable selection, the physical understanding of the system must be advanced to develop variables that better capture the complex relationship between local- to- synoptic scale meteorological forcing, geomorphology, and any resultant non-linearities inherent to dynamic estuarine and coastal systems.

Acknowledgements: This work was funded, in part, by the U.S. Defense Threat Reduction Agency (DTRA). Special thanks to CAPT David Ruth (PhD; PMP) and Dr. Gavin Taylor for their support and their expert advice.