Homework #1 **CE 6338 August 22, 2022**

= =

Historical Developments in Hydrology

Prepare a 2 page paper on an interesting figure important to the development of the hydrologic sciences. Choose from those presented in class or choose your own. Focus on hydrologic and hydraulic research and professional background (not personal).

Include reference list (citations not necessary).

Due: August 29, 2022

CE 6099
Advisor: Guirong (Grace) Yan
Committee: Daoru (Frank) Han and Xiasong Du
Author: Ronald Adomako
Date: July 29, 2022

Develop an AI System to Predict Surge and Wave Height (Surrogate Model)

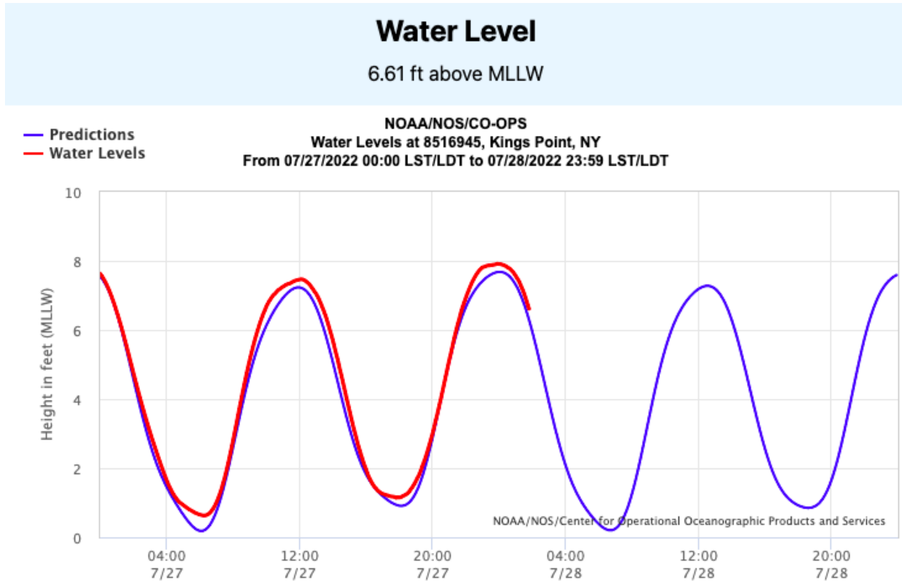
Abstract
Tropical storms have a negative economic impact on coastal areas. Evacuation procedures occur forty-eight hours or more before a storm surge impacts a coast. Current research shows that a storm tide can be predicted with accuracy for the first hour. Our research advances work that has been done to accurately predict storm tide heights after the first hour. The Long Short Term Method (LSTM) for machine learning has proven effective for predictions in the domain of storm tides and the case of then encoder-decoder implementation even more so (Bai and Xu 2021). The aim is to increase accuracy training on the parameters of direction, speed, size, and air temperature.

1. Research Background and Motivation

The significance in creating an artificial intelligent system to predict storm tide is for both accuracy of water level from storm surge, the wave heights and the duration of the storm. A storm tide is the combination of an astronomical tide and storm surge during an event. Contemporary research predicts storm tides accurately for the first hour. More accurate measurements for the duration of a storm will allow for better assessment of the economic impact of a storm. Additionally, predicting long-term storm behavior may better inform city officials of how to go about evacuation procedures.

Existing approaches to measuring storm tides have included various traditional, in-situ instruments. These instruments are tide stations, high water marks, and pressure sensors[1]. For tide stations, they are in areas sheltered from waves and able to measure "still water". Tide stations offer real time data and traditionally are the most reliable in-situ instrument. The con to tide stations is that they are limited in the number of locations and often fail due to power outages during the height of an event in the case of catastrophic events. Tide station data are delivered in a gauge fashion where an observer can track surge heights for any given day.[1]

[1] NOAA, Tides and Currents, https://tidesandcurrents.noaa.gov/map/index.html?id=8516945



High water marks may be robust in capturing the max surge height, but alone does not differentiate between "still water", astronomical tide, nor storm surge. Primarily, it is assumed that their markings represent storm tide since its datum is relative to a mean sea level to account for "still water" prior to an event. Because high water marks are observed on site after a catastrophic event by surveyors, data is not available in real-time and the data, via forensic evidence, is available for a limited time.

In the case of pressure sensors, both storm surge height and the time in which it occurred are available. In practice they are installed in locations of highest expected surge locations prior to an event. However, traditional pressure sensors do not provide real-time data and must be retrieved post event if at all retrievable from a site. Without real-time data, an ensemble of pressure sensors will include the effects of waves, not just the height of the storm surge.

Naturally, these traditions are used together to provide better assessment of the actual behavior of a storm tide. Together, these form a baseline of how the height of storm tides are measured. For predictions, empirical analysis is too complicated and there are not too many events to track, especially since tropical storms and hurricanes of the same class are limited due to the rarity of the events and the location in which they occur.

Contemporary research in machine learning has done well to track heights of a storm tide during an event and has done well in computing, that is, handling large data calculations. Furthermore, machine learning research has shown increasing accuracy for the early portion of the event, the first hour, and has optimized runtime and data memory for predictions.[2] LSTM has proven to be the most effective and its encoder-decoder implementation even more so. The encoder-decoder implementation allows for compression of large data sets and unpacking values when needed. LSTM encoder-decoder increases computational efficiency.

[2] Accurate storm surge forecasting using the encoder–decoder long short term memory recurrent neural network by Long-Hu Bai and Hang Xu (2021).

Previous research tested these machine learning algorithms on a limited asset class. As mentioned, tropical storms and hurricanes are rare events and location matters. Data must be separated by regions in which they occur– not all storm tide events can be trained together for a model. Categories matter for their sampling and the more data one has the better the model.

In contrast, having access to a larger dataset will both increase accuracy over a longer period and create a more robust model. The benefit to this machine learning model is that you train it once and the model is ready to be used on a new set of inputs in a shorter time since the algorithm is already cached.

2. Research Objective(s)

The aim is to increase accuracy of predictions of storm surge heights. The input parameters are direction, speed, size, and air temperature. The primary research objective is to improve the predictions accuracy over longer periods. This can be done with a sensitivity analysis. A sensitivity analysis would help us discern the relative importance of each input parameter and which grouping of the input parameters is optimal.

3. Research Approach

LSTM works well for time series problems that have a degree of predictability. In addition to predicting significant wave height, related fields of success for LSTM predictions include wake flows, blood glucose, and nonlinear dynamics in pattern formation. It is worth noting that for storm tides, convolution neural networks (CNN) trained on image data is an alternative approach. Because the data is numeric and non-image based, The image based CNN is out of scope for our research.

Data

We split the data into categories based on the geographical region of the storm tide. We extract the direction, size, air temperature, barometric pressure of the center of the storm, water level, significant wave height, and storm duration. The data covers forty years of storm tide events.

Training and test sets

We split the data into training and test sets. The first experiment is to use all the data we have available with two-thirds allocated to training and one-third allocated to tests (validation).

Training LSTM with encoder-decoder

The model is trained with the input parameters of direction, size, air temperature, and barometric pressure of the center of the storm and the output parameters of storm surge height. Each cell has an output of the surge height at the end of a delta-t (change in time). We discretize our model into n cells. The duration of the storm is divided by n cells of equal time. Since each cell is an interdependent unit of an adjacent cell, the discrete storm surge height of a cell plays a role in the input for the next cell . This forms a continuum for the changing heights for the storm surge during an event. The following figure illustrates a flow diagram of the LSTM encoder-decoder.

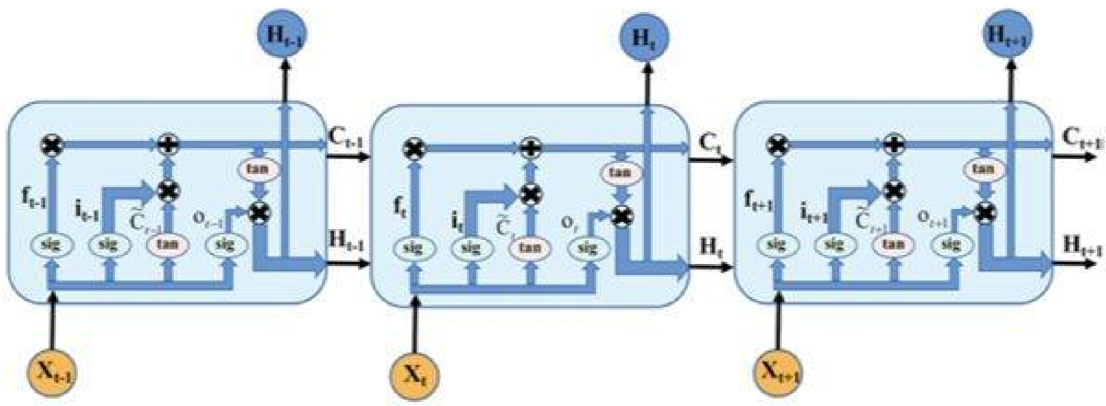


Image borrowed from Long-Hu and Bai (2021)

The first and last cell are terminal cells. We assume a surge height of zero in the first cell and we stop our predictions at our last cell. The *f-gate* is what the cell forgets. The *o-gate* is the encoded output. The decoded output is a back-propagation of the previous to account for probability weights when doing a current cell prediction.

For the LSTM, we employ a coded algorithm in an attempt to visualize what happens in the black box.[3]

```
def LSTMCELL(prev_ct, prev_ht, input):
    combine = prev_ht + input
    ft = forget_layer(combine)
    candidate = candidate_layer(combine)
    it = input_layer(combine)
    Ct = prev_ct * ft + candidate * it
    ot = output_layer(combine)
    ht = ot * tanh(Ct)
    return ht, Ct



ct = [0, 0, 0]
ht = [0, 0, 0]

for input in inputs:
    ct, ht = LSTMCELL(ct, ht, input)
```
[3]

Sensitivity Analysis

We run various combinations of the input parameters using binomial permutation. The optimal solution will be the minimum set of input parameters that still achieve accurate results.

Results and Assessment

In keeping with previous research, we assess the root mean square error (RMSE), the mean squared error (MSE), the correlation coefficient (CC), and the mean absolute error (MAE). But we introduce a percentage error of the difference to measure whether the MAE is meaningful. The test runs for *n* equal to 10, 50, and 100 cells.

[3] Michael Phi (2018), Medium: Illustrated Guide to LSTM's and GRU's: A step by step explanation

5. References

Long-Hu Bai and Hang Xu (2022). Accurate storm surge forecasting using the encoder–decoder long short term memory recurrent neural network

National Hurricane Center, Storm Surge Unit. [accessed July 4, 2022] Introduction to Storm Surge. https://www.nhc.noaa.gov/surge/surge_intro.pdf

Faria T. Zahura, Jonathan L. Goodall, Jeffrey M. Sadler, Yawen Shen, Mohamed M. Morsy, Madhur Behl (2021). Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community.

M. Reza Hashemi, Malcolm L. Spaulding, Alex Shaw, Hamed Farhadi, and Matt Lewis (2016). An efficient artificial intelligence model for prediction of tropical storm surge.

Sung Hyup You and Jang-Won Seo (2009). Natural Hazards 51, 97–114. Storm surge prediction using an artificial neural network model and cluster analysis. https://doi.org/10.1007/s11069-009-9396-x

Jayaram Veeramony, Andrew Condon, and Maartan Van Ormondt (2017). Forecasting Storm Surge and Inundation: Model Validation.

Lleyton Arrington (2021). Medium: Analytics Vidhya. No LSTMs Can't Predict Stock Prices.

Martin Gauch, Frederik Kratzert, Daniel Klotz, Grey Nearing, Jimmy Lin, and Sepp Hochreiter (2020). Rainfall–Runoff Prediction at Multiple Timescales with a Single Long Short-Term Memory Network.

Michael Phi (2018), Medium: Illustrated Guide to LSTM's and GRU's: A step by step explanation medium.com/towards-data-science/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21