Project Proposal / CS565
Filling gaps of the GIW area time-series using Artificial Intelligence methods

Junehyeong Park

I. Introduction

The GIWs (Geographically Isolated Wetlands), one of the types of wetlands, were defined by landscape position as “wetlands with no apparent surface water connection to perennial rivers and streams, estuaries, or the ocean” [1]. These GIWs were surrounded by dry land, and the wetlands along intermittent streams connected to perennial streams were designated as non-isolated. To understand the interactions between wetlands and climate, surface later flows, and groundwater, neglecting large interventions of tidal effects or inundation from nearby perennial streams, GIWs are currently being surveyed.

There are no time-series data of the area or water amounts of wetlands not only officially published but also in research articles. Therefore, to analyze using temporal information of wetland, researchers should get in-situ information in their own. Or, only stationary analysis could be done. In this regard, my current research has focused on generating time-series maps of GIW by GIS works on the time-series of satellite observation [2] and the GIW map [3]. Now, I have generated time-series of GIW area for each sub-region in GA.

The main objective of my research is to represent climate controls on GIWs seasonally and spatially. It would propagate to the analysis for evaluating seasonal reactions, resiliency against drought events, spatial differences of sensitivity, and finally finding causations of disappeared wetlands. However, there is a problem that the generated time-series of GIW area has too much blanks because of the characteristics of satellite observation. To conduct researches mentioned above, the quality control procedure is needed. According to that, in this project, I will build some models to filling gaps with dependent variable as GIW area time-series and explanatory variables as climate records [4].

The dependent variable, artificial GIW are time-series, will be divided into available timesteps to use and blanks. During the available timesteps, GIW area and other variables will be used to fit the models. Some models, including ANN and RNN, will be adopted to compare the performance of simulation and prediction. Because there are no true values for GIW area, it is hard to find the answer model, but I need to compare models specifically for selecting the most appropriate (or the least error occurred) model.

II. Data

A. Research area

In this research, North Dakota and Georgia were selected as representative states of Northward and Southward of U.S.. The regions spatially divided by the NLDAS grids to use climate information of NLDAS forcing data. Even though this report does not include any results using forcing information yet, this work is regarded as a primary research that will use climate information to analyze what will cause the changes of wetlands, so that the target regions are divided in NLDAS grids.

Among numerous grids in North Dakota and Georgia, 4 grids are selected that have less observation failures during 380 months (Mar 1984 Oct 2015) compared to other grids, respectively. In the Fig. 1, selected 8 grids are colored in red. Their identification names are ’x206y174’, ’x213y176’, ’x207y171’, and ’x220y169’ for ND, and ’x349y54’, ’x349y55’, ’x338y51’, and ’x339y50’ for GA, respectively.

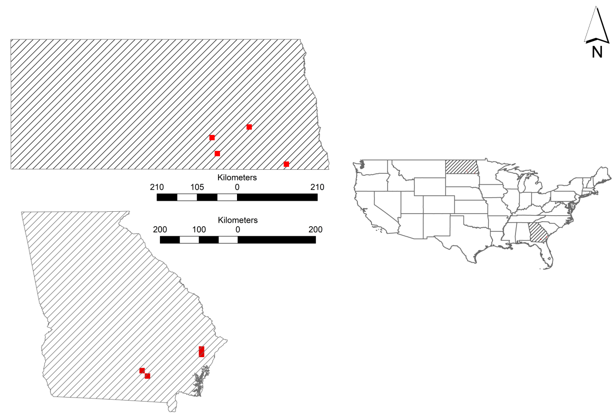

Fig. 1. Selected regions for GA and ND

B. Global Surface Water

Global Surface Water (GSW) is a Landsat-based globally spatial information about the location and persistence of surface water (inland and coastal) developed by [2]. The GSW includes various types of data such as global surface water mapping layers, monthly water recurrence, monthly water history, and yearly water classification history. The monthly water history data is a basic result that includes raster pixels with 0 (no data), 1 (no water), and 2 (water) integers with 30 m resolution for the globe. Because it is a long-term (Mar 1984 Oct 2015), high-resolution satellite data, the basic result itself has some limitations such as observation failures due to could or others, or lack of information during winter even though there is no snow or ice happened. Other types include additional results of analysis based on the basic result.

Because of the uniqueness that has serial and spatial distribution of surface water presence in 30 m resolution, The global surface water data was used for many researches including wetland researches. For example, Krapu [5] used the global surface water for identifying wetlands consolidation in the North Dakota Prairie Pothole Region. In this research, the monthly water history data of the global surface water imported from the Google Earth Engine was used as a clue for defining there was a surface water body and an input data that was inserted into the algorithm matching surface water body with GIW polygons.

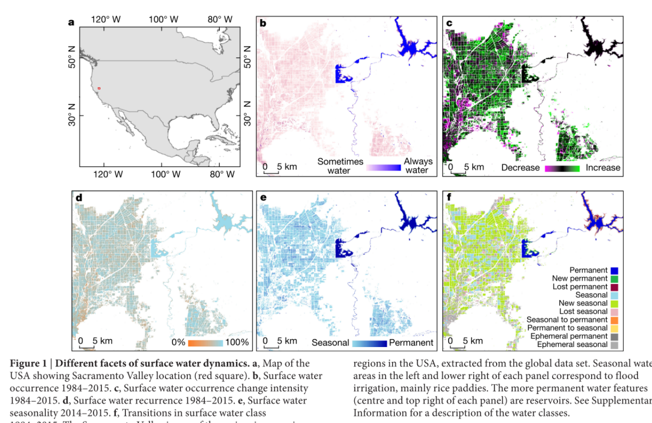

Fig. 2. An example figure of the GSW imported from [2]

C. National Wetlands Inventory and Geographically Isolated Wetlands

National Wetlands Inventory (NWI) [6] is a spatial data presented by the U.S. Fish and Wildlife Service that includes spatial distribution of all kinds of wetlands in the format such as marine, estuarine, riverine, lacustrine, and palustrine systems. The NWI data is regarded as a standard in wetland researches which sets the target region in CONUS.

GIW was delineated in several ways using various logics, tools and date, but Lane [3] showed the most recent and representative way to delineate GIW features from the NWI spatial data. In this research, the spatial data of GIW in ND and GA received from Dr. Lane was used.

In this regard, this research also considered the NWI data and the GIW as the standard definition and delineations of wetlands and GIWs. In Fig. 3, NWI (green) and GIW (blue) features are plotted in the same region (North Dakota) to look at in a comparative view. There are wetlands in North Dakota as we know, so there are so many green features in the Fig. 3. Among these green features, rivers and lakes are also included. However, the blue features, which represents the GIWs, do not include rivers or lakes and only includes GIWs that are apart from such large surface water bodies.

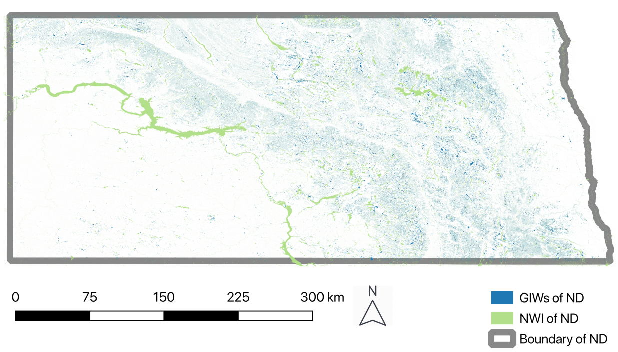

Fig. 3. NWI and GIW for North Dakota

D. Methodology to delineate GIWs from GSW maps

With hopeful assumptions, the water pixels in GSW raster data that have the value 2 can be categorized into GIW if the location of the pixel is on the GIW features. However, it is hard to follow that because of following reasons. First, if we use just pixels that is located on the GIW features, we can't capture the cases that wetlands were enlarging or merging with other nearby GIWs. Second, the spatial resolution of the GSW data is 30 m and GIWs are very small, there can be some unexpected cases to avoid. For example, because of the uncertainty of satellite data, the GSW pixels can represents a different place some meters apart from the accurate place. Thus, in this regard, this research generated following two steps with a time-step: 1) find water pixels (value 2) of the GSW map that touch any water bodies of GIW. 2) find all water pixels of the GSW map that is located nearby the selected pixel in the step 1. 3) repeat step 2 until there are no more water pixels. 4) repeat from step 1 to 3 for all pixels in the GSW map.

However, another problem can be occurred when the surface water of a GIW was merged with any large lakes or rivers for a short time during floods and the instant case became the case as the representative observation of the month. Then, the time-series of GSW pixels that regarded as GIWs would have extraordinary values. To solve this problem, another process was inserted earlier than mentioned above. The additional process is that 1) generate a map that represents NWI but not GIW (i.e. lakes and rivers), 2) find any water pixels of the GSW map that touch any water bodies of "NWI - GIW", 3) find all water pixels of the GSW map that is located nearby the selected pixel in the step 2, 4) repeat step 3 until there are no more water pixels, 5) repeat from step 2 to 4 for all pixels in the GSW map, and finally 6) define all water pixels of GSW map detected during step 5 as "no water" to neglect.

Therefore, the whole process from the raw data to the results can be rephrase as below. Figure 4 shows sample "NWI - GIW" map (0 (black) indicates no wetlands, 1 (gray) indicates NWI wetlands but not GIWs, and 2 (white) indicates GIWs), GSW map (0 (black) indicates no data, 1 (gray) indicates no water, and 2 (white) indicates water pixels), and finally a result map (0 (black) for nothing and 1 (white) for GIWs). The code for this main algorithm was written in python and attached as "Appendix" at the end of the document.

Step 1: clip a GSW map for a time-step with boundary of a NLDAS grid

Step 2: rasterize the NWI and GIW feature map with boundary of a NLDAS grid and also generate "NWI - GIW" raster map

Step 3: find water pixels of the GSW map that touch any water bodies of "NWI - GIW" and their neighbors, and neglect them.

- Step 3-1: find any water pixel of the GSW map that touches any water bodies of "NWI - GIW".
- Step 3-2: find all water pixels of the GSW map that is located nearby the selected water pixel in the step 3-1.
- Step 3-3: repeat the step 3-2 until there are no more nearby water pixels.
- Step 3-4: repeat the steps from 3-1 to 3-3 for all water pixels in the GSw map.
- Step 3-5: change values of the selected pixels in the step 3-4 from 2 (water exists) to 1 (no water).

Step 4: find water pixels of the GSW map that touch any water bodies of GIW and their neighbors, and define the number of selected GSW water pixels as the area of GIWs in the region of NLDAS grid during the time-step.

- Step 4-1: find any water pixel of the GSW map that touches any water bodies of GIW.
- Step 4-2: find all water pixels of the GSW map that is located nearby the selected water pixel in the step 4-1.
- Step 4-3: repeat the step 4-2 until there are no more nearby water pixels.
- Step 4-4: repeat the steps from 4-1 to 4-3 for all water pixels in the GSw map.
- Step 4-5: count the number of the selected pixels in the step 4-4 and define the number as the area of GIWs in the region of NLDAS grid during the time-step.

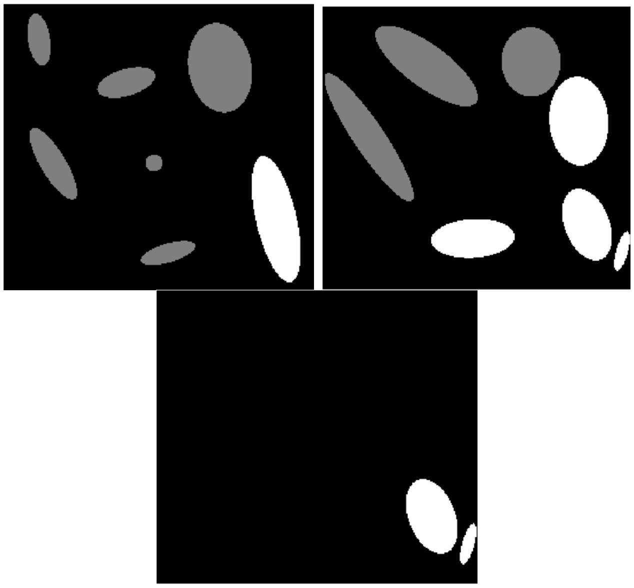

Fig. 4. An artificial sample maps for NWI - GIW (top-left), GSW (top-right), and the result of algorithm explained above (bottom)

E. Artificial Neural Networks

Artificial Neural Networks (ANN) is a method that learning relationships using neurons located on the hidden layers. Even though it is a kind of a set of fitted equations like regressions, because of complex structure and meaning less nodes (each nodes cannot have identified physical meanings that input or output variables can have), ANN would be suggested for the simulation or re-generation of data but cannot be used for sensitivity analysis or understanding relationships. In this case, surroundings such as climate variables or observed discharge rates nearby target regions would be used as input variables and the area of GIW would be used as output variable. The months when the area of GIW is not available due to the fraction of observation failure is high or observation is absence, and the months will be used as prediction timesteps. The specific configurations and results will be included in the final draft.

F. Recurrent Neural Networks

Recurrent Neural Networks (RNN) is an advanced method of ANNs dealing with time-series data. This method uses the current state as output variable, and uses the last state as input variable with other variables. It would be applied when we assume the variable has memory so that auto-correlation is important to simulate the variable. The area of GIW is one of them that even though there was a storm, the result would be different when the last state was extremely dry or wet enough. Therefore, the RNN, especially LSTM, will be applied to the data to tran the model and predict GIW area at the gaps. The specific conficurations and results will be included in the final draft.

References

[1] Grey R. Evenson et al. “Geographically isolated wetlands and watershed hydrology: a modified model analysis”. In: *Journal of Hydrology* 529.P1 (2015), pp. 240–256. DOI: https://doi.org/10.1016/j.jhydrol.2015.07.039.

[2] Jean-François Pekel et al. “High-resolution mapping of global surface water and its long-term changes”. In: *Nature* 540.7633 (2016), pp. 418–422. DOI: https://doi.org/10.1038/nature20584.

[3] Charles R. Lane and Ellen D’Amico. “Identification of Putative Geographically Isolated Wetlands of the Conterminous United States”. In: *Journal of the American Water Resources Association* 52.3 (2016), pp. 705–722. DOI: https://doi.org/10.1111/1752-1688.12421.

[4] Youlong Xia et al. “Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products”. In: *Journal of Geophysical Research: Atmospheres* 117 (2012), p. D03109. DOI: https://doi.org/10.1029/2011JD016048.

[5] Christopher Krapu, Mukesh Kumar, and Mark Borsuk. “Identifying wetland consolidation using remote sensing in the North Dakota Prairie Pothole Region”. In: *Water Resources Research* 54.10 (2018), pp. 7478–7494. DOI: https://doi.org/10.1002/2018WRR023338.

[6] U.S. Fish and Wildlife Service. *National Wetlands Inventory*. 2019.