STATE-WIDE WETLAND INVENTORY MAP OF MINNESOTA USING MULTI-SOURCE AND MULTI-TEMPORAL REMOTE SENSING DATA

Victor Igwe a, *, Bahram Salehi b, Masoud Mahdianpari b

a Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry (SUNY-ESF), NY 13210, USA - vicigwe@syr.edu, bsalehi@esf.edu
b C-CORE and Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John’s, NL A1B 3X5, Canada - masoud.mahdianpari@c-core.ca

Commission III, WG III/10

KEY WORDS: Wetlands, remote sensing, multi-source data, google earth engine, Minnesota, random forest, object-based image analysis.

ABSTRACT:

Carbon sequestration coupled with flood mitigation and other functions of wetlands, such as water filtration, coastal protection, biodiversity, and providing recreational spots, make wetland mapping and monitoring important for different countries. Google Earth Engine (GEE) cloud computing platform is becoming a very important tool for lots of environmental studies as it provides a suite of tools and access to data that facilitate large-scale environmental monitoring projects through its powerful parallel processing capabilities. In this study, we use GEE to access multi-source remote sensing datasets and implement an object-based image analysis, and random forest algorithm for the classification of wetlands in the state of Minnesota. Emergent, forested, and scrub-shrub wetland classes, water, as well as urban, forest, and agriculture land cover types were classified using Sentinel-2, Sentinel-1, USGS 3D Elevation Program 10-meter DEM, and gridded soil data. NDVI, EVI, BSI, NDBI, and NDWI spectral indices were calculated from Sentinel-2 imagery, VV and VH polarization channels, and their ratio, as well as span parameters, were calculated from Sentinel-1 imagery, and slope and aspect features were extracted from DEM. Simple Non-Iterative Clustering (SNIC), Gray-Level Co-occurrence Matrix (GLCM), Principal Components Analysis (PCA), and random forest algorithms were implemented to classify wetlands from the GEE platform. Emergent wetlands, water, urban, and agriculture classes performed well with producer accuracies greater than 90%. Sentinel-1, DEM, and soil datasets improve the identification of wetland classes and highlight the importance of multi-source approaches for wetland mapping.

1. INTRODUCTION

Wetlands are regions where soils are temporarily or permanently saturated by water with plant community-dependent and adapted to these conditions (Federal Geographic Data Committee, 2013; U.S. Army Corps of Engineers, 1987). The role of wetlands in global climate and carbon ecosystems has sparked attention in its study and advocacy for its management, monitoring, and restoration. Although wetland ecosystems contribute significant amounts of methane, (Whalen, 2005) estimates at about 25% of total emissions, the carbon sequestration function of wetlands and the decay of methane in the atmosphere make them net carbon sinks (Mitsch et al., 2013) and hence are valuable ecosystems in the global carbon cycle. Wetlands are also helpful in mitigating flooding and reducing potential damage from flooding, they can store water and release it slowly over time (Steve et al., 2019). Carbon sequestration coupled with flood mitigation and other functions of wetlands such as water filtration, coastal protection, biodiversity, and recreation (Corcoran et al., 2013; Mahdianpari et al., 2020) make wetland mapping and monitoring important in this modern day.

Wetlands are landscapes of transition between land and water, which are marked by the presence of hydrophytic vegetation and saturated soils. The transitional characteristic of wetlands adds more complexity to wetlands mapping (Mahdianpari et al., 2020). Hierarchical classification systems have been designed to circumvent these challenges. The Wetland Classification Standard FGDC-STD-004-2013 established by the Federal Geographic Data Committee (Cowardin et al., 1979) is the standard for the National Wetland Inventory in the United States (Federal Geographic Data Committee, 2013). These groupings have been made based on plant community composition, soil morphology, and site wetness indicators. The Cowardin classification system is shown in Figure 1.

Methods of mapping using remote sensing technology have been adapted to different classification schemes over the years. Remote sensing data sources and processing techniques provide an upgrade to traditional methods of mapping these ecosystems. Satellite remote sensing especially presents great potential for mapping wetlands due to their ability to cover large spatial scales in very short periods. Improvements in the spatial resolution of satellite sensors have also improved the ability to map wetland ecosystems with increased accuracy and precision. (Mahdianpari et al., 2020) analyzing the 40-year trend in remote sensing of wetland reported that the accuracy of wetland mapping increased as the spatial resolution of the satellite imagery increased. The ubiquity and explosion of different sources of earth observation data has also increased our ability to identify different wetland classes. In recent years, multi-source approaches to wetland mapping have received great attention.
(Battaglia et al., 2021; Corcoran et al., 2013; Kloiber et al., 2015; Mahdianpari et al., 2021). Using Landsat 5 Thematic Mapper optical imagery, PALSAR L-band radar, RADARSAT-2 C-band radar, US Geological Survey (USGS) National Elevation Dataset topographic data, and soil data from US Department of Agriculture (USDA) Soil Survey Geographic Database (SSURGO), (Corcoran et al., 2013) identified key combinations that could be useful for differentiating wetlands classes.

Some of the array of algorithms available through GEE are Simple Non-Iterative Clustering (SNIC) and random forest algorithms. SNIC is a key technique in object-based image analysis, which is fast replacing pixel-based classification methods. Advantages of object-based image classification schemes include reducing the salt-and-pepper effect on high-resolution imagery and improving classification accuracy (Tassi & Vizzari, 2020). Random forest has been shown to perform well for integrating multi-source datasets for the classification of wetlands (Corcoran et al., 2013; Mahdianpari et al., 2021). SNIC and random forest algorithms were used in this study for image classification.

This study focuses on the large-scale mapping of wetlands in the state of Minnesota. The first wetland inventory mapping for Minnesota was carried out around the 1980s (Steve et al., 2019). Dwindling funding for wetland mapping impacted the ability to update inventory, major state-wide inventory update was only carried out by the Minnesota Department of Natural Resources (DNR) in 2019 after funding from Environment and Natural Resources Trust Fund (Steve et al., 2019). This highlights the need to utilize satellite remote sensing sources for large-scale wetland inventory mapping, which have cost advantages over aerial imagery and cover larger spatial scales in shorter periods, reducing the costs of wetland mapping projects and offering the ability to monitor wetlands constantly over short temporal scales. These maps will be invaluable to government agencies, non-profit organizations, and other institutions that require regularly updated wetland inventory information for decision making and policy planning. This study aims to adapt the established methodology for wetland remote sensing to map wetlands on a large scale using multi-source datasets.

2. STUDY AREA

The study is focused on the state of Minnesota which totals about 225,000 Km² in area, 91% of which is land. According to the Cowardin classification scheme in the state, predominant wetland systems are the riverine, lacustrine, and palustrine systems (Steve et al., 2019). Forested wetlands are the largest wetland in Minnesota, most of the wetlands in Minnesota are in the north-eastern region of the state, and about half of the wetlands in the state have been lost since 1850 (Steve et al., 2019). Average annual precipitation in the study area ranges from 18 to 32 inches, with the most precipitation occurring in the summer season. Figure 2 shows a satellite image view of Minnesota.

3. DATA AND METHODS

3.1 Datasets

3.1.1 Sentinel 1: This study used images from the European Space Agency’s Sentinel-1 twin satellites. Aboard both Sentinel-1 satellites is a C-band Synthetic Aperture Radar instrument observing the earth day and night, in all weather conditions with a 6-day revisit period (European Space Agency, 2014). The ability of SAR instruments to penetrate cloud cover and its 24-hour per day earth monitoring capability makes it suitable for mapping wetlands, especially to compensate for some limitations.

Figure 1. Cowardin Classification System

(Corcoran et al., 2013) empirically observed that red band, near-infrared band, middle infrared band and derived Normalized difference Vegetation Index from optical satellite imagery, elevation, and curvature from topographic data, horizontal-vertical (HV) polarization of L-band Radar, and hydric soil data are key variables for differentiating between upland, water, and wetland classes. For classifying different wetland types, C-band radar and other spectral bands from satellite imagery are useful in addition to the other input variables. (Mahdianpari et al., 2020) also observed an increased accuracy trend when LiDAR, optical, and SAR data were combined for wetland classification. Modern-day challenges in wetland classification lie in adapting existing functionality for large-scale mapping. Hardware and software limitations affect the large-scale production of land cover maps (Shafizadeh-Moghadam et al., 2021). Advances in high-performance computing and cloud computing technology address these limitations (Tamininia et al., 2020). Parallel processing capabilities mean that computationally intensive tasks can be distributed across different units to save time and improve efficiency. Google Earth Engine (GEE), a cloud-based platform for processing geographic related information (Tamininia et al., 2020) is increasingly being used in recent land cover and wetland mapping studies (Mahdianpari et al., 2019, 2021; Shafizadeh-Moghadam et al., 2021; Tassi & Vizzari, 2020; Valenti et al., 2020). GEE makes multi-source remote sensing workflows easier providing access to multi-petabytes of earth observation data and complementary functionality for implementing algorithms and visualizing results (Shafizadeh-Moghadam et al., 2021; Tamininia et al., 2020).
of optical satellite imagery, VV+VH polarization and ratio and span parameters were calculated and used as input variables for the random forest classifier. This dataset was directly accessed through the GEE interface.

3.1.2 Sentinel 2: This study used images from European Space Agency’s Sentinel-2 satellites. The optical instrument samples 13 spectral bands in the visible, near-infrared, and shortwave infrared wavelengths; with 4 bands (B2, B3, B4, and B8) in 10m resolution, 6 bands (B5, B6, B7, B8A, B11, and B12) in 20m resolution and 3 bands (B1, B9, and B10) in 60 m resolution (European Space Agency, 2015). It has a 5-day revisit period (European Space Agency, 2015). Bands B2, B3, B4, B6, B8, and B11 corresponding to blue, green, red, red edge 2, Near-infrared, and shortwave infrared 1 of the Level-2A orthorectified atmospherically corrected surface reflectance product were used as inputs for our model. Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Bare Soil Index (BSI), Soil Adjusted Vegetation Index (SAVI), Normalized Difference Vegetation Index (NDWI), Normalized Difference Building Index (NDBI) spectral were calculated from spectral bands of the sentinel-2 image and used as input variables for the model. This dataset was directly retrieved from the GEE interface.

3.1.3 U.S. Geological Survey, 3DEP 10-Meter Resolution Digital Elevation Model: In 2016, the 3D Elevation Program (3DEP) run by the U.S. Geological Survey National Geospatial Program was established to provide high-resolution elevation data for the United States. In this study, we use the seamless 1/3 arc-second 3DEP product with a resolution of approximately 10 meters. Slope and aspect parameters were calculated and used as input variables for the model. This dataset was directly assessed through the GEE interface.

3.1.4 US FWS National Wetlands Inventory: The National Wetland Inventory (NWI) managed by the US Fish and wildlife service (FWS) was used to download wetland data for the study. NWI data for Minnesota was downloaded through the NWI wetlands mapper tool. The Wetland layer from the NWI data was used to generate test and training samples for the study. Test and training samples generated were uploaded to the GEE platform.

3.1.5 National Land Cover Database (NLCD): National Land Cover Database is a collection of land cover products created by The U.S. Geological Survey (USGS) in collaboration with the Multi-Resolution Land Characteristics (MRLC) consortium. Land cover product was used to generate test and training samples for upland classes.

3.1.6 Gridded Soil Data: Gridded soil data for Minnesota was downloaded from the Gridded Soil Survey Geographic (gSSURGO) Database of the United States Department of Agriculture, Natural Resources Conservation Service. Raster data for hydric soil was extracted from the database and manually uploaded to the GEE platform.

4. METHODS

4.1 Test and train data preparation

To generate training samples for the machine learning model, the wetlands layer from NWI was overlaid on sentinel-2 imagery. Samples were picked at class level following the Cowardin classification system; wetland classes were grouped into emergent, forested and scrub-shrub categories. Aquatic bed, Unconsolidated bottom, and Unconsolidated shore classes were put into a unified water class due to their spectral similarity. Similarly, upland classes were generated from the NLCD layer to prevent overfitting in the model. Three large groups corresponding to the forest, agriculture, and urban were chosen based on spectral similarity of the upland classes; subclasses such as mixed forest, evergreen forest, deciduous forest were grouped under the forest class, barren and developed land were grouped under the urban category, and grassland, pasture, and cultivated crops were merged into the agriculture class. Table 1 shows the number of samples generated for training and testing the model. Samples were split in a 70:30 ratio in GEE.

| TYPE          | CLASS          | TOTAL | TRAIN | TEST |
|---------------|----------------|-------|-------|------|
| WETLAND       | Water          | 933   | 669   | 264  |
|               | Emergent wetland | 658   | 457   | 201  |
|               | Forested wetland | 630   | 463   | 167  |
|               | Scrub-shrub wetland | 686   | 463   | 223  |
| UPLAND        | Urban          | 410   | 273   | 137  |
|               | Forest         | 429   | 317   | 112  |
|               | Agriculture     | 401   | 276   | 125  |

Table 1. Wetland and upland test and train points
Satellite image processing

Sentinel-2 Optical imagery, Sentinel-1 C-band Radar, and USGS 3DEP topographic data used for classification were retrieved from the GEE data catalog. Sentinel-2 Level 2A products, atmospherically corrected surface reflectance, for the 2017-2020 period was used to create a mosaic for the entire study area after filtering scenes with more than 20% cloud cover and applying a cloud mask; the 4 years is used to generate cloud-free scenes for the mosaic. Similarly, a 2017-2020 Sentinel-1 radar imagery mosaic was created for the study area. Image collection containing Sentinel-1 imagery was also queried to retrieve images acquired in interferometric wide imaging mode, images with a resolution of 10 meters, images acquired in ascending orbit, and images in VV-VH polarization. Sentinel-1 images available on GEE have been pre-processed, applying thermal noise correction, radiometric calibration, and terrain correction.

Image Classification and process flow

As summarized in Figure 4 image classification was carried out but applying object-based image classification and random forest algorithm on input variables. Simple Non-iterative Clustering algorithm was applied to the red, green, blue, and Near-infrared bands of the Sentinel-2 data to segment the imagery into objects of connected pixels called superpixels. Grey level co-occurrence matrix (GLCM) was applied to red, green, and Near-infrared bands to extract textural information which was subsequently passed through a principal component analysis algorithm to reduce the dimensionality of textural outputs. These outputs together with all bands of sentinel-2 imagery, sentinel-1 imagery, USGS 3DEP DEM data, gridded soil data and derived parameters including spectral indices, ratio and span of the radar image, and slope and aspect of DEM data were used as inputs for the random forest classifier. Different scenarios were empirically tested to quantify the impacts of multi-source remote sensing on wetland classification. Classification for all the scenarios was carried out on a scale of 10 meters over the entire study area. Workflow for image classification was adapted from a procedure developed by (Tassi & Vizzari, 2020).

Accuracy assessment

30% of the data generated from the train-test split was used to assess the accuracy of the random forest classification. The overall accuracy of the model and ability of the classifier to identify classes wetlands, defined during training, from test data was recorded through the producer’s accuracy metric. Different combinations of data were also assessed to observe the influence of adding various data sources. Figure 4 shows the distribution of training samples in the study area. Table 1 summarizes the total number of test and train samples.

RESULTS

Tables 2, 3, and 4 show the confusion matrices of the model for three of the scenarios tested. Using only sentinel-2 imagery with spectral indices derived from input bands was able to identify water and urban classes with high accuracy. This is due to the large spectral dissimilarity between both classes and other classes being mapped. However, using the optical source alone did poorly identifying wetland classes and the forested upland with producer accuracies of 82%, 71%, 62%, and 59% for emergent, forested, scrub-shrub wetlands, and forested upland lasses, respectively. The inability of the classifier to distinguish between forested wetlands, scrub-shrub wetlands, and forested upland classes is due to the ecological similarity of these classes. Adding Sentinel-1 C-band radar image improved the ability of the classifier to identify wetlands and distinguish between forested wetlands, scrub-shrub, and forested upland, with producer accuracy of the emergent wetland class jumping by 8%, forested wetland by 5%, scrub-shrub by 10%, and forested wetland by 11%. Overall accuracy improved by 6%, showing the significance of combining radar imagery with optical imagery for wetland classification. DEM and soil datasets were also significant data sources for the classifier and were key in discriminating between the forested and scrub-shrub wetlands classes and the forested upland class. Producer accuracies of these classes increased by 7%, 7%, and 9%, respectively, after adding these ancillary data sources. The overall accuracy of the best performing combination resulted was 89%. Figure 5 shows the final map of the best-performing combination.
This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.

6. CONCLUSION

One of the main goals of this study is to show the potential of producing large-scale wetland inventory maps from major satellite sources in order to improve the periodicity of the production of such inventory maps. GEE is a very vital tool in this process making it possible to bypass challenges associated with large-scale map production. Providing access to preprocessed data by running simple queries on its interface to the data catalog makes for very efficient workflows. Parallel processing of this cloud computing platform allowed for the methodology to be applied to a large area without having to divide into smaller processing units while achieving very good accuracy.

Utilizing multiple data sources is seen to improve the ability of the classifier to identify and distinguish between wetland classes. Combining optical, radar, DEM, and soil datasets was key to training the classifier to discriminate better ecologically similar forested, scrub-shrub wetland, and forested upland classes.

Further improvements can be made to the classification model by using polygons as training data, and integrating additional ancillary data sources, especially from climate datasets to increase the overall accuracy and ability of the random forest classifier in discriminating wetland classes. Also, the possibility...
of training deep learning classifiers for large-scale classification should be explored.

![Figure 5. Output classification of random forest classifier](image)

7. REFERENCES

Bataglia, M., Banks, S. N., Behnamian, A., Bourgeau-Chavez, L., Brisco, B., Corcoran, J., Chen, Z., Huberty, B., Klassen, J., Knight, J., Morin, P., Murnaghan, K., Pelletier, K., & White, L. (2021). Multi-Source EO for Dynamic Wetland Mapping and Monitoring in the Great Lakes Basin. Remote Sensing, 13. https://doi.org/10.3390/rs13040599

Corcoran, J. M., Knight, J. F., & Gallant, A. L. (2013). Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota. Remote Sensing, 5(7), 3212–3238. https://doi.org/10.3390/rs5073212

Cowardin, L. M., Carter, V., Golet, F. C., & Laroe, E. T. (1979). Classification of Wetlands & Deepwater Habitats of the U. S. DIANE Publishing.

European Space Agency. (2014). Sentinel-1 SAR User Guide. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar

European Space Agency. (2015). Sentinel-2 MSI User Guide. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi.

Federal Geographic Data Committee. (2013). Classification of wetlands and deepwater habitats of the United States. FGDC-STD-004-2013. Second Edition. Wetlands Subcommittee, Federal Geographic Data Committee and U.S. Fish and Wildlife Service, Washington, DC.

Kloiber, S. M., Macleod, R. D., Smith, A. J., Knight, J. F., & Huberty, B. J. (2015). A Semi-Automated, Multi-Source Data Fusion Update of a Wetland Inventory for East-Central Minnesota, USA. Wetlands, 35(2), 335–348. https://doi.org/10.1007/s1157-014-0621-3

Mahdianpari, M., Brisco, B., Granger, J., Mohammadimanesh, F., Salehi, B., Homayouni, S., & Bourgeau-Chavez, L. (2021). The Third Generation of Pan-Canadian Wetland Map at 10 m Resolution Using Multisource Earth Observation Data on Cloud Computing Platform. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 8789–8803. https://doi.org/10.1109/JSTARS.2021.3105645

Mahdianpari, M., Granger, J. E., Mohammadimanesh, F., Salehi, B., Brisco, B., Homayouni, S., Gill, E., Huberty, B., & Lang, M. (2020). Meta-Analysis of Wetland Classification Using Remote Sensing: A Systematic Review of a 40-Year Trend in North America. Remote Sensing, 12(11), 1882. https://doi.org/10.3390/rs12111882

Mitsch, W. J., Bernal, B., Nahlik, A. M., Mander, Ú., Zhang, L., Anderson, C. J., Jorgensen, S. E., & Brix, H. (2013). Wetlands, carbon, and climate change. Landscape Ecology, 28(4), 583–597. https://doi.org/10.1007/s10109-012-0758-8

Shafizadeh-Moghadam, H., Khazaie, M., Alavigian, S. K., & Weng, Q. (2021). Google Earth Engine for large-scale land use and land cover mapping: An object-based classification approach using spectral, textural and topographical factors. GIScience & Remote Sensing, 58(6), 914–928. https://doi.org/10.1080/15481603.2021.1947623

Steve, K. M., Doug, N. J., & Andrea, B. L. (2019). Minnesota Wetland Inventory: User Guide and Summary Statistics. Minnesota Department of Natural Resources, St.Paul, MN, 68.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS Journal of Photogrammetry and Remote Sensing, 164, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001

Tassi, A., & Vizzari, M. (2020). Object-Oriented ULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. Remote Sensing, 12(22), 3776. https://doi.org/10.3390/rs12223776

U.S. Army Corps of Engineers, E. L. (1987). Corps of Engineers Wetlands Delineation Manual. 143.

Valenti, V. L., Carcelen, E. C., Lange, K., Russo, N. J., & Chapman, B. (2020). Leveraging Google Earth Engine User Interface for Semiautomated Wetland Classification in the Great Lakes Basin at 10 m With Optical and Radar Geospatial Datasets. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 6008–6018. https://doi.org/10.1109/JSTARS.2020.3023901

Whalen, S. c. (2005). Biogeochemistry of Methane Exchange between Natural Wetlands and the Atmosphere. Environmental Engineering Science, 22(1), 73–94. https://doi.org/10.1089/ees.2005.22.73

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.

https://doi.org/10.5194/issprs-annals-V-3-2022-411-2022 | © Author(s) 2022. CC BY 4.0 License. 416