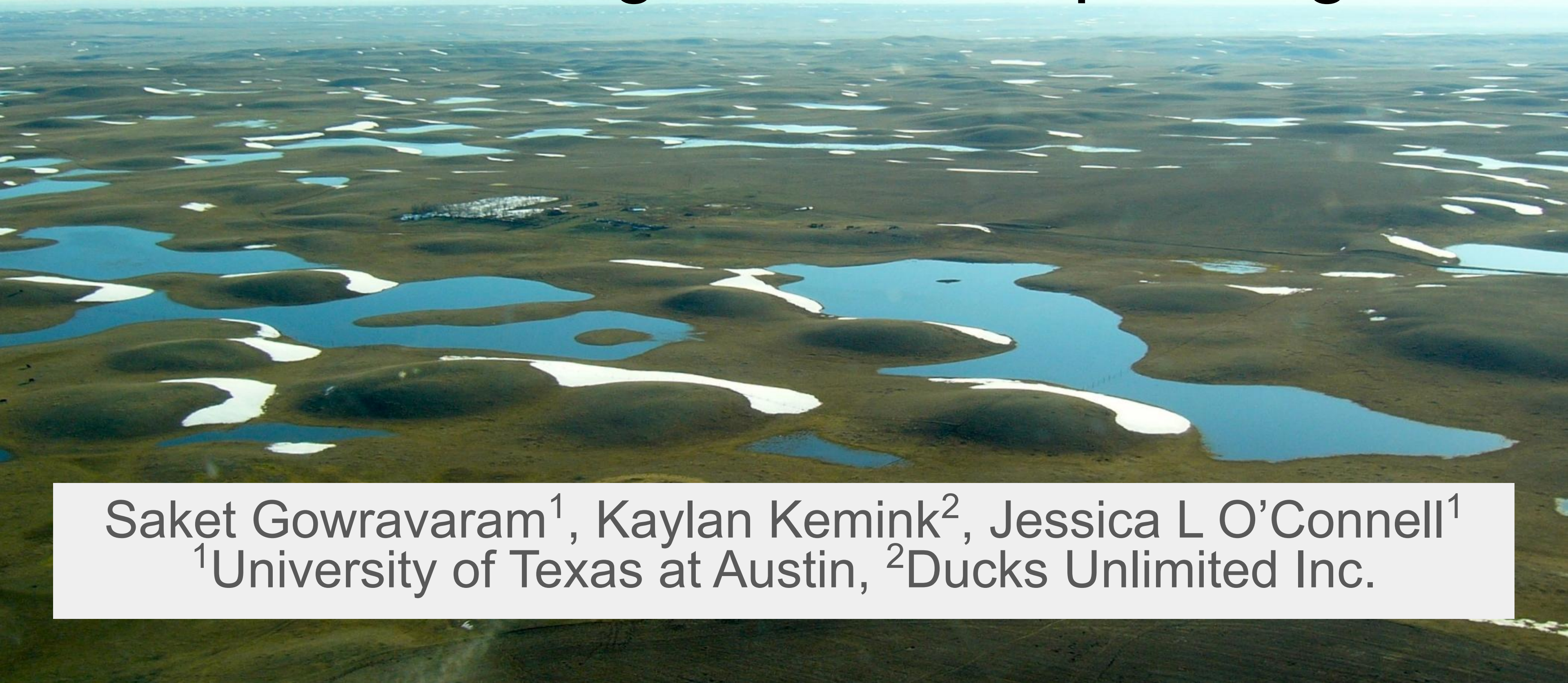
Automating measurement of Prairie Pothole Region wetland ponding



Saket Gowravaram[1], Kaylan Kemink[2], Jessica L O’Connell[1]
[1]University of Texas at Austin, [2]Ducks Unlimited Inc.

Thanks to project partners:

Prairie Pothole Joint Venture

US Fish and Wildlife Service/HAPET

Ducks Unlimited

Introduction

Goals

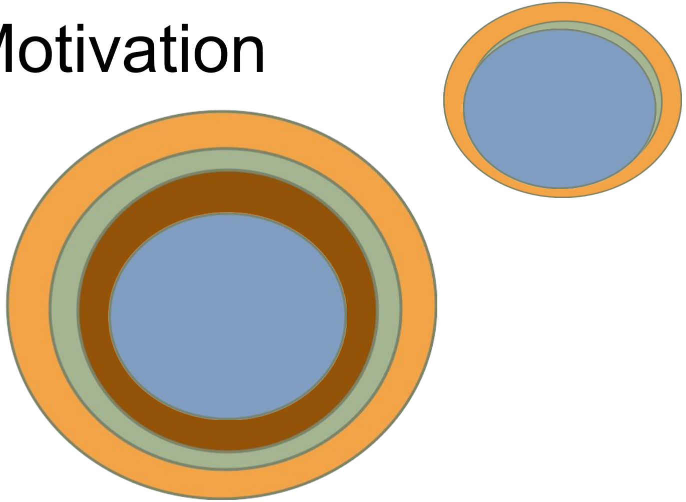
Develop an automated and shareable algorithm for mapping surface water in depressional Prairie Pothole wetlands from free and open satellite data

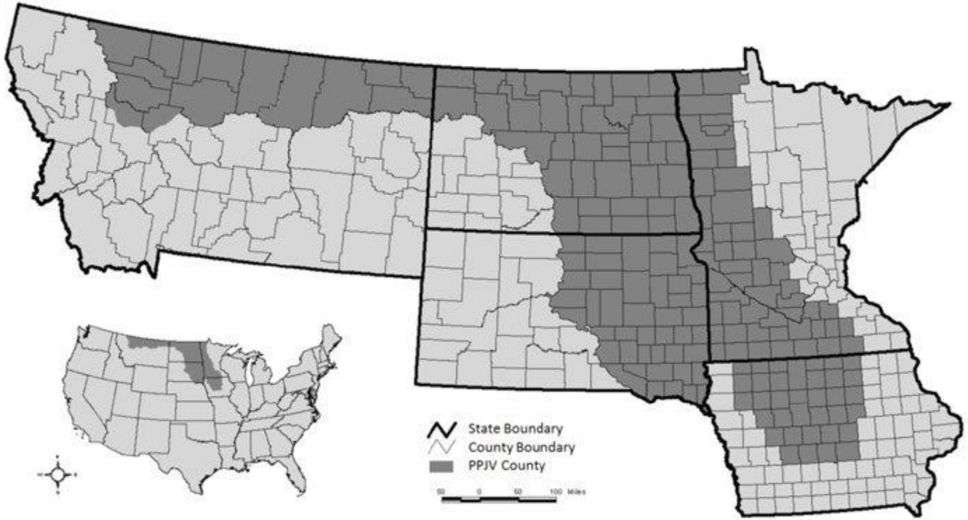
Train and test the algorithm against US Fish and Wildlife HAPET aerial survey data

Create a framework that accomplishes the essential tasks of creating surface water geospatial layers from high quality source data

Share and network with potential end-users in partnership with the Prairie Pothole Joint Venture and Ducks Unlimited

Motivation





Enhance planning ability of the conservation and management community by providing freely available open source wetland surface water predictions

Map surface water across the range of conditions experienced by PPR wetlands:
- shallow and turbid water, submerged aquatics, emergent vegetation, moist soils, wet meadows

Progress since last presentation

Based on user feedback from our last presentation, we:

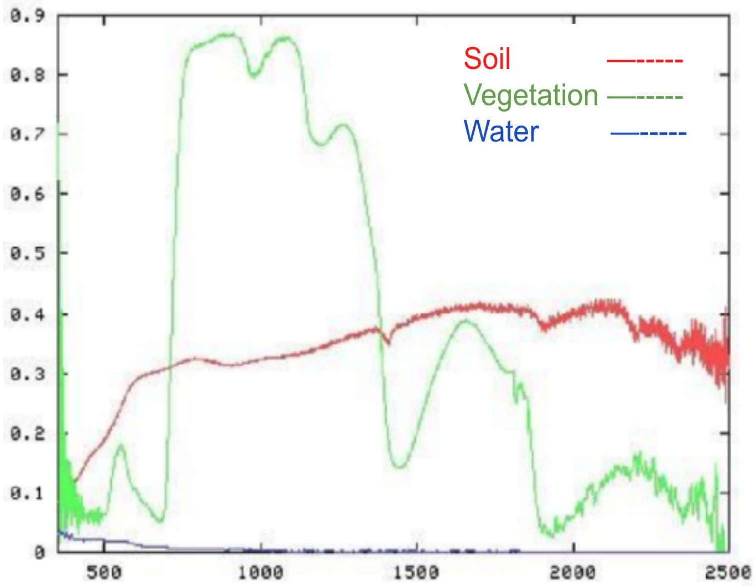
1. Updated our model training set to include both wetter (2016) and dryer (2020) years for pair (May) and brood (July/August) data
2. Ceased development on a web app and continued development on shareable code, allowing us to provide a) higher quality products that take longer to produce and b) flexibility for code knowledgeable end-users to develop their own workflows

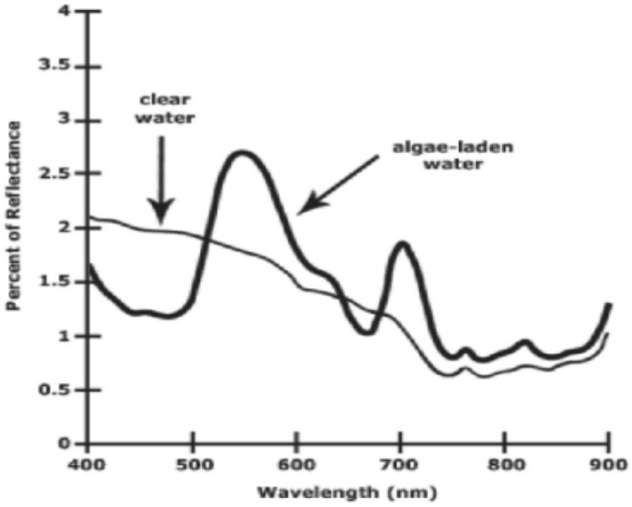
Remote sensing background

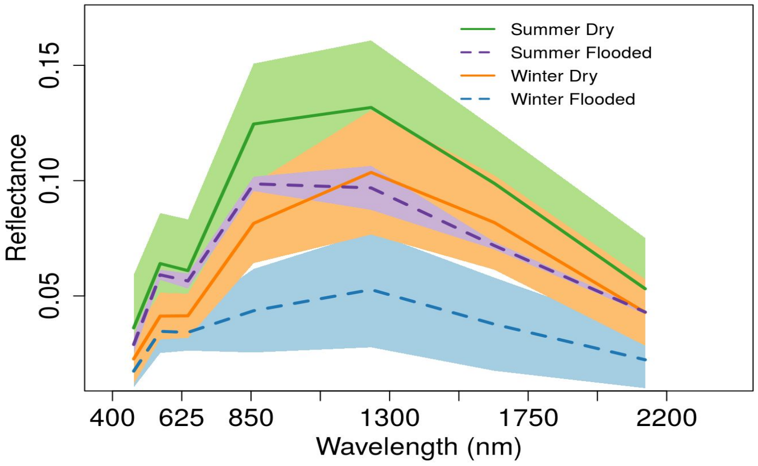
Spectral reflectance is the reflectance of light off an object

Water has a unique reflectance signature and we can use indices such as NDVI/NDWI to separate deep water from other classes

Classifying wetland surface water is not straightforward and benefits from training data across the range of landscape conditions







Machine learning background

Surface water classification is a supervised classification problem that requires training data as ground truth

Models are reliable only within the context of the ground-truth data

Training vs testing of models

- Training error is irrelevant: machine learning notoriously overfits training data
- Testing data needs to be truly independent

It is important to fine-tune a selected ML model to optimally suit the problem and the available training data

Wetland surface water classification and mapping

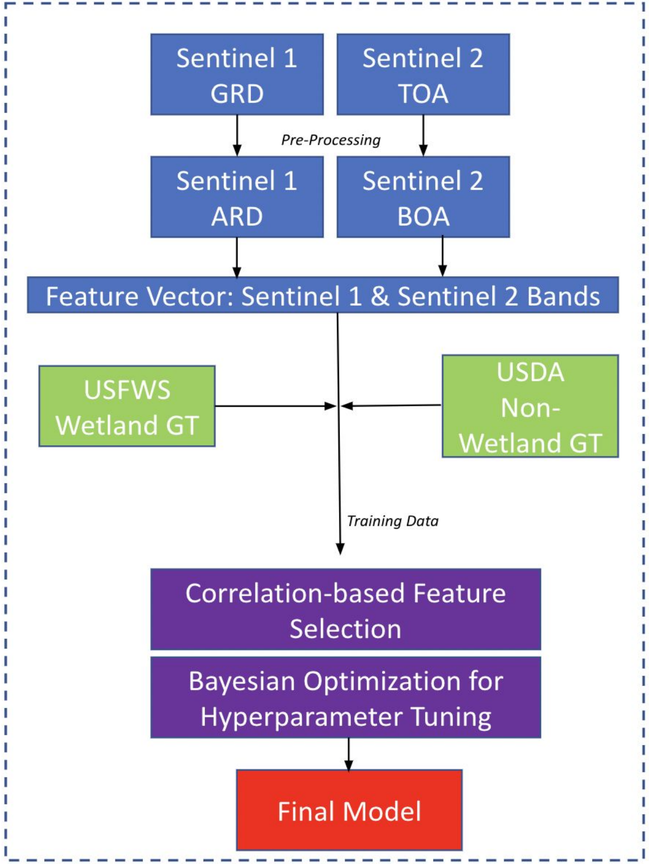
Model description

Random Forest Model, developed on Google Earth Engine

Sentinel-1 SAR 10-m: 5-day returns; provide insights during clouds and increases temporal resolution. **Preprocessing:** noise removal, speckle filter removal, terrain normalization;

Sentinel-2 MSI reflectance (10-m bands): 6 day returns; includes bands from visible to NIR spectrum. **Preprocessing:** atmospheric correction, TOA to BOA

Before model building and data classification, it is important to perform the preprocessing corrections described above



Model description

Machine learning model:

Ensemble learning provided by Random Forest

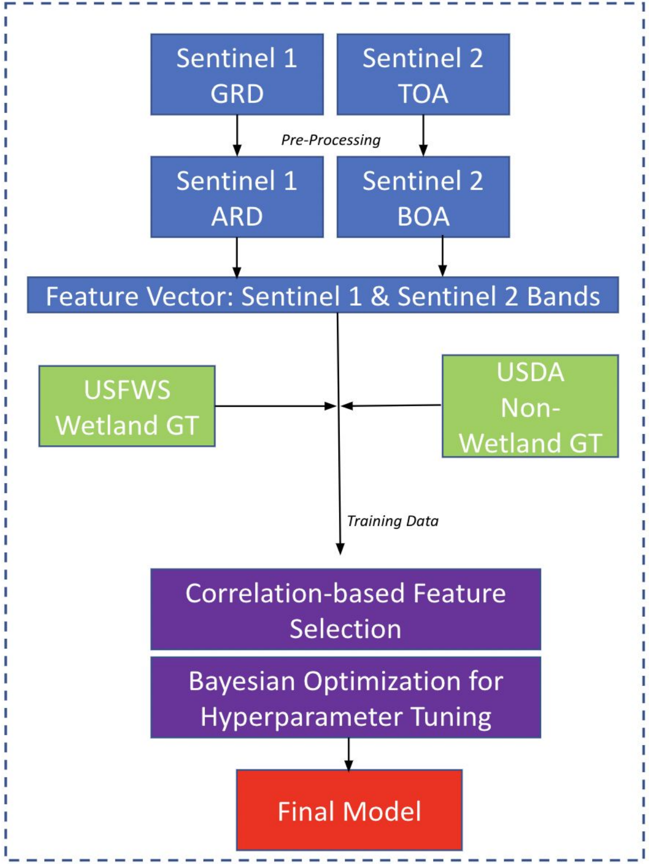
Model optimization methods: correlation-based feature selection (feature reduction and optimal feature selection) and bayesian optimized hyperparameter tuning (tuning model parameters)

Ground truth data:

Wet labels: USFWS aerial data

Non-Wet labels: USDA cropland data layer

Goal: a minimum 2-week prediction window to allow for potential future hydroperiod estimation

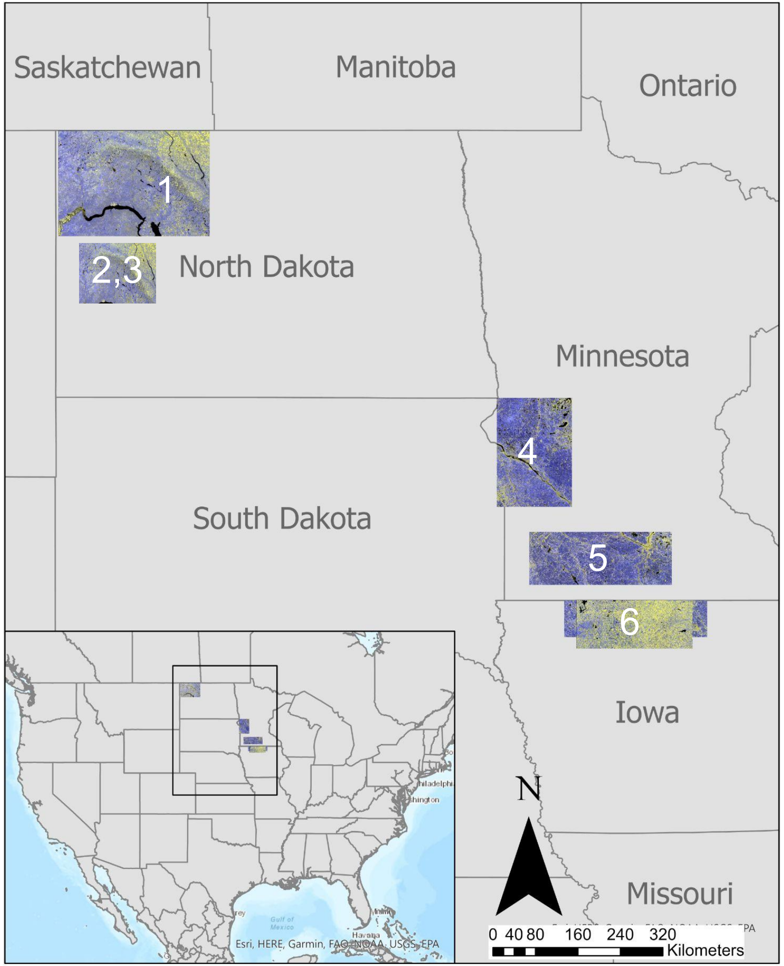


Model training and validation

Training: ND box 1 2016 May (pair ponds) and IA 2020 May-Aug

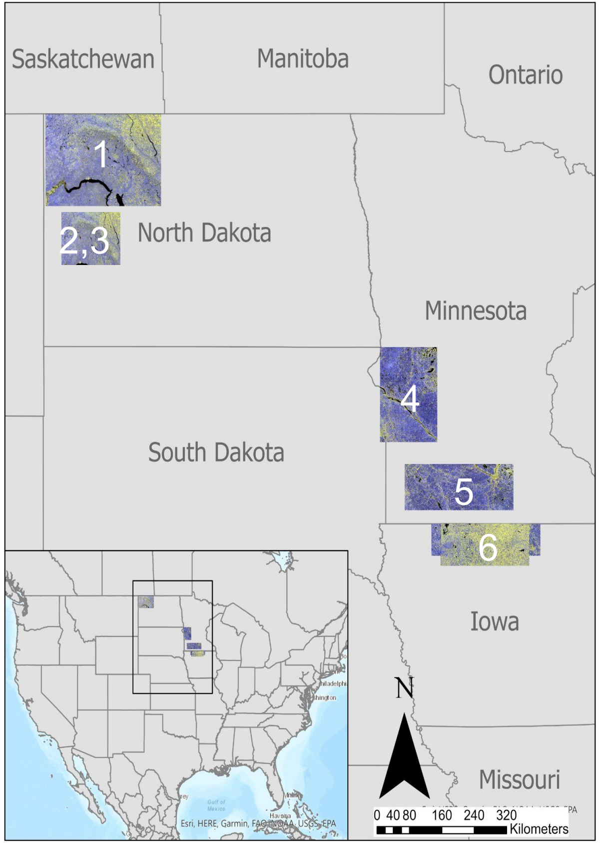
Testing:

Test 1: ND 2016 Jul/Aug (brood ponds)
Test 2: ND 2016 Jul/Aug (brood ponds)
Test 3: ND 2017 July/Aug (brood ponds)
Test 4 & 5: MN 2019, May-Aug
Test 6: IA 2019, May-Aug



Results and Discussion

Results



Test set		N	Sensitivity	Specificity	Accuracy
1.	ND May 2016	1334	0.98	0.94	0.95
2.	ND July 2016	2350	0.99	0.95	0.97
3.	ND July 2017	1666	0.97	0.98	0.98
4.	MN-1 2019	1031	0.70	0.93	0.93
5.	MN-2 2019	1990	0.65	0.96	0.80
6.	IA 2019	1271	0.77	0.92	0.89

Accuracy shows the overall classification effectiveness for novel data

Sensitivity is ability to detect true surface water

Specificity is the reduction of false positives

Model discussion

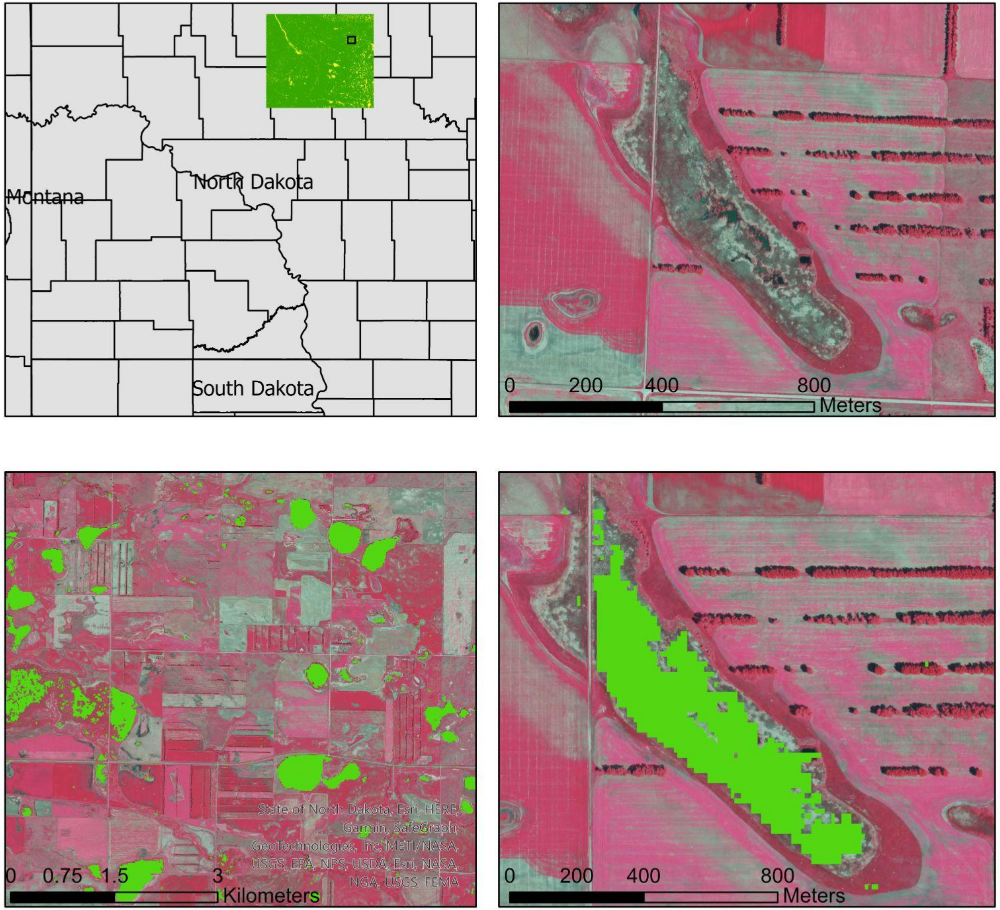
Reliability– Independent model tests suggest accuracies from 80-98%

Estimates are 2-wk averages via data fusion: variable number of observations within window

Model trained against breeding season data (May-August) and has not been tested for other seasons

Model tested against depressional wetland surface water (not rivers, ditches, etc)

Post-processing against road/urban, cropland layers may be useful. We're looking at adding this step into the code, but for now, this on the end-user.



Model and product sharing

Model documentation and links are hosted at:

https://www.landscapemodeling.net/surface_water.html

There you will find:

a) Links to the code
b) A written end-user guide explaining how to use the code
c) Examples images showcasing model results
d) A video tutorial with a code deployment walk-through (forthcoming)
e) A link to publications explaining how the model was developed and documenting accuracy and results (forthcoming)

Live code walk-through: 2-steps

1) Image correction: Sentinel-1(S1) SAR and Sentinel-2 (S2) top of atmosphere 10-m reflectance bands are filtered and corrected to “analysis-ready” data

   https://code.earthengine.google.com/50f81ba8045a115a04c4235e5bbcbf68

   Inputs: your Google username (allows you to output results to your Google account), a Region of Interest (ROI), and a date range, ideally spanning a 2-wk or > interval, and encompassing May-August time periods from Aug. 2015 forward

   Output: corrected S1 SAR and S2 Surface reflectance, added to your Google Earth Engine Assets, for use in step 2

2) Surface water prediction: Create surface water predictions from the Step 1 “analysis-ready” data

   https://code.earthengine.google.com/487a6efe646dc66e668117e387182d8b

   Inputs: Files from Step 1 as GEE assets in your own GEE account

   Output: a predicted surface water geotiff layer, saved to your Google Drive folder

Some tips for best results

1) You need to create a accounts for 1) Google, and 2) Google Earth Engine
2) Make sure you have enough room for output files in your GEE assets and your Google Drive
3) Feel free to save the code we provided in your scripts folder and make it your own. You won't affect our source script.
4) To create the script outputs, you need to provide your username, date, and ROI info), "run" the script, and then remember to look at your task folder and execute the tasks
5) Spatial considerations: Larger areas take more memory; If you have a free GEE account, you can't predict areas much larger than our default ROI which adds 225 GB to your assets
6) Temporal considerations: Only try to predict dates where Sentinel exists (2015-present); May-August and short time windows will provide the best predictions
7) File name considerations: GEE won't overwrite existing files in your assets; Make sure the file names in Step 1 are unique and change the input names for step 2 to match these files
8) Step 1 especially takes awhile to run, and time to run is greater with greater temporal and spatial scale; Scripts run faster when others aren't using GEE; Default ROI runtime was 9 hrs at last test

Questions?