Wetland Intrinsic Potential Tool: Identifying forested wetlands through lidar-derived machine learning

University of Washington
Meghan Halabisky, Monika Moskal, Luke Rogers

TerrainWorks
Daniel Miller, Lee Benda, Kevin Andras
Problem Statement: Identifying wetlands

In Washington State available wetland inventories are often out-of-date and have high errors of omission (especially in forested and agricultural areas).
Where are the wetlands on forest lands?
Where are the wetlands on forest lands?
Key technology used: LiDAR - Light Detection and Ranging

- Active airborne laser scanner
- Returns are points with X, Y and Z coordinates
- LiDAR Products:
  - Ground model
  - Canopy Surface model
  - Slope
  - Intensity image
**Phase 1:** Develop a tool for mapping hydrological and geomorphological controls on wetland occurrence.

The ‘Wetlands Intrinsic Potential’ (WIP) tool uses digital elevation models (LiDAR) and may incorporate other digital data, including soils, geology, and multi-spectral imagery. (Luke Rogers (UW), TerrainWorks)

**Phase 2:** Use field data on wetland locations to evaluate methods developed in Phase 1, and to develop new machine learning models mapping probability of wetland occurrence. (Meghan Halabisky, TerrainWorks).
WIP Tool Phase 2: Project objectives

1.) Identify key variables used to predict wetlands in the PNW

2.) Collect sample training of wetland and non-wetland locations

4.) Test machine learning methods – random forest models

5.) Develop an ArcGIS tool that is flexible and can be used by anyone as screening tool
METHODS
Study Areas

Model development:
• Puyallup watershed

Tested transferability of model on:
• Mashel watershed
• Coulter Creek-Kitsap peninsula
• Hoh watershed
Literature Review:

East coast, Midwest, & E. WA have had success mapping wetlands:

- Topographic wetness index
- Lidar intensity
- Leaf-off imagery
- Depth-to-water index
- Rule based approach v. random forest method

Identify Key Variables
1.) Topographic features

Topographic Indices:
- plan curvature,
- profile curvature,
- gradient (slope)
- DEV

Plan Curvature (across slope)

Profile Curvature (along slope)

DEV = (elevation - mean elevation)/standard deviation elevation
Topographic attributes can be measured over different length scales to highlight landforms of different sizes.
2.) Hydrologic Modelling - TWI

Martin Kopecký, Martin Macek, Jan Wild,
Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition,
2.) Hydrologic Modelling - Depth-to-water index

3.) Spectral indices - NDVI
4.) Vector datasets – soils, geology, other wetland inventories
Collect training and validation data collection
Total points = 2,417

GRTS sample design:

Puyallup
• 1,270 point photo interpreted
• 101 assessed in the field

Mashel:
• 94 points photo interpreted
• 74 assessed in field

WIP Tool sample design:

Coulter Creek:
410 points photo interpreted.
Spent 5 days in the field.
36 assessed in the field.

Hoh watershed:
360 points photo interpreted.
Spent 5 days in the field.
145 assessed in the field
• Random forests are built from decision trees
• Place observations into classes by making binary decisions on their features

• In this study:
  • Observations: Training points in the study area
  • Classes: wetland/upland
  • Features: data from input rasters

Image: Tony Yiu
Random forests are built from decision trees

• Random forests generate 100s-1000s of decision trees, built on unique subsets of observations and features in the training data.

• Classifications are predicted by taking a vote of all trees in the model.

```
Observation

Slope < 10%?

  Yes

  Yes

  Catchment > 2mi²?

    Yes

    Wetland

  No

  No

  Upland

  Upland

  Upland
```

23
New Observation

Classify with all trees

74% wetland
26% upland

Score: 0.74
Class: Wet
RESULTS
Results – Wetland probability for Puyallup watershed

• Model outputs = Probability raster, 0 - 1 likelihood

• Using a cutoff of 0.5 = 4 x more wetland area in forested areas than the NWI
NWI
Overall accuracy = 88.1%
Error of Commission = 2.1%
Error of Omission = 41.8%

WIP model
Overall accuracy = 96.6%
Error of Commission = 4.3%
Error of Omission = 8.0%
Random Forest Model

Variable Importance

Feature importance for full random forest model for Puyallup. The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest. Variables at the top contributed the most definition in the random forest model.
A model trained on a similar area can be transferred to that area.

Overall accuracy = 97%
Error of omission = 16%

Overall accuracy = 96%
Error of omission = 21%
However, need to build a new model for areas that are in a different ecoregion – Easy to do!

Coulter Creek – Kitsap peninsula

Hoh watershed
Conclusion

• The WIP tool identifies wetlands missed in existing wetland inventories
• These may be wetlands that are hard to identify in aerial imagery alone.
• The model can be improved as new input data layers are identified as important.
• Can be used for improving sampling efficiency
• Can be used to screen for potential wetlands – can lower the cutoff or raise the cutoff.
• WIP model performs better when field data is used, but works very well with NWI training data (available everywhere).
Limitations of the WIP Tool:

The WIP tool provides an improvement on identifying wetland locations in forested areas, but does not delineate wetland borders or classify wetland types. For any policy or management application, the WIP tool is best used as an initial screening for follow-up on the ground.

There are several limitations of the WIP tool:
1.) We did not use a jurisdictional wetland definition.

2.) While in theory, the WIP tool should effectively map wetlands in Eastern Washington, none of our study areas for this project were located in Eastern Washington.
Limitations of the WIP Tool (Cont...):

4.) The WIP tool may not provide useful results for slope wetlands and these wetlands will likely be missed in any WIP tool product. We did not have adequate training data locations of slope wetlands to train our model and therefore we could not test out the effectiveness of mapping slope wetlands using the WIP tool.

5.) The WIP tool is based on topographic features and surface water flow models. It does not account for well-drained soils. Certain areas may identify strongly as wetlands, but in fact be false positives due to underlying geology and soil types.
6.) The WIP tool was created primarily for forested wetlands. It may be useful for other non-forested areas, but this was not the focus of this research, and therefore it has not been assessed.

7.) The WIP tool may not produce useful results for areas with constructed human modification of water flows (i.e. drains, ditches) as these are not mapped as part of the lidar-derived hydrologic flow models used as inputs to the WIP tool.
Updates : Continuing to improve WIP tool

Increasing data in the Hoh watershed
- 800 additional points (1/4 visited in the field)
- Exploring additional soils layers
- Extending to the Colville watershed (eastside).
- Running for the Snohomish county
Some of the uses of the tool:

• Mapping wetlands on forestlands – screening tool
• Used for sampling design for forested wetlands.
• Identifying wetlands to consider how they might mitigate drought and stream permanence by recharging groundwater and storing water in hotter summer months. (Tulalip tribe)
• Currently using the WIP tool for a NASA study to map below ground carbon sequestration on forestlands along the wet to dry gradient, and specifically in forested wetlands.
Questions

Meghan Halabisky, halabisk@uw.edu