Wiscland-2 Quarterly Project Update 1

Prepared by Wiscland-2 Project Team
Contact: Howard Veregin
University of Wisconsin-Madison
howard.veregin@wisc.edu

March 23, 2015

Executive Summary

The objective of the Wiscland-2 project is to update the Wisconsin statewide land cover map to show the land cover of the state circa 2010. Since beginning Phase 2, the project team has focused on synthesizing the existing land cover sample datasets, assessing field collection requirements, developing online tools for efficient data submission and monitoring, performing photo-interpretation, and automating image processing tasks. In the upcoming quarter, planned tasks include continuing to augment and refine the samples database and image classification runs for Level 1 classes.

Introduction

The purpose of this document is to update the Wiscland-2 Land Cover Guidance Team on the progress of the project since the completion of the pilot project (Phase 1). The initial pilot project for Wiscland-2 was completed in October, 2014 with an updated (final) pilot project report submitted to the Land Cover Guidance Team in December, 2015. The primary work tasks since the pilot project update have been focused on:

- Identifying and acquiring available data layers from Landsat satellite imagery and ancillary (topographic, soil) datasets.
- Developing processing routines for data extraction, image classification, and cross-validation.
- Synthesizing and performing quality assurance of available sample data from existing land cover datasets.
- Identifying deficiencies in the sample database and assessing field collection needs for the upcoming field season.
- Developing web applications to aid field collection efforts.
- Establishing procedures for the field data to be submitted and ingested into the training database.

Results

Thirteen Landsat scenes (footprints) are needed to cover the state of Wisconsin. Over 100 images spanning dates from 2009-2014 have been acquired for each scene. Topographic and soil information including topographic wetness index (TWI), slope, available water supply (AWS), and soil organic carbon (SOC) layers were created using soil and digital elevation model (DEM) data.

The 21 land cover datasets provided by the WDNR that are suitable for both training and testing data were synthesized into a sample database by the Production Project Team. General steps included:

1. Crosswalk relevant samples to Wiscland-2 classification scheme.
    a. Samples are crosswalked explicitly through lookup tables for most datasets. For forest inventory data, a series of decision rules were written which use information on forest composition to meet Wiscland-2 class definitions.
2. Subset sample to meet quality criteria.
    a. All polygons sampled must meet or exceed a 2 acre minimum size. Other datasets are subset according to year of collection as appropriate.
3. Sample polygons to points.
    a. Polygons are sampled along a regular grid at 100m intervals. Small and/or irregularly shaped polygons are sampled at the centroid.
4. Merging datasets and QA/QC.
    a. Samples are screened for overlap. Any points falling on the same Landsat pixel that have conflicting class labels are flagged and removed, as are duplicate points.

Approximately 500,000 unique training/testing data points are suitable for all levels of the classification. However, the sample is unevenly distributed across the state and across classes.

To address the sample bias issue, we identified gaps in the sample database and with input from DNR staff, assessed the additional field needs and collection locations for each class across the state. Field needs were supplied to the DNR in February, 2015. Across the entire state of Wisconsin, approximately 16,000 additional data points are needed to meet the minimum requirements (100 points per class per Landsat footprint) for the Wiscland-2 classification scheme.

To facilitate efficient and standardized collection and submission of the field data from DNR staff to the Production Project Team, a field collection form and online submission application was developed. The Land Cover Field Reporter (LCFR) application will allow DNR staff to submit their field collected data to be ingested into the training point database on a weekly basis. This will allow for an ongoing, updated count of the needs for each class and each county by DNR staff and the Production Project Team and assist in efficient assignment of DNR staff to regions that require collection.

For the classes not requiring field visits (i.e. urban classes, barren, water), the Production Project Team has designed a sampling scheme and begun photo-interpreting samples using aerial imagery to address deficiencies in the samples database. Approximately 50% of the interpretation is completed for the state. Sample data from Wiscland-1 is also being reviewed and will supplement the samples where the land cover can be verified through aerial imagery.

Workflow automation and tools were developed to handle image preprocessing, including cloud masking, calculating vegetation indices and annual metrics, and image stacking. Tools to streamline the classification, including sampling data from the samples database for training and

testing, extracting sample data from the imagery, and cross-validating several classification algorithms were produced. Computing resources have been set up locally as well as through the UW Center for High Throughput Computing (CHTC).

Image preprocessing and cross-validation trials for the Level 1 classes have been tested and implemented using one Landsat footprint where sample data is most widespread. Cross-validation tests have been completed for this footprint to test performance of different atmospheric correction and cloud masking algorithms, and to identify informative features for the classification.

Conclusions

The pilot project provided proof of concept to extend the methods and lessons learned in order to classify land cover statewide. The next major task and obstacle is the timely collection and integration of field samples required for image classification. Currently, there are no footprints with complete sample data for any level of classification 1-4. As the field data is collected and imported into the database, it will allow the Production Project Team to better understand and identify which classes may be difficult to classify with high accuracy, as well as allowing for testing of possible solutions to low-accuracy outcomes. The Production Project Team is therefore currently focusing efforts on assisting the DNR with field campaign preparation, augmenting the samples database through photo-interpretation of new and existing sample points, and streamlining workflows and automation to make future image processing efficient. While cross-validation trials are beginning where feasible, most classifications require additional training data to meet the classification scheme and sufficient accuracy.

Current

The Production Project Team is currently focusing efforts on:

- Completing photo-interpretation for the new urban, water, and barren areas.
- Validating existing Wiscland-1 sample data.
- Assisting DNR with preparing for field collection.
- Finalizing the LCFR tool and developing a field collection status application for progress monitoring.
- Running cross-validation trials to compare performance of available atmospheric correction, cloud masking, classification, and segmentation algorithms.

Upcoming

There are several priority tasks currently being worked on for the next quarter:

- Documentation development and training of DNR field staff for the field collection season.
- Finalizing field collection status application to allow ‘real time’ monitoring and direction of the field collection effort.
- Field collection of the necessary training data by DNR staff.
- Feature selection processes to identify the most informative layers of the >1000 bands available for each footprint.

- Final selection of the classification and image segmentation algorithms.
- Statewide level 1 classification runs and beginning the iterative classification process.
- Development of workflows to integrate CHTC resources into the classification processes.