FOREST SERVICE CONTRIBUTIONS TO THE NATIONAL LAND COVER DATABASE (NLCD): TREE CANOPY COVER PRODUCTION

Bonnie Ruefenacht, Robert Benton, Vicky Johnson, Tanushree Biswas, Craig Baker, Mark Finco, Kevin Megown, John Coulston, Ken Winterberger, Mark Riley[1]

Abstract—A tree canopy cover (TCC) layer is one of three elements in the National Land Cover Database (NLCD) 2011 suite of nationwide geospatial data layers. In 2010, the USDA Forest Service (USFS) committed to creating the TCC layer as a member of the Multi-Resolution Land Cover (MRLC) consortium. A general methodology for creating the TCC layer was reported at the 2012 FIA Symposium in Knoxville, Tennessee by several USFS researchers. Since that time, remote sensing specialists at the USFS Remote Sensing Applications Center (RSAC) have translated those methods into a process capable of being implemented over the Contiguous United States, Coastal Alaska, Hawaii, Puerto Rico, and the US Virgin Islands.

This paper describes the products produced by the NLCD 2011 TCC team, the challenges encountered, and the solutions devised while creating this Landsat grained map over the entire nation. The NLCD TCC 2011 was produced in two forms. The first is called the analytical dataset and is intended primarily for purposes of research and analysis. This dataset has two data layers, which are a per pixel estimate of tree canopy cover and a per pixel estimate of standard error. The second form of NCLD TCC 2011 is called the cartographic dataset and is intended primarily as an image backdrop or map display. This dataset consists of a single layer – tree canopy cover – that is a statistically masked version of the analytical dataset. Both versions of the NLCD TCC dataset are distributed through the MRLC NLCD website (http://www.mrlc.gov).

INTRODUCTION

The first NLCD products were prepared by the Earth Observation and Science (EROS) Center. The latest version of the NLCD percent canopy cover dataset was prepared by the USDA Forest Service (USFS) and the Remote Sensing Applications Center (RSAC). The NLCD products are available for free at http://www.mrlc.gov/ (last accessed 6 Jul 2015) and are downloaded more than 400 times per month.

METHODS

There are 456 Landsat WRS2 path/rows for the conterminous United States (CONUS), 51 for coastal Alaska (AK), 10 for Hawaii (HI), and 6 for Puerto Rico (PR) and the United States Virgin Islands (USVI). For all these path/rows except for HI, PR, and USVI, all Landsat 5 images with less than 70 percent cloud cover acquired during the growing season for the years 2009-2011 were downloaded from glovis.usgs.gov (last accessed: 6 Jul 2015). For HI, PR, and USVI, Landsat 8 images for the

[1] Senior Remote Sensing Analyst (BR), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center, 2222 West 2300 South, Salt Lake City, UT 84119; Project Manager/ Remote Sensing Analyst (RB), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center; Remote Sensing Analyst (VJ), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center; Remote Sensing Analyst (TB), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center; Remote Sensing Analyst (CB), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center; Principal Investigator (MF), RedCastle Resources, Inc. working onsite at the Remote Sensing Applications Center; Program Leader (KM), Remote Sensing Applications Center; Research Scientist (JC), Southern Research Station, USDA Forest Service; Forester (Retired) (KW), Formerly with Pacific Northwest Research Station, USDA Forest Service; Remote Sensing Program Lead (MR), Region 6, USDA Forest Service. BR is corresponding author: to contact, call (801) 975-3828 or e-mail at bruefenacht@fs.fed.us

2013-2015 growing season were used. The individual Landsat scenes for each path/row were combined into a single median composite for each path/row.

For each of the Landsat median composites, NDMI (normalized difference moisture index), NDVI (normalized difference vegetation index), and Tassel Cap (Baig and others 2014, Crist and Cicone 1984) images were created to use as explanatory variables. Other explanatory variables available for modeling but not necessarily used included elevation data, elevation derivatives, ecoregions, NLCD 2001 land cover, NLCD 2001 TCC, pixel coordinates, soils data, climate data, and geology data. Using 3 by 3 windows, focal standard deviation images were created for the Landsat data, Landsat derivatives, elevation data, and elevation derivatives.

There were 63,010 (CONUS), 1,884 (coastal AK), 737 (PR and USVI), and 1,385 (HI) USDA Forest Service Inventory and Analysis (FIA) plot locations used to collect the response data. A circle with a radius of 43.9 m was placed over each FIA plot center. Each circle contained a 109-dot grid, which was oriented 15 degrees east of true north, with each dot separated by 8 m. Photo-interpreters evaluated each dot as being either tree or not tree. For each plot, percent TCC was calculated from these dot counts.

Because the spatial resolution of the response data was approximately 90 m, focal means using 3 by 3 windows were created for all of the 30 m explanatory variables except for the focal standard deviation images, large-scale images, and thematic datasets, which included ecoregions, NLCD 2001 land cover, pixel coordinates, soils data, climate data, and geology data. For thematic datasets that were not large-scale such as NLCD 2001 land cover, focal majority algorithms using 3 by 3 windows were used.

The algorithm used to model TCC was random forest as implemented in R 3.02 (Liaw and Wiener 2002, R Core Team 2013). Selected explanatory variables along with the response data were used to train the random forest model. The random forest model was applied to the original datasets.

The number of trees used in the random forest algorithm was 500, which means for every pixel, 500 TCC predictions were generated. The final TCC estimate for each pixel was the mean of these 500 predictions. Standard errors for each pixel were derived from these 500 predictions.

To create the cartographic NLCD 2011 TCC dataset, 500 random forest models were created using a portion of the data that was used to create the TCC dataset. The portion of the data not used was applied to the random forest models to obtained predicted TCC values, which were used to derive t-values: (predicted TCC – observed TCC)/standard error. The derived t-values were multiplied by the standard errors of the NLCD 2011 TCC dataset. If the TCC value was less than this product, the TCC value was forced to 0.

RESULTS AND DISCUSSION

There were many challenges and learning experiences encountered while creating the NLCD 2011 TCC product. One of the first challenges was to efficiently process over 7,000 Landsat scenes. Elements of this challenge included how to remove clouds and shadows, how to deal with banding effects caused by Landsat 7 SLC-off gaps, and how to condense individual Landsat scenes for a path/row into a single image for a path/row. FMASK (Zhu and Woodcock 2012), a cloud and shadow masking program for Landsat, was released in January 2012, which helped with the cloud and shadow removal problem. FMASK, is not perfect and there were occasions when manual editing of clouds and shadows was necessary especially in coastal Alaska and the mountain regions of the southwest US. The Landsat 7 SLC-off gap banding problem was solved by not using Landsat 7. We developed a median composite technique (Ruefenacht in review) to condense the individual Landsat scenes into a single image. Additionally, an automatic processing system was developed, which was instrumental in being able to process the volume of data in a timely manner.

Another challenge was the shifting of Landsat scenes between path/rows. We built a national grid system based upon the NLCD 2001 TCC layer and anchored all Landsat scenes to this national grid.

Originally, all of the explanatory data were used for the TCC modeling. However, some of the explanatory data, such as ecoregions and the focal standard deviation images, caused artificial boundaries in the TCC product. Thus, we carefully selected a subset of the explanatory variables for use in the TCC modeling.

There were other challenges encountered in creating the NLCD 2011 TCC dataset that were not solvable. For instance, we did our best to select and process Landsat scenes to avoid visible boundaries or seamlines between overlapping Landsat path/rows. Some seamlines are visible in the TCC data, but they are not prevalent. The influence of terrain shadowing can also be observed in the TCC data. Finally, in Alaska, clear-cut areas quickly regenerate into extremely thick stands of trees with 100 percent canopy cover. These areas were difficult to represent accurately.

ACKNOWLEDGMENTS

The authors thank the IT staff of the Remote Sensing Applications Center for their help with various computer software and hardware issues.

LITERATURE CITED

Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. 2014. Derivation of a tasseled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sensing Letters. 5(5):423-431.

Crist, E.P.; Cicone, R.C. 1984. A physically-based transformation of thematic mapper data-the TM tasseled cap. IEEE Transactions on Geoscience and Remote Sensing. GE-22(3):256-263.

Liaw, A.; Wiener, M. 2002. Classification and regression by randomForest. R News. 2(3):18-22.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ruefenacht, B. In review. Using Landsat-5 TM composites to model percent tree canopy cover. Photogrammetric Engineering and Remote Sensing.

Zhu, Z.; Woodcock, C.E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment. 118(2012):83-94.