A new dataset for quantifying surface water dynamics in the North American high northern latitudes

Introduction

Characterizing the continental surface of the Arctic and Boreal regions of the world is critical to understanding components of the ecosystem as well as the system as a whole. These landscapes are dominated by a heterogeneous land surface dotted with lakes, ponds and wetlands (Carroll et al. 2011; Downing et al. 2006; Smith et al. 2005). As a land cover type, water plays a significant role in measurements such as surface energy balance, carbon cycling, temperature, and soil moisture, among many others.

The upcoming Arctic and Boreal Vulnerability Experiment (ABoVE) http://above.nasa.gov/ will conduct multi-disciplinary studies to explore and quantify the linkages between all of the individual components of the ecosystem. One common thread among the components of the ecosystem is the need for maps of the location, extent and dynamics of surface water (Kasischke et al. 2011). Water represents a transport vehicle for nutrients and other solutes, provides habitat for flora and fauna, and has a controlling effect on surface energy exchange yet it remains one of the least characterized features of the study region.

Satellite remote sensing data have been used to create maps of the location of persistent water bodies at large spatial scales (Carroll et al. 2009; Lehner and Doll 2004). These maps represent a snapshot in time of water location and extent. Global and regional land cover maps usually have water as a land cover class but this can often be applied after the classification using a static water mask (Bartholome and Belward 2005; Friedl et al. 2010; Homer et al. 2007). Additional work has been done using multiple satellite images or time series to show changes in the extent of water, mostly on a local basis or at moderate to coarse resolution (Carroll et al. 2011; Karlsson et al. 2012; Pavelsky and Smith 2008; Pietroniro et al. 1999; Smith et al. 2005). All of these studies have been limited in spatial or temporal resolution (or both).

The current work utilizes a time series of Landsat data to produce maps of surface water extent covering the entire Boreal and Tundra regions of North America for three epochs 1991, 2001, 2011 at 30m spatial resolution. Each of the maps has been generated with at least 3 years of ice-free data and will enable future work for the ABoVE campaign in determining the dynamic changes in small water bodies that dominate the region. With continental scale maps of surface water extent at a decadal time step we can evaluate change in surface water in a consistent way across the region in the same way as previous land cover change phenomena have been examined (Hansen et al. 2010; Zhan et al. 2002).

Here we present the first results of this new dataset showing the methodology, evaluation of results, and initial examples of how the product can be used to detect and monitor change in lakes in the Arctic and Boreal regions of North America.

Region of interest and data inputs

The primary region of study for this project is the ABoVE study region shown in figure 1. During the course of the project additional compute resources became available that enabled the expansion of the study region to include all of Canada and Alaska. This expansion allowed the inclusion of dynamic areas in the southeast Boreal region and to complete the continental map.

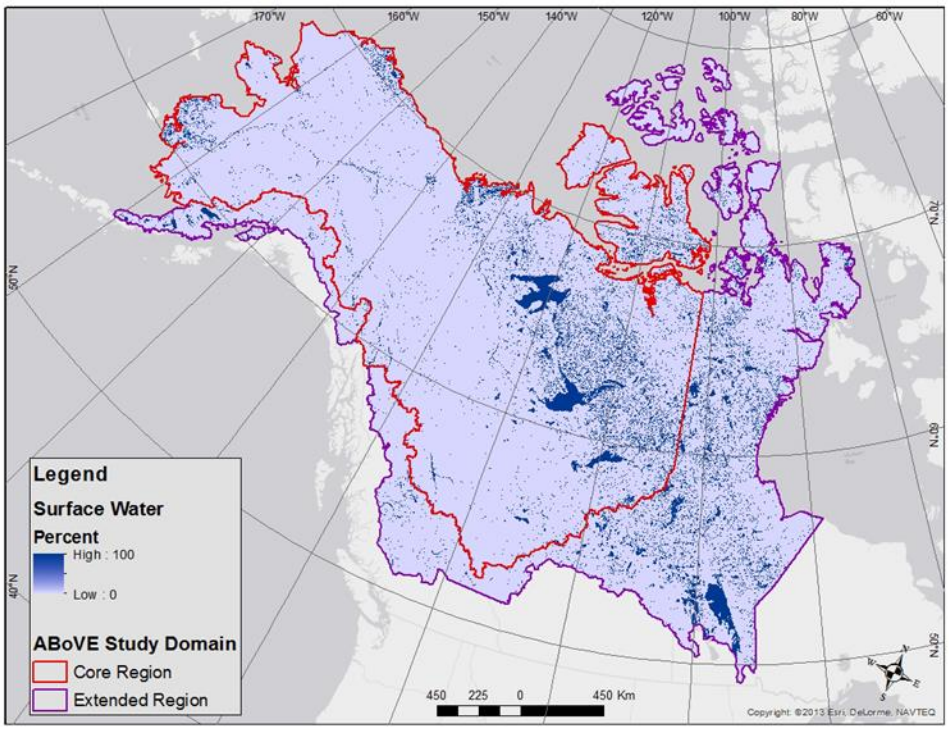

Figure 1 ABoVE study area in northern North America overlain with a subset of the Global 250m Water Map.

Data used for this project included Landsat 4 and 5 Thematic Mapper (TM) data and Landsat 7 Enhanced Thematic Mapper (ETM+). The data from 2000 to present for Landsat 5 and 7 is of good quality and the coverage is quite good with a minimum of 30 individual dates per scene for most scenes. The data for the 1990 – 1992 epoch is primarily from Landsat 5 supplemented with Landsat 4 in Alaska where data density is thinner. Data acquisition strategies in the early 1990's resulted in a limited archive of data over Alaska at this time. This presented a challenge to get enough observations to complete the map and will be discussed later.

Methods

The Landsat data were acquired from the US Geological Survey Eros Data Center through the ESPA https://espa.cr.usgs.gov/ ordering interface. Data acquired here has been processed to convert the radiance to surface reflectance. Conversion to surface reflectance is essential to generating an automated classification process that can be applied across scenes.

Unlike most land cover processing efforts using Landsat data, we acquire and process all available Landsat scenes regardless of cloud coverage. This decision is based on the experience with MODIS data where even a scene that is 80% - 90% cloudy can have clear observations of certain areas. By using all of the available observations we can increase the total density of the observations in any given area to get a consensus result. This method is enhanced in the Arctic due to the high degree of overlap between adjacent Landsat path/rows. By processing all of the available overlapping scenes it becomes possible to increase the total number of observations, which allows us to minimize the impacts of clouds and weather related phenomena such as drought and flood. Over 110,000 individual Landsat scenes have been processed to date.

High performance compute resources

The 110,000 Landsat scenes that comprise the raw data for this project result in total data volume in excess of 180 TB. In addition, processing time averaged 7 – 8 minutes per scene. The total amount of processing time on a single machine would turn out to be approximately 9 months. Since this project was funded through the NASA Terrestrial Ecology program as a pre-ABoVE data set, resources that were being developed for the ABoVE field campaign were available for use in processing and storing the input data. The ABoVE Science Cloud (ASC) has been created as a computing resource for use by the ABoVE science team for the duration of the field campaign. This resource consists of a data storage capacity greater than 2 Petabytes surrounded by a large amount of compute capacity (over 960 CPU cores) in a fully virtualized environment. Much like commercially available cloud computing resources, users request access to the system and request a system configuration and a number of individual processing nodes. The same system configuration is applied to each of the virtual processing nodes and the user is able to distribute their processing about the nodes. A full description of this system can be found online http://above.nasa.gov/science_cloud.html .

We utilized this system to store the raw data and to process through to final products. Using this system we were able to process all of the data in approximately 7 – 8 weeks instead of the projected 9 months if we processed on one desktop workstation. The availability of this resource allowed us to expand the spatial coverage beyond the ABoVE study domain to include most of Canada. In addition, this system offers the possibility of reprocessing the entire dataset if necessary to improve the results or to correct any issues that are identified in post processing or by users after release of the product. With current resources we anticipate that we could reprocess the entire data record in approximately 10 days.

Automated process

The processing scheme is completed in four steps: classification per scene, compilation of thematic layers per scene, compilation per tile, and post processing evaluation for error identification. The complete processing flow is shown in figure 2. All processing was performed using open source software (primarily gdal and python) and resources within a Linux Centos operating system.

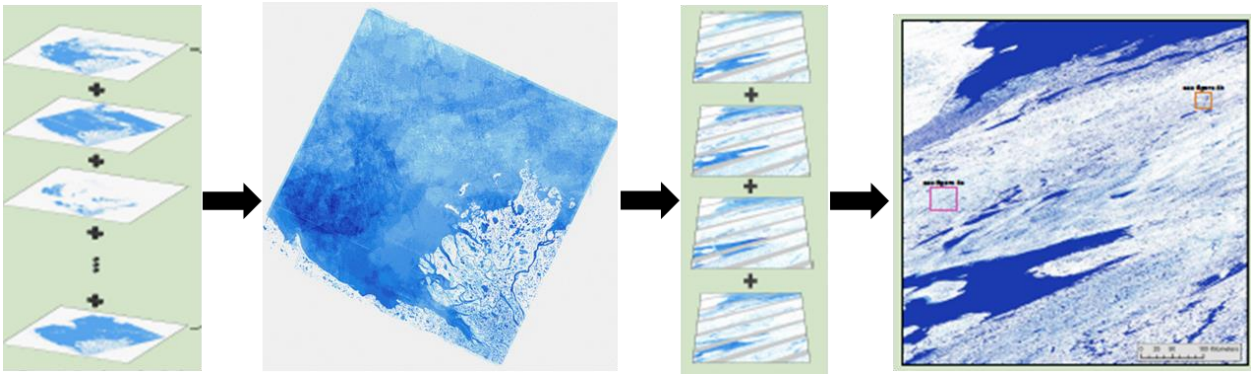

Figure 2 Processing work flow for the generation of the ABoVE water maps from Landsat scenes to ABoVE tiles.

Classification per scene

Each Landsat scene that was acquired was processed through in image classification scheme to generate an output with 5 classes: Water, Land, Snow/Ice, Cloud, Cloud Shadow. Quality assurance flags in the surface reflectance files provide an indication of cloudy and cloud shadow pixels and were used in the classification process. The image classification is a decision tree (Breiman et al. 1984; Kim and Loh 2001) classification system. This process was modeled after the approach taken to generate the Global 250m Water Map with Moderate Resolution Imaging Spectro-radiometer (MODIS) data (Carroll et al. 2009) and a host of other land cover mapping efforts. The resulting image classification has 5 classes (water, land, cloud, shadow, snow/ice). The algorithm is designed to be conservative in the labeling of water because we want to minimize the errors of commission (i.e. labeling land as water). The overall goal is to generate a map of the nominal extent of water for a given epoch, where nominal is neither the maximum nor the minimum but rather a representative extent for that time period.

Compilation per scene

The classification results have 5 classes, as described above. From this we extract thematic layers (i.e. water or land) and create summary results for a particular path/row using all of the available dates within an epoch. This result is a value from 0 to “n” where “n” is the total number of dates that were available for that path/row for that particular theme. This process can be repeated for each theme as necessary to produce the desired result.

Compilation per tile

Ultimately the goal of the project is to generate a unified product for the entire study region. To do this the individual path/row results need to be combined into a common projection and grid.

In coordination with the ABoVE project office and other pre-ABoVE projects it was decided to generate a grid based on the Canada Alber's Equal Area projection with equal sized "tiles" modeled after the MODIS labeling convention with horizontal and vertical offsets (figure 3). To take advantage of the overlapping of the path/rows it was necessary to mosaic with sum per pixel in order to not lose information. This was performed with intermediate steps to map the individual scenes into a series of tiles with non-overlapping scenes and then sum the resulting tiles.

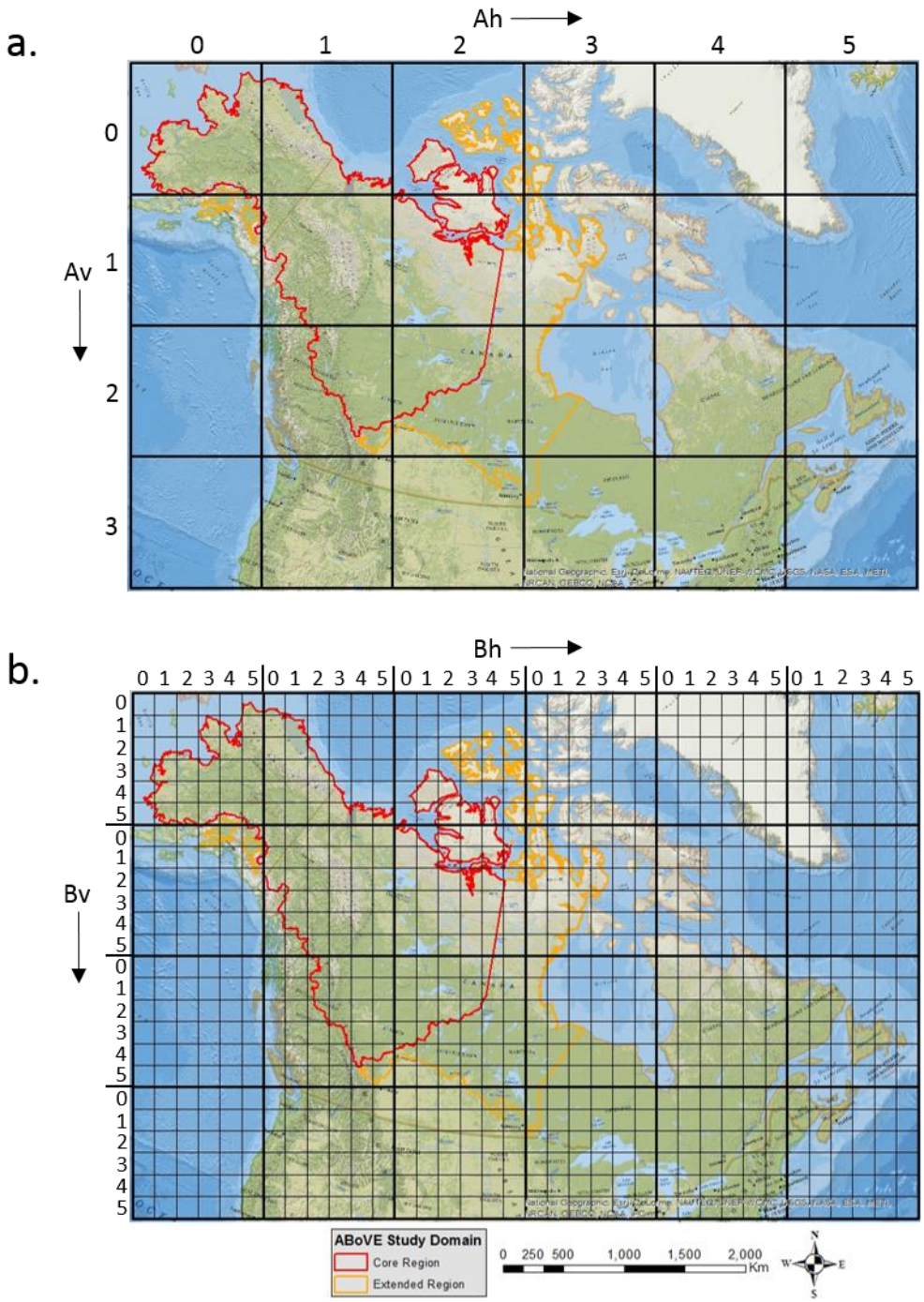

Figure 3 Proposed ABoVE standard reference grid for use with large area products. Based conceptually on the MODIS tiling scheme but utilizing Canada Alber's Equal Area projection.

Post processing evaluation

With the results compiled into cohesive and contiguous spatial units we went through systematically and evaluated the results. We found a variety of issues, as expected, ranging from geolocation shifts of up to 3 km in individual scenes, as well as false identification of water caused by terrain shadow. The geolocation shifted individual dates were identified and removed

from the results and the tile recompiled. A process is currently underway to address the false positives created by terrain shadow were masked using a custom algorithm using sensor geometry in conjunction with solar azimuth and topography data. The evaluation is an ongoing process and will be renewed when the data are released to the public, expected in Summer 2015.

References

Bartholome, E., & Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing, 26*, 1959-1977

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*. New York: Chapman & Hall

Carroll, M.L., Townshend, J.R., DiMiceli, C.M., Noojipady, P., & Sohlberg, R.A. (2009). A new global raster water mask at 250 m resolution. *International Journal of Digital Earth, 2*, 291-308

Carroll, M.L., Townshend, J.R.G., DiMiceli, C.M., Loboda, T., & Sohlberg, R.A. (2011). Shrinking lakes of the Arctic: Spatial relationships and trajectory of change. *Geophysical Research Letters, 38*

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M., & Middelburg, J.J. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography, 51*, 2388-2397

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*, 168-182

Hansen, M.C., Stehman, S.V., & Potapov, P.V. (2010). Quantification of global gross forest cover loss. *Proc Natl Acad Sci U S A, 107*, 8650-8655

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., & Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing, 73*, 337-341

Karlsson, J., Lyon, S., & Destouni, G. (2012). Thermokarst lake, hydrological flow and water balance indicators of permafrost change in Western Siberia. *Journal of Hydrology, in press*

Kasischke, E.S., Goetz, S.J., Kimball, J.S., & Mack, M.M. (2011). The Arctic-Boreal Vulnerability Experiment (ABoVE): A Concise Plan for a NASA-Sponsored Field Campaign. In. http://cce.nasa.gov/terrestrial_ecology/pdfs/ABoVE Final Report.pdf

Kim, H., & Loh, W.Y. (2001). Classification trees with unbiased multiway splits. *Journal of the American Statistical Association, 96*, 589-604

Lehner, B., & Doll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology, 296*, 1-22

Pavelsky, T.M., & Smith, L.C. (2008). Remote sensing of hydrologic recharge in the Peace-Athabasca Delta, Canada. *Geophysical Research Letters, 35*

Pietroniro, A., Prowse, T., & Peters, D. (1999). Hydrologic assessment of an inland fresh water delta using multi-temporal satellite remote sensing. *Hydrological Processes, 13*

Smith, L.C., Sheng, Y., MacDonald, G.M., & Hinzman, L.D. (2005). Disappearing Arctic lakes. *Science, 308*, 1429

Zhan, X., Sohlberg, R.A., Townshend, J.R.G., DiMiceli, C., Carroll, M.L., Eastman, J.C., Hansen, M.C., & DeFries, R.S. (2002). Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment, 83*, 336-350