GLanCE
Global Land Cover mapping and Estimation

(B11I-2286) An approach for collecting and evaluating land cover training data using time series of Landsat data

Konrad Turlej, Curtis E Woodcock, Katelyn Tarrio, Yingtong Zhang, Paulo Arturo Arevalo, Eric Bullock, Mark A Friedl
Boston University - Department of Earth and Environment

BOSTON UNIVERSITY NASA USGS science for a changing world

Highlights

- We present a new automated method for efficient training data collection for mapping land cover and land cover change at regional and continental scales
- Our approach relies on statistical analysis of the spectral and temporal characteristics of the land cover and selecting the most representative examples for existing land cover variations
- Our approach can be used for augmenting already existing training data sets or generating new ones from scratch
- We present the potential of our method by augmenting already existing training data set for the area of the North-Eastern US initially collected via visual interpretation (Fig. 1)

Methods

Our automated approach consists the following steps:

1. Collect the time series of Landsat data for a random sample of locations within a given region of interest
2. Perform change detection for the sampled locations and model the intro- and intra-annual patterns of surface reflectance for the land cover via CCDC algorithm
3. Apply the K-Means clustering analysis to the CCDC models' parameters to find the optimal number of land cover variations (K variable) and to group them
4. Select the most representative examples of the land cover variations defined by the clustering analysis (Fig. 2). We based the selection on: a) the lowest proximity of the sample from the clusters' centers (Euclidean distance), b) the highest similarity of a sample to the other samples within each cluster (silhouette score), and c) the length of the time series for which the sample's sample model remained constant
5. Optionally, select only the training samples that represent the land cover variations which are not represented by already existing training data.

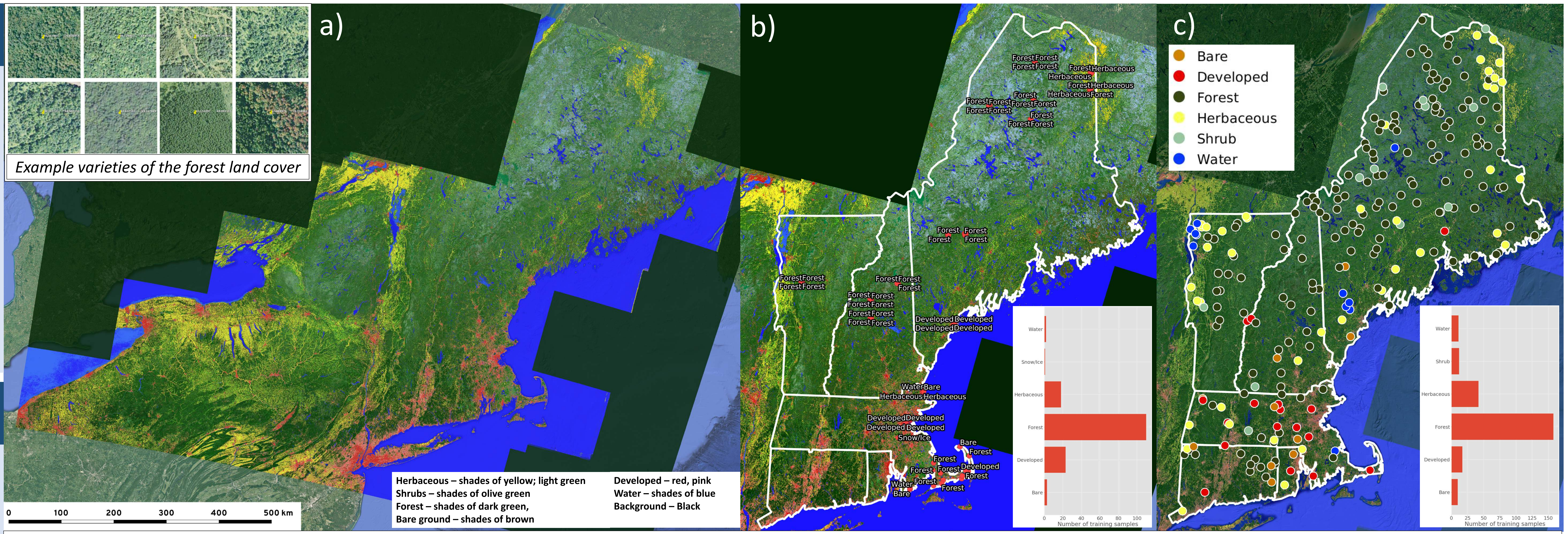


Figure 1. The results: a) Visualization of the clustering analysis for the North-Eastern US. We estimated that 50 unique variations of land cover can be found in the area and collected 5 best samples for each variety. The clusters are based on intro- and intra-annual patterns of reflectance; b) A set of 157 initial training samples defined via visual interpretation. The samples are spatially clumped and cover only few varieties of the existing land cover; c) The generated training data set consisting 250 samples. The generated samples cover all varieties of the existing land cover and their spatial distribution is more even comparing to the manually generated samples.

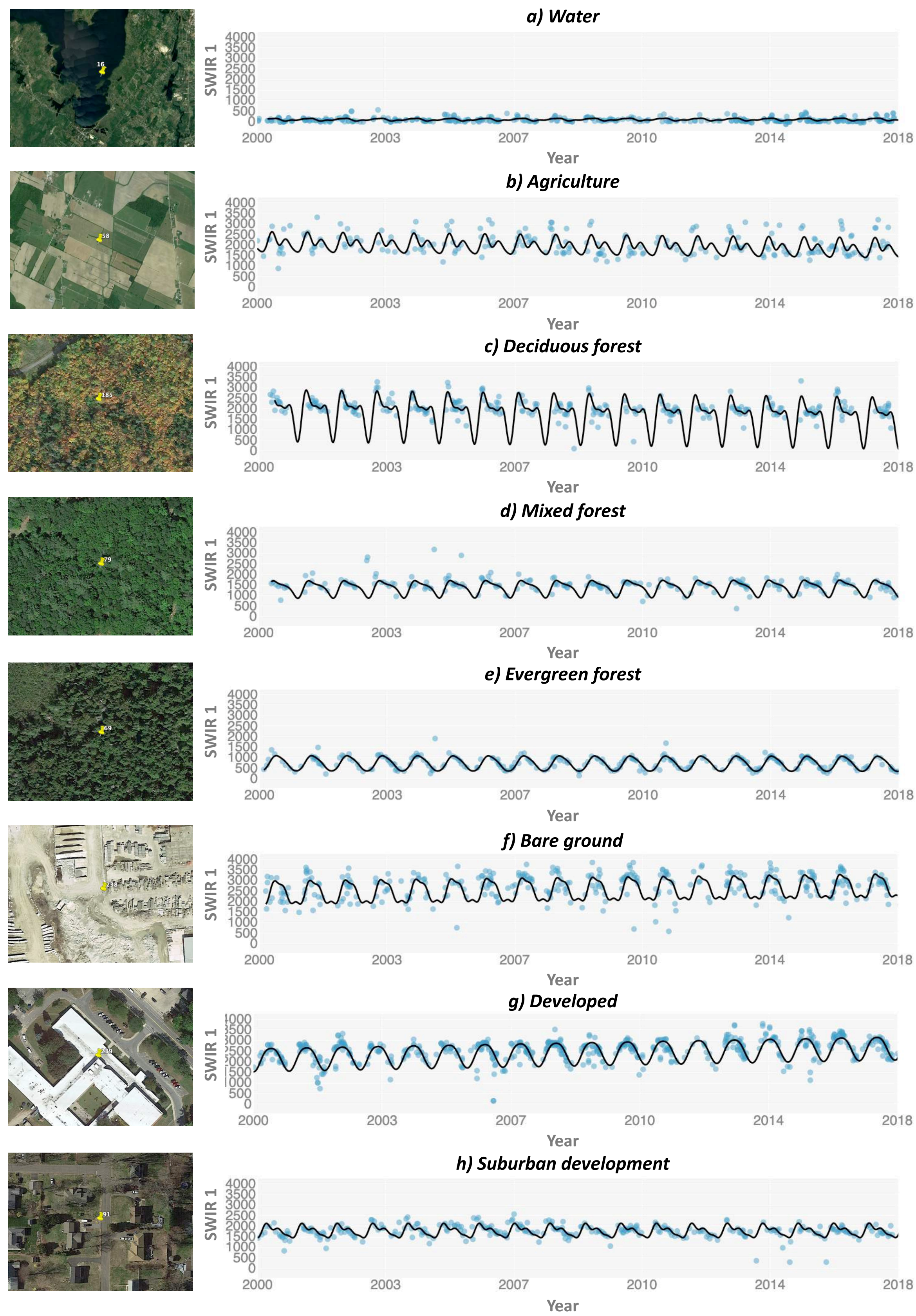
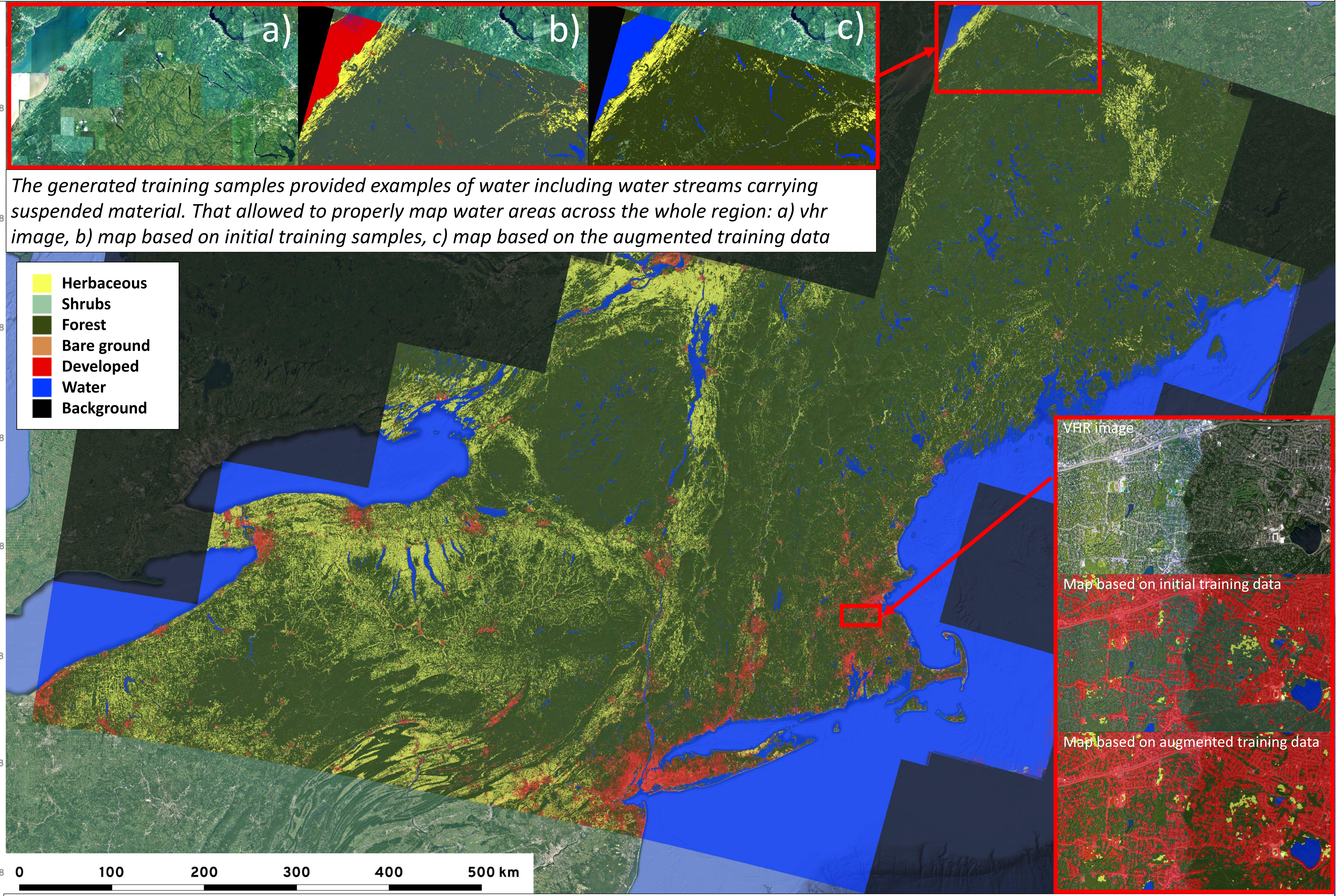


Figure 2. Examples of the modelled patterns of surface reflectance for various types of land cover occurring in the study area



The generated training samples provided examples of water including water streams carrying suspended material. That allowed to properly map water areas across the whole region: a) vhr image, b) map based on initial training samples, c) map based on the augmented training data

Figure 3. The map of land cover based on the final augmented set of training samples. The best improvements are visible in the areas where the land cover is mixed. In example, the sub-urban type of development is often misclassified as forest due to high presence of tree vegetation. The training samples provided by our method considerably improved the delineation of the this type of development and thus its representation in our map.

Conclusions

- Our approach efficiently locates the optimal training samples that represents the varieties of the land cover present in the mapped area (Fig. 2)
- The addition of the generated samples to the initial training data set improved the land cover classification especially in the areas representing mixed land cover e.g. suburbs (Fig. 3)
- The selection of the samples is based purely on statistical analysis and captures the variations of land cover that can be easily overlooked during visual interpretation

Contact info

Konrad Turlej – Postdoctoral Associate
Boston University - Department of Earth and Environment
E-mail: kturlej@bu.edu

Acknowledgements:

The work was supported by NASA MEaSURES grant # 80NSSC18K0994 P00001
"A Moderate Spatial Resolution Data Record of 21st Century Global Land Cover, Land Use and Land Cover Change"