Good Practices for Assessing Accuracy and Estimating Area of Land Change

Pontus Olofsson[a,*], Giles M. Foody[b], Martin Herold[c], Stephen V. Stehman[d], Curtis E. Woodcock[a] and Michael A. Wulder[e]

a. *Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, 02215 Boston, MA, USA*

b. *School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

c. *Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 Wageningen, The Netherlands*

d. *Department of Forest and Natural Resources Management, State University of New York, 1 Forestry Drive, Syracuse, NY 13210, USA*

e. *Canadian Forest Service, Natural Resources Canada, Victoria, BC, 12 V8Z 1M5, Canada*

*Corresponding author
Email olofsson@bu.edu
Tel +1-617-353-9734

- Aim to provide practitioners with hands-on recommendations for designing accuracy assessments of change maps, and how to use accuracy information for estimating area and uncertainty
- “Good” (not “best”) practices provided – recommendations satisfy accepted scientific practice
- Practitioners can point to us if assessment designs questioned (i.e., aim to transfer some of the responsibility from practitioners to us)
- Recommendations for all steps of the process are given including (I) sampling design, (II) response design and (III) analysis; (IV) example provided to illustrate recommended workflow

- Draft in final stage of in-house review
- Aim to submit paper to peer-reviewed journal this or the following week

- Additional paper that provides clear examples:

Olofsson, P., Foody, G.M., Stehman, S.V., and C.E. Woodcock, 2013. Making better use of accuracy data when mapping land change. *Remote Sensing of Environment*, 129:122-131.

I. Sampling Design: Good Practices

- Implement a probability sampling design to provide a rigorous foundation to sampling inference
- Document and quantify deviations from probability sampling design
- Choose a sampling design on the basis of specified accuracy objectives and desirable design criteria – both simple random and systemic selection protocols are acceptable options
- Stratify by map class to reduce standard errors of class-specific accuracy estimates
- Use cluster sampling only if it provides a substantial cost savings or if the objectives require a cluster unit for the assessment
- Evaluate the total sample size using the standard error formula and a specified target standard error

- Allocate sample to strata such that sample size for rare change classes are increased (~100) to achieve an acceptable standard error for estimated user’s accuracies and allocate the remaining sample size roughly proportional to the area occupied by the common classes
- *Stratified random sampling* using the map classification to define strata is a simple, but generally applicable design that will typically satisfy most accuracy and area estimation objectives and desirable design criteria

II. Response Design: Good Practices

- The overhead cost required for field visits is likely not justified
- The reference data should provide sufficient temporal representation consistent with the change period of the map
- Reference data should be of higher quality than the data used for creating the map, or if using the same source, the process of creating the reference classification should be more accurate than the process of creating the map
- Data from the Landsat open archive in combination with high spatial resolution imagery provide a low-cost and often useful source of reference data (national photo archives, satellite photo archives (e.g., Kompsat), and the collections available through Google Earth™ are possible high resolution imagery sources)
- Specify protocols for accounting for uncertainty in assigning the reference classifications
- Assign each sample unit a primary and secondary label (secondary not required if highly confident)

- Include an interpreter specified confidence for each reference label (e.g., high, medium, or low confidence)
- Ideally have three interpreters label each unit sampled, and implement protocols to ensure consistency among individual interpreters or teams of interpreters
- Specify a protocol for defining agreement between the map and reference classifications that will lead to an error matrix expressed in terms of proportion of area

III. Analysis: Good Practices

- Report the error matrix in terms of estimated area proportions
- Report the area (or proportion of area) of each class as determined from the map
- Report user’s accuracy (or commission error), producer’s accuracy (or omission error), and overall accuracy (Equations 1-3)
- Avoid use of the kappa coefficient of agreement for reporting accuracy of land change maps
- Estimate the area of each class according to the classification determined from the reference data
- Use estimators of accuracy and area that are unbiased or consistent
- For simple random, systematic, and stratified random sampling when the map classes are defined as strata, use stratified estimators of accuracy and a stratified estimator of area

- Quantify sampling variability of the accuracy and area estimates by reporting standard errors or confidence intervals
- Use design-based inference to define estimator properties and to quantify uncertainty
- Assess the impact of reference data uncertainty on the accuracy and area estimates

Where are we and is there a next step?

Basically done at this stage – minor edits and submission, followed by likely requests for revision

Are there other issues that would benefit?

One possibility in the REDD+ context might be some standard change products

Designing a Global Reference Database for Assessing Accuracy of Land Cover Maps

Pontus Olofsson, Curtis E. Woodcock, Christopher Holden, Mark A. Friedl and Damien Sulla-Menashe
Dept of Geography and Environment, Boston University
675 Commonwealth Avenue, Boston, MA 02215, USA

Stephen V. Stehman
Dept of Forest and Natural Resources Management, State University of New York, 1 Forestry Drive, Syracuse, NY 13210, USA

Martin Herold
Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3 , Wageningen, The Netherlands

Chandra Giri and Tom Loveland
USGS

Corresponding authors
olofsson@bu.edu
curtis@bu.edu
svstehma@syr.edu
martin.herold@wur.nl

GOFC-GOLD
Global Observation of Forest and Land Cover Dynamics

Goals

- Provide reference data that could be used to help evaluate the accuracy of any LCCS compliant land cover maps (and probably others as well!)
- Provide an accuracy assessment framework that can be augmented regionally or thematically
- Promote the importance of validation activities
- Ultimately, facilitate the process of improving efforts to develop synthetic land cover maps (using multiple land cover maps to produce a new single “improved” land cover map)

Sample Design

- Applicable to assess accuracy for a variety of land-cover maps - independent of existing land-cover maps
- Ability to increase sample size for rare land-cover types
- Ability to enlarge (augment) sample size in targeted regions or land-cover classes
- A random stratified sampling meets these criteria
- Each sample is a 5x5 km block

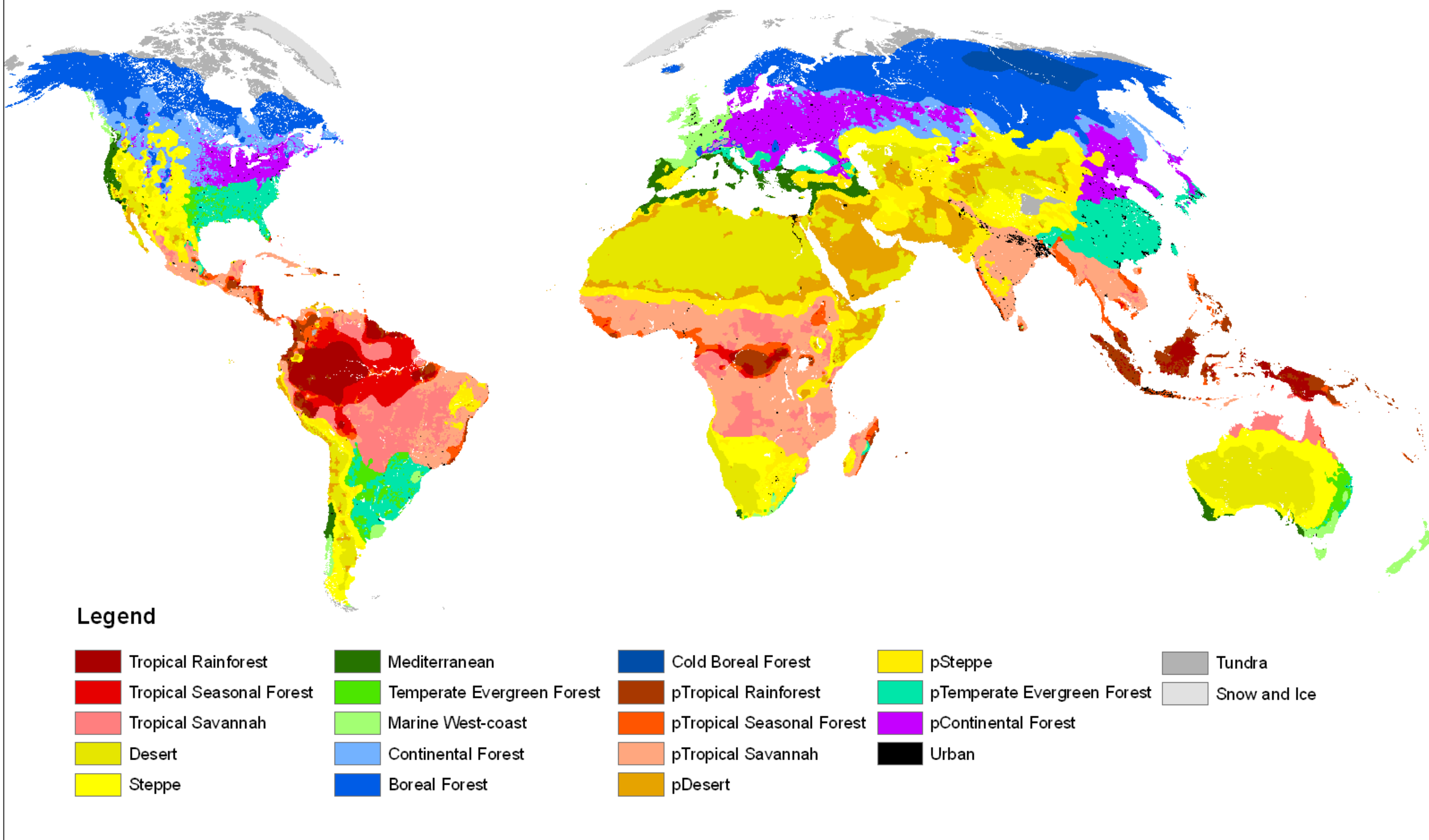
Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J.D., Friedl, M.A. and Herold, M., 2012. A global land cover validation dataset, I: Fundamental design principles. *International Journal of Remote Sensing*. 13(18):5768-5788.

Stehman, S. V., Olofsson, P., Woodcock, C. E., Herold, M., Friedl, M. A., 2012. A global land cover validation dataset, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing*. (33)12:6975-6993.

The Köppen climate system and population data – basis of the stratification:

- Köppen map based on station data – manual editing of climatic borders
- Original Köppen map (Peel et al., 2007) – 32 classes – collapsed to 13 classes
- Intersected by population data: 5 persons/km^2 => 26 populated and unpopulated classes
- Collapsed to 20 classes and areas of > 1000 persons/km^2 as a 21st class (built up)
- Areas with water removed by intersection of land/water mask

Strata



Legend
Tropical Rainforest
Tropical Seasonal Forest
Tropical Savannah
Desert
Steppe
Mediterranean
Temperate Evergreen Forest
Marine West-coast
Continental Forest
Boreal Forest
Cold Boreal Forest
pTropical Rainforest
pTropical Seasonal Forest
pTropical Savannah
pDesert
pSteppe
pTemperate Evergreen Forest
pContinental Forest
Urban
Tundra
Snow and Ice

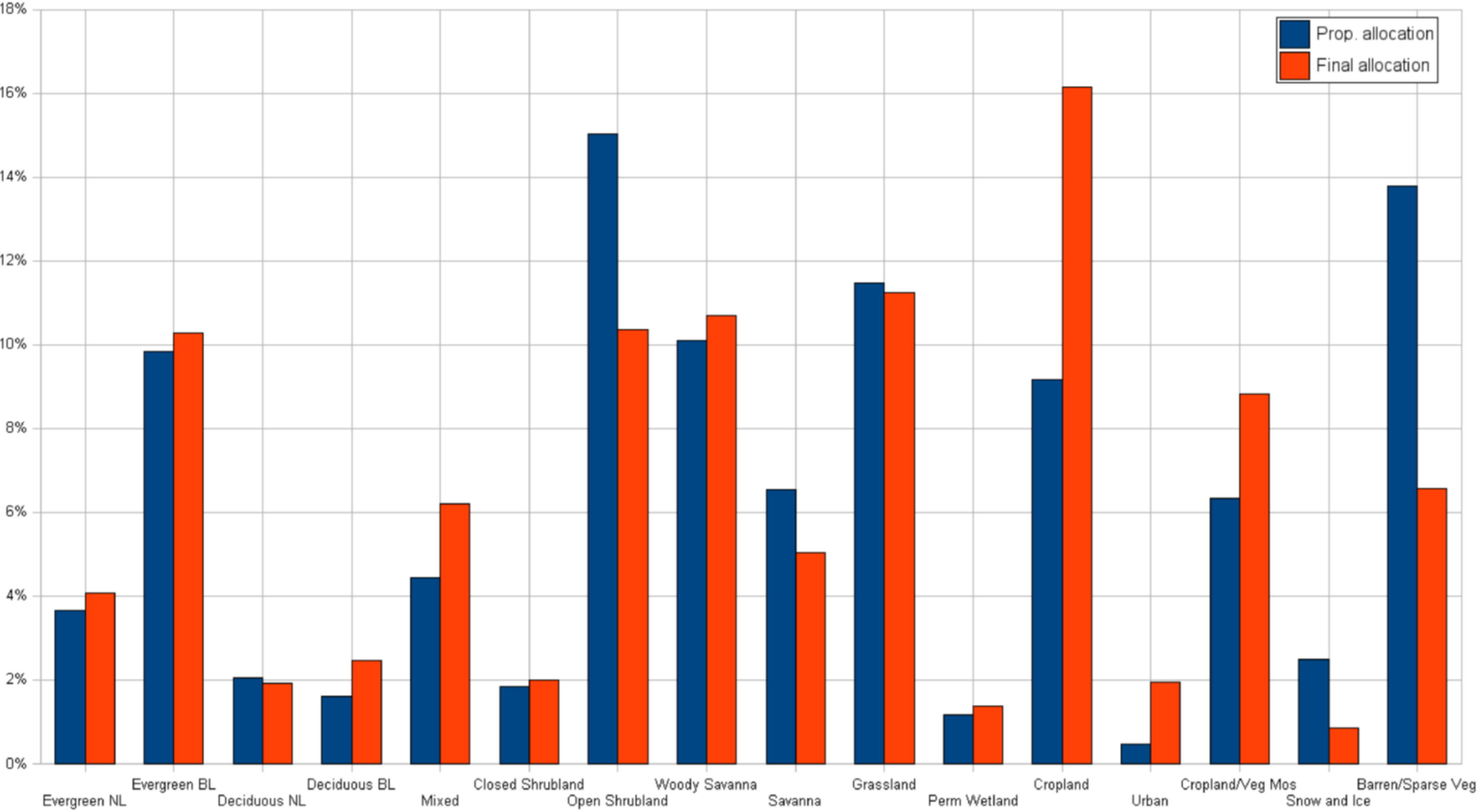
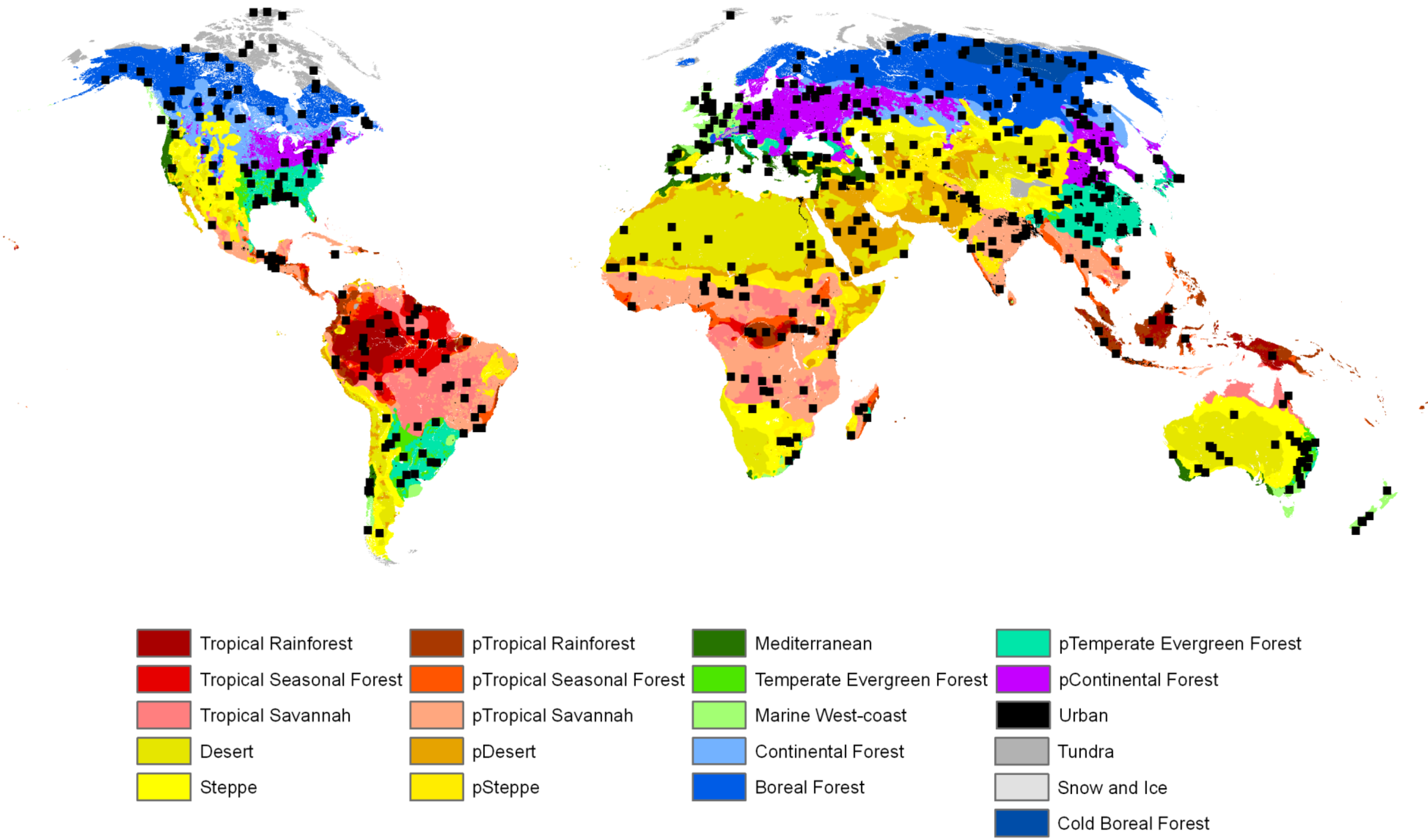

18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
Prop. allocation
Final allocation
Evergreen NL
Evergreen BL
Deciduous NL
Deciduous BL
Mixed
Closed Shrubland
Open Shrubland
Woody Savanna
Savanna
Grassland
Perm Wetland
Cropland
Urban
Cropland/Veg Mos
Snow and Ice
Barren/Sparse Veg

Table 3: The final strata, its global distribution and allocation. A "p" in front of the strata name denotes populated version.

No.	Strata	Distr.	Prop. all.	Final all.
1	Tropical rainforest	2.4%	12	10
2	Tropical seasonal forest	2.0%	10	10
3	Savannah	5.0%	25	15
4	Desert	14.4%	72	20
5	Steppe	8.3%	41	20
6	Mediterranean	1.6%	8	25
7	Temperate evergreen forest	1.2%	6	25
8	Marine west-coast	1.6%	8	25
9	Continental forest	4.3%	22	30
10	Boreal forest	12.7%	63	50
11	Cold boreal forest	1.2%	6	10
12	Tundra	3.3%	17	10
13	Frost	1.2%	6	0
14	pTropical rainforest	2.2%	11	15
15	pTropical seasonal forest	1.9%	10	10
16	pTropical savannah	11.0%	55	40
17	pDesert	6.0%	30	25
18	pSteppe	7.0%	35	35
19	pTemperate evergreen forest	5.2%	26	40
20	pContinental forest	6.7%	34	50
21	Urban	0.6%	3	35

Locations of sample sites



Response Design

- Land cover legend and workflow for reference map production defined
- Object-based analysis of 0.5m pan-sharpened data (mostly from Digital Globe)
- Legend based on the following required classes

Tree

1. Broadleaf evergreen
2. Broadleaf deciduous herbaceous
3. Needleaf evergreen
4. Needleleaf deciduous
5. Shrub

Herbaceous

6. Row crop
7. Other

Non-vegetated land

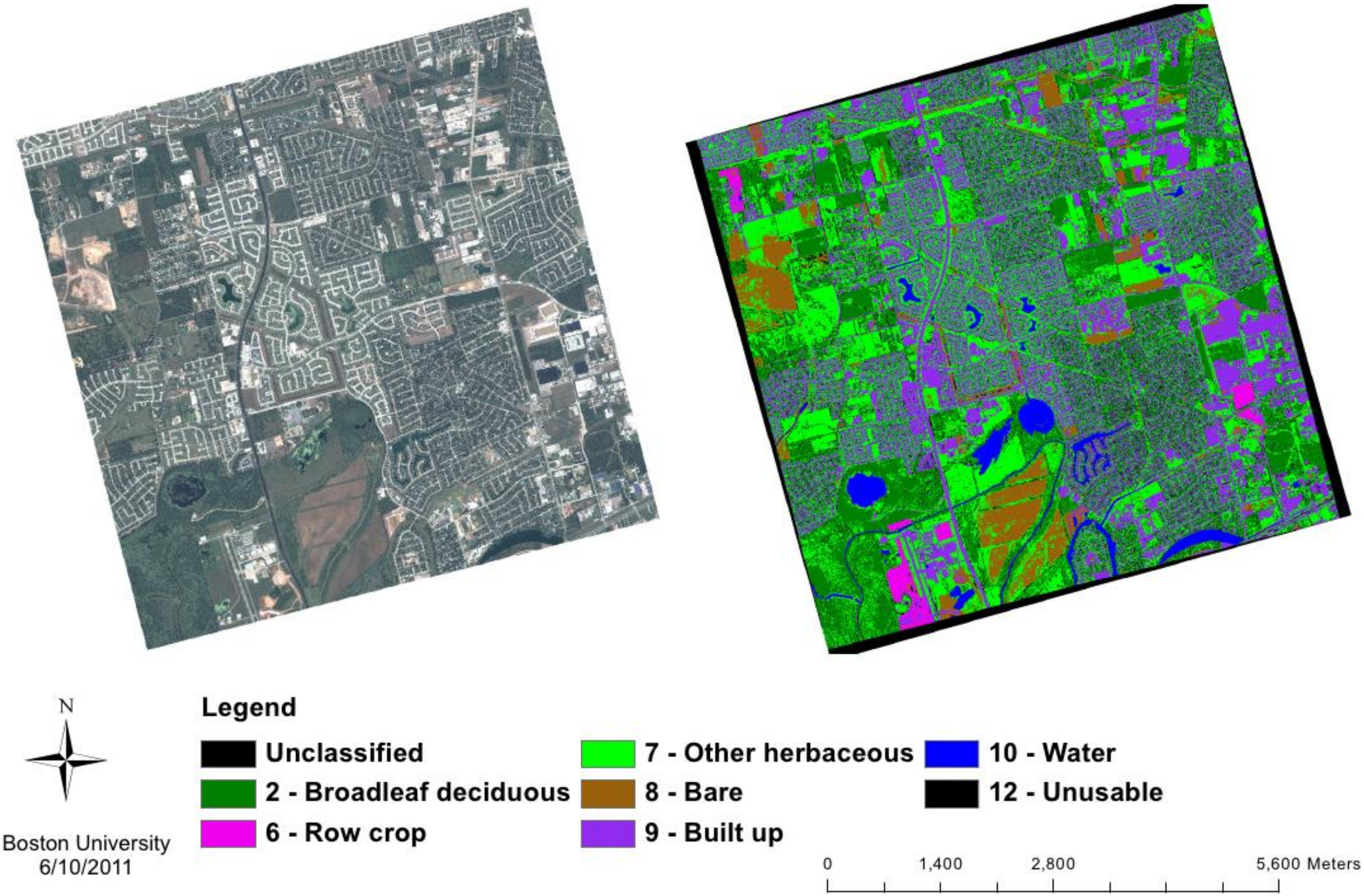
8. Bare
9. Built-up
10. Water
11. Snow/Ice

Other

12. Unusable

Site 272 - Sugar Land, TX

Scene acquired: 5-25-2005
Sensor: Quickbird



Site 272 - Sugar Land, TX

Scene acquired: 5-25-2005
Sensor: Quickbird

N

Legend

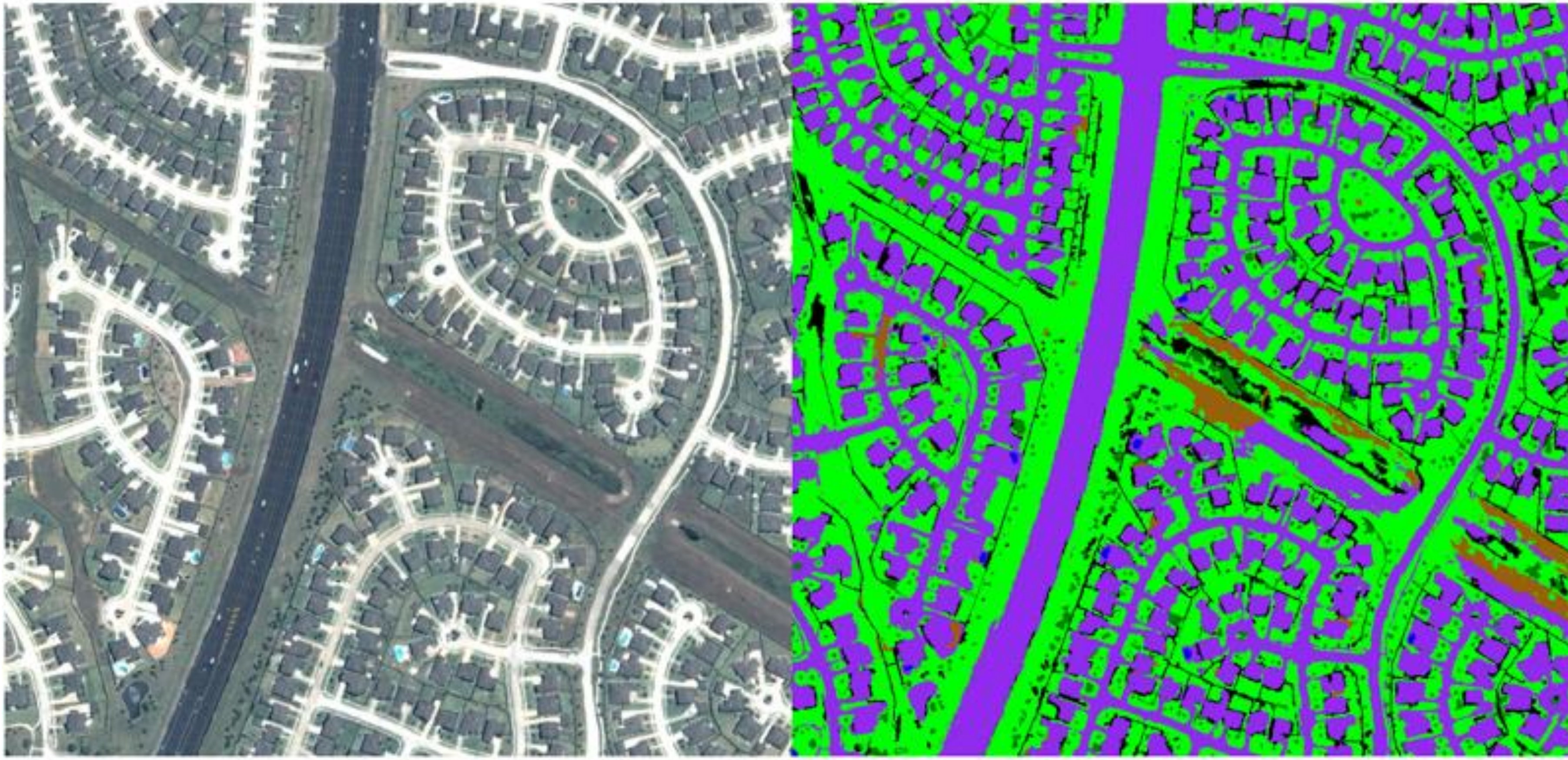
Unclassified	7 - Other herbaceous
2 - Broadleaf deciduous	8 - Bare
6 - Row crop	9 - Built up
10 - Water	12 - Unusable

Boston University
6/10/2011

0 175 350 700 Meters

Site 272 - Sugar Land, TX

Scene acquired: 5-25-2005
Sensor: Quickbird



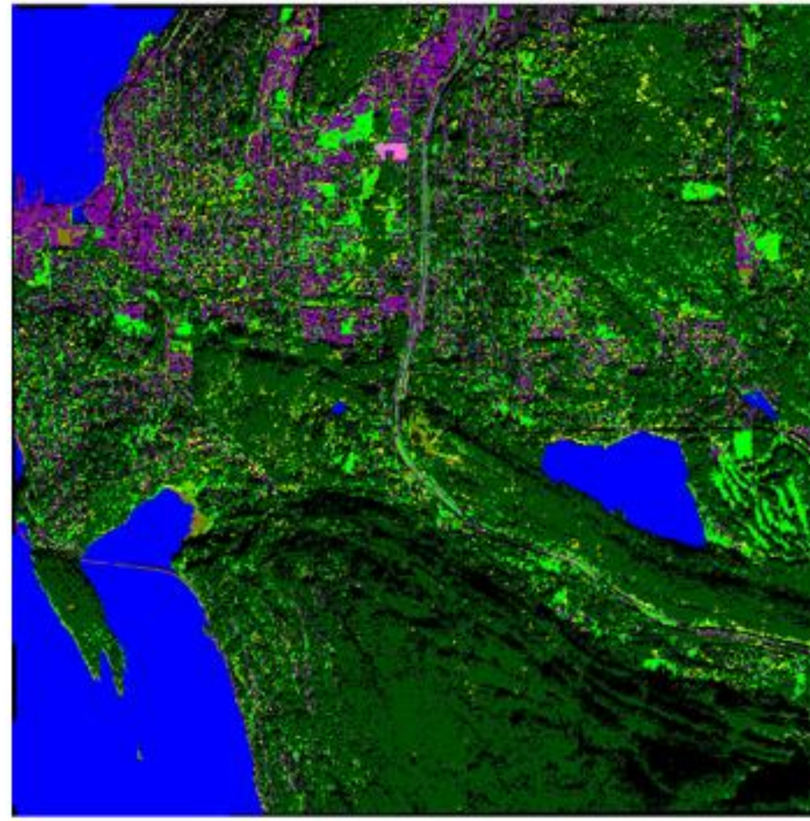
Legend

Unclassified	7 - Other herbaceous	10 - Water
2 - Broadleaf deciduous	8 - Bare	12 - Unusable
6 - Row crop	9 - Built up	

Boston University
6/10/2011

Site 128 - Bellingham, WA

Scene acquired: 10-26-2007
Sensor: Quickbird



Legend

Unclassified	6 - Row crop	10 - Water
2 - Broadleaf deciduous	7 - Other herbaceous	12 - Unusable
3 - Needleleaf evergreen	8 - Bare	
5 - Shrub	9 - Built up	

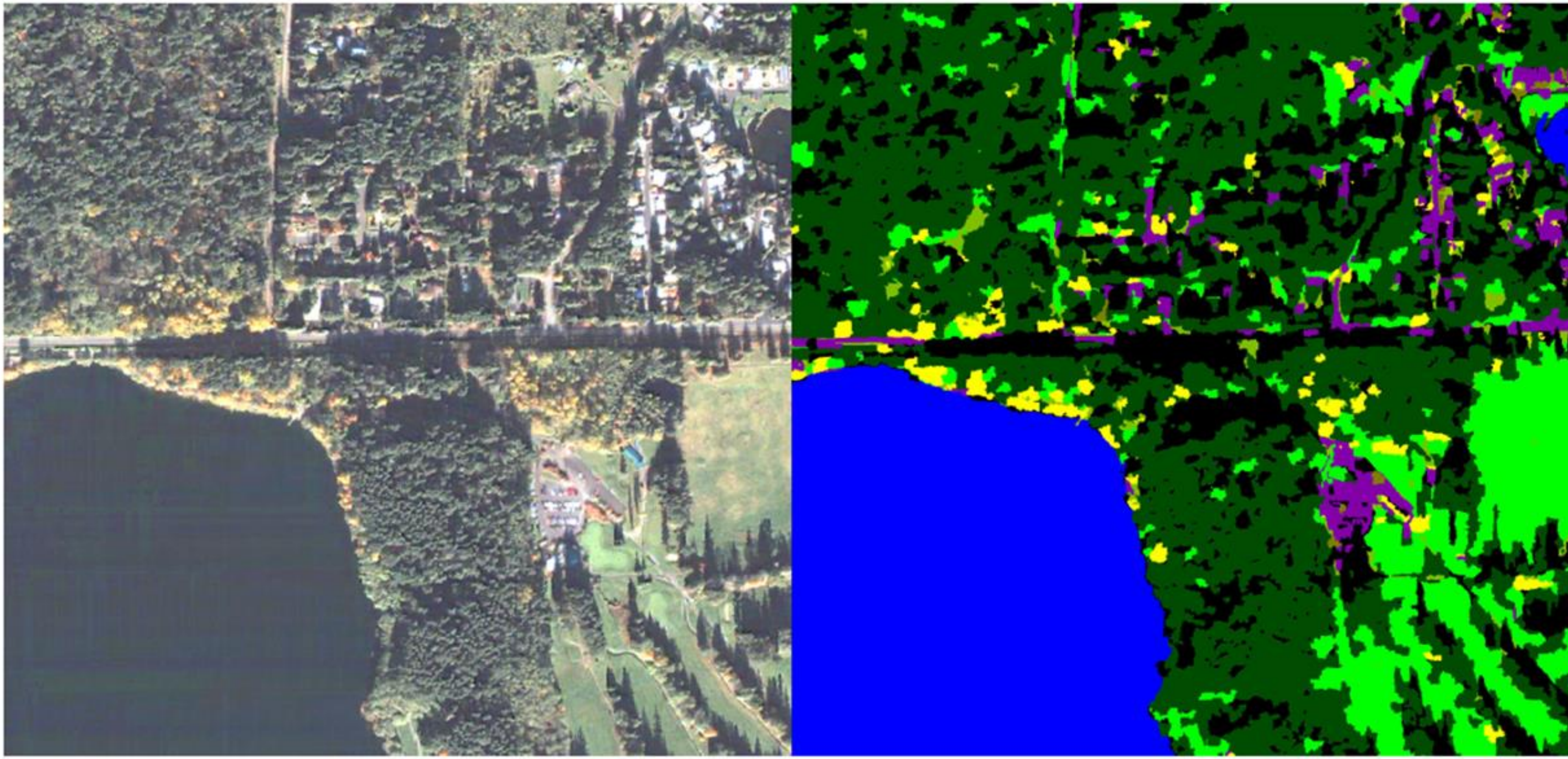
N

Boston University
6/10/2011

0 1,125 2,250 4,500 Meters

Site 128 - Bellingham, WA

Scene acquired: 10-26-2007
Sensor: Quickbird



Legend

Unclassified	6 - Row crop	10 - Water
2 - Broadleaf deciduous	7 - Other herbaceous	12 - Unusable
3 - Needleleaf evergreen	8 - Bare	
5 - Shrub	9 - Built up	

N

0 150 300 600 Meters

Boston University
6/10/2011

USGS Effort (Chandra Giri, Tom Loveland et al.,)

Simplified response design: fewer classes (forest, other veg, water, bare)

Classification of the 2.5m multispectral data

225 sites done (all of western hemisphere as well as a 98 site representative sample!

Current status

- Sampling and response design defined – sampling designed published
- A prioritized subset of 98 images has been identified
- 33 reference maps created as of March 2013

Next Steps

- Find funding!
- Recruit more help for map production
- Write paper on response design
- Continue working on the analysis design
- Upcoming workshop at Wageningen University this summer (Nandika)