Overview of MODIS-based mapping of NELDA Land Cover and Approaches to its Validation

Alessandro Baccini, Mark Friedl, Curtis Woodcock

abaccini@bu.edu

Department of Geography

Center for Remote Sensing

Boston University

http://geography.bu.edu/landcover/

MODIS Regional Land Cover

Objectives:

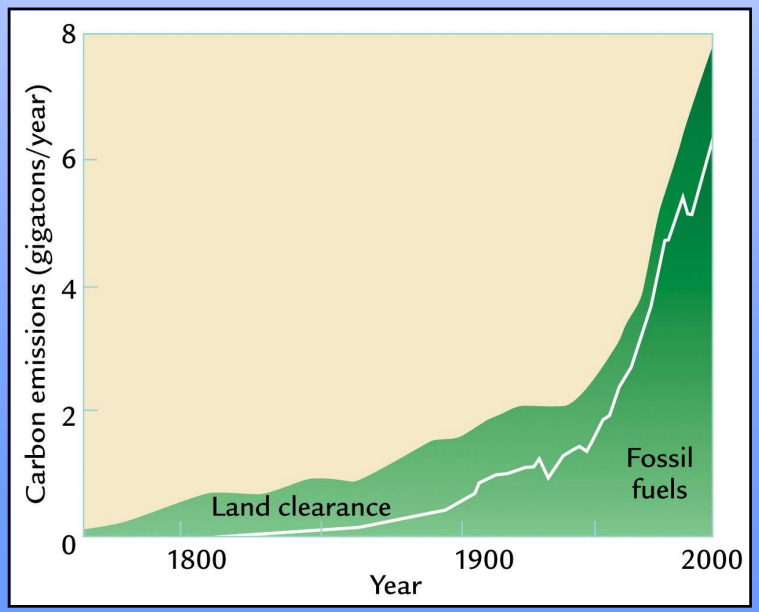
- *Combine regional expertise and existing products in the mapping procedure*
- *Improve and update (circa 2005) of Northern Euroasia land cover characterization*
- *Develop a new legend consistent with FAO LCCS*

What is Land Cover?

- Generalized classification of the biophysical conditions at the Earth's land surface
- *Three key dimensions*
  - Natural vegetation
  - Barren and unvegetated land areas
  - Developed/Human modified land areas

Why Land Cover?

- Global Change Perspective
  - Land conversion and land use by humans represent the largest single mechanism of environmental change
    - Carbon storage/release
    - Biodiversity
    - Land resources & food security
    - Hydrology and water resources
    - Etc……



Outline

- ***Introduction and Context***
- ***MODIS Land Cover Mapping***
  - Description of data sets
  - Classification methods
    - Post-Processing
  - “Validation”
- New Legend LCCS compliant

MOD12Q1: What Is It?

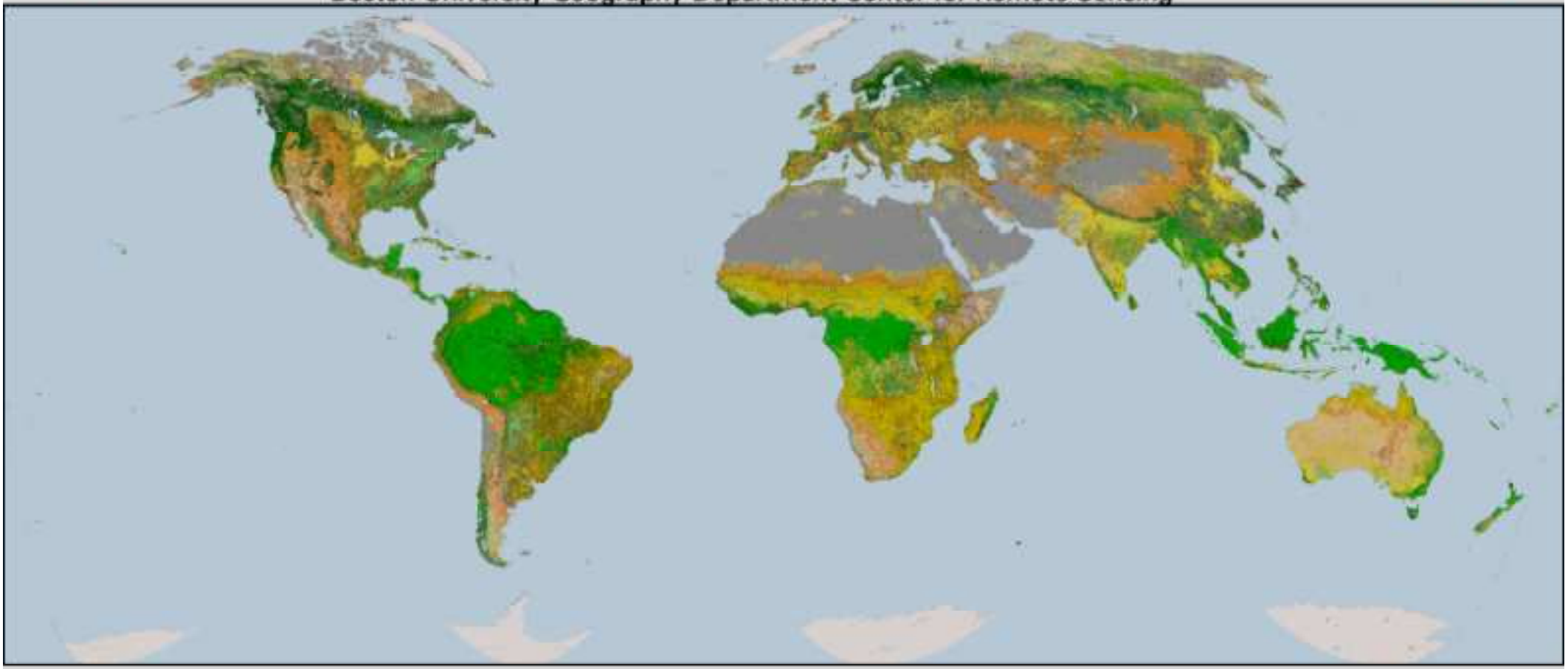
- Land Cover Types
  - IGBP, UMD, LAI/FPAR, BGC, CLM
  - 1 km
- Confidences
  - Classification confidence (percent scale) for each pixel
- Secondary IGBP Label
  - For IGBP, a secondary class label for each pixel

IGBP Land Cover Units (17)

(Primary Layer)

- Natural Vegetation (11)
  - Evergreen Needleleaf Forests
  - Evergreen Broadleaf Forests
  - Deciduous Needleleaf Forests
  - Deciduous Broadleaf Forests
  - Mixed Forests
  - Closed Shrublands
  - Open Shrublands
  - Woody Savannas
  - Savannas
  - Grasslands
  - Permanent Wetlands
- Developed and Mosaic Lands (3)
  - Croplands
  - Urban and Built-Up Lands
  - Cropland/Natural Vegetation Mosaics
- Nonvegetated Lands (3)
  - Snow and Ice
  - Barren
  - Water Bodies

NASA EOS MOD12Q1 2001001 V004 SDS01
International Geosphere-Biosphere Programme Land Cover Classes
Derived from MODIS Satellite Imagery Acquired 1/1/2001 - 12/31/2001
Boston University Geography Department Center for Remote Sensing



- 0 Water
- 1 Evergreen Needleleaf Forest
- 2 Evergreen Broadleaf Forest
- 3 Deciduous Needleleaf Forest
- 4 Deciduous Broadleaf Forest
- 5 Mixed Forests
- 6 Closed Shrublands
- 7 Open Shrublands
- 8 Woody Savannas
- 9 Savannas
- 10 Grasslands
- 11 Permanent Wetlands
- 12 Croplands
- 13 Urban and Built-Up
- 14 Cropland/Natural Vegetation Mosaic
- 15 Snow and Ice
- 16 Barren or Sparsely Vegetated
- 254 Unclassified

NASA

EOS

Global Land Cover Classification Methods

Three main components

1. Exploits spectral and temporal information from MODIS
2. Robust, repeatable classification algorithm
3. Requires extensive, high quality training site data base (STEP)

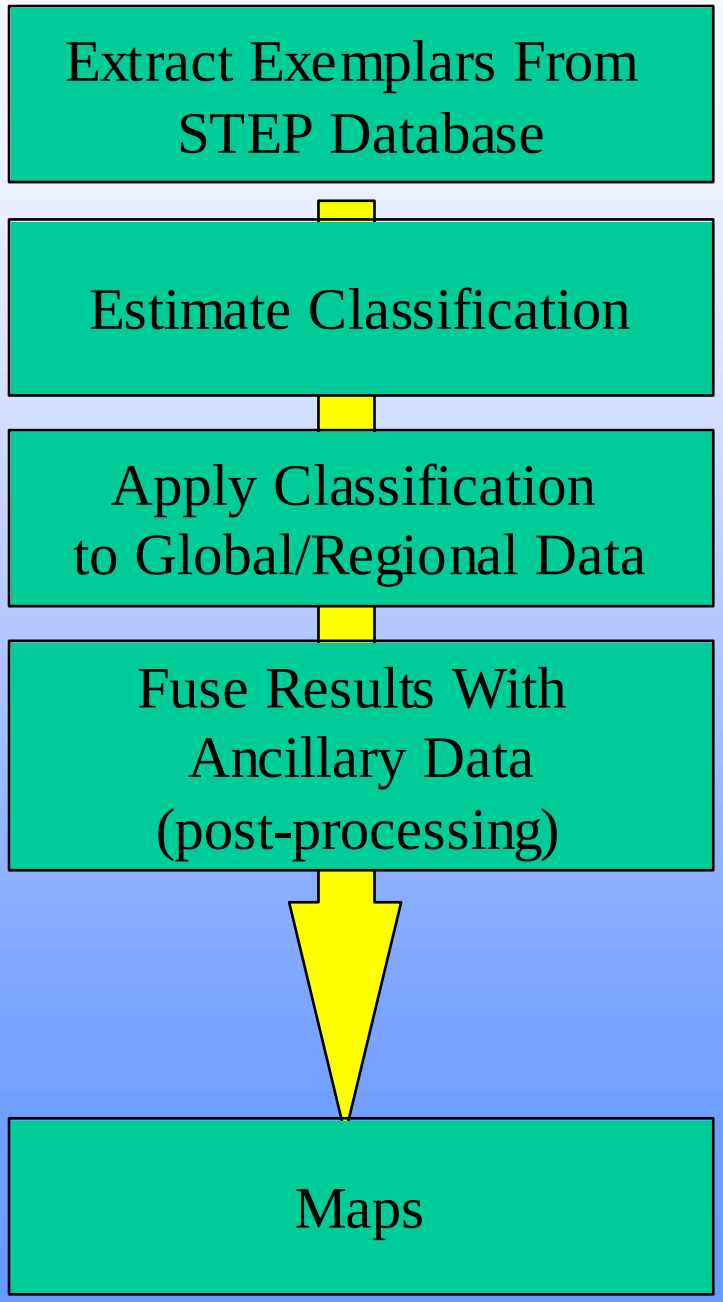
Data

- MODIS Data
  - 32-day Normalized BRDF-Adjusted Reflectances (NBARs) assembled over one year of observations
  - 7 spectral bands, 0.4–2.1 μm, similar to Landsat
  - 32-day Enhanced Vegetation Index (EVI)
- Training Data
  - 2130 training sites delineated from high resolution satellite imagery (largely Landsat)

Inputs and Classification Flow

(Friedl et al. 2002; RSE)

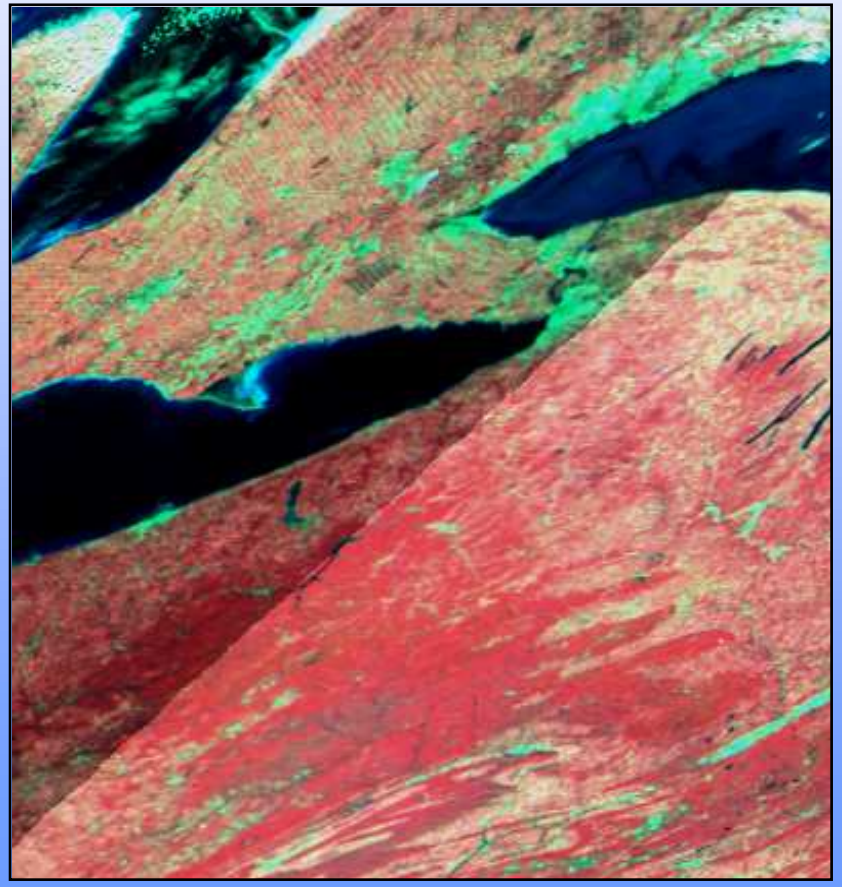
- Features From MODIS:
  - *Temporal and spectral information*
  - *12 (annual) 32-day composites*
- Surface Reflectance (NBAR)
  - *View-angle corrected surface reflectance*
  - *7 land bands*
- Enhanced Vegetation Index (EVI)
  - *Computed from NBARs*
- Annual Metrics
  - *Min, max, mean for each band*

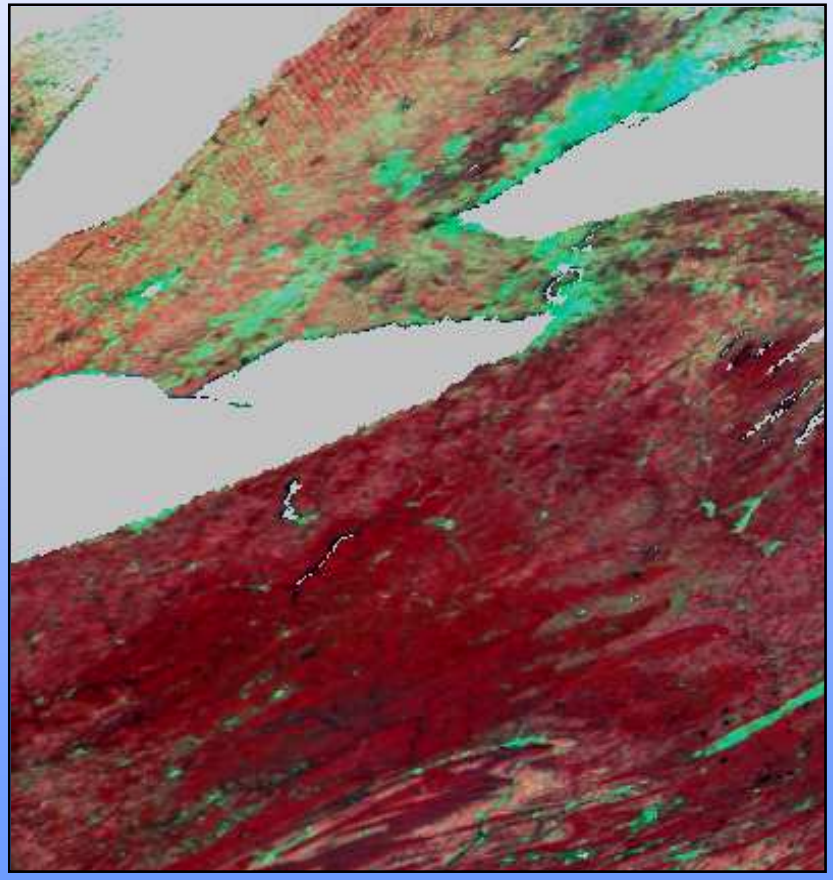


Key Input Used for Classification: NADIR, BRDF-Adjusted Reflectance

(Schaaf et al., 2002; RSE)

Removes artifacts associated with variable view geometry





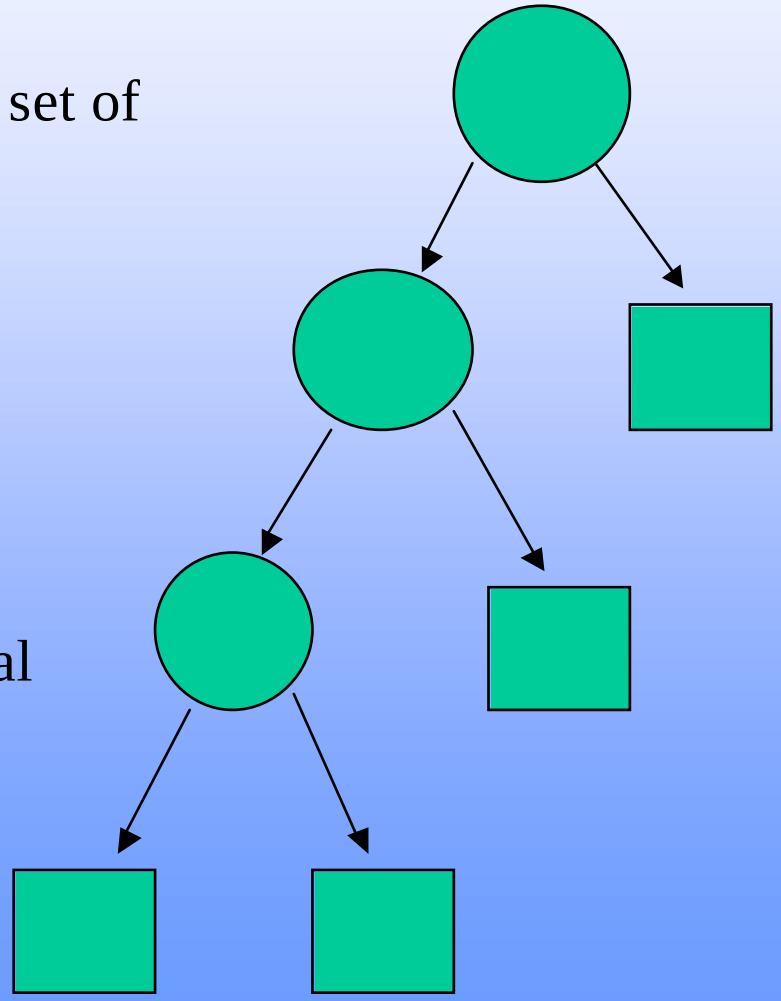
Classification Algorithm

- Decision Tree
  - ***C4.5: Univariate Decision Tree***
  - ***Nonparametric***
  - ***Boosting***
- Provides robust, repeatable results
- Relies heavily on input training database

Decision Tree Classification

(Friedl and Brodley, 1997; RSE)

- Goal:
  - Optimal prediction of class labels from a set of feature values
- Basic approach
  - Supervised learning using training data
- Key attributes:
  - Nonparametric
  - Able to handle noisy or missing features
  - Adept at capturing non-linear, hierarchical patterns



Optimizing Classification: Boosting

(McIver and Friedl, IEEE TGARS 2001)

- Estimate multiple trees
  - At each iteration, re-weight sample to focus on difficult cases
- Final classification
  - Accuracy weighted vote across multiple trees

Basic Algorithm

1. Initialize $w(i)^t=1/N$
2. At each iteration:
   1. $\varepsilon^t = \sum w(i)$ for incorrect predictions
   2. $w(i)^{t+1} = w^t(i)\ \varepsilon^t / (1-\varepsilon^t)$
3. Re-estimate tree
4. Weight for each tree
   - $B = \varepsilon^t / (1-\varepsilon^t)$

- Where $w(i)^t$ = weight for the i^{th} case in iteration t, and N is the total number of cases

Post-Classification Processing

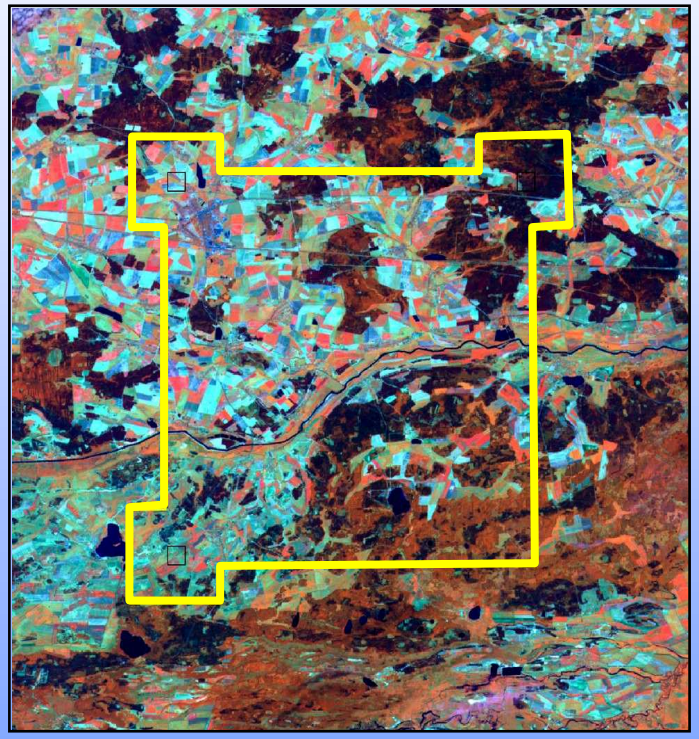
(McIver and Friedl 2002, RSE)

- Application of Prior Probabilities
  - Global priors to remove training site class distribution biases
  - Moving-window priors from earlier products
  - Use of external maps of prior probabilities to resolve confusions
    - Agriculture/natural vegetation confusion in some regions
    - Use of city lights DMSP data to enhance urban class accuracy
- Filling of Cloud-Covered Pixels from Earlier Maps
  - Use of previous year product when there are not sufficient values to classify a pixel with confidence

Training Sites—STEP Database

(Muchoney et al., 1999; PERS)

- STEP:
  - System for Terrestrial Ecosystem Parameterization
  - Interpreted from Landsat & ancillary data
- Key STEP Parameters
  - Life form, cover fraction, leaf type, phenology, elevation, moisture regime, disturbance
  - Simple description of site and type



A confidence site near Pinsk, Belarus (20 x 20 km)

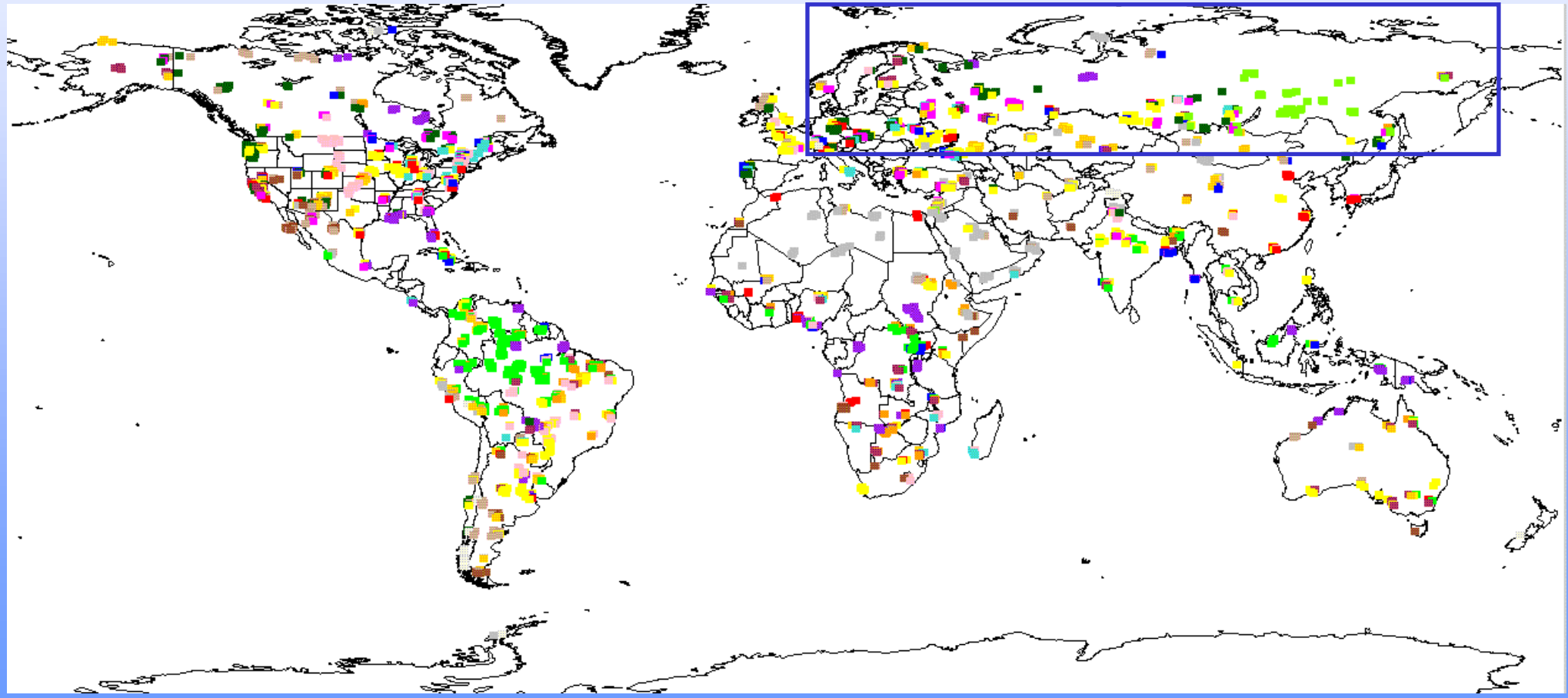
IGBP Land Cover Units (17)

(Primary Layer)

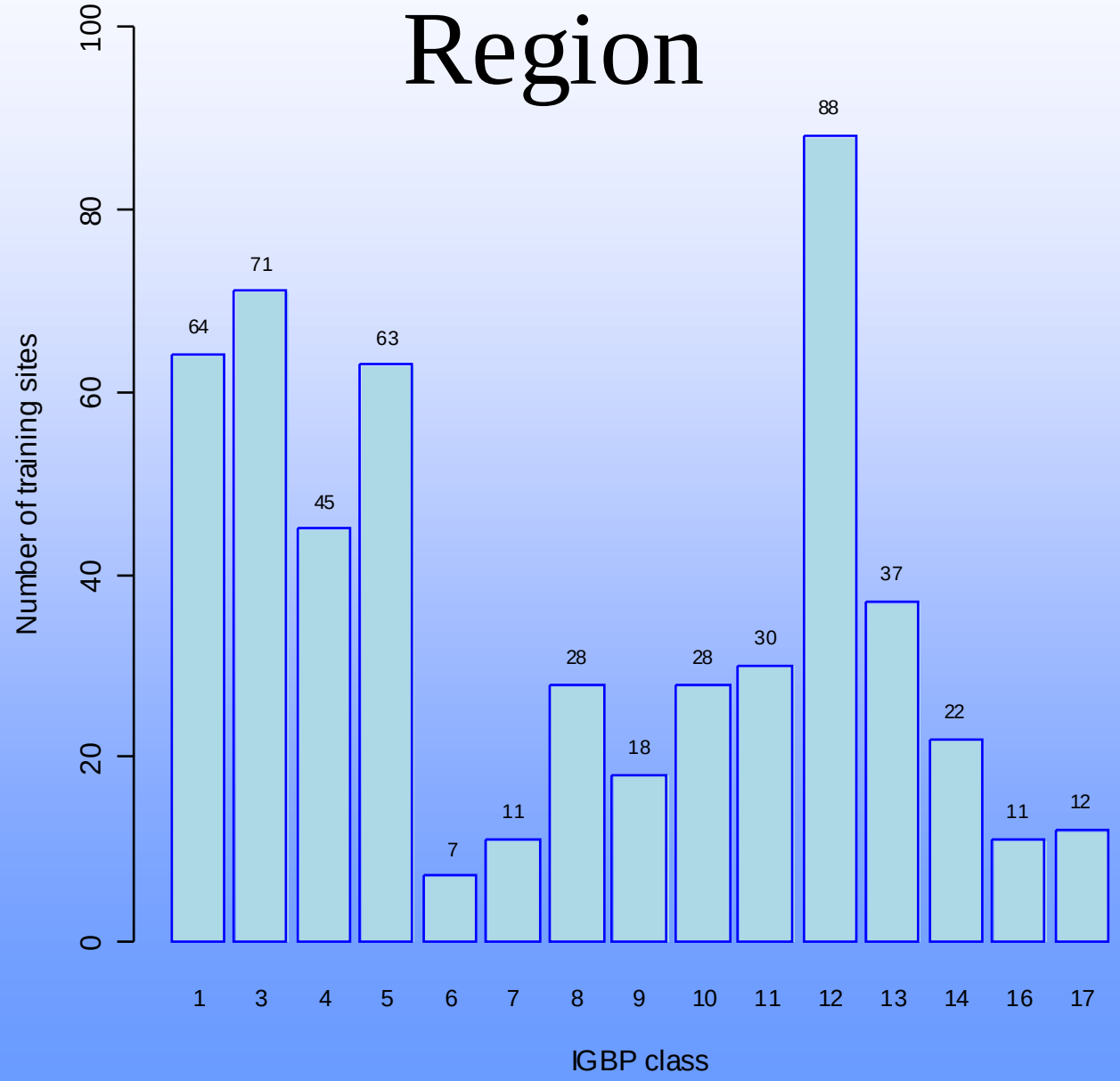
- Natural Vegetation (11)
  - Evergreen Needleleaf Forests
  - Evergreen Broadleaf Forests
  - Deciduous Needleleaf Forests
  - Deciduous Broadleaf Forests
  - Mixed Forests
  - Closed Shrublands
  - Open Shrublands
  - Woody Savannas
  - Savannas
  - Grasslands
  - Permanent Wetlands
- Developed and Mosaic Lands (3)
  - Croplands
  - Urban and Built-Up Lands
  - Cropland/Natural Vegetation Mosaics
- Nonvegetated Lands (3)
  - Snow and Ice
  - Barren
  - Water Bodies

Global Sampling and STEP Maintenance

- Live (!!) Database: currently ~2300 sites globally

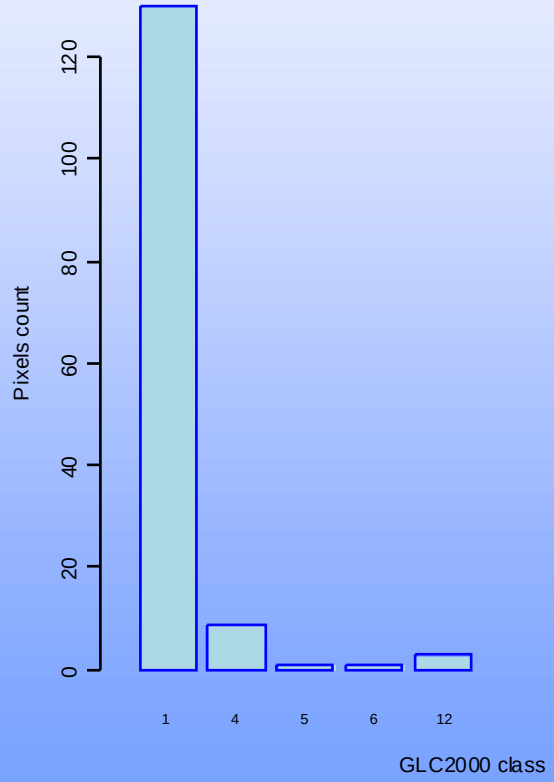


STEP Training Sites in Nelda Region

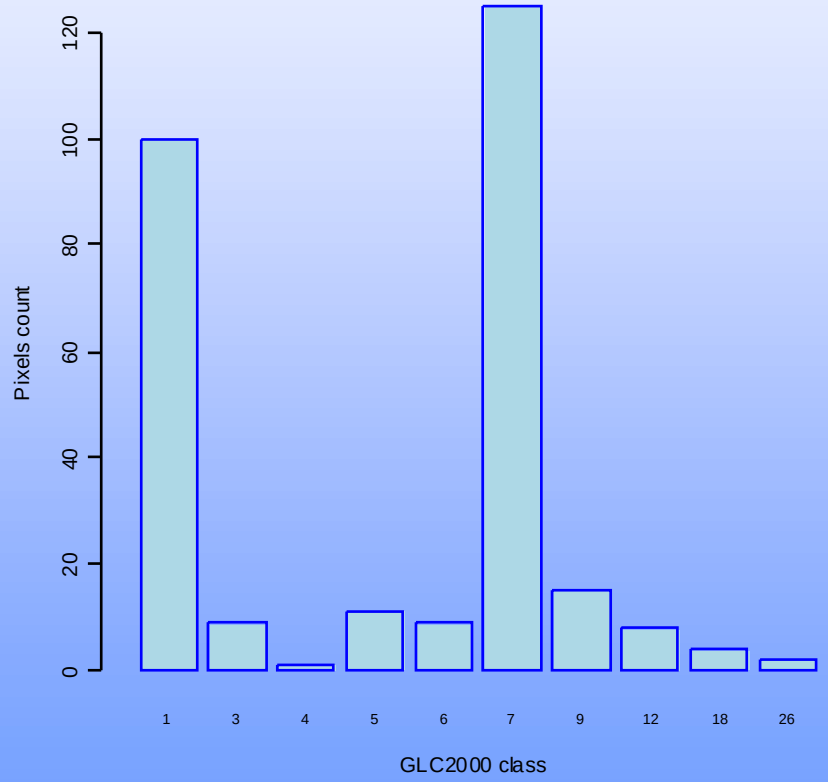


IGBP site label and GLC2000

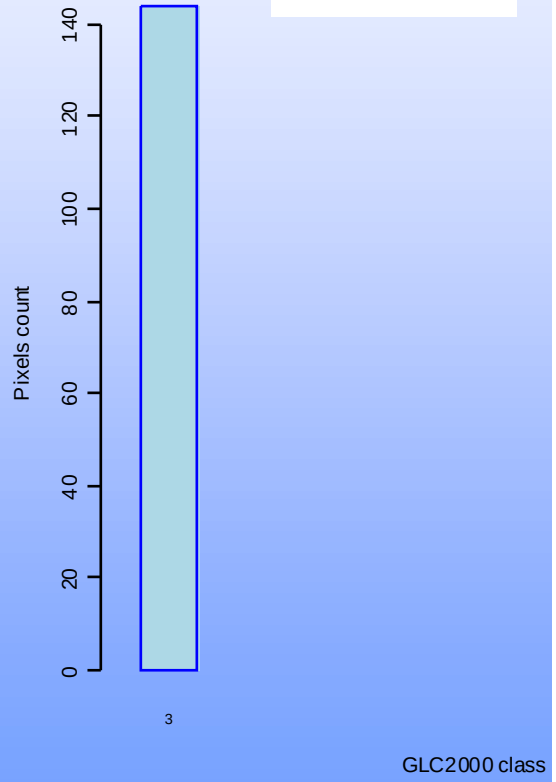
EGN 1



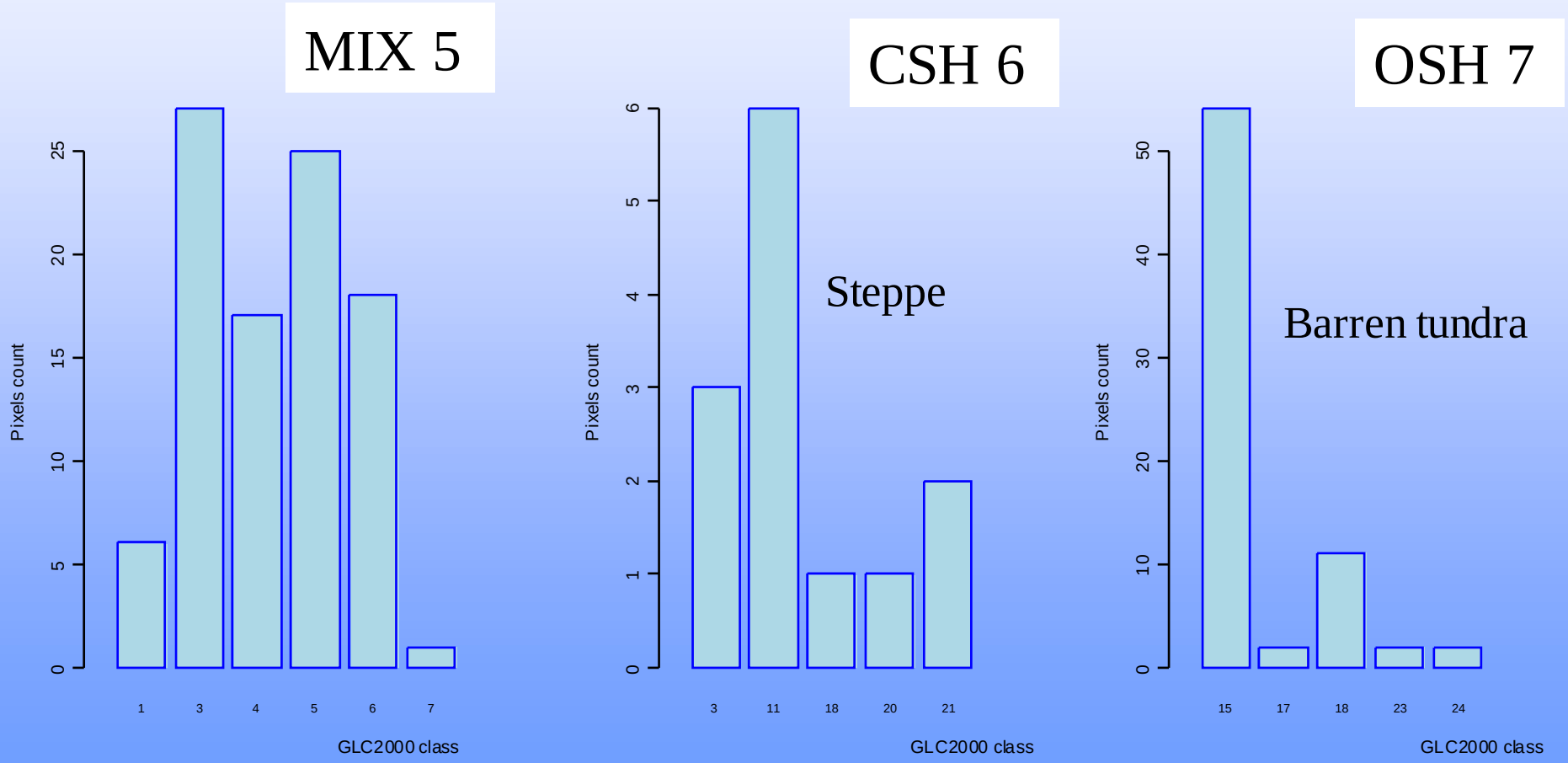
DEN 3



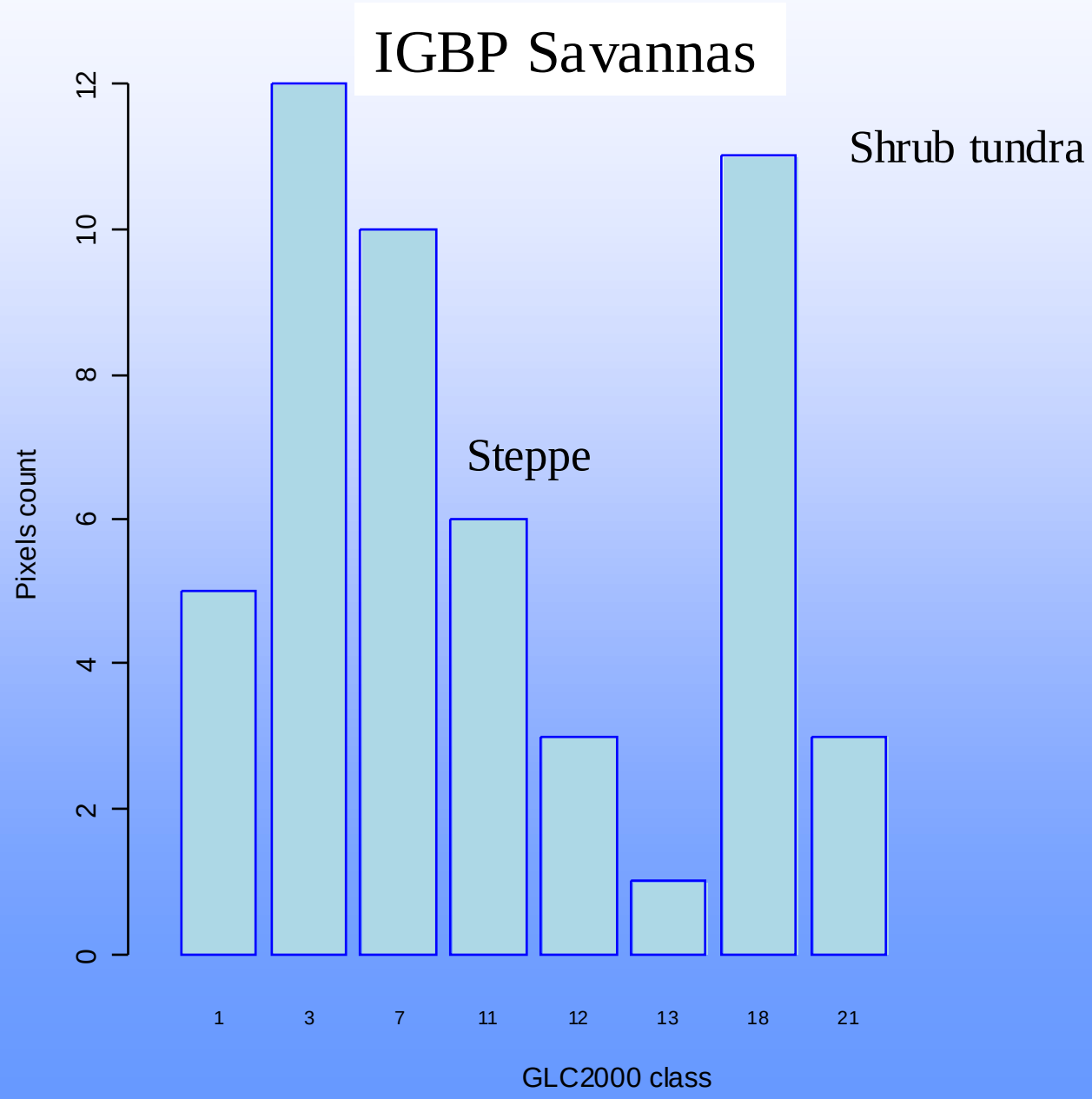
DEB 4



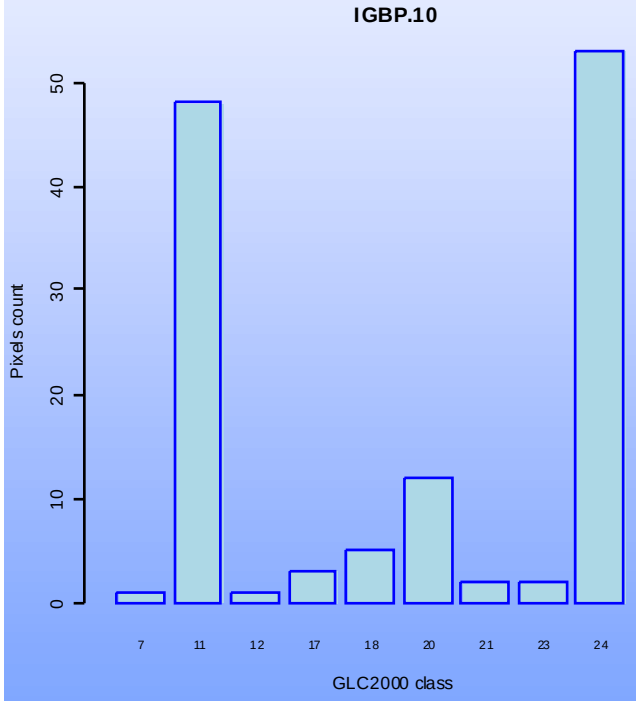
IGBP site label and GLC2000

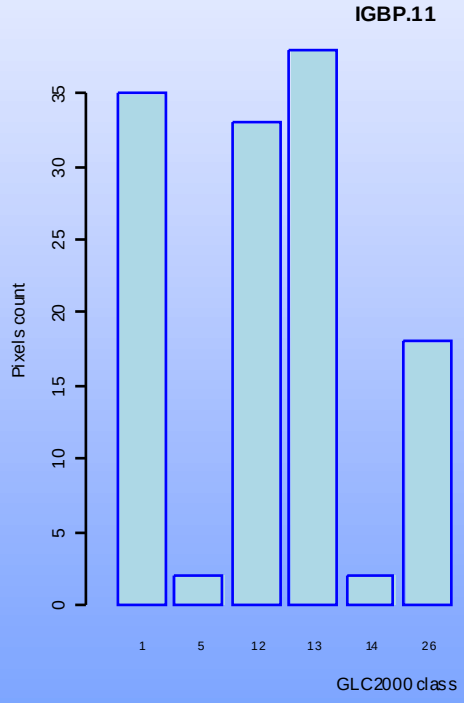


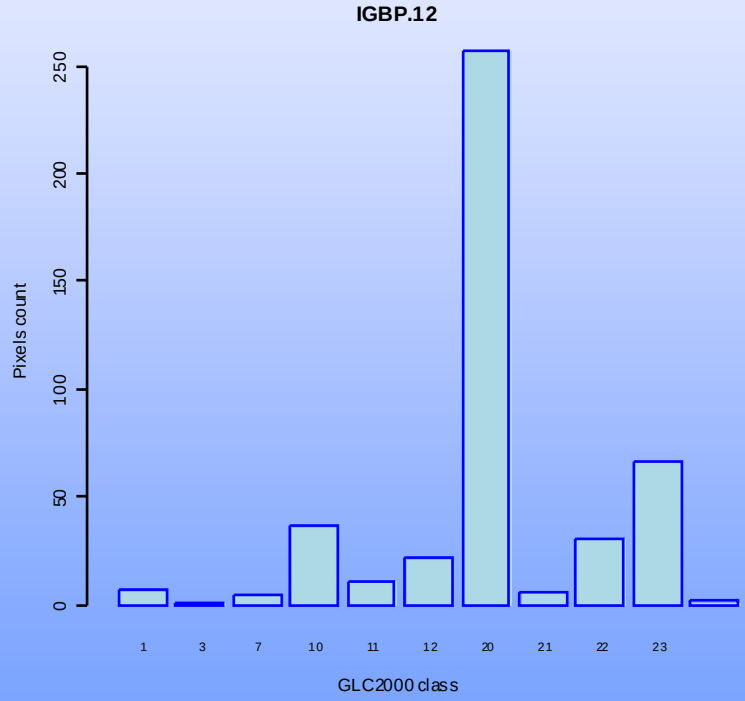
IGBP site label and GLC2000



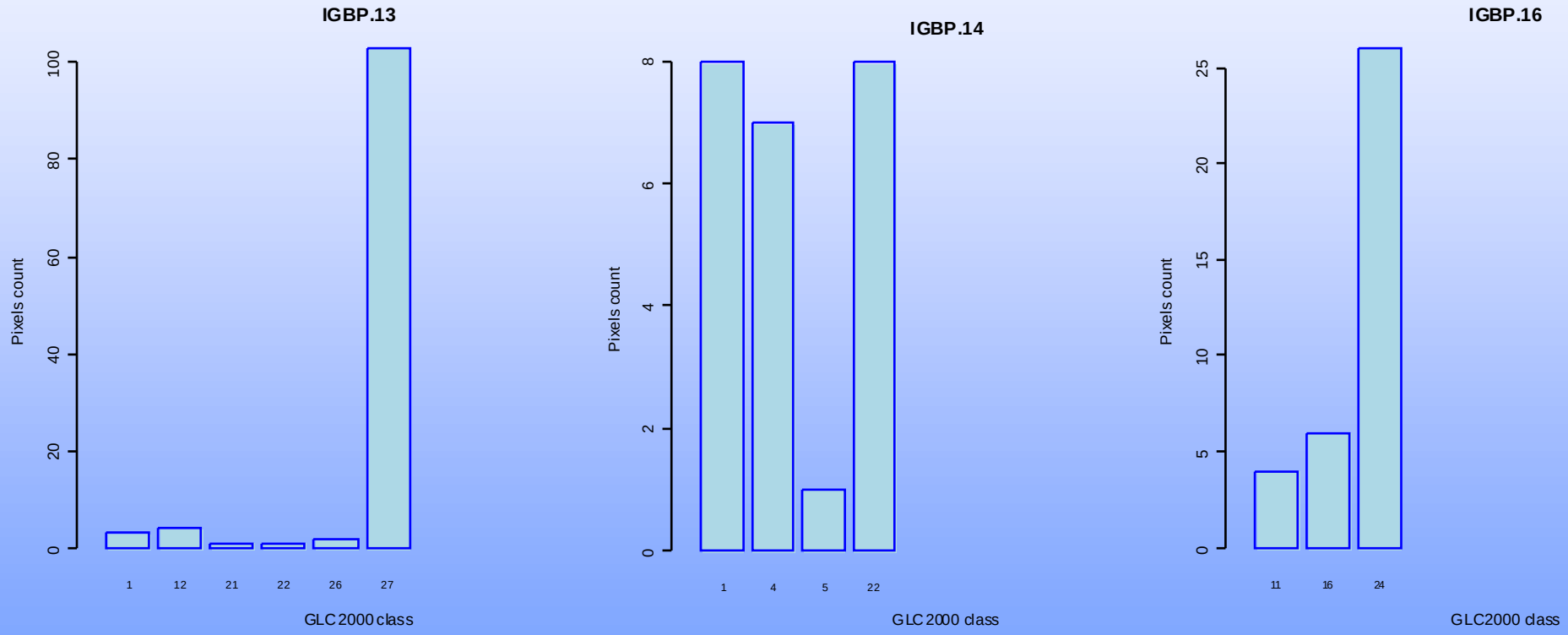
IGBP sites label and GLC2000







IGBP sites label and GLC2000



Proposed NELDA Land Cover Legend

***Baseline Legend*[1]**

Tree Dominated

Needleleaved

- Evergreen
  - Closed[2]
  - Open[3]
- Deciduous
  - Closed
  - Open

Broadleaved

- Evergreen
  - Closed
  - Open
- Deciduous
  - Closed
  - Open

Mixed

- Closed
- Open

Possible Additional Distinctions

Cover Detail
Mortality (yes/no)
Species
Wetland Trees (yes/no)
Understory Characteristics
Managed Plantation (Tree Farm/Orchard)

[1] The assumption is to use high resolution imagery (20 – 50 meters) and minimum mapping unit 1 – 2 hectares
[2] Closed >(> 65) %
[3] Open (65-15)%

Proposed NELDA Land Cover Legend

Shrub Dominated

Closed
Broadleaved
Open

Closed
Needleleaved
Open

Closed
Mixed
Open

Possible Additional Distinctions

Species
Wetland Shrubs (yes/no)
Leaf Longevity – Deciduous or Evergreen
Tundra (yes/no)
Trees < 15 % Present/not Present (Trees < 5 %)
Managed Plantations (Vineyard, for example)
Tree Regeneration (yes/no)

Proposed NELDA Land Cover Legend

Baseline Legend	***Possible Additional Distinctions***
Herbaceous Dominated	
Closed	Species (grasses, lichens, mosses, etc) Wetland Herb (yes/no) Tundra (yes/no) Pasture (yes/no)
Open	
Urban	Vegetation Dominated (Vegetation Cover > 50 %) Non-Vegetation Dominated (Vegetation Cover < 50 %)
Bare Areas	
Permanent Snow and Ice	
Water	

NELDA to LCCS

LC	LPos	LCCCode	LC	LCCLeve	LCCOwnI	LCCOwnDescr	LCCLabel	MapCode
1Forest	2	20092	0	A3A10B2XXD2E1			Needleleaved Evergreen Trees	1
1Forest	2	20093	0	A3A10B2XXD2E2			Needleleaved Deciduous Trees	2
1Forest	2	20089	0	A3A10B2XXD1E1			Broadleaved Evergreen Trees	3
1Forest	2	20090	0	A3A10B2XXD1E2			Broadleaved Deciduous Trees	4
2Woodlan	2	20134	0	A3A11B2XXD2E1			Needleleaved Evergreen Woodlar	5
2Woodlan	2	20135	0	A3A11B2XXD2E2			Needleleaved Deciduous Woodlar	6
2Woodlan	2	20131	0	A3A11B2XXD1E1			Broadleaved Evergreen Woodland	7
2Woodlan	2	20132	0	A3A11B2XXD1E2			Broadleaved Deciduous Woodlanc	8
3Thicket	2	20151	0	A4A10B3XXD1			Broadleaved Shrubs Close	9
3Thicket	2	20154	0	A4A10B3XXD2			Needleleaved Shrubs Closed	10
4Shrublan	2	20172	0	A4A11B3XXD1			Broadleaved Shrubland	11
4Shrublan	2	20175	0	A4A11B3XXD2			Needleleaved Shrubland	12
5Grasslan	1	20026	0	A2A10B4			Closed Herbaceous Vegetation	13
5Grasslan	1	20037	0	A2A11			Herbaceous Open Vegetation	14
1BuiltUp A	1	5003-9	0	A4-A13			Urban Area(s)	15
1NaturalW	1	8002	0	A1B1		Present > 11 months	PerennialNaturalWaterbodies	19
0Dichotom	1	0011	0	B16			Bare Area(s)	20
2Snow	1	8006	0	A2B1		Present > 11 months	PerennialSnow	22
3Ice	1	8009	0	A3B1		Present > 11 months	PerennialIce	23
1Forest	2	20092(2)[Z3]	0	A3A10B2X	T.N.E.C.M	Presence of dead trees (mortality	Needleleaved Evergreen Trees	24
1Forest	2	20092(2)[Z4]	0	A3A10B2X	T.N.E.C.B	Presence of Bog/Wetland	Needleleaved Evergreen Trees	25

“Validation” Efforts

- Issues
  - Lack of probability sample
  - Mixed pixel problem in coarse resolution data
  - Ambiguous class definitions
  - Spectral separation of classes (can we actually distinguish them with MODIS?)
- Approaches
  - Independent assessments (Warren Cohen, OSU; Bigfoot)
    - NELDA sites for validation
  - Cross validation of STEP database Independent evaluation/assessment activities (independent evaluators)
  - Model-based assessment (confidences)

Cross Validation

(Strahler, 2003; http://geography.bu.edu)

- Cross-Validation Procedure
  - Exploits STEP database
  - Hide 10 percent of training sites, classify with remaining 90 percent; repeat ten times for ten unique sets of all sites
  - Provides “confusion matrix” based on unseen pixels where whole training site is unseen
  - Not a stratified random sample, but a indication of accuracy

Summary

- MODIS Decision Tree
- Add new examples from NELDA sites to the STEP database
- Review and change STEP polygons labels
- Finalize NELDA legend