POLITECNICO
MILANO 1863

A FREE AND OPEN SOURCE TOOL TO ASSESS THE ACCURACY OF LAND COVER MAPS: IMPLEMENTATION AND APPLICATION TO LOMBARDY REGION (ITALY)

Maria Antonia Brovelli, Gorica Bratic, Monia Elisa Molinari

Developments, Technologies and Applications in Remote Sensing
May 7-10, 2018 Beijing, China

Context

- Importance of Land Use and Land Cover (LULC) maps for environmental studies and applications, e.g. biodiversity, natural resources, climate change
- Rapidly increasing number of high-resolution LULC datasets due to the continuous advances in Remote Sensing sensors and mapping technologies
- Increasing availability of global open LULC datasets
- Importance of LULC maps classification accuracy, a key factor to evaluate their suitability for the various applications where they are exploited

LULC & Validation

- The accuracy assessment of digital remotely-sensed data started around 1975:

  I. simple visual checkup: “looking good” requirement

  II. non-site-specific assessment approach: simply comparison of land use classes areal extent for classified and ground truth datasets

  III. error/confusion matrix technique and its derived accuracy measures

“The standard descriptive reporting tool for accuracy assessment of remotely sensed data” [1]

[1] Lunetta, R.S., Lyon, J.G., 2004. Remote Sensing and GIS Accuracy Assessment, CRC Press, Boca Raton, FL

Error matrix & FOSS GIS

- Error matrix enables the comparison of two sources of spatial information, thus represents a key tool for Geographical Information System (GIS) software

- FOSS GIS software currently focus on the simple error matrix computation and they provide very few indexes:
  - **QGIS** (*Accuracy Assessment Plugin*): user's and producer's accuracies, allocation and quantity disagreements
  - **GRASS GIS** (*r.kappa module*): overall accuracy, user's and producer's accuracies, Kappa

Development of a new FOSS tool that can be easily integrated into GIS systems and enables users to automatically calculate all the statistics based on confusion matrix proposed by literature

Outline

- Literature review
- Low resolution bias analysis
- Implementation
- Application
  - Datasets
  - Data Processing
  - Results
- Conclusions
- Further work

Literature review: error matrix

	Reference (ground truth) map				
Classified (comparison) map	Class	j=1	j=2	…	j=q
	i=1	n_{11}	n_{12}	…	n_{1q}
	i=2	n_{21}	n_{22}	…	n_{2q}
	…	…	…	…	…
	i=q	n_{q1}	n_{q2}	…	n_{qq}

Literature review: indices

- Most commonly used:
  - Overall accuracy (P0)
  - Producer's accuracy (PA) [1- Omission error]
  - User's Accuracy (UA) [1- Commission error]
- Derived from P0, PA, UA
  - Average of user's accuracy (AUA) or of producer's accuracy (APA)
  - Combined user's (CAU) or producer's accuracy (CAP)
  - Hellden's mean accuracy (MAH)
  - Short's mean accuracy (MAS)
  - Classification success index (CSI) and its variations Group Success Index (GCSI) and Individual classification success index (ICSI)
- Margfit

Literature review: indices

- Derived from information theory
  - Average mutual information (AMI) and different ways of normalizing it (NMIa – arithmetic mean, NMIg – geometric mean)
- Kappa and kappa-like indexes
  - Standard kappa index (K)
  - Conditional kappa(Kc)
  - Weighted kappa(Kw)
  - Tau(τ)
  - Aickins alpha(α)
  - Ground truth index (GT)
- Indexes of disagreement
  - Quantity disagreement
  - Allocation disagreement

Low-resolution bias analysis

- It is applied on the ground truth map (higher resolution)
- It quantifies error due to the loss of information while resampling from high to lower resolution
- It is only applicable to the binary maps so it requires reclassification
- The results are given by means of Pareto frontier
- Tool created by Python programming language

FOSS tool: implementation

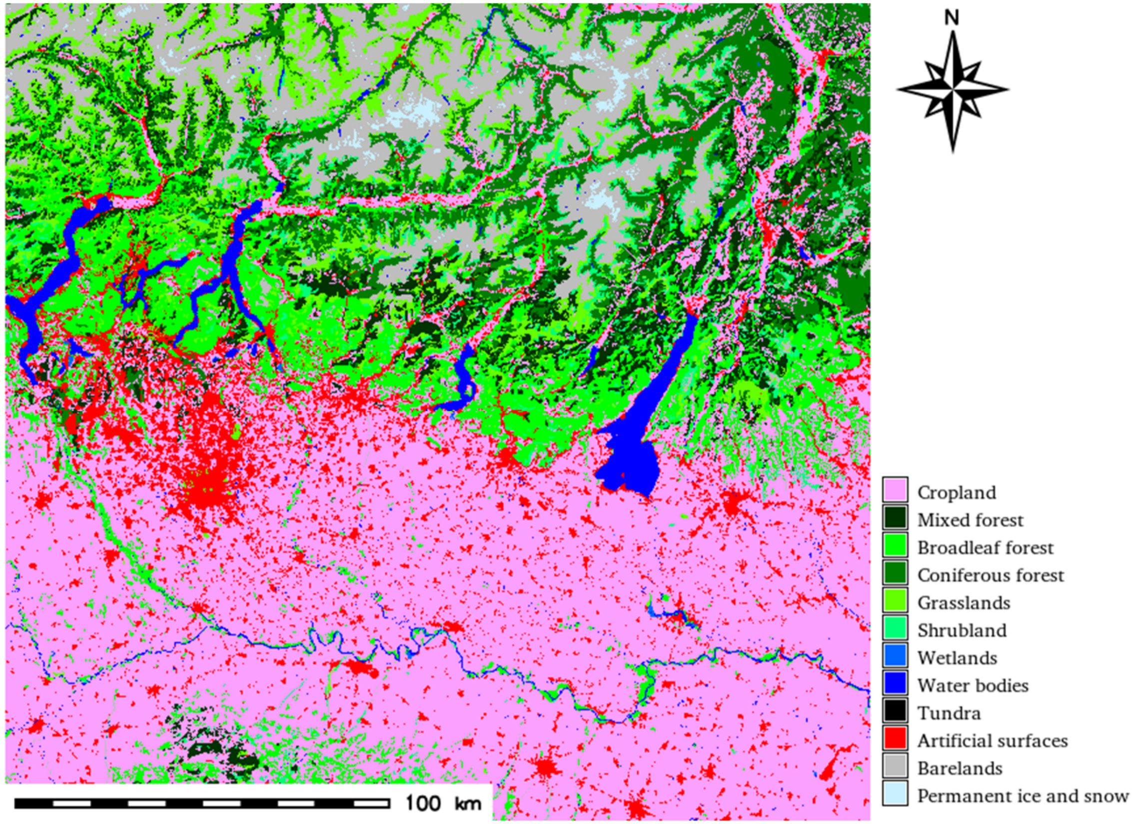
- Stand-alone tool in Python Programming language based on Numpy and Pandas libraries
  - Input data: error matrix
  - Output data: csv with accuracy measures

- GRASS GIS script based on GRASS Python scripting libraries:
  - Input data: classified and reference datasets
  - Output data: csv with accuracy measures

Application: classified datasets

GlobeLand30 (GL30)

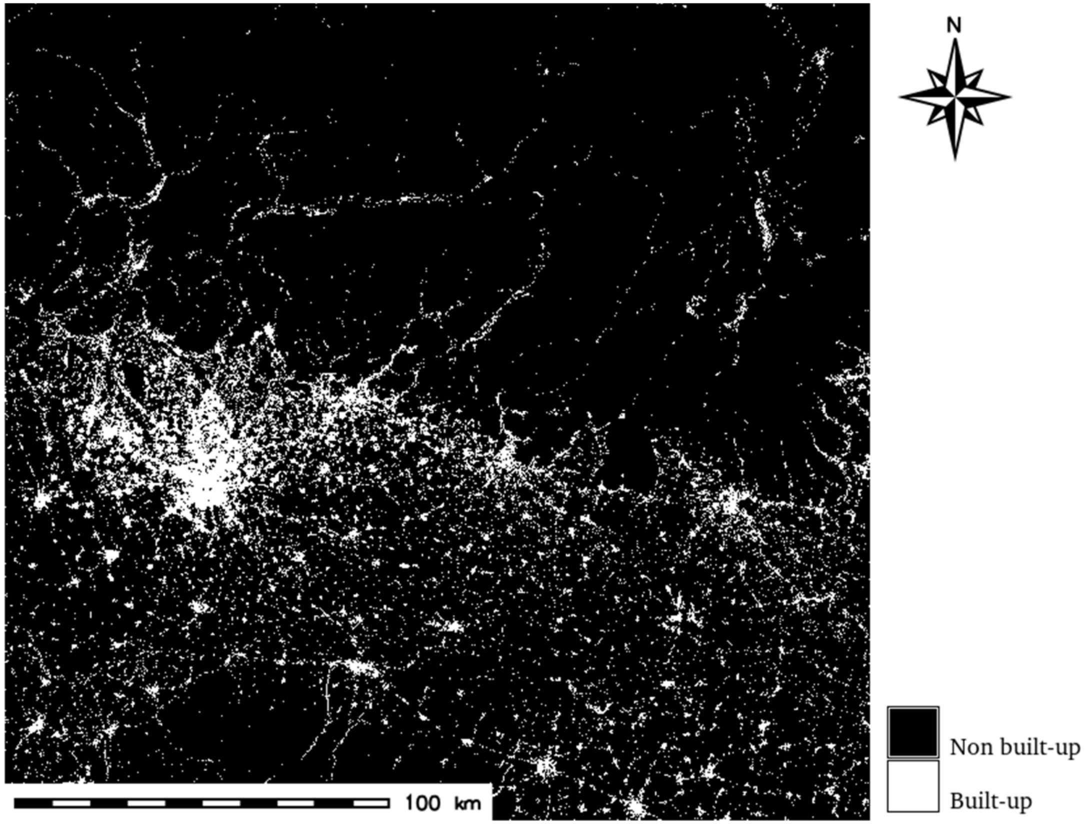
- Product of National Geomatics Center of China (NGCC)
- 12 land cover categories
- 30 m of spatial resolution
- Reference year: 2010



Application: classified datasets

Global Human Settlement (GHS)

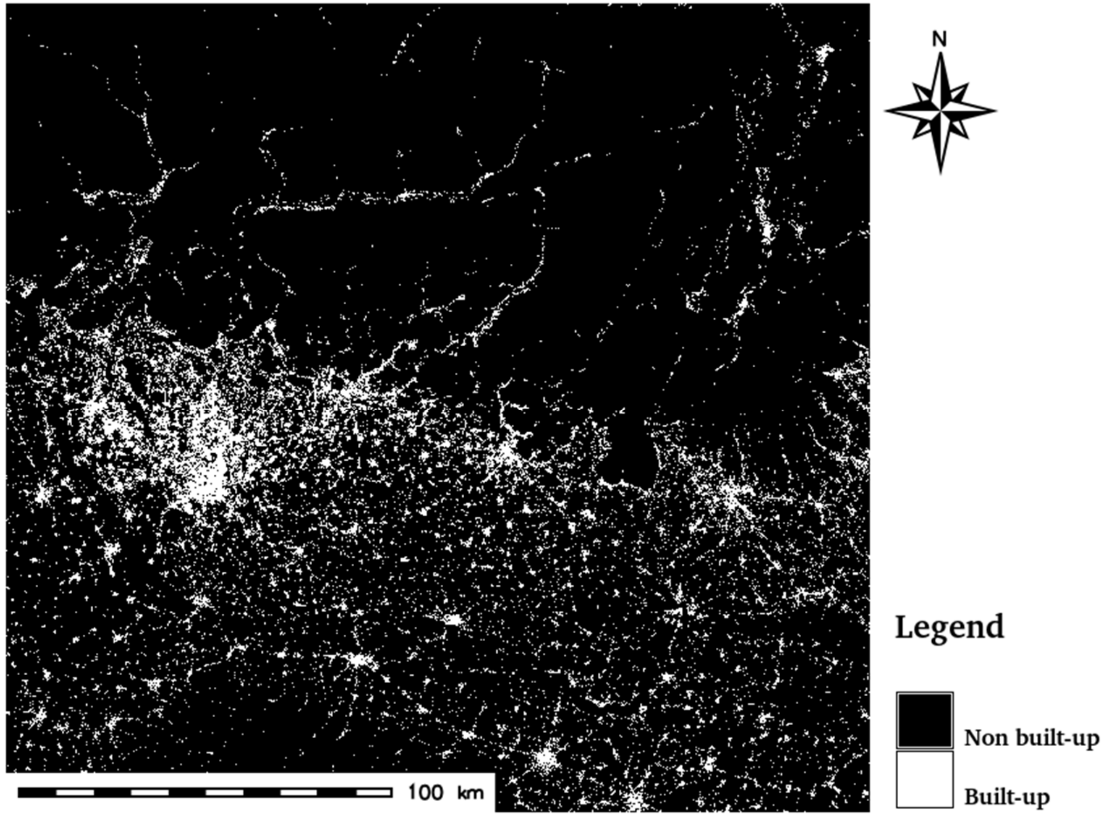
- Product of EU Joint Research Center
- 2 categories: built-up, no built-up
- 38 m spatial resolution
- Reference year: 2014



Application: classified datasets

Global Urban Footprint (GUF)

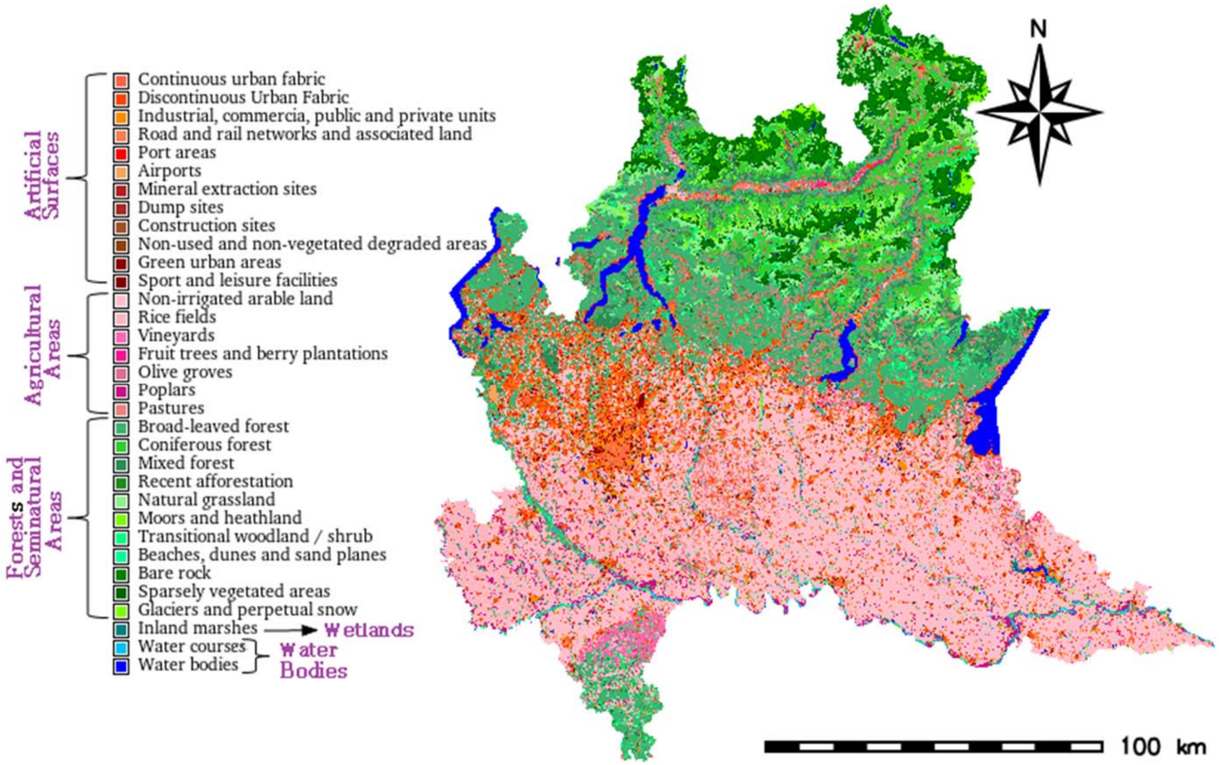
- Product of German Aerospace Center (DLR)
- 2 categories: built-up, no built-up
- 12 m of spatial resolution
- Reference year: 2011



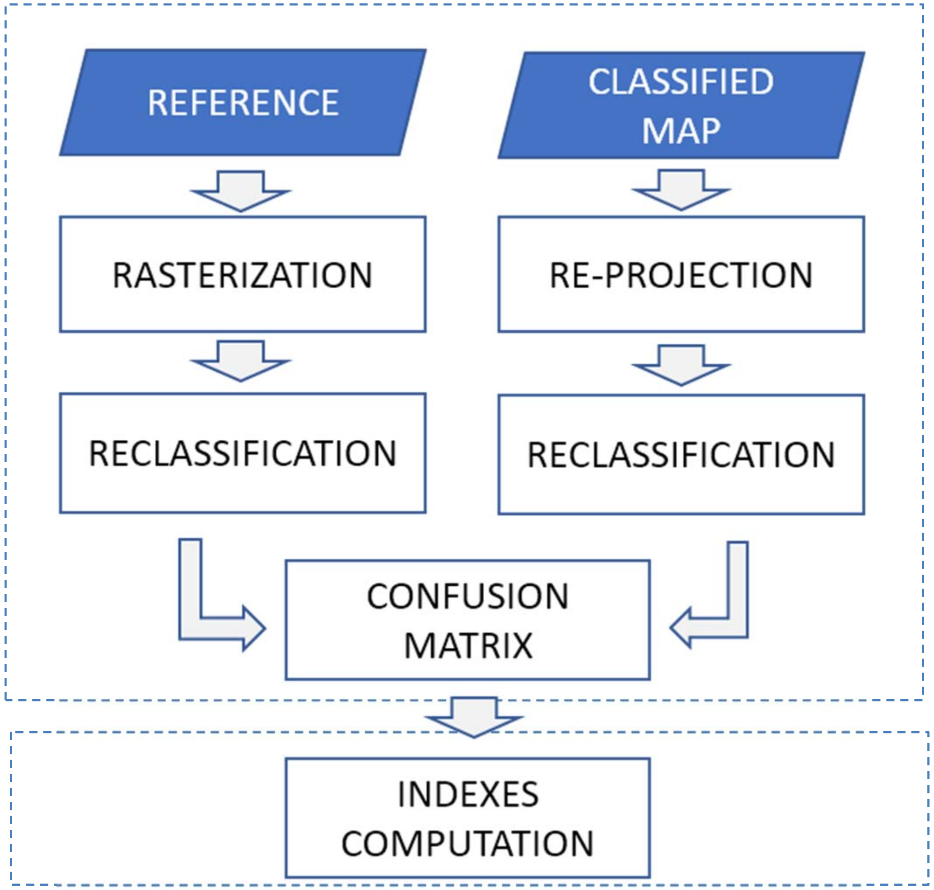
Application: reference dataset

DUSAF 4.0

- Database of land use of Lombardy Region
- 5 level hierarchical classification system
- Scale 1:10,000
- Reference year: 2012



Application: processing



GRASS GIS

QGIS

python™

Application: GL30 reclassification process

GL30-11 case study

The classes of DUSAF have been reclassified according to GL30 thematic legend based on eleven classes

DUSAF classes	GLOBELAND30 classes
Agricultural Areas	Cultivated land
Mixed Forest	Mixed forest
Broad-leaved forest, Recent afforestation	Broadleaf forest
Coniferous forest	Coniferous forest
Natural Grassland	Grasslands
Moors and heathland, Transitional woodland/shrub	Shrublands
Wetlands	Wetlands
Artificial Areas	Artificial surfaces
Beaches, dunes and sand planes, Bare Rock, Sparsely vegetated areas	Bare lands
Water Bodies	Water
Glaciers and perpetual snow	Permanent snow and ice

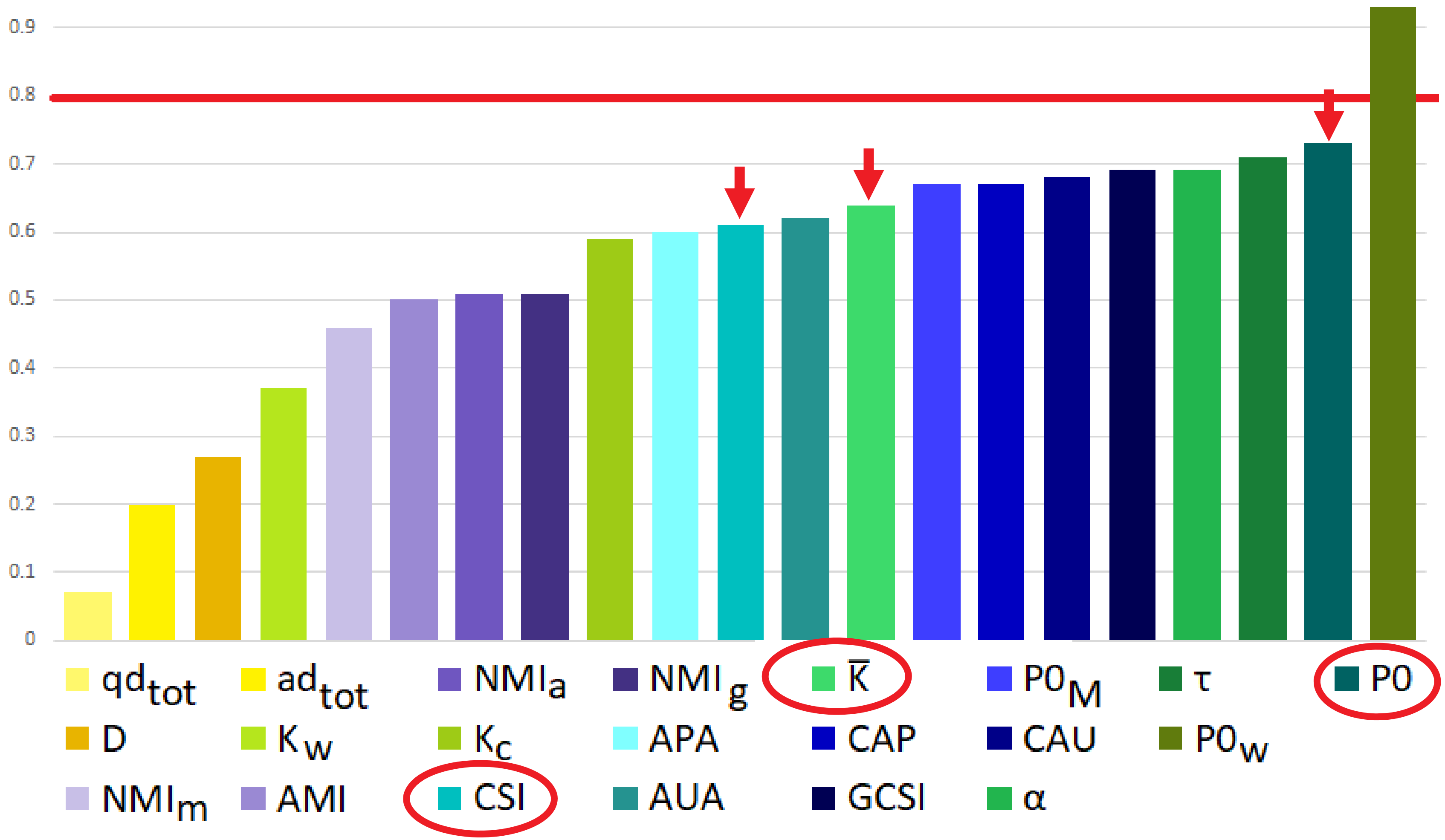
Application: GL30 reclassification process

GL30-5 case study

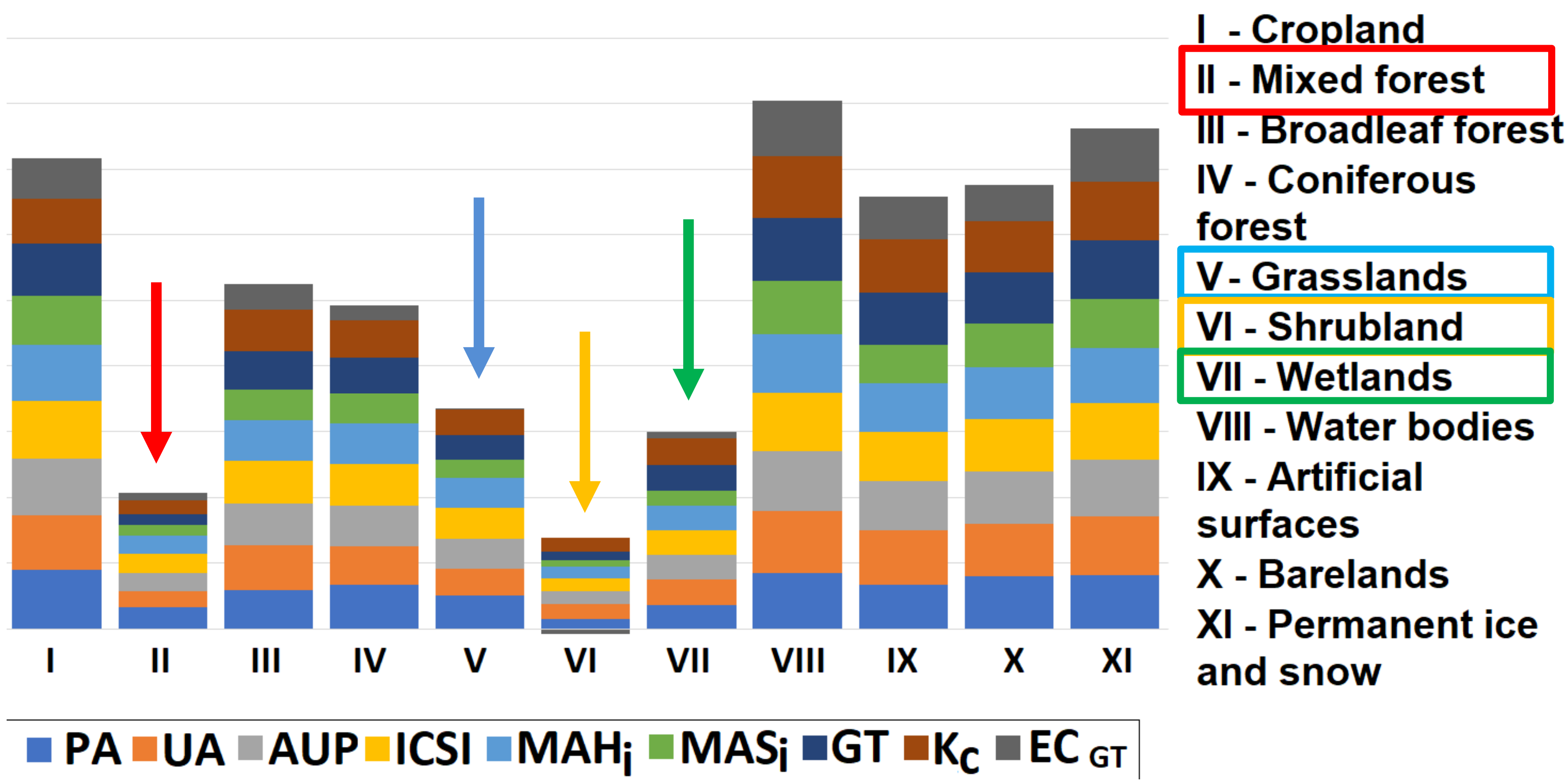
The GL30 has been reclassified according to the five first-level classes of DUSAF

DUSAF classes	GLOBELAND30 classes
Artificial surfaces	Artificial surfaces
Agricultural areas	Cultivated land
Forest and semi natural areas	Broadleaf forest, Coniferous forest, Mixed forest, Grasslands, Shrublands, Bare lands, Permanent snow and ice
Wetlands	Wetlands
Water bodies	Water

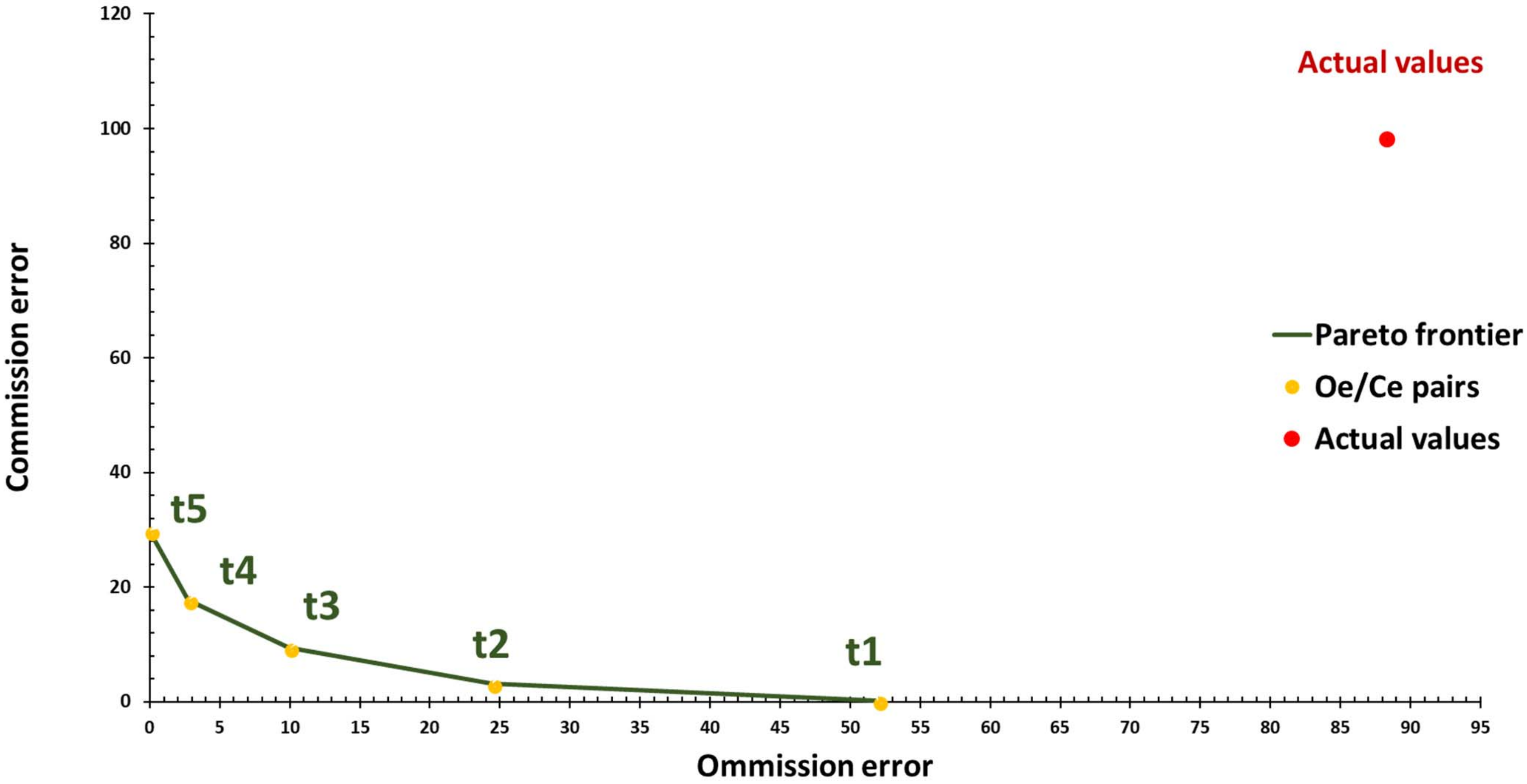
Application: results GL30-11 (global indexes)



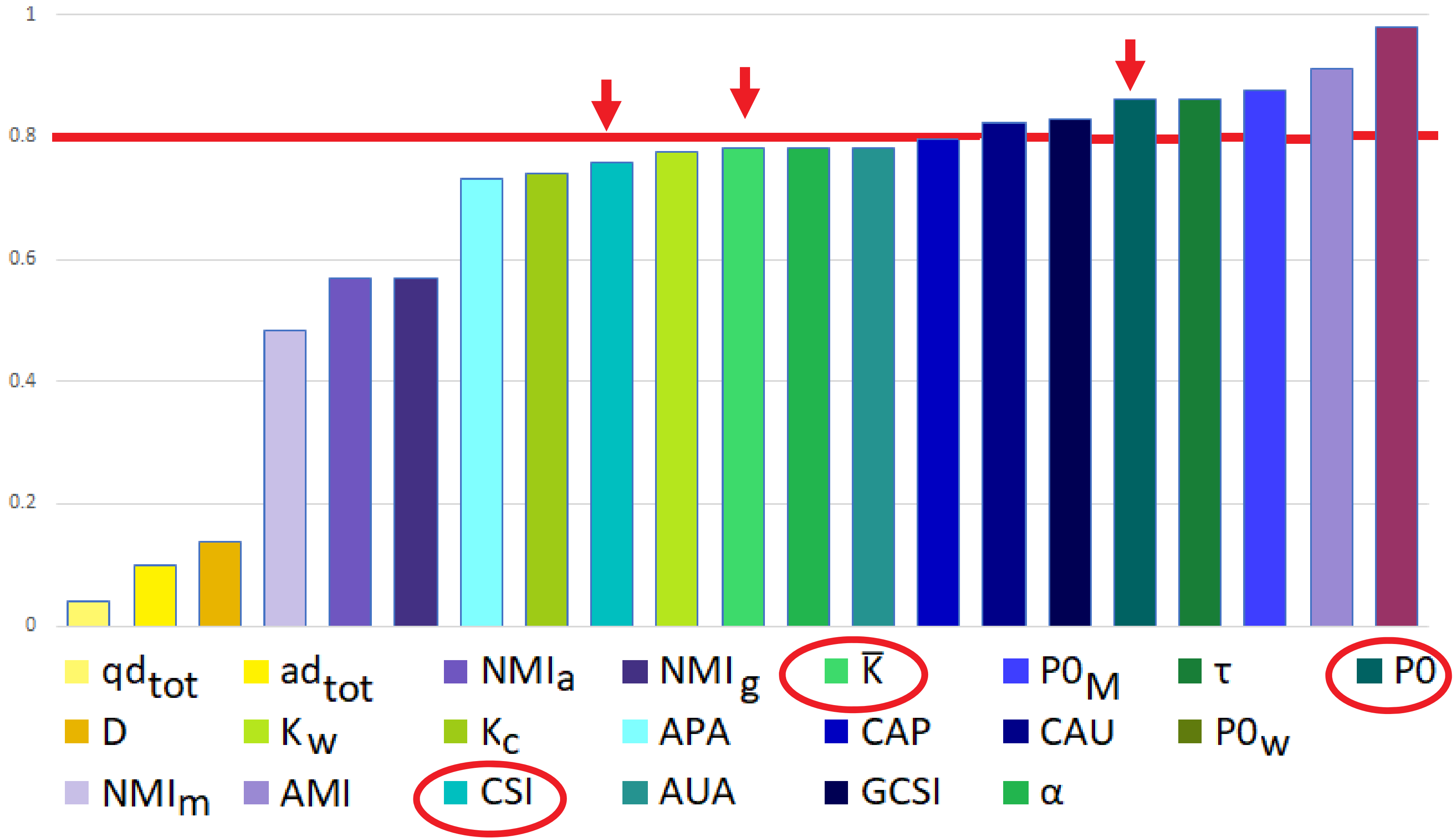
Application: results GL30-11 (per-classes indexes)



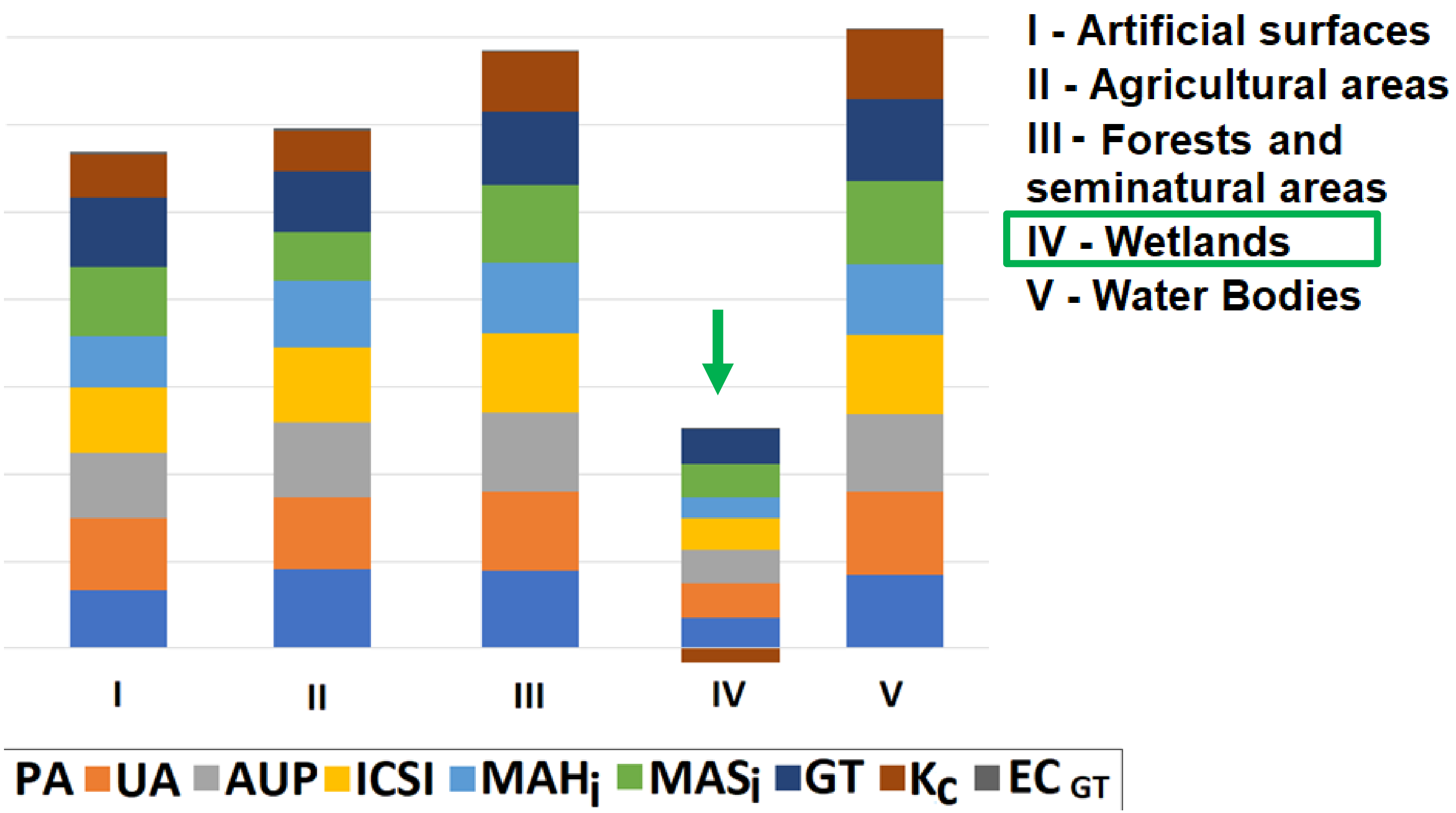
Low-resolution bias-Shrubland class



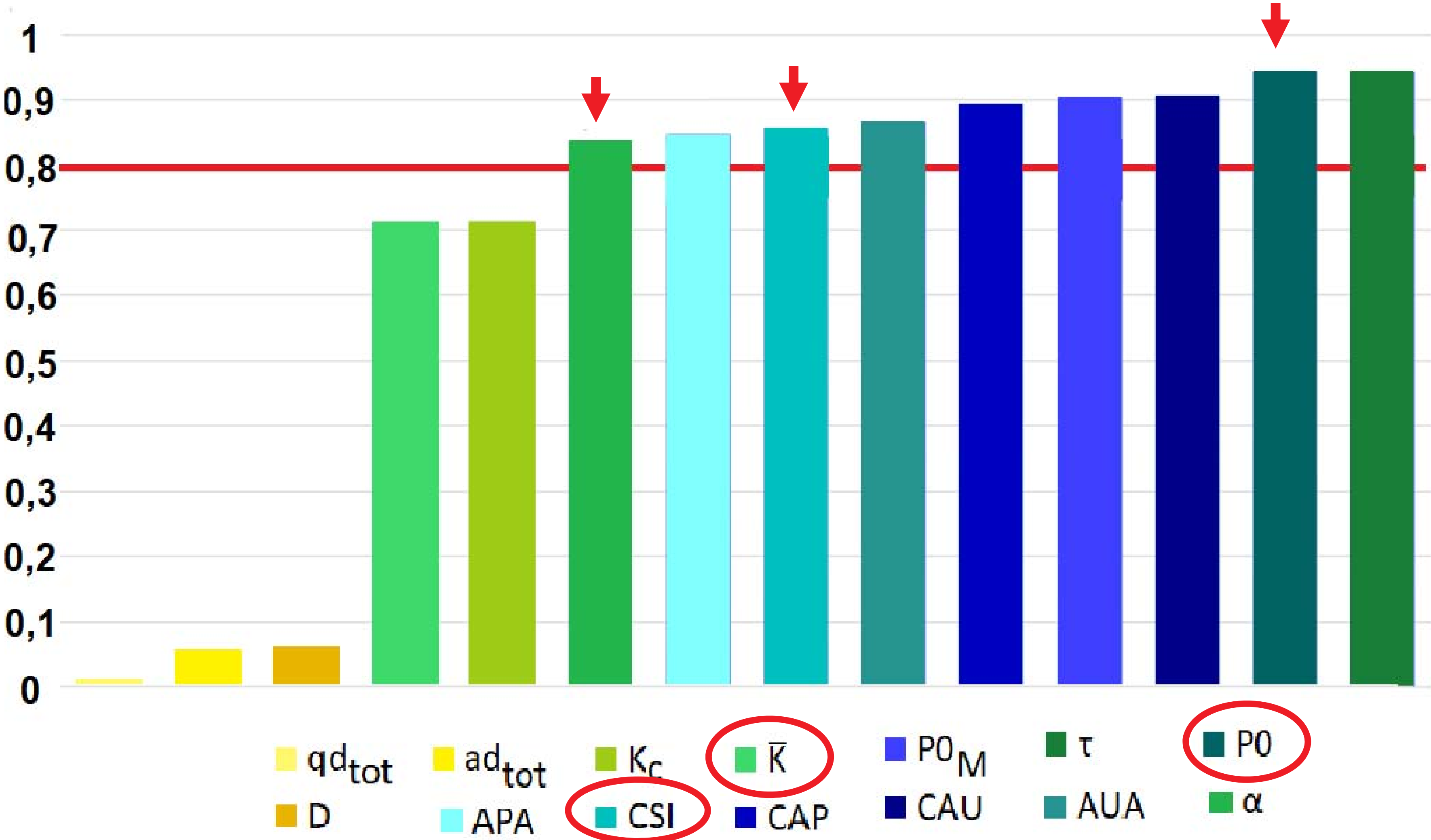
Application: results GL30-5 (global indexes)



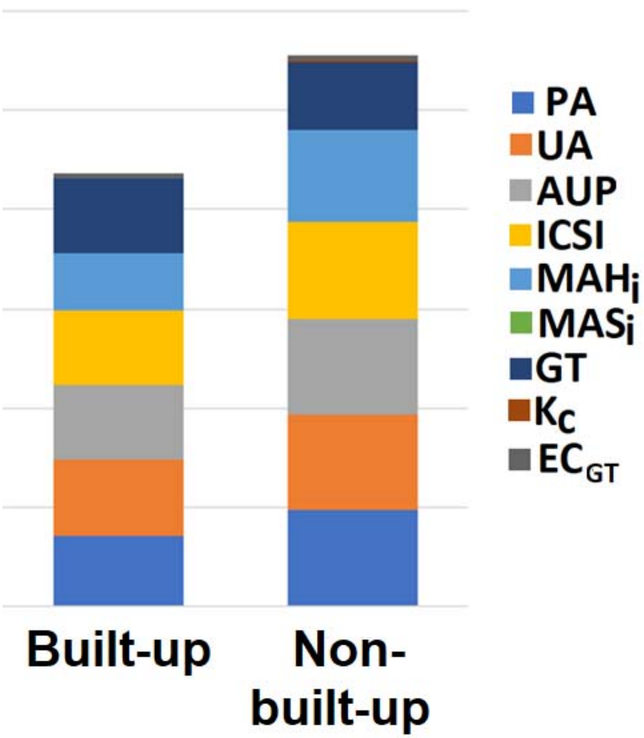
Application: results GL30-5 (per-classes indexes)



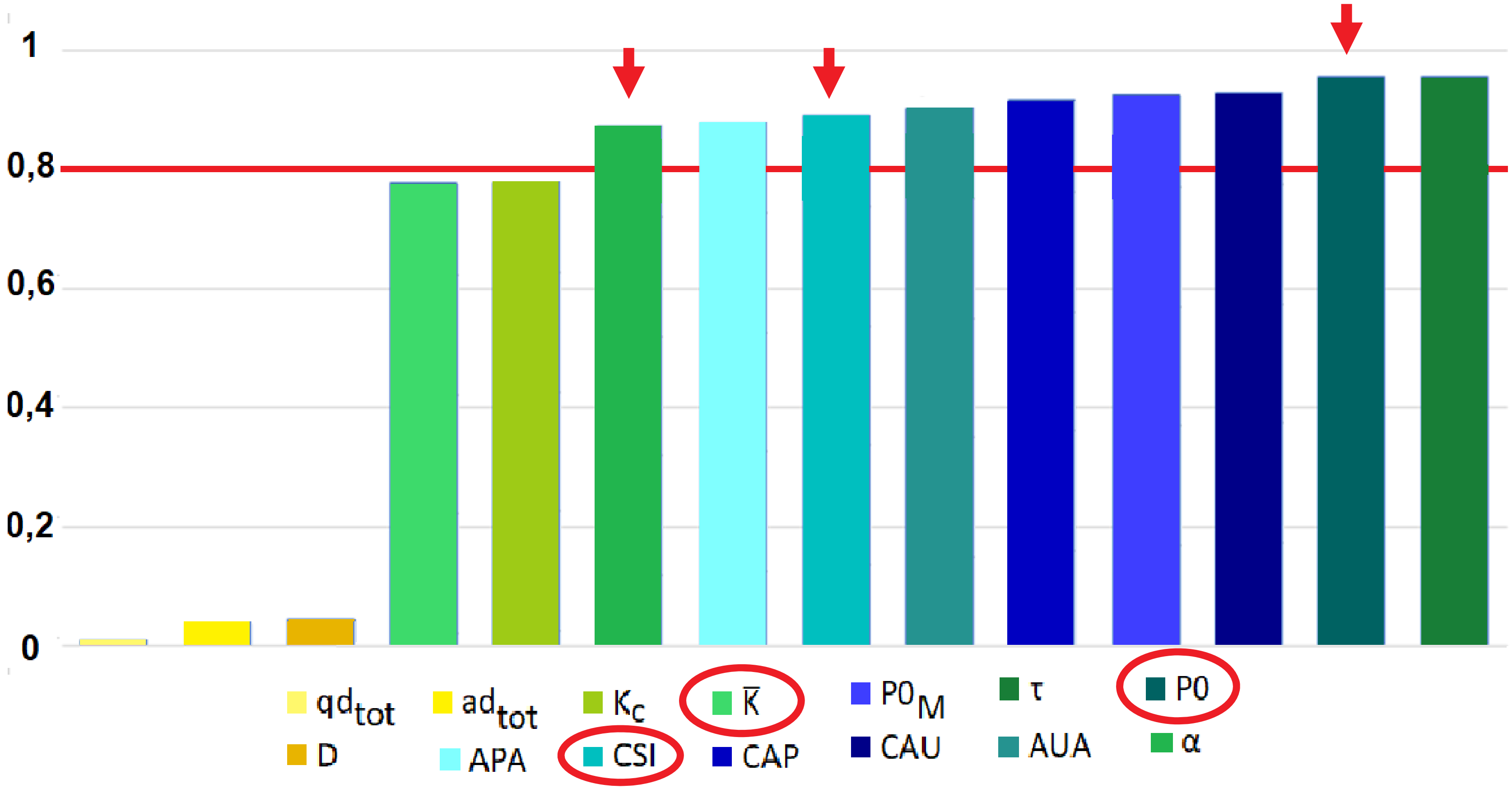
Application: results GHS (global indexes)



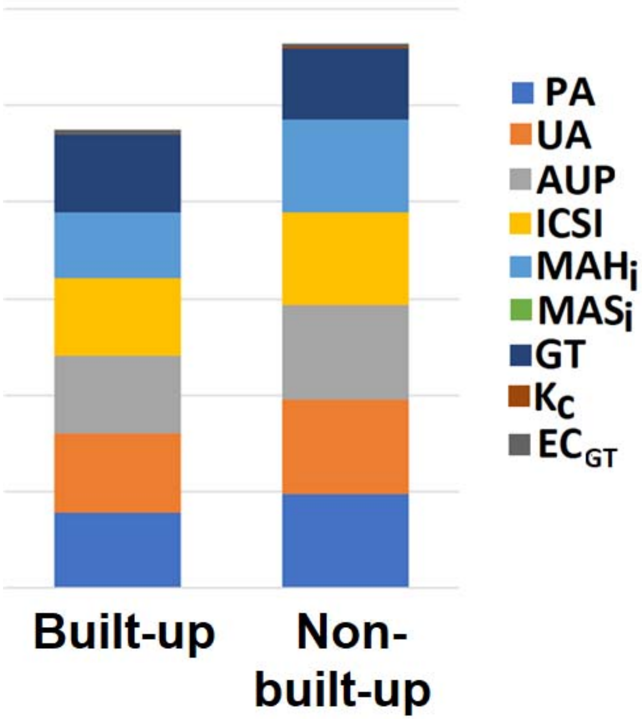
Application: results GHS (per-classes indexes)



Application: results GUF (global indexes)



Application: results GUF (per-classes indexes)



Conclusions

- A detailed investigation about confusion matrix-derived indexes was performed and a Python FOSS module was implemented
- The tool was successfully applied to evaluate the classification accuracy of three high-resolution LULC datasets (GL30, GHS,GUF)
  - very satisfactory accuracy is obtained for GUF and GHS built-up datasets
  - high overall accuracy is obtained by considering the GL30-5 case study
  - accuracy decreases if a more detailed thematic legend is considered (GL30-11 case study), especially for the classes related to vegetation

Further work

- Extension of the tool with additional functions able to detect any patterns of error in discrepancies between the LULC (Join counts, Moran's I…)
- Development of a Plugin for QGIS, with the purpose of creating a user-friendly Graphical User Interface and widen its usage among users, professionals and researchers
- The plugin will be published for the three workshops "High-Resolution Land Cover Inter-comparison and Validation" (ISPRS Capacity Building Initiative 2018) which will be held in Dar Es Salaam (1/9/2018), Nairobi (3/9/2018) and Delft (October 2018, Technical Commission IV Symposium)

isprs information from imagery

POLITECNICO MILANO 1863

NGCC 国家基础地理信息中心 National Geomatics Center of China

WORLD BANK

REGIONAL CENTRE FOR MAPPING OF RESOURCES FOR DEVELOPMENT

Thank you for the attention

Maria Antonia Brovelli

Politecnico di Milano
Department of Civil and Environmental Engineering
Piazza Leonardo da Vinci, 32 - 20133 Milano (Italy)

maria.brovelli@polimi.it

This work was partially supported by URBAN-GEO BIG DATA, a Project of National Interest (PRIN) funded by the Italian Ministry of Education, University and Research (MIUR) – id. 20159CNLW8.