D2.4 Report on Accuracy assessment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805

Georgios Spanos, Aristotle University of Thessaloniki

Conference "Carbon sequestration potential of Marginal Lands in Europe", 13/12/2021

Authors

- Anastasios Stamnas
- Abdulrashid Hassan
- Georgios Spanos
- Zoi Touloudi

Editors

- Luis A. Ruiz
- Juan Pedro Carbonell-Rivera
- Jesús Torralba Pérez
- Elke Krätzschmar
- Charalampos Georgiadis

Contents

- Aim and Objectives
- Background
- Methodology development
- Assessment techniques
- Evaluation Metrics
- Results
- Discussion points
- Land Cover analysis
- Conclusions

Aim and Objectives

- Accuracy assessment of the methodology of Task 2.3
    - Definition of the statistical accuracy assessment methods
    - Development of an accuracy assessment methodology of MLs detection/classification through:
        - Stratified random sampling
        - Area-based assessment

Background

- Accuracy assessment is critical in spatial investigation projects:
  - Need to self-evaluate and learn from errors
  - Need to quantitatively compare methods and algorithms

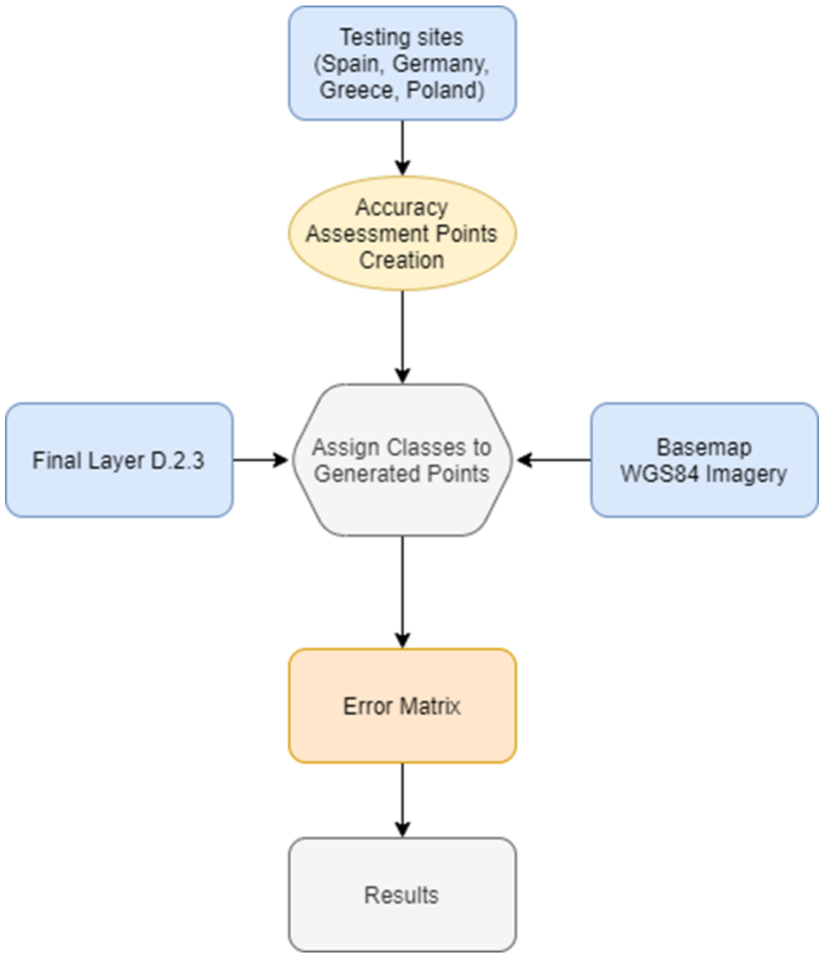
- No good-fit for all accuracy assessment, as no single map producing technique exists

Sources: Congalton 2001, Congalton & Green 2019

Initial approach

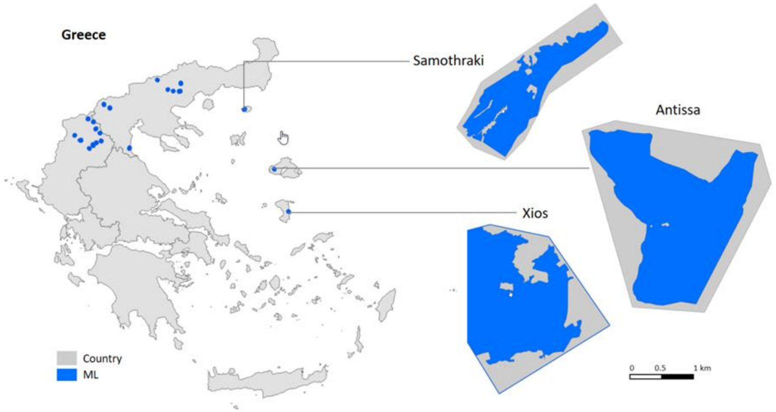
- Assessment of the Final ML suitability map
- Stratified random sampling
- Visual interpretation

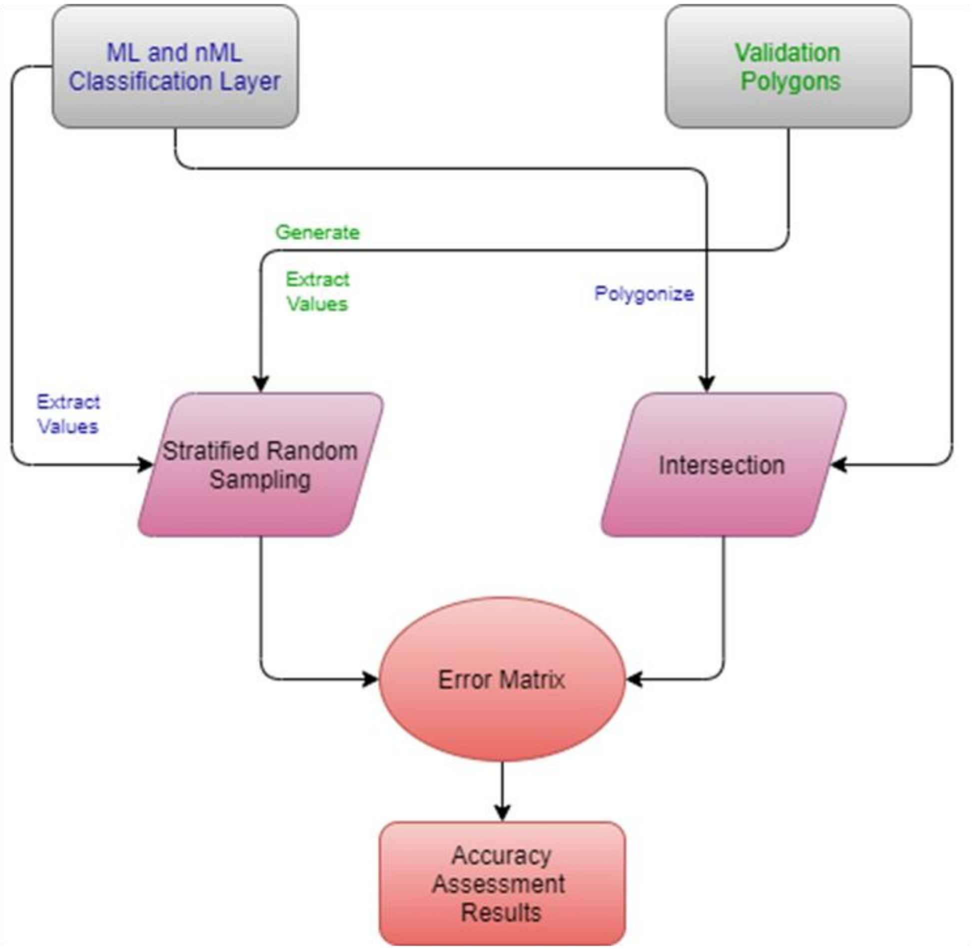
- Tedious
- Possibility of bias and uncertainty



Methodology workflow

- Assessment of the ML Hard Thresholds map
- Project partners provided reference polygons





Assessment techniques

Point-based assessment

	ML		nML	
	Area [ha]	Allocated Points	Area [ha]	Allocated Points
Greece	7988	7988	5274	5274
Spain	1649	1648	2199	2199
Germany	352	352	20,913	20,913
Poland	539	539	2463	2463
Merged	10,529	10,529	30,849	30,849

Area-based assessment

- No sampling – all of the reference area is exploited
- Validation polygons superimposed on Classified image
  - Intersection algorithm

Evaluation Metrics

- **Overall Accuracy (OA)** $OA = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} * 100$
- **User's accuracy (UA)** error of commission [%] = 100% – User's Accuracy[%]
- **Producer's accuracy (PA)** error of omission[%] = 100% – Producer's Accuracy [%]
- **Error rate (ERR) or misclassification rate** $ERR = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}$
- **F1-score** $F1\text{-}score = 2 * \frac{\mathrm{PREC} * \mathrm{REC}}{\mathrm{PREC} + \mathrm{REC}}$
- **Matthew's correlation coefficient (MCC)** $MCC = \frac{\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})\,(\mathrm{TP} + \mathrm{FN})\,(\mathrm{TN} + \mathrm{FP})\,(\mathrm{TN} + \mathrm{FN})}}$
- **Kappa** $Kappa = \frac{\mathrm{accuracy} - \mathrm{expAccuracy}}{1 - \mathrm{expAccuracy}}$

Sources: Congalton 1991, Banko 1998, Bradley 1997, Sokolova et al., 2006, Matthews, 1975, Jensen, 1996; Sim & Wright, 2005

Results

Metric of accuracy	Greece		Spain		Germany		Poland		Merged	
	Point-based	Area-based	Point-based	Area-based	Point-based	Area-based	Point-based	Area-based	Point-based	Area-based
OA (%)	71.52	70.75	82.87	83.42	60.61	59.79	90.97	90.56	67.98	67.73
UA (%)	77.73	76.86	77.47	77.98	3.62	3.53	92.41	90.74	42.69	42.40
PA (%)	73.89	73.58	84.66	85.45	90.06	88.60	54.17	52.79	75.36	74.87
F1-SCORE	75.76	75.19	80.90	81.54	6.96	6.78	68.30	66.75	54.50	54.14
ERR (%)	28.48	29.25	17.13	16.58	39.84	40.21	9.03	9.44	32.02	32.27
KAPPA	0.41	0.40	0.65	0.67	0.04	0.04	0.64	0.62	0.33	0.32
MCC	0.41	0.40	0.66	0.67	0.13	0.13	0.67	0.65	0.36	0.35

Results

Comparison of predicted and reference ML classes:

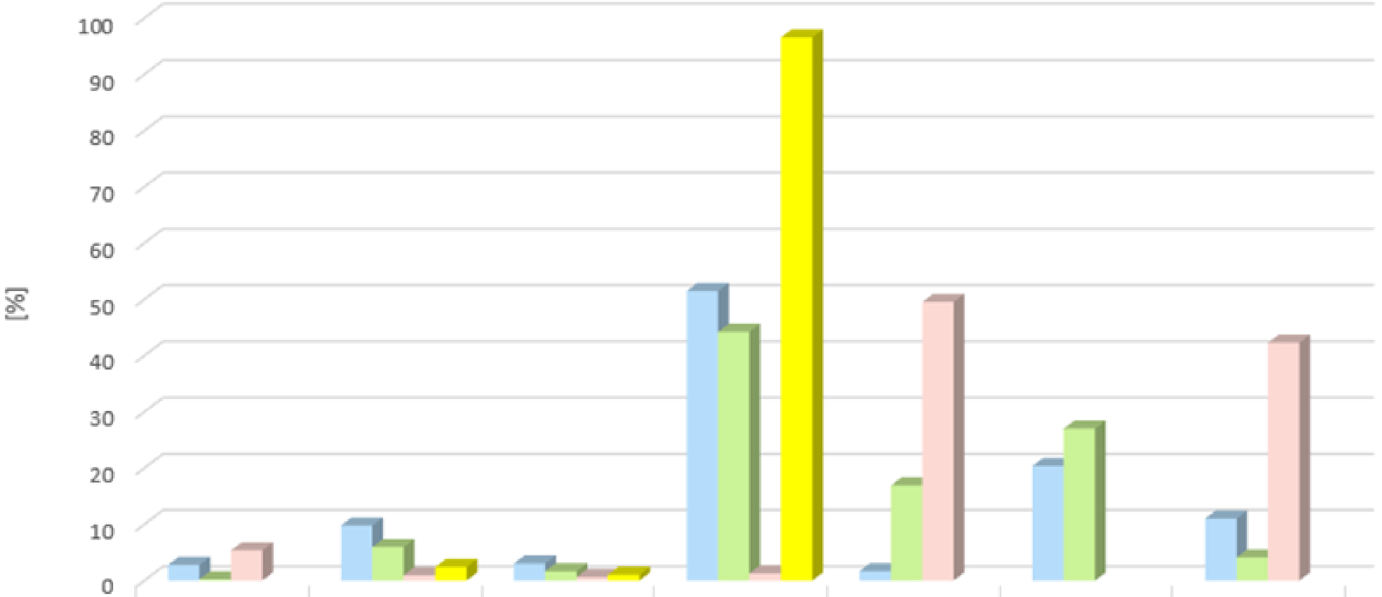
Area of ML in ha	Greece	Spain	Germany	Poland	Merged
Predicted (area-based)	7,646	1,807	8,820	313	18,589
Reference	7,987	1,649	351	538	10,529

Discussion points

- Hard Layers and S2GLC not 100% accurate
- Detection methodology based on 2017 – validation data on 2021
- Different methods are used by different experts in each country in acquiring data
- Large differences in the amount of validation data areas provided by each country
- Class imbalance

Land Cover Analysis

- TP ML samples overlaid on the S2GLC map
- Identify which types of land cover are associated with MLs



	Artificial surfaces and constructions	Broadleaf tree cover	Coniferous tree cover	Herbaceous vegetation	Moors and Heathland	Sclerophyllous vegetation	Natural material surfaces
Greece	2.78	9.76	3.02	51.45	1.66	20.34	11.00
Spain	0.21	5.95	1.65	44.20	16.91	27.01	4.08
Germany	5.36	0.95	0.63	1.26	49.53		42.27
Poland		2.40	1.03	96.58			

Conclusions

- Poland manifested the highest overall accuracy in both assessment techniques followed by Spain, Greece and Germany
- For Greece, Spain and Poland majority of the ML areas were over the Herbaceous LC class
- Germany over the Moors and Heathland, and Natural Material Surfaces land cover classes
- Differences in accuracy measures among countries can be balanced by deriving a standard procedure, and/or
- by field measurements of the test areas.

References

[1] Banko, G. (1998). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory. IR-98-081.

[2] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[3] Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37(1), 35–46.

[4] Congalton, R. G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. International Journal of Wildland Fire, 10(4), 321. https://doi.org/10.1071/WF01031

[5] Congalton, R. G., & Green, K. (2019). Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

[6] Jensen, J. R. (1996). Introductory digital image processing: a remote sensing perspective. (Issue Ed. 2). Prentice-Hall Inc.

[7] Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure, 405(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

[8] Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Physical Therapy, 85(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

[9] Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. AAAI Workshop - Technical Report, WS-06-06, 24–29. https://doi.org/10.1007/11941439_114

Thank you for your attention!

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805

Georgios Spanos, geospan@gmail.com