

Supported by:



Federal Ministry
for the Environment, Nature Conservation
and Nuclear Safety



Federal Agency for
Nature Conservation



EOCap4Africa

10 Validation / Accuracy Assessment



INES Ruhengeri
Institute of Applied Sciences



Learning objectives



Understand the principles and importance of accuracy assessment in spatial analysis.

Differentiate between various accuracy metrics and their applications.

Implement accuracy assessment techniques using QGIS and R.

Critically evaluate classification results and understand sources of errors.

Accuracy assessments



Why assess accuracy?

- Ensures reliability and credibility of spatial data analyses
- Identifies and quantifies errors in classification outputs

Common Applications

- Land cover classification
- Change detection studies
- Environmental monitoring

Accuracy assessments



Can answer the following questions:

- What is the overall accuracy of the classification map?
- What is the accuracy of single classes in the classification map?
- Which classes are misclassified and confused with each other?

Understanding classification errors



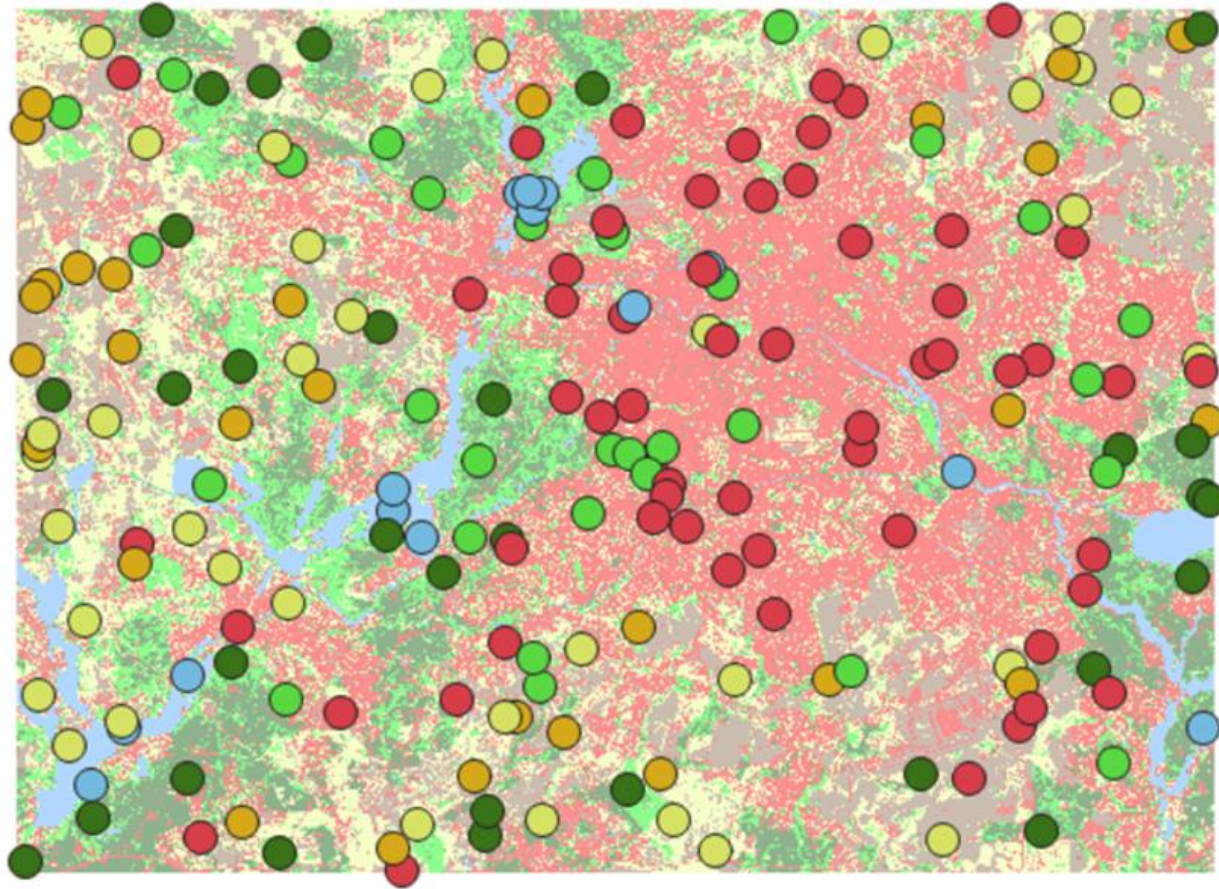
Sources of errors

- Data quality issues (e.g., sensor errors, atmospheric conditions).
- Mislabeling in training data
- Algorithm limitations

Impacts of errors

- Misinterpretation of spatial phenomena
- Erroneous decision-making based on flawed data

Validation data



Validation data

Point samples with class labels representing the 'reality' for the following land cover classes:

- Urban (built-up and non built-up)
- Grass & Crops
- Broadleaf trees
- Coniferous trees
- Soil (incl. harvested cropland)
- Water

(Humboldt Universität zu Berlin n.d.)

Validation data collection



Collection methods

- Field surveys and in-situ measurements.
- High-resolution imagery interpretation
- Existing reliable datasets

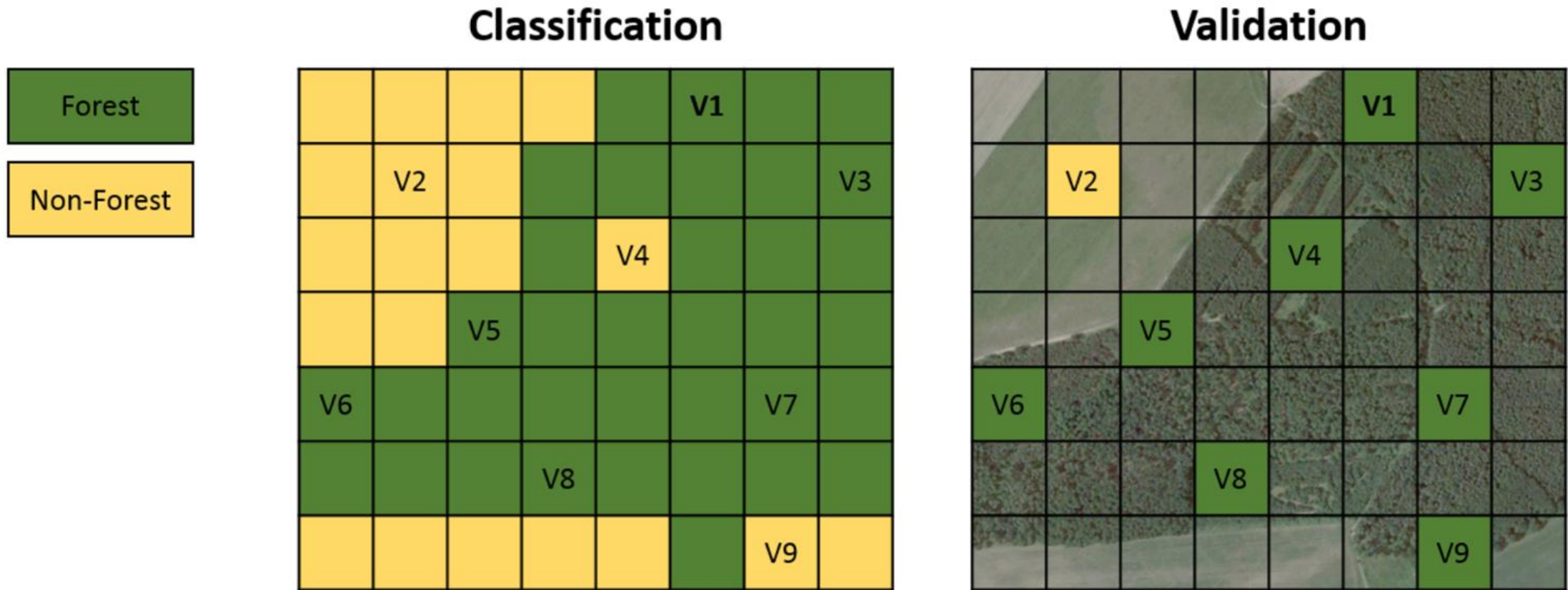
Sampling techniques

- Random sampling
- Stratified random sampling
- Systematic sampling



Validation method: Confusion matrix

Provides a detailed breakdown of classification performance, for example:

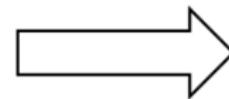
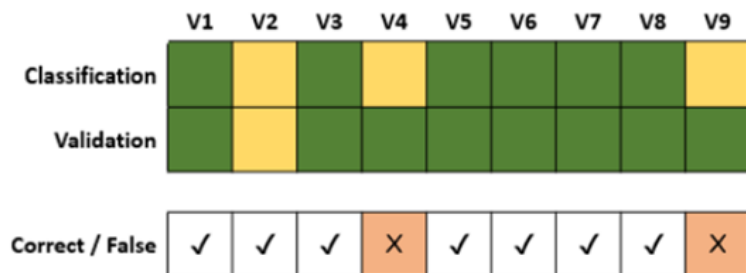


(Humboldt Universität zu Berlin n.d.)



Validation method: Confusion matrix

- The table that compares predicted classifications to actual classes
- Rows represent actual classes
- Columns represent predicted classes
- Diagonal elements indicate correct classifications; off-diagonal elements indicate misclassifications



		Validation		
		Forest	Non-For.	<i>Sum</i>
Class.	Forest	6	0	6
	Non-For.	2	1	3
	<i>Sum</i>	8	1	9

(Humboldt Universität zu Berlin n.d.)



Accuracy metrics derived from confusion matrix

Overall accuracy

- Proportion of correctly classified instances out of the total instances

		Validation		<i>Sum</i>
		Forest	Non-For.	
Class.	Forest	6	0	6
	Non-For.	2	1	3
<i>Sum</i>		8	1	9

$$OA = \frac{6 + 1}{9} = 0.78 = 78\%$$

Overall Accuracy (OA)

$$OA = \frac{\text{Number of correctly classified samples}}{\text{Number of samples}}$$

(Humboldt Universität zu Berlin n.d.)



Accuracy metrics derived from confusion matrix

Producer's accuracy (recall):

- Likelihood that a given land cover of an area on the ground is classified as such

		Validation			UA	Com. Err.
		Forest	Non-For.	Sum		
Class.	Forest	6	0	6	100%	0%
	Non-For.	2	1	3	33%	67%
	Sum	8	1	9		
	PA	75%	100%			
	Om. Err.	25%	0%			

Producer's Accuracy (PA) of class c

$$PA_c = \frac{\text{Number of correctly classified samples of } c}{\text{Sum of samples with true label of } c}$$

Omission error (Om. Err.) of class c

$$Om. Err._c = 100\% - PA_c$$

$$PA_{\text{Forest}} = \frac{6}{8} = 0.75 = 75\%$$

$$Om. Err._{\text{Forest}} = 100\% - 75\% = 25\%$$

(Humboldt Universität zu Berlin n.d.)



Accuracy metrics derived from confusion matrix

User's accuracy (precision):

- Probability that a pixel classified into a given category actually represents that category on the ground

		Validation		Sum	UA	Com. Err.
		Forest	Non-For.			
Class.	Forest	6	0	6	100%	0%
	Non-For.	2	1	3	33%	67%
	Sum	8	1	9		

User's Accuracy (UA) of class c

$$UA_c = \frac{\text{Number of correctly classified samples of } c}{\text{Sum of samples classified as } c}$$

Commission error (Com. Err.) of class c

$$\text{Com. Err.}_c = 100\% - UA_c$$

$$UA_{\text{Forest}} = \frac{6}{6} = 1 = 100\%$$

$$\text{Com.Err.}_{\text{Forest}} = 100\% - 100\% = 0\%$$

(Humboldt Universität zu Berlin n.d.)



Task

Fill out the table using the formulas on the slides before:

Classification	Wetland	Savanna	Urban	Agriculture	Water	Sum	UA [%]	Com.Err. [%]
Wetland	12	3	0	0	1	16	75%	25
Savanna	2	28	1	3	0	34	82.3	17.7
Urban	0	0	15	2	0	17	88.23	11.77
Agriculture	0	2	4	21	0	27	77.77	22.22
Water	0	0	0	0	10	10	100	0
Sum	14	33	20	26	11	104		
PA [%]	85.7	85	75	81	91			
Om. Err. [%]								

Task



Complete an accuracy assessment in R and in QGIS using the land cover classification from lecture 9d



Summary & key takeaways

Accuracy assessment verifies how well your classification matches reality

The **confusion matrix** is the main tool to evaluate classification performance

Key metrics: **Overall Accuracy, User's Accuracy, Producer's Accuracy**, and related error rates

QGIS (SCP plugin) and R (caret, terra) both support practical and reproducible accuracy checks

Reliable results depend on **good validation data** and thoughtful sampling design

Sources



Humboldt-Universität zu Berlin. (n.d.). *9. Accuracy assessment*. Retrieved March 31, 2025, from https://pages.cms.hu-berlin.de/EOL/geors/S09_Accuracy_assessment.html

Supported by:



Federal Ministry
for the Environment, Nature Conservation
and Nuclear Safety



Federal Agency for
Nature Conservation



Thank you for your attention!

Dr. Insa Otte, Hanna Schulten
(on behalf of the EOCap4Africa Team)
and colleagues

insa.otte@uni-wuerzburg.de



MARTIN-LUTHER-UNIVERSITÄT



HALLE-WITTENBERG



INES Ruhengeri
Institute of Applied Sciences

