Map Accuracy Assessment

EON-Workshop

Paul Magdon

University of Applied Sciences and Arts, Göttingen

12/09-16/09 2022

Topics and learning goals

Content:

- Importance and purposes of accuracy assessments
- Historic development of techniques to assess the accuracy of thematic maps
- Current good practice recommendations

Student learning outcomes:

After the meeting students should :

- attain a basic understanding of the statistical concepts used for assessing the accuracy of thematic map products.
- know three sample designs to collect reference data.
- be able to compile and interpret an error matrix following good practice standards.

Topics and learning goals

Content:

- Importance and purposes of accuracy assessments
- Historic development of techniques to assess the accuracy of thematic maps
- Current good practice recommendations

Student learning outcomes:

After the meeting students should :

- attain a basic understanding of the statistical concepts used for assessing the accuracy of thematic map products.
- know three sample designs to collect reference data.
- be able to compile and interpret an error matrix following good practice standards.

Forest Cover Maps of Costa Rica

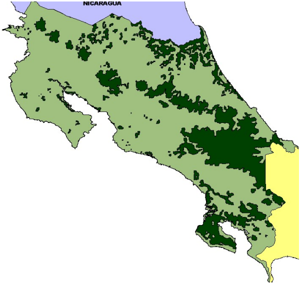

Fig.: Proarca,1992 (Landsat TM)

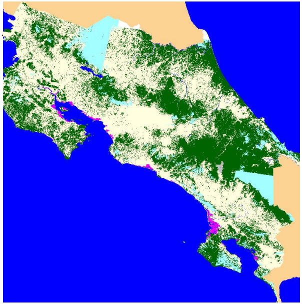

Fig.: Fonafifo,1996/1997 (Landsat TM)

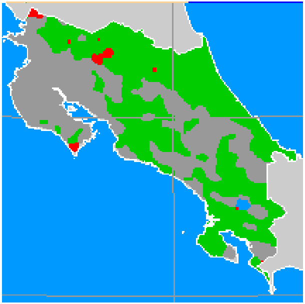

Fig.: FAO FRA 2000 (NOAA AVHRR)

Example provided by Kleinn, C. mod. Magdon, P. (2020)

Forest Cover Maps of Costa Rica

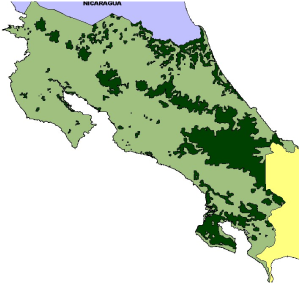

Fig.: Proarca,1992 (Landsat TM)

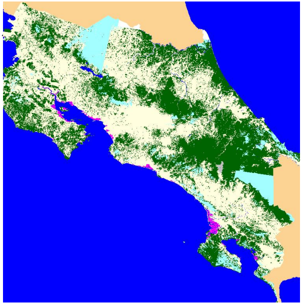

Fig.: Fonafifo,1996/1997 (Landsat TM)

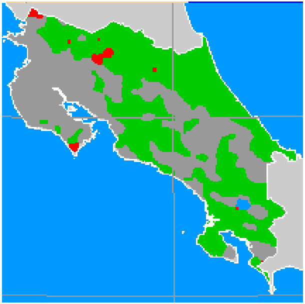

Fig.: FAO FRA 2000 (NOAA AVHRR)

Example provided by Kleinn, C. mod. Magdon, P. (2020)

Forest Cover Maps of Costa Rica

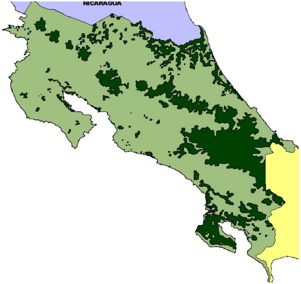

Fig.: Proarca,1992 (Landsat TM)

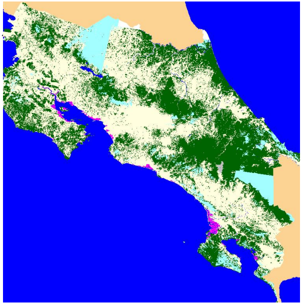

Fig.: Fonafifo,1996/1997 (Landsat TM)

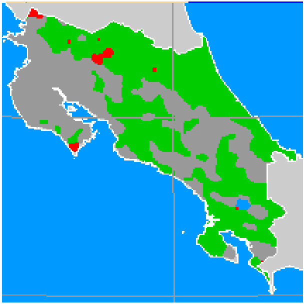

Fig.: FAO FRA 2000 (NOAA AVHRR)

Example provided by Kleinn, C. mod. Magdon, P. (2020)

Forest Cover Maps of Costa Rica

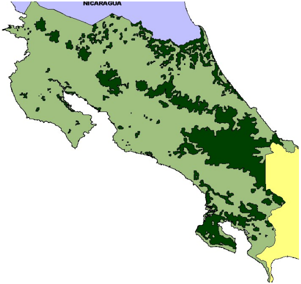

Fig.: Proarca,1992 (Landsat TM)

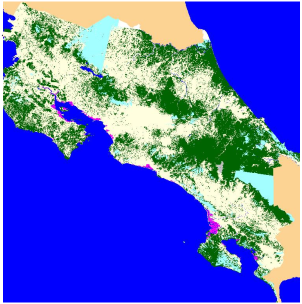

Fig.: Fonafifo,1996/1997 (Landsat TM)

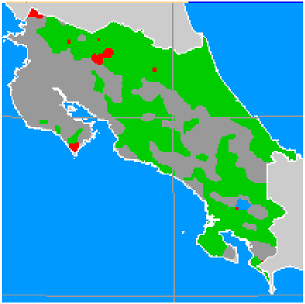

Fig.: FAO FRA 2000 (NOAA AVHRR)

Example provided by Kleinn, C. mod. Magdon, P. (2020)

Accuracy Assessment of Thematic Maps

Maps depicting land cover and other thematic classes are central to numerous scientific and practical applications:

- They are often created by remote sensing image classification
- They can be considered generalized models of the Earth surface
- Different images, classifiers and training datasets produce different maps/models for the same area

When interpreting maps it must be acknowledged that they provide imperfect information with varying quality

Accuracy Assessment of Thematic Maps

Maps depicting land cover and other thematic classes are central to numerous scientific and practical applications:

- They are often created by remote sensing image classification
- They can be considered generalized models of the Earth surface
- Different images, classifiers and training datasets produce different maps/models for the same area

When interpreting maps it must be acknowledged that they provide imperfect information with varying quality

How can the accuracy of a map be assessed?

Error sources in land cover mapping

- Misidentification of land cover classes
  - related to the definition of land cover classes
  - related to erroneous interpretation / classification
- Misregistration of reference and map
  - if the spatial co-registration of reference and map data has a low quality the assessment of thematic accuracy is hampered
- Mixed Pixels
  - relevant if spatial resolution is larger then the objects of interest

Historical stages of accuracy assessments

Visual control

- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control

- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent

- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control

- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent

- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific

- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control
- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent
- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific
- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Error Matrix
- Error matrix is constructed by comparing reference and map at specific locations
- Various accuracy statistics are calculated from the error matrix

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control
- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent
- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific
- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Error Matrix
- Error matrix is constructed by comparing reference and map at specific locations
- Various accuracy statistics are calculated from the error matrix

Samp-ling
- probability sampling designs are used
- standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control	• The map was visually checked and compared to other maps • Quality of the map was judge based on subjective impressions
Extent	• Quality of the map was evaluated by comparing the area covered by the thematic classes to other references • Not site specific, No measures of spatial accuracy
Site specific	• Thematic map was compared to reference data at specific locations • First accuracy metrics (e.g. overall accuracy)
Error Matrix	• Error matrix is constructed by comparing reference and map at specific locations • Various accuracy statistics are calculated from the error matrix
Samp-ling	• probability sampling designs are used • standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control	• The map was visually checked and compared to other maps • Quality of the map was judge based on subjective impressions
Extent	• Quality of the map was evaluated by comparing the area covered by the thematic classes to other references • Not site specific, No measures of spatial accuracy
Site specific	• Thematic map was compared to reference data at specific locations • First accuracy metrics (e.g. overall accuracy)
Error Matrix	• Error matrix is constructed by comparing reference and map at specific locations • Various accuracy statistics are calculated from the error matrix
Samp-ling	• probability sampling designs are used • standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control	• The map was visually checked and compared to other maps • Quality of the map was judge based on subjective impressions
Extent	• Quality of the map was evaluated by comparing the area covered by the thematic classes to other references • Not site specific, No measures of spatial accuracy
Site specific	• Thematic map was compared to reference data at specific locations • First accuracy metrics (e.g. overall accuracy)
Error Matrix	• Error matrix is constructed by comparing reference and map at specific locations • Various accuracy statistics are calculated from the error matrix
Samp-ling	• probability sampling designs are used • standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control
- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent
- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific
- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Error Matrix
- Error matrix is constructed by comparing reference and map at specific locations
- Various accuracy statistics are calculated from the error matrix

Samp-ling
- probability sampling designs are used
- standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control
- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent
- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific
- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Error Matrix
- Error matrix is constructed by comparing reference and map at specific locations
- Various accuracy statistics are calculated from the error matrix

Samp-ling
- probability sampling designs are used
- standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Historical stages of accuracy assessments

Visual control

- The map was visually checked and compared to other maps
- Quality of the map was judge based on subjective impressions

Extent

- Quality of the map was evaluated by comparing the area covered by the thematic classes to other references
- Not site specific, No measures of spatial accuracy

Site specific

- Thematic map was compared to reference data at specific locations
- First accuracy metrics (e.g. overall accuracy)

Error Matrix

- Error matrix is constructed by comparing reference and map at specific locations
- Various accuracy statistics are calculated from the error matrix

Samp-ling

- probability sampling designs are used
- standard errors and confidence intervals are reported

Fig.: Source: Congalton (1994) mod. Magdon, P. (2020)

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
  - Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
  - Probability sampling design for design-based estimators
  - reporting standard errors

(3) Quality assured
  - Evaluate the quality of reference data

(4) Reliable
  - small standard errors
  - high quality of the reference data

(5) Transparent
  - report all details positive or negative

(6) Reproducible
  - clearly describe the sample, plot and estimator design
  - provide the raw error matrix

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
   - Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
   - Probability sampling design for design-based estimators
   - reporting standard errors

(3) Quality assured
   - Evaluate the quality of reference data

(4) Reliable
   - small standard errors
   - high quality of the reference data

(5) Transparent
   - report all details positive or negative

(6) Reproducible
   - clearly describe the sample, plot and estimator design
   - provide the raw error matrix

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
  - Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
  - Probability sampling design for design-based estimators
  - reporting standard errors

(3) Quality assured
  - Evaluate the quality of reference data

(4) Reliable
  - small standard errors
  - high quality of the reference data

(5) Transparent
  - report all details positive or negative

(6) Reproducible
  - clearly describe the sample, plot and estimator design
  - provide the raw error matrix

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
   - Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
   - Probability sampling design for design-based estimators
   - reporting standard errors

(3) Quality assured
   - Evaluate the quality of reference data

(4) Reliable
   - small standard errors
   - high quality of the reference data

(5) Transparent
   - report all details positive or negative

(6) Reproducible
   - clearly describe the sample, plot and estimator design
   - provide the raw error matrix

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
- Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
- Probability sampling design for design-based estimators
- reporting standard errors

(3) Quality assured
- Evaluate the quality of reference data

(4) Reliable
- small standard errors
- high quality of the reference data

(5) Transparent
- report all details positive or negative

(6) Reproducible
- clearly describe the sample, plot and estimator design
- provide the raw error matrix

Criteria for good practices

Stehman & Foody (2019) identified the following criteria to evaluate the quality of accuracy assessments:

(1) Map relevant
   - Error matrix must reflect the area proportions of the study region

(2) Statistically rigorous
   - Probability sampling design for design-based estimators
   - reporting standard errors

(3) Quality assured
   - Evaluate the quality of reference data

(4) Reliable
   - small standard errors
   - high quality of the reference data

(5) Transparent
   - report all details positive or negative

(6) Reproducible
   - clearly describe the sample, plot and estimator design
   - provide the raw error matrix

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban						
Cropland						
Grassland						
Forest						
Water						
Total						

Collection of validation data

General Approach:

- For selected locations in the map the true class is determined
- The true class is compared against the classification result / map
- The comparison is done using a confusion / error matrix
- Statistics to quantify different aspects of the map accuracy are derived from the error matrix

Compilation of the error matrix

Can we use the training data set for compiling the error matrix?

- the classification can be regarded a model
- testing of a model must not be done with the data that was used to fit the model
- we are not interested how good the train data is classified but we want to know how good the model classifies the map pixels
- the error rate of the training data is be no means a good predictor for the error rate of the classification!

In principle there are two options:

(1) data splitting/cross-validation

(2) collection of an independent validation data set

Compilation of the error matrix

Can we use the training data set for compiling the error matrix?

- the classification can be regarded a model
- testing of a model must not be done with the data that was used to fit the model
- we are not interested how good the train data is classified but we want to know how good the model classifies the map pixels
- the error rate of the training data is be no means a good predictor for the error rate of the classification!

In principle there are two options:

(1) data splitting/cross-validation

(2) collection of an independent validation data set

Compilation of the error matrix

Can we use the training data set for compiling the error matrix?

- the classification can be regarded a model
- testing of a model must not be done with the data that was used to fit the model
- we are not interested how good the train data is classified but we want to know how good the model classifies the map pixels
- the error rate of the training data is be no means a good predictor for the error rate of the classification!

In principle there are two options:

(1) data splitting/cross-validation

(2) collection of an independent validation data set

Cross-Validation

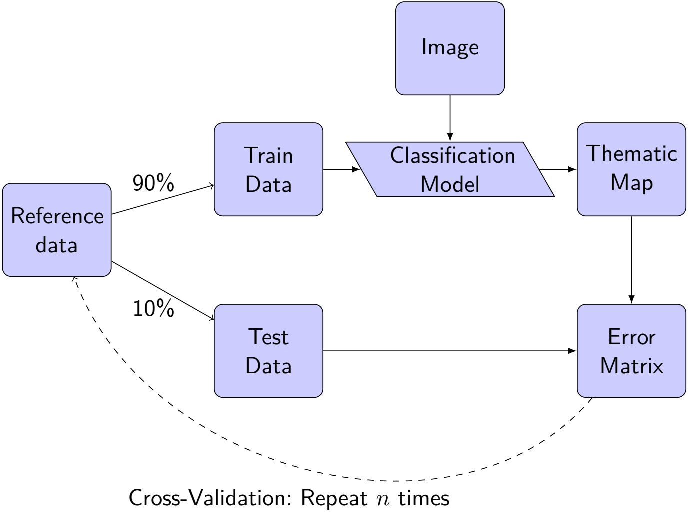

Fig.: Compilation of the error matrix using random cross-validation. Magdon, P. (2021)

Cross-Validation

If the amount of reference data is limited and a split into test and train data cannot be done data splitting techniques are applied.
Cross-validation:

- divide reference data set randomly into k folds
- repeatedly train models on $k-1$ folds
- test against held back test data

Often random cross-validation over estimates the overall accuracy as:

- train data is selected such, that easy to identify and clearly separable pixels are used
- cross-validation assumes that training data is independent, which is often not the case due to spatial auto-correlation of neighboring pixels

Meyer et al. (2019) present an approach for *spatial cross-validation*, which can be used to reduce the effect of spatial auto-correlation.

Independent validation data

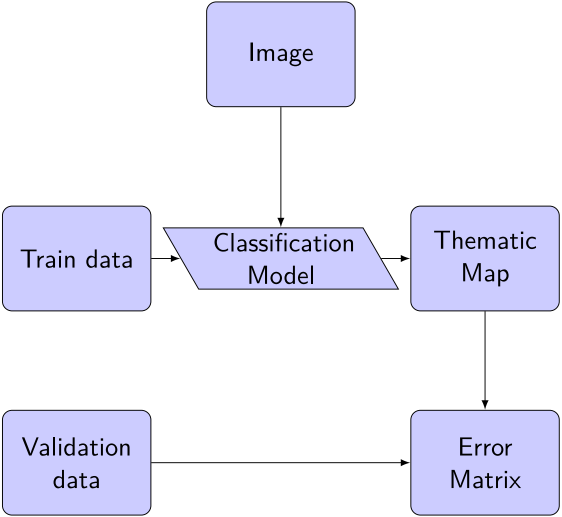

Fig.: Compilation of the error matrix with independent train and validation data sets. Magdon, P. (2021) CC-BY-SA

Collection of independent reference data

Following the good practice guidelines a design-based sampling scheme needs to be implemented for the collection of validation data.

There are three design elements to be considered:

- Plot / Response design
- Sampling design
- Estimators design

Response / Plot Design

The response design defines how a decision on agreement between the predicted map class a and the observed reference class is made (Stehman & Foody, 2019).

The following aspects of the response design can be controlled:

- spatial unit of the assessment
- data source of reference information (field visits, high resolution images, maps)
- labelling protocol
- protocol how thematic agreement will be defined

Sample design

The following sample designs are used:

- *(non-probability / representative /subjective sampling)*
- systematic sample designs
- random sampling

Stratification

To gain efficiency and to ensure sufficient number of samples per class, stratified designs with the thematic map classes defining the strata, are recommended. The allocation of the samples to the strata can be done by:

- equal allocation→Each class has the same number of samples
- proportional allocation→Samples are allocated according to the map proportion of the thematic classes
- optimal allocation→Samples are allocated such that one or multiple aspects of accuracy assessments is optimized (e.g area estimates)

Sample design

The following sample designs are used:

- *(non-probability / representative /subjective sampling)*
- systematic sample designs
- random sampling

Stratification

To gain efficiency and to ensure sufficient number of samples per class, stratified designs with the thematic map classes defining the strata, are recommended. The allocation of the samples to the strata can be done by:

- equal allocation→Each class has the same number of samples
- proportional allocation→Samples are allocated according to the map proportion of the thematic classes
- optimal allocation→Samples are allocated such that one or multiple aspects of accuracy assessments is optimized (e.g area estimates)

Sample design

How many samples do we need?

$$n = \frac{z^2 p(1-p)}{d^2} \tag{1}$$

With $z = 1.96$ for a 95% confidence interval. d is the desired half-width of the confidence interval, and p is the expected overall accuracy.

Example: For a land cover classification where we expect a classification accuracy of 80% we want to estimate the overall accuracy with a precision of $\pm$ 10% at a confidence interval of 95%:

$$n = \frac{1.96^2 0.8(1-0.8)}{0.1^2}$$

$$n = \frac{1.96^2 * 0.8 * 0.2}{0.01}$$

$$n \approx 62$$

Sample design

How many samples do we need?

$$n = \frac{z^2 p(1-p)}{d^2} \tag{1}$$

With $z = 1.96$ for a 95% confidence interval. d is the desired half-width of the confidence interval, and p is the expected overall accuracy.

Example: For a land cover classification where we expect a classification accuracy of 80% we want to estimate the overall accuracy with a precision of $\pm$ 10% at a confidence interval of 95%:

$$n = \frac{1.96^2 0.8(1-0.8)}{0.1^2}$$

$$n = \frac{1.96^2 * 0.8 * 0.2}{0.01}$$

$$n \approx 62$$

Sample design

How many samples do we need?

$$n = \frac{z^2 p(1-p)}{d^2} \tag{1}$$

With $z = 1.96$ for a 95% confidence interval. d is the desired half-width of the confidence interval, and p is the expected overall accuracy.

Example: For a land cover classification where we expect a classification accuracy of 80% we want to estimate the overall accuracy with a precision of $\pm$ 10% at a confidence interval of 95%:

$$n = \frac{1.96^2 0.8(1-0.8)}{0.1^2}$$

$$n = \frac{1.96^2 * 0.8 * 0.2}{0.01}$$

$$n \approx 62$$

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban						
Cropland						
Grassland						
Forest						
Water						
Total						

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban	*150*	*22*	*0*	*13*	*7*	*192*
Cropland	*0*	*730*	*91*	*113*	*12*	*946*
Grassland	*23*	*120*	*320*	*54*	*4*	*531*
Forest	*3*	*18*	*14*	*350*	*14*	*401*
Water	*23*	*12*	*2*	*2*	*350*	*387*
Total	*209*	*902*	*427*	*532*	*387*	*1900*

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban	*150*	*22*	*0*	*13*	*7*	*192*
Cropland	*0*	*730*	*91*	*113*	*12*	*946*
Grassland	*23*	*120*	*320*	*54*	*4*	*531*
Forest	*3*	*18*	*14*	*350*	*14*	*401*
Water	*23*	*12*	*2*	*2*	*350*	*387*
Total	*209*	*902*	*427*	*532*	*387*	*1900*

1. **O**verall **A**ccuracy:

$$OA = \frac{\text{correctly classified points}}{\text{total number of points}}$$

$$OA = \frac{1900}{2457} = 0.773$$

$$OA = 77.3\%$$

How many reference points are correctly classified?

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban	*150*	*22*	*0*	*13*	*7*	*192*
Cropland	*0*	*730*	*91*	*113*	*12*	*946*
Grassland	*23*	*120*	*320*	*54*	*4*	*531*
Forest	*3*	*18*	*14*	*350*	*14*	*401*
Water	*23*	*12*	*2*	*2*	*350*	*387*
Total	*209*	*902*	*427*	*532*	*387*	*1900*

2. **U**ser **A**ccuracy:

$$UA = \frac{\text{correctly classified points of a class}}{\text{total number of map points of a class}}$$

$$UA_{Forest} = \frac{350}{532} = 0.6579$$

$$UA_{Forest} = 65.8\%$$

Which proportion of map pixel are correctly classified in the selected class?

Error Matrix

Classified Map

Reference	Urban	Cropland	Grassland	Forest	Water	Total
Urban	*150*	*22*	*0*	*13*	*7*	*192*
Cropland	*0*	*730*	*91*	*113*	*12*	*946*
Grassland	*23*	*120*	*320*	*54*	*4*	*531*
Forest	*3*	*18*	*14*	*350*	*14*	*401*
Water	*23*	*12*	*2*	*2*	*350*	*387*
Total	*209*	*902*	*427*	*532*	*387*	*1900*

3. **P**roducer **A**ccuracy:

$$PA = \frac{\text{correctly classified points of a class}}{\text{number of reference points in a class}}$$

$$PA_{Forest} = \frac{350}{401} = 0.8728$$

$$PA_{Forest} = 87.3\%$$

Which proportion of reference points are correctly classified in the selected class?

Error Matrix

Tab.: Error matrix for a classification with five thematic classes where rows (i) represent the map classification and columns (j) represent the reference classification; p_{ij} is the proportion of the total reference points with in the cell ij; PA = Producer Accuracy, UA = User Accuracy.

Map	Urban	Crop	Grass	Forest	Water	Total	UA
Urban	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{1+}	p_{11}/p_{1+}
Crop	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{2+}	p_{22}/p_{1+}
Grass	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{3+}	p_{33}/p_{1+}
Forest	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}	p_{4+}	p_{44}/p_{1+}
Water	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{4+}	p_{55}/p_{1+}
Total	p_{+1}	p_{+2}	p_{+3}	p_{+4}	p_{+5}	1	
PA	p_{11}/p_{+1}	p_{22}/p_{+2}	$p_{33}/p+3$	p_{44}/p_{+4}	p_{55}/p_{+5}		

Error Matrix

Map	Urban	Crop	Grass	Forest	Water	Total	UA
Urban	150	22	0	13	7	192	0.781
Crop	0	730	91	113	12	946	0.772
Grass	33	120	320	54	4	531	0.603
Forest	23	12	2	350	14	401	0.873
Water	3	18	14	2	350	387	0.904
Total	209	902	427	532	532	2457	
PA	0.718	0.809	0.749	0.658	0.904		

Tab.: Error matrix expressed in proportions p_{ij}.

Map	Urban	Crop	Grass	Forest	Water	Total	UA
Urban	0.061	0.009	0.000	0.005	0.003	0.078	0.781
Crop	0.000	0.297	0.037	0.046	0.005	0.385	0.772
Grass	0.013	0.049	0.130	0.022	0.002	0.216	0.603
Forest	0.009	0.005	0.001	0.142	0.006	0.163	0.873
Water	0.001	0.007	0.006	0.001	0.142	0.158	0.904
Total	0.085	0.367	0.174	0.217	0.158		
PA	0.718	0.809	0.749	0.658	0.904		

Accuracy Measures

Overall Accuracy:

$$O = \sum_{i=1}^{c} p_{ii} \quad (2)$$

Proportion of area correctly classified.

User Accuracy (UA)

$$UA_i = p_{ii}/pi+ \quad (3)$$

Accuracy conditional on the area mapped of class i.

Producer Accuracy (UA)

$$PA_j = p_{ii}/p+j \quad (4)$$

Accuracy conditional the reference area of class j.

If stratified designs are used the strata weight (w_i) needs to be considered:

$$O = \sum_{i=1}^{c} w_i p_{ii} \quad (5)$$

Accuracy Assessment Report

What to report?

- Sample Design
  - How was the reference data selected?
- Response Design
  - How was the reference label assigned?
- Estimator Design
  - Which accuracy measures are used?
  - Which estimators are used (Strata weights?)?
- Accuracy of the reference data
  - What is the accuracy of the reference data?
- The raw error matrix
  - The raw matrix should be provided

Accuracy Assessment Report

What to report?

- Sample Design
    - How was the reference data selected?
- Response Design
    - How was the reference label assigned?
- Estimator Design
    - Which accuracy measures are used?
    - Which estimators are used (Strata weights?)?
- Accuracy of the reference data
    - What is the accuracy of the reference data?
- The raw error matrix
    - The raw matrix should be provided

Take Home Message

- A rigorous assessment of the map quality is imperious if we want to base decisions on map information.
- Without, each map remains nothing more than a pretty picture representing simply one possible untested hypothesised model (Strahler et al., 2006; McRoberts, 2011; Stehman & Foody, 2019).
- Design-based sampling guaranties unbiased estimates by design!

References

Congalton, R. (1994). Accuracy assessment of remotely sensed data: future needs and directions. In *Proceedings of Pecora*, volume 12 (pp. 383–388).

McRoberts, R. E. (2011). Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sensing of Environment*, 115(2), 715–724.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411(March), 108815.

Stehman, S. V. & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, 231(April), 111199.

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., Mayaux, P., Morisette, J. T., Stehman, S. V., & Woodcock, C. E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *European Communities, Luxembourg*, 51(4).