

Annex 7: Test of Competence

After selecting the top three tenderers we will ask those tenderers to carry out a test of competence. Each tenderer will be asked to generate the results of five different markers. Together with the marker results we also expect a number of quality indicators. This will include a confusion matrix, together with a score for sensitivity, specificity, precision and an F1-score for each marker. For crop classification we will also ask for the user- and producer accuracy.

Markers

The following markers are requested:

- Crop classification (end of season 2019)
- Mowing (from 15th of March until the 1st of October 2019)
- Homogeneity (grassland) (throughout 2019)
- Ploughing, bare soil (throughout 2019)
- Catch crop detection (starting around August 2019)
 - Detection harvest of main crop
 - Increase biomass/presence catch crop
 - Decrease biomass (harvest, ploughing, mechanical consolidation) catch crop

The test of competence will be performed for 2 specific areas in the Netherlands over the year 2019. Only the use of Sentinel 1 and 2 data or open data that is equivalent in terms of spatial resolution is allowed. The tenderers will receive the following data for performing the test of competence:

- Of these 2 areas: the shapefile of all declared crop parcels in 2019 for these two specific areas. Parcels which (should) have a catch crop will also be indicated in this file.
- For entire country: all declared parcels for 2019. This files contains the geometry and crop code of each declared parcel. This data can for example be used as training data for the crop classification.
- A set of Ground Truth data which can be used for training and validation for the markers other than crop classification.

The table below indicates for each marker what must be monitored, which ground truth must be used to validate, what the validation method must be and what the validation criteria are.

Marker	What to monitor	Validation criteria	Validation method
Crop classification	crop type	1st or 2nd detected crop equal to GT	f-score, users- and producers accuracy confusion matrix as mentioned below (TN, FN, FP, TP) for each marker
Mowing	mowing period(s)	Deviation mowing period no more than 1 week	
Homogeneity grass	homogeneity of the parcel	Overlap parcels GT	
Ploughing, detection of bare soil	Bare soil period(s)	Deviation bare soil period no more than 1 week	
Catch crop	Catch crop period	Deviation catch crop period no more than 1 week	

Table 1 - overview of markers, what has to be monitored, how to validate

In order to make an easy comparison between the different tenders we have created a delivery format for the marker results. Given the short timeframe we will also except different formats or multiple files. However, please keep in mind that it has to be possible to calculate the required statistics.

In table 2 we give a description of the required output.

Fieldname	Type	Description
FID	Text	Unique ID of the parcel
CLASS	Numeric	Indication if a particular marker was used for this parcel Yes = 1 No = 0
MOWING	Numeric	Indication if a particular marker was used for this parcel Yes = 1 No = 0
HOMOGENEITY	Numeric	Indication if a particular marker was used for this parcel Yes = 1 No = 0
BARE_SOIL	Numeric	Indication if a particular marker was used for this parcel Yes = 1 No = 0
CATCHCROP	Numeric	Indication if a particular marker was used for this parcel Yes = 1 No = 0
CROP	Text	Declared crop code
CROP_DESCR	Text	Description of the crop
CROP_CLASS_1	Numeric	Probability score found for highest scoring crop (e.g. 0.80)
CROP_FOUND_1	Text	First crop code found
CROP_CLASS_2	Numeric	Probability score found for second highest scoring crop
CROP_FOUND_2	Tekst	Second crop code found
CROP_CLASS_3	Numeric	Probability score found for third highest scoring crop
CROP_FOUND_3	Tekst	Third crop code found
S1_PIX	Numeric	Number of Sentinel 1 pixels used, if used
S2_PIX	Numeric	Number of Sentinel 2 pixels used, if used
TOT_MOW	Numeric	Total number of mowing events. Note that NEA is expecting no more than five mowing events. And no mowing events before the 15 th of March or after the 1 st of October.
MOW_START_DATE_1	Timestamp	Detected start date for a mowing event
MOW_END_DATE_1	Timestamp	Detected end data for a mowing event
MOW_PROB_1	Numeric	Probability of the mowing event.
MOW_START_DATE_2	Timestamp	Detected start date for a mowing event
MOW_END_DATE_2	Timestamp	Detected end data for a mowing event
MOW_PROB_2	Numeric	Probability of the mowing event.
MOW_START_DATE_3	Timestamp	Detected start date for a mowing event
MOW_END_DATE_3	Timestamp	Detected end data for a mowing event
MOW_PROB_3	Numeric	Probability of the mowing event.
MOW_START_DATE_4	Timestamp	Detected start date for a mowing event
MOW_END_DATE_4	Timestamp	Detected end data for a mowing event
MOW_PROB_4	Numeric	Probability of the mowing event.
MOW_START_DATE_5	Timestamp	Detected start date for a mowing event
MOW_END_DATE_5	Timestamp	Detected end data for a mowing event
MOW_PROB_5	Numeric	Probability of the mowing event.
BARE_SOIL_NUM	Numeric	Number of detected bare soil events.
BARE_SOIL_START_DATE_1	Timestamp	Detected start date for a mowing event
BARE_SOIL_END_DATE_1	Timestamp	Detected end data for a mowing event
BARE_SOIL_PROB_1	Numeric	Probability of the mowing event.
BARE_SOIL_START_DATE_2	Timestamp	Detected start date for a mowing event
BARE_SOIL_END_DATE_2	Timestamp	Detected end data for a mowing event

BARE_SOIL_PROB_2	Numeric	Probability of the mowing event.
HOMOGENEITY_GRASS	Numeric	Indication if the parcel is homogeneous Yes = 1 No = 0
CATCH_CROP_PRESENT	Numeric	Indication if a catch crop is present Yes = 1 No = 0
CATCH_CROP_START_DATE	Timestamp	Detected start date of the catch crop
CATCH_CROP_END_DATE	Timestamp	Detected end date of the catch crop
RESULT	Numeric	Indication if the parcel is monitorable Yes = 1 No = 0
REASON_RESULT	Text	If the RESULT = 0 an explanation why: Not enough pixels, parcel too small, etc.

Table 2: delivery format for the marker results

Deliverables:

We expect the following deliverables:

- A shapefile or GPKG containing all parcels and marker results
- For each event marker an Excel file containing a confusion matrix similar to the one explained below.
- For crop classification we expect an Excel file containing information about the overall accuracy, kappa, producer's accuracy, user's accuracy, F-Score for max. the 10 main crops.

In order to be able to calculate some of the statistics we will provide the tenderers with ground truth data.

Assessment criteria:

For crop classification we expect the following outputs:

Validation value	Description	Formula
Overall Accuracy (OA)	It gives the overall accuracy of the classification	$\frac{\text{Well classified parcels}}{\text{Total validation parcels}}$
Kappa	It gives the indication on how much better the classification is compared to a random model	$\frac{OA - \text{random accuracy}}{1 - \text{random accuracy}}$
Producer's accuracy (by crop type)	It gives for each crop type the chance that a crop which is that crop type in reality is classified as such	$\frac{\text{Well classified parcels}}{\text{Parcels in reality}}$
User's accuracy (by crop type)	It gives for each crop type the chance that the classified crop on the map is this crop type in reality	$\frac{\text{Well classified parcels}}{\text{Classified parcels}}$
F-Score (by croptype)	It combines both producer's and user's accuracies in a single value	$2 * \frac{PA * UA}{PA + UA}$

For each event marker a confusion matrix like the one presented below is expected to be delivered.

		Predicted		
		Negative (0)	Positive (1)	
Actual	Negative (0)	True Negative TN	False Positive FP (Type I – error)	Specificity $= \frac{TN}{TN + FP}$
	Positive (1)	False Negative FN (Type II – error)	True Positive TP	Recall, Sensitivity, True Positive Rate (TPR) $= \frac{TP}{TP + FN}$
		Accuracy $= \frac{TP + TN}{TP + TN + FP + FN}$	Precision, Positive predictive value (PPV) $= \frac{TP}{TP + FP}$	F1-Score $= 2 \times \frac{Recall \times Precision}{Recall + Precision}$

All marker results will be assessed based on the thresholds presented in table 3. In order to be able to judge all results in the same way we NEA will use a separate ground truth dataset. Meeting the thresholds will earn the Tenderer points. If the thresholds are not met the Tenderer will earn no points. Points will also be awarded for the presence of each marker, the correct calculation of the statistics and the correct validation methodology. Also see chapter 5.7 of the Tender Document.

marker	Quality criteria
Crop classification	overall accuracy $\geq 95\%$
Mowing events	FN $\leq 20\%$, FP $\leq 20\%$
Homogeneity of grassland	FN $\leq 20\%$, FP $\leq 20\%$
Plowing/bare soil	FN $\leq 20\%$, FP $\leq 20\%$
Harvest main crop	FN $\leq 20\%$, FP $\leq 20\%$
Catch crop	FN $\leq 20\%$, FP $\leq 20\%$

Table 3: quality thresholds for each marker