Supplementary Materials: Outlier detection at the parcel-level in wheat and rapeseed crops using multispectral and SAR time series

Florian Mouret[1,2,*], Mohanad Albughdadi [1], Sylvie Duthoit [1], Denis Kouamé [3], Guillaume Rieu [1] and Jean-Yves Tourneret [2]

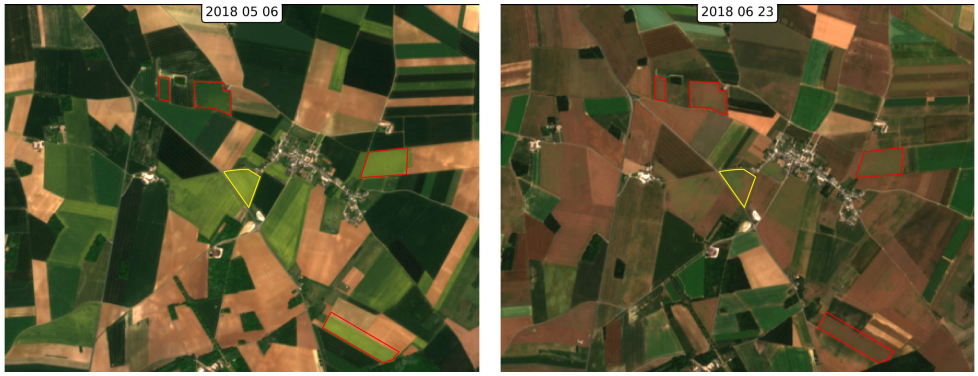
1. Complementary examples of outlier parcels

This section shows some anomalies identified by agronomic experts, provided to illustrate the labeling process of this study in complement of the examples already provided in the main document (Figs. 3 to 7). In what follows, NDVI time series are mostly represented because their interpretations are generally easier and straightforward but the agronomic expert had at his disposal all the features (S1 and S2) and all the multispectral images acquired within a growing season.

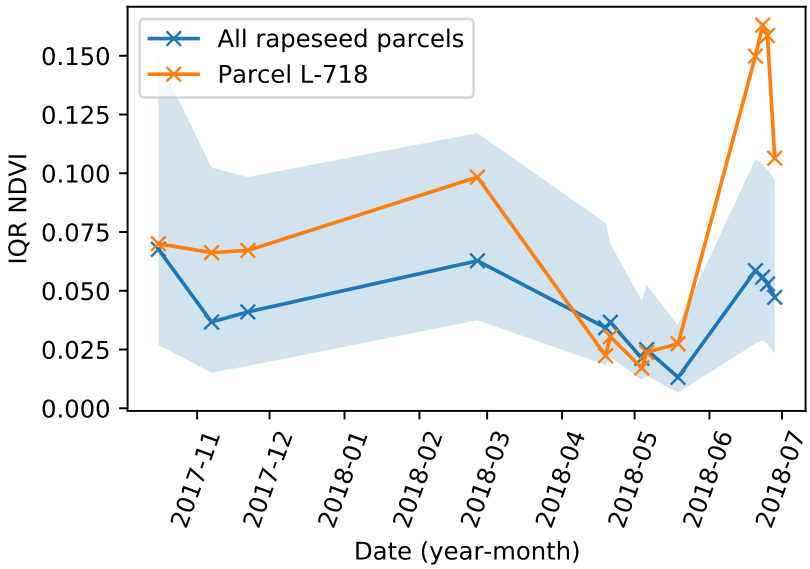
1.1. Heterogeneity

In Fig. 3 of the main document, a case of global heterogeneity was presented. Figure S1 depicts an example of heterogeneity where two different parts are clearly visible. This example shows that it is difficult to decide if wrong boundaries were provided, if two different varieties of rapeseed were sowed or if soil differences led to heterogeneity. Figure S2 shows an example of heterogeneity after senescence, where the parcel is perfectly homogeneous in May whereas it presents some heterogeneity in June. As explained in the paper, this phenomenon is likely to be related to differences in the soil water content and could be of interest for inter season analysis. Figure S2(b) shows the IQR NDVI time series of the same parcel (in yellow) and shows that late heterogeneity can be visible and that its IQR value is greater than 90% of the rest of the data. The same observations can be made for early heterogeneity. In order to compare the analyzed parcel to the rest of the data, the median, the 10th and the 90th percentiles of the whole dataset were displayed (similarly to a boxplot visualization). This representation allows the agronomic expert to know if the observed parcel has indicator values higher (or lower) than 90% of the data.

Figure S1. A rapeseed crop parcel (yellow boundaries) affected by a two-part heterogeneity. The left image was acquired in February 2018 and the right image in April 2018.



(a)



(b)

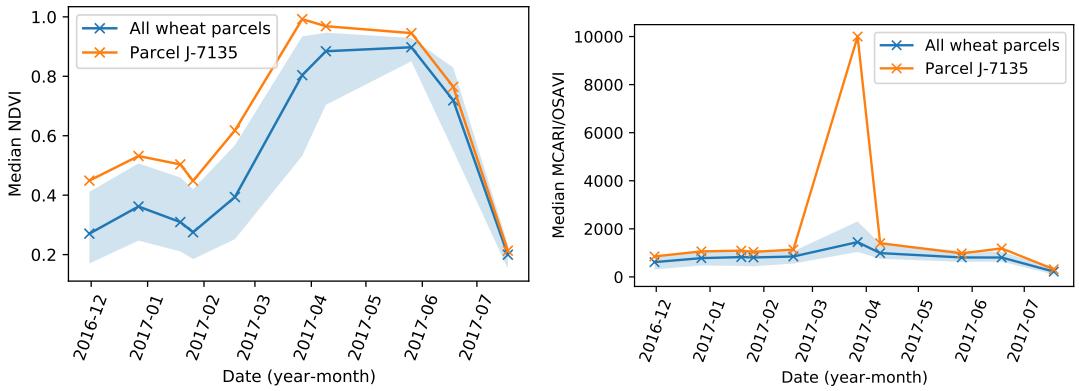
Figure S2. (a) A rapeseed crop parcel (yellow boundaries) affected by a heterogeneity after senescence. The left image was acquired in May and the right image in June. (b) Associated Interquartile Range (IQR) of the parcel NDVI time series. The blue line is the median value of the whole dataset. The blue area is filled between the 10th and 90th percentiles.

1.2. Growth anomalies

Growth anomalies are a bit more difficult to analyze when looking at a single composite multispectral image and visualizing the corresponding time series is in general more relevant. Examples of late growth affecting rapeseed and wheat parcels are shown in Figure 8 and 15 of the main document.

We noticed that in the S2 images acquired in March 27 2017, a small amount of vigorous wheat parcels have the majority of their red pixels equal to zero, causing extreme values of the S2 features as illustrated in Figure S3. This issue was caused by the MAJA processing[1] and could be easily fixed using another processing chain like Sen2core or with a threshold for the red band. Since this phenomenon affected only a small amount of vigorous parcels, it did not change the quality of the detection results. Figure S4 shows the S2 true color composite of the parcel analyzed in Figure S3, which is darker than its neighbors.

[1] https://labo.obs-mip.fr/multitemp/using-ndvi-with-atmospherically-corrected-data/, online accessed 10 March 2020

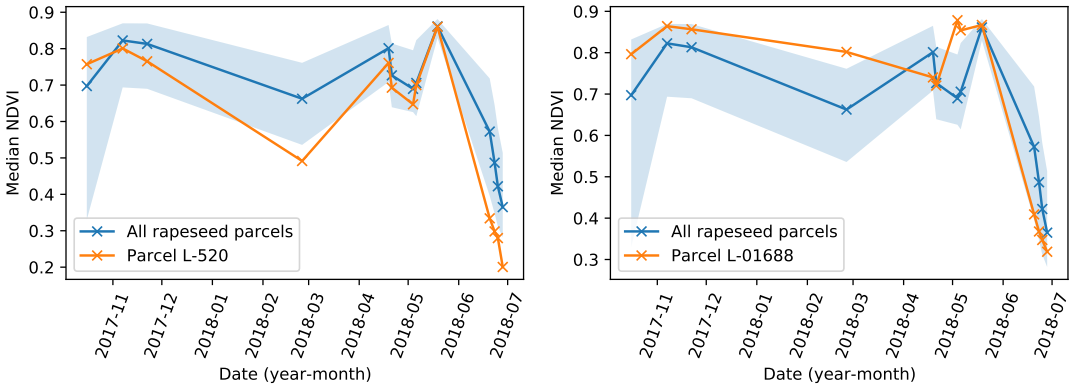


(a) **(b)**

Figure S3. Example of time series subjected to a red channel problem in March 27 for a wheat parcel: (a) median NDVI and (b) median MCARI/OSAVI. The blue line is the median value of the whole dataset. The blue area is filled between the 10th and 90th percentiles. The orange line is a specific parcel analyzed.

Figure S4. A rapeseed crop parcel affected by a red channel problem (yellow boundaries), which is also highly vigorous. A parcel with a late growth can be observed at the bottom of the image (triangle red boundaries).

Figure S5.a displays an example of early senescence. As mentioned in the paper this type of anomaly is more subtle than a global heterogeneity or a late growth problem because it occurs only at the end of growing season. In this particular example, a low median NDVI is also observed in February. It could be a problem during winter but it is challenging to confirm abnormality with only one value of NDVI (S1 time series are valuable in that case). Figure S5.b is an example of early flowering. This case is also subtle and was not observed frequently. In general, the early flowering category is also affected by early senescence.



(a) **(b)**

Figure S5. Time series of median NDVI for a rapeseed parcel presenting sign of (a) early senescence and (b) early flowering. The blue line is the median value of the whole dataset. The blue area is filled between the 10th and 90th percentiles. The orange line is a specific parcel analyzed.

1.3. Non-agronomic anomalies

Figure S6 displays examples of two different types of anomalies. The yellow parcel is an example of a parcel affected by shadows. A case of a too small parcel is also visible in the image, the particular form of the boundaries causes difficulties to filter the parcel only using its area.

Figure S6. Rapeseed parcels: the parcel with yellow boundaries is affected by shadow caused by the trees located next to the parcel. Also, at the bottom a too small parcel is visible.

2. Complementary information about SAR images and their anomalies

A strong correlation between SAR and plant vigor (late growth / vigorous crop) was observed in this study. Figure S7 illustrates the effect of late growth on S1 features. Figure S8 shows that SAR images are not always affected by heterogeneity within the crop parcel, as highlighted in the main document. Heterogeneity is detectable with SAR features when the crop structure is affected, which is undestrandable considering the nature of the sensor.

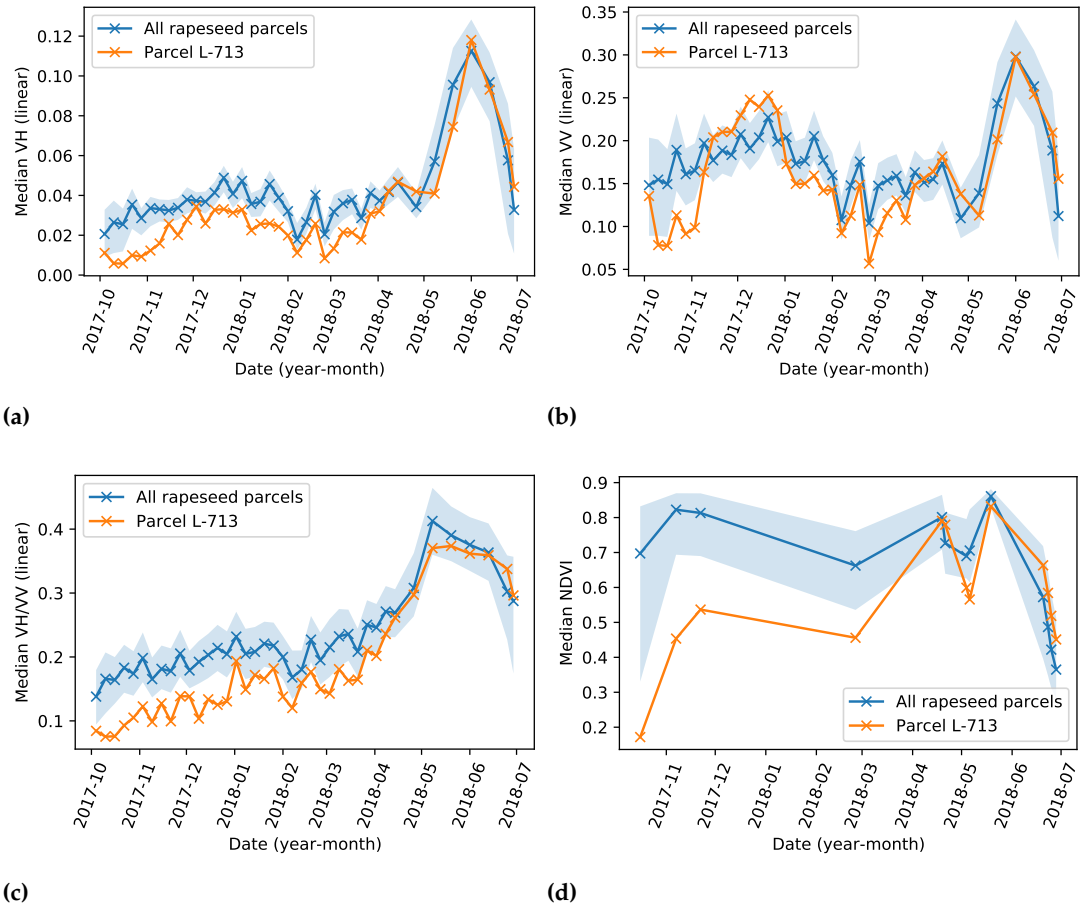


Figure S7. Time series of median SAR features (VV, VH, VH/VV) and median NDVI for a rapeseed parcel. The blue line is the median value of the whole dataset. The blue area is filled between the 10th and 90th percentiles. The orange line is a specific parcel analyzed.(a): median VH, (b): median VV, (c): median ratio VH/VV, (d): median NDVI

Figure S8. Example of a parcel of rapeseed crop (yellow boundaries) where heterogeneity occurs almost during the complete season. Some other parcels show some signs of heterogeneity too. Left: true color S2 image acquired in May, right: composite SAR image (green channel is VV polarization and red channel is VH polarization) acquired in May with multi-temporal speckle filtering.

3. Hyperparameter tuning of the outlier detection algorithms

This section summarizes the different hyperparameter tunings that have been considered for each algorithm leading to the values reported in Table 1.

For the LoOP algorithm, the number of nearest neighbors knn was fixed by grid search to $k = 701$. The extent parameter of LoOP was fixed to $\lambda = 2$ as recommended in Constantinou [1].

For the OC-SVM algorithm, an efficient heuristic [2,3] consists of estimating the parameter σ as the median of the pairwise Euclidean distances between vectors from the learning set $\mathcal{X}$, denoted as median$(\text{dist}(\mathcal{X}))$. This estimator of σ provided good results without a need to a manual tuning for each new dataset.

The IF algorithm was used with a number of iTrees equal to $n_{trees} = 1000$ and a subsampling fixed to $\text{n}_{\text{samples}} = 256$ as in the original paper [4]. Changing these two parameters did not have a significant effect on the results, which is a crucial advantage compared to the other algorithms.

The parameters of the AE were tuned by grid search. We considered a classical structure similar to the one proposed in the Python library for outlier detection PyOD [5]: 4 hidden layers with respectively 64, 32, 32 and 64 neurons. A Relu activation function was used for all layers except for the output using a sigmoidal function. The regularization parameter of the layers (referred to as "activity regularizer" in Keras) was set to 10^{-3}. Note that this specific regularization significantly improved the detection results, contrary to changes in the network structure (e.g., number of neurons).

Because they are using distances, the OC-SVM, LoOP and AE algorithms also require a normalization in order to have input features in the interval $[0, 1]$, while this step is not mandatory when using the IF algorithm.

Table 1: Hyperparameters used in the different algorithms

Algorithm	Hyperparameter	Value
IF	n_{trees}	1000
	$\text{n}_{\text{samples}}$	256
LoOP	k	701
	λ	2
AE	hidden neurons	64, 32, 32, 64
	output regularization	10^{-3}
OC-SVM	σ	median$(\text{dist}(\mathcal{X}))$

4. Complementary results

4.1. Effect of the algorithm used for crop anomaly detection

The performance of the different outlier detection algorithms (AE, LoOP, OC-SVM and IF) was first tested for a complete season analysis. A first experiment was made by computing precision vs. outlier ratio curves for the different algorithms. The OC-SVM algorithm does not provide a unique outlier score associated with each parcel as it depends on the maximum amount of outliers defined by the parameter ν. For the other algorithms, an anomaly score is attributed to each parcel and it is then possible to choose the outlier ratio by sorting the parcels according to their anomaly score (in order to select an appropriate percentage of parcels to be detected). Figure S9 shows that all the outlier detection algorithms provide similar precision for outlier ratios lower than 30% with a majority of relevant anomalies detected (more than 90%), confirming that multiple methods can lead to similar accuracy. The outlier ratio is plotted up to 0.5 to highlight the differences when detecting a large amount of outliers, which is possible for a complete growing season as various transient anomalies can occur. IF and AE perform slightly better overall with a higher AUC and the IF outlier score is more relevant for outlier ratio greater than 30%.

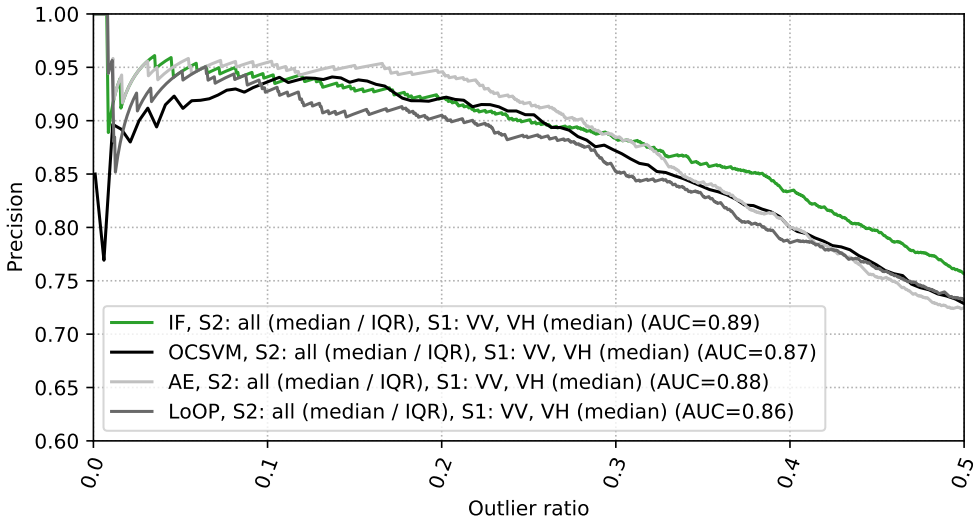


Figure S9. Precision vs. outlier ratio for a complete growing season analysis of the rapeseed parcels. Various algorithms are compared using all S1 and S2 features.

To analyze potential differences within the anomaly categories detected, the 4 algorithms were run with an outlier ratio fixed to 20%. The percentages of the detected parcels within each category for LoOP, OC-SVM, AE and IF are displayed in Figure S10. Overall, all the algorithms detect a majority of heterogeneity (34% for IF) and late growth (25% for IF) anomalies, which can be understood as these anomalies generally affect the complete growing season. It also appears that almost all the "wrong type" and "wrong shape" of the dataset are detected, which seems also logical as these anomalies strongly affect the parcel time series. The histograms obtained with LoOP and IF are very similar, whereas OC-SVM and AE seems to detect less heterogeneity anomalies (44 heterogeneous parcels detected by IF are not detected by OC-SVM) in favor of anomalies related to delay in growth (late growth, vigorous crop, senescence anomalies). It is reasonable to detect heterogeneity issues having a global impact on the crop season before senescence problems, which occurred only at the end of the season. To that extent, results obtained with the LoOP and IF algorithms are more interesting.

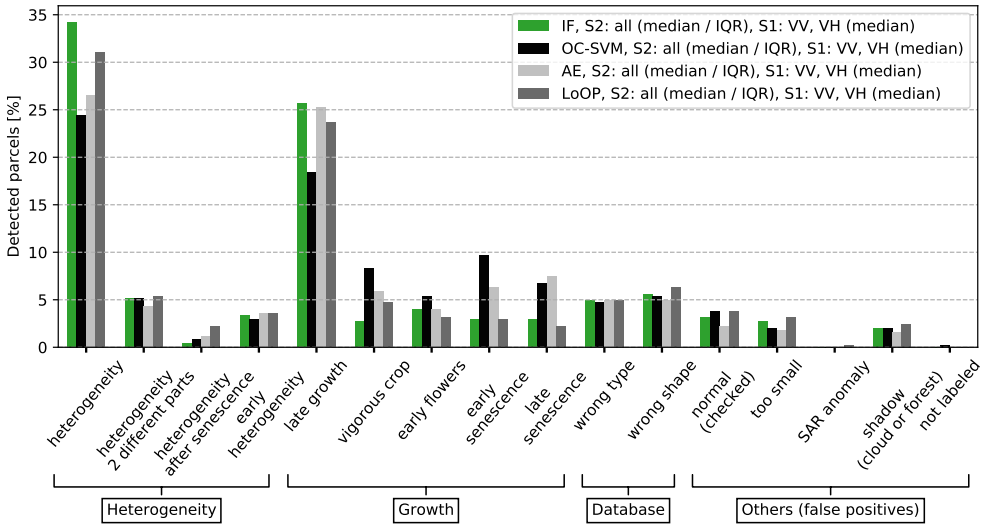


Figure S10. 100×(Number of detected parcels in each category / Number of detected parcels). Rapeseed parcels are analyzed with various outlier detection algorithm and with an outlier ratio equal to 20%.

4.2. Effect of the feature set used for crop anomaly detection

Figure S11 shows the precision against outlier ratio for a selection of the best feature sets resulting from S1 and S2 data using the IF algorithm. The best AUC is obtained when using all S1 and S2 features jointly. Note that S1 data allow a larger amount of true positives to be detected accurately. However, when using S2 features only (or even only NDVI), a similar precision is reached for outlier ratios lower than 30% (even if the percentage of detected parcels in each category is not identical, see details below). Using all S2 features instead of NDVI only increases slightly the precision of the results for outlier ratio higher than 40%, confirming that NDVI is relevant to characterize efficiently a growing season but that some anomalies are better described using all 5 VIs. Finally, a lower AUC is obtained when using NDVI and S1 data when compared to the 5 VIs combined with VH and VV.

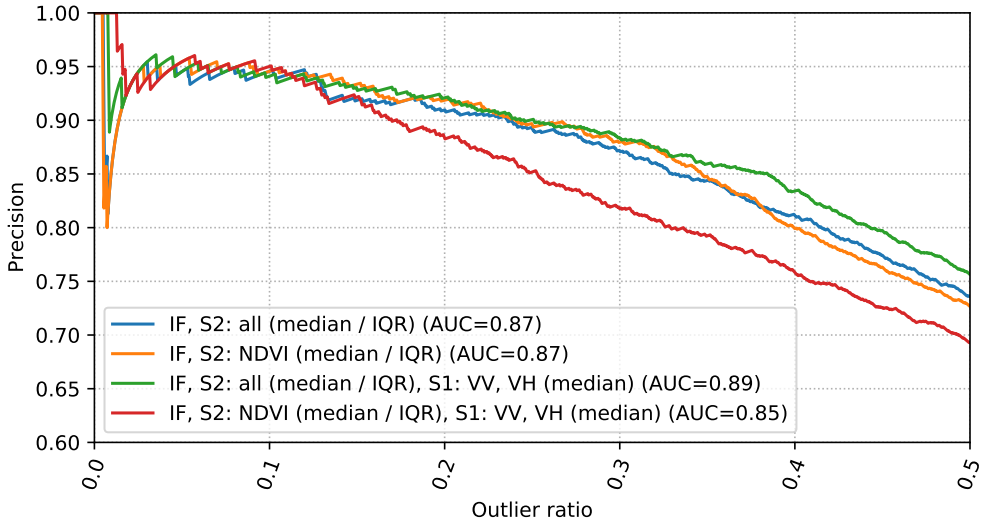


Figure S11. Precision vs. outlier ratio for a complete growing season analysis of the rapeseed parcels. Various sets of features using the IF algorithm are compared.

The percentages of detected parcels within each category for an outlier ratio of 20% are depicted in Figure S12. Again, the most frequent detected anomalies are due to heterogeneity and late growth. More growth anomalies (late/vigorous growth) and less heterogeneity are detected when S1 data is used. Results obtained when using all S2

features or NDVI only are close to each other in this example. However, the subsets of detected parcels by each configuration are not identical (55 parcels detected by one set of features are not detected by the other). Note that more false positives are detected when using NDVI only and S1 features, which leads to a precision of 89.4% whereas it is close to 93% for the other feature sets. The results of this section confirm the relevance of using median and IQR of all 5 S2 features jointly with median of VV and VH S1 features for detecting anomalous crop development.

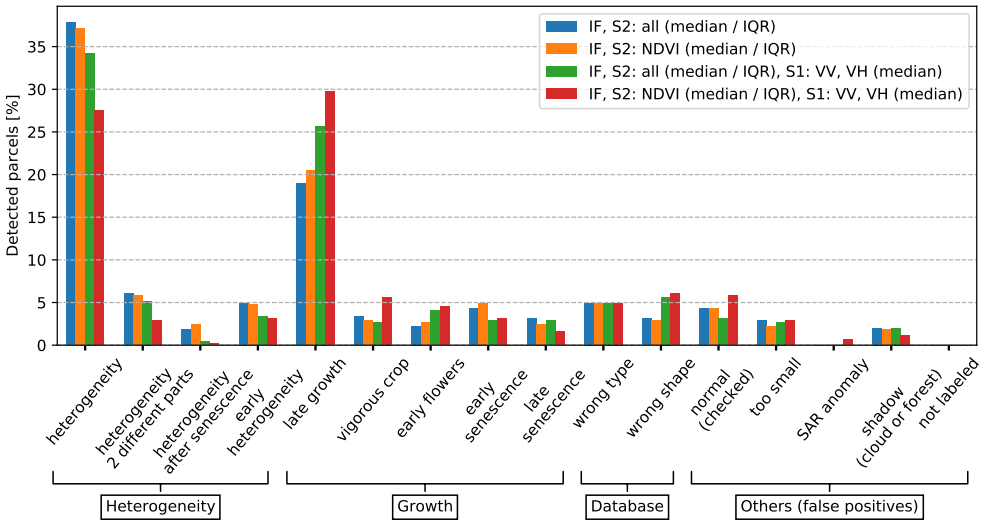


Figure S12. 100×(Number of detected parcels in each category / Number of detected parcels). Various sets of features are compared with the IF algorithm and an outlier ratio equal to 20% for a complete growing season analysis (rapeseed crops).

4.3. Effect of the outlier ratio

Three experiments were run using the median and IQR statistics derived from S2 images, the median statistics derived from S1 images and the IF algorithm, varying the outlier ratio in $\{0.1, 0.2, 0.3\}$. The percentages of detected parcels in the different anomaly categories for each of these experiments are depicted in Figure S13. For an outlier ratio of 10%, the detected anomalies are mostly concentrated in wrong types, late growth and global heterogeneity which is relevant and confirms the observations made in the main document of this study. Moreover, for this outlier ratio, 45% of the detected parcels belong to the category referred to as "global heterogeneity", which is coherent since this type of anomaly is (generally) strongly affecting the crop development of the parcels. Increasing the outlier ratio allows anomalies affecting smaller time periods of the season to be detected, such as early flowering and senescence problems in accordance to the observation made during labeling. For an outlier ratio of 30%, much more false positives are detected (parcels labeled as normal). These results show that the IF algorithm provides a relevant anomaly score since more severe anomalies have higher anomaly scores. Moreover, because the score given by IF is computed only once, there is no need to run the algorithm several times when changing the outlier ratio and the amount of parcel to be detected can be easily adapted to the users' needs. Finally, for a generic analysis, choosing an outlier ratio of 20% is a good balance between the precision of the detection results and the amount of parcel to be detected.

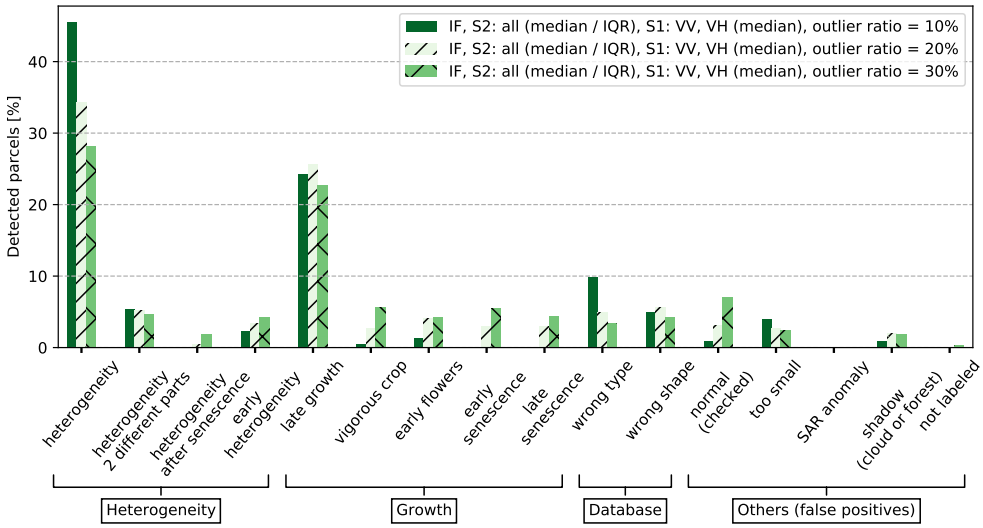


Figure S13. 100×(Number of detected parcels in each category / Number of detected parcels). Various outlier ratio are tested with the same set of features and the IF algorithm for a complete growing season analysis (rapeseed crops).

4.4. Effect of adding new statistics for S2 data

All the previous experiments were conducted using the median and IQR of S2 data as statistics computed at the parcel-level. This section investigates two new statistics, namely the skewness and kurtosis (*i.e.*, the normalized third and fourth order moments of the features). Figure S14 shows the precision vs. outlier ratio when using the IF algorithm and these two additional statistics computed from S2 images to detect anomalies in rapeseed parcels. All the parcels are labeled for outlier ratios that are at least smaller than 10% (less tests were made with skewness and kurtosis statistics as poor results were obtained). It can be observed in this figure that even for an outlier ratio lower than 5%, using skewness and kurtosis statistics leads to a significant difference in the precision results. One issue encountered when using these new statistics is the detection of too subtle anomalies that are not always related to agronomic anomalies. Using the median only is also tested but provides a lower average precision score. This analysis confirms the importance of

IQR statistics, which allows a larger number of relevant anomalies to be detected, and in particular heterogeneity problems. This section showed that using median and IQR statistics of S2 features computed at the parcel level is recommended for crop monitoring.

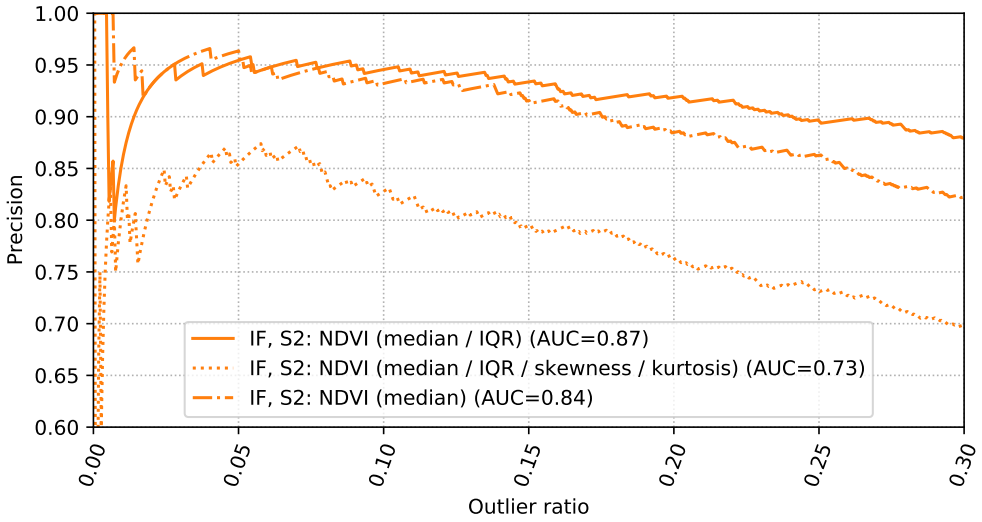


Figure S14. Precision vs. outlier ratio for a complete growing season analysis of the rapeseed parcels. Various statistics of the NDVI are compared using the IF algorithm.

4.5. Comparing various S1 feature sets

Different sets of S1 features have been tested by changing the S1 processing chain presented in Figure 4 of the main document. Results obtained for a complete growing season analysis of the rapeseed crop with different S1 features are presented in Figure S15. Best results are obtained with the S1 features used in the main document.

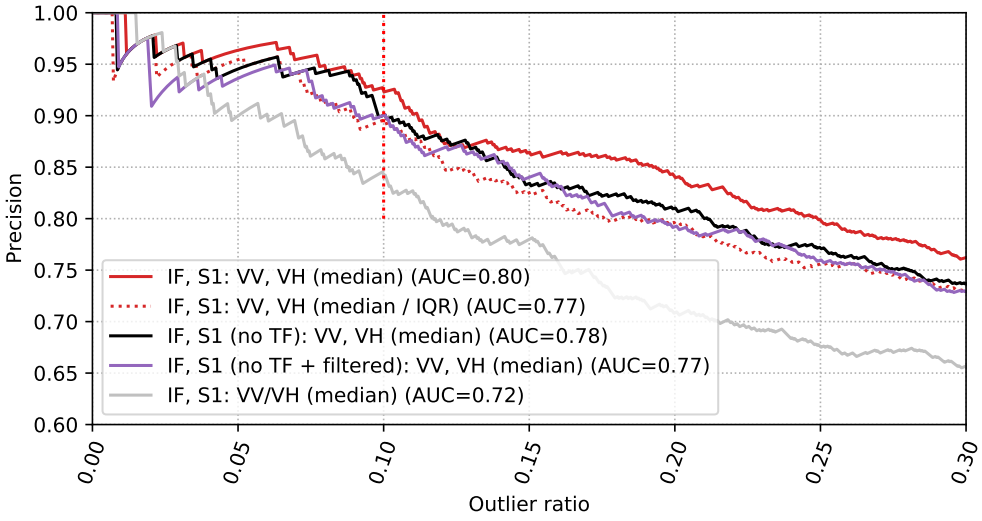


Figure S15. Precision vs. outlier ratio for a complete growing season analysis of the rapeseed parcels. Various sets of S1 features are compared using the IF algorithm. "No TF" means that the Terrain Flattening operation was removed from the S1 processing chain. "Filtered" means that a multi-temporal speckle filtering step was added to the processing chain.

4.6. Effect of missing S2 images

Two scenarios were investigated to evaluate the effect of missing S2 images.

- Scenario 1: the proposed approach was investigated using 6 S2 images instead of 13 to analyze the influence of a reduced amount of S2 images through the season. Only 1 image out of 2 was considered for the detection (the first S2 image was not used, the second S2 image was used and so on). Precision vs. outlier ratio curves are presented in Figure S16, where it can be observed that the proposed method is robust to missing S2 images.

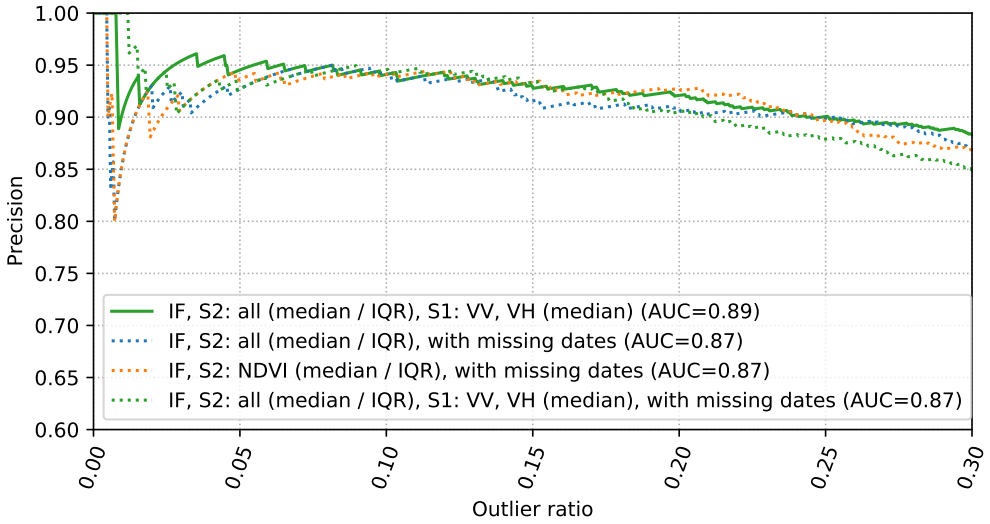


Figure S16. Precision vs. outlier ratio for a complete season analysis of the rapeseed dataset. Missing dates means that only 1 S2 image out of 2 was taken (6 S2 images instead of 13).

- Scenario 2: another experiment was conducted to evaluate the effect of missing S2 images during the first part of the growing season (e.g, more clouds during winter). Precisely, we consider only 7 dates of S2 data between May and June that are used jointly with all S1 images. Precision vs. outlier ratio curves are presented in Figure S17. In that case, using S1 images improve significantly the precision of the results. The reason is that using S1 features allows the algorithm to detect almost the same amount of late growth crops when compared to using a complete season of S2 images which is understandable since S1 data are well suited to detect growth anomalies. These results confirm the interest of using S1 features as a complement to S2 sparse time series.

4.7. Mid-season analysis

A mid-season analysis (using only dates before February) was conducted for multiple reasons detailed in the main manuscript of this study (see Section 3.5). A first experiment was made with the best sets of features selected in Section 4.2 for a complete season analysis using rapeseed parcels. Results displayed in Figure S18 show that even with a small number of images, many agronomic anomalies are detected (best precision=87.7% for an outlier ratio equal to 20%). This confirms the previous results found in the case of missing S2 images. Figure S18 also shows that the best results are again obtained using all S1 and S2 features jointly with a higher average precision since more actual anomalies are detected for larger outlier ratios (*e.g.*, the precision is 5% better for an outlier ratio fixed to 30%).

The impact of a mid-season analysis regarding the different categories of detected anomalies is depicted in Figure S19. In this case, almost no senescence problems are detected, which is easy to understand. Even with only 3 S2 images acquired between October and December, most other agronomic anomalies are detected by the algorithm. A mid-season analysis is able to detect more late growth anomalies and fewer heterogeneous

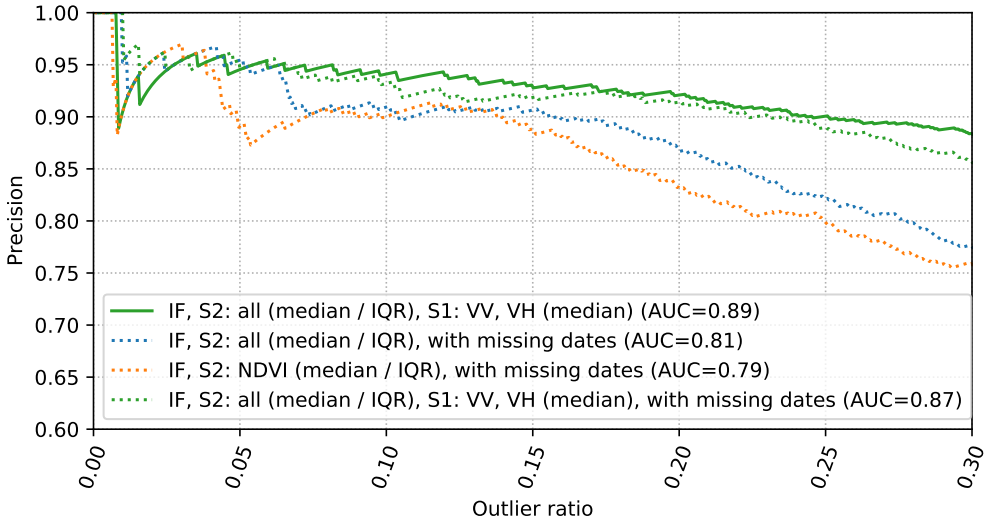


Figure S17. Precision vs. outlier ratio for complete season analysis of the rapeseed dataset. Missing dates means that only the S2 images acquired after April were used (7 images).

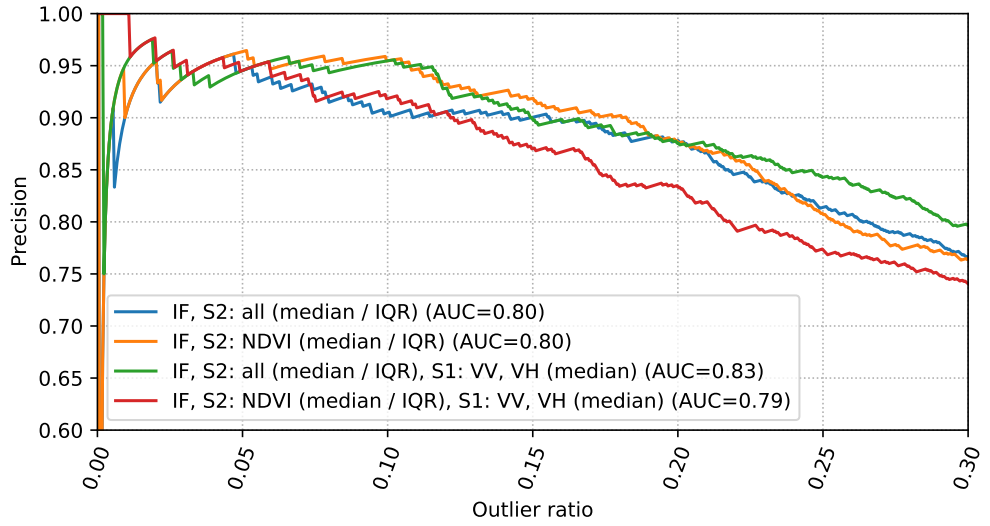


Figure S18. Precision vs. outlier ratio for a mid-season analysis of rapeseed parcels (all images available before February). Various sets of features are compared using the IF algorithm.

parcels because late growth is impacting mostly the beginning of the season (especially for rapeseed crops). Finally, more false positives are detected with a mid-season analysis, which can be understood since the amount of potential anomalies to be detected is lower.

Complementary results for a mid season analysis are briefly presented in what follows since they confirm the observations made for a complete growing season analysis. The IF algorithm provides overall better results (AUC=0.83) and is more robust to changes. The AE performs slightly worse than IF (AUC=0.81), especially for outlier ratios greater than 20%. OCSV (AUC=0.79) and LoOP (AUC=0.77) perform significantly worse in this case. These differences in performance can be explained by the fact that the parameters of OC-SVM, LoOP and AE algorithms are more difficult to tune compared to the IF algorithm. Regarding the influence of the outlier ratio, as for a complete season analysis increasing its value logically leads to detect more subtle anomalies (i.e., affecting a limited time interval) and more false positives, which confirms the relevance of the anomaly score given by IF. Almost no early heterogeneity and vigorous crop is detected with an outlier ratio of 10%. Early heterogeneity is a more subtle anomaly than global heterogeneity, which confirms separation between these two categories. Finally, when using S1 data only, the detection results obtained for an outlier ratio of 10% are still accurate with a precision equal to 89.6%.

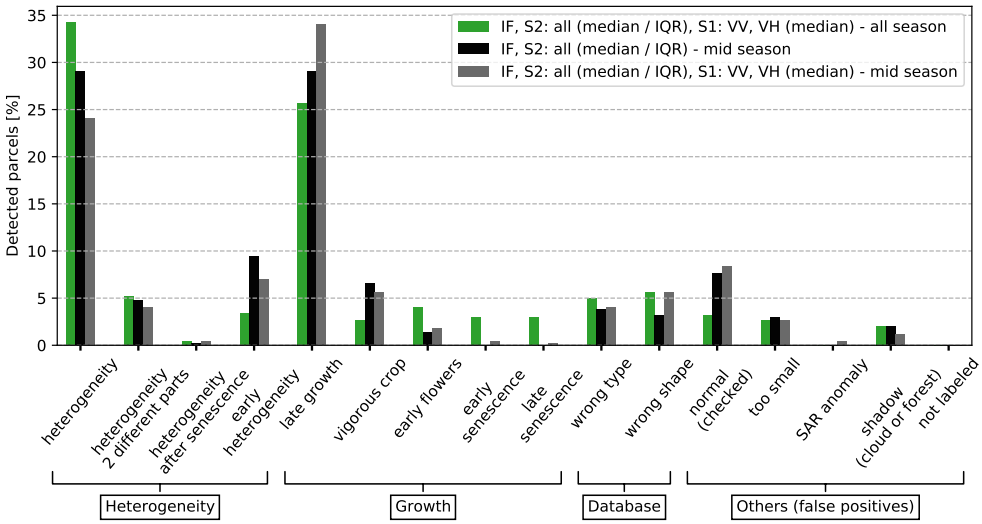


Figure S19. 100×(Number of detected parcels in each category / Number of detected parcels). Results obtained for a mid season analysis (before February) and a complete growing season analysis are compared for a outlier ratio equal to 10% in the rapeseed dataset.

These results confirm that S1 images are adapted to an early season analysis, especially thanks to an easier detection of late growth problems.

4.8. Robustness to changes in parcel boundaries

The robustness of the proposed method to changes in the parcel boundaries was validated using another parcel delineation system for the rapeseed growing season. To that extent, 2118 parcel delineations resulting from the French Land Parcel Identification System (LPIS) was considered. The French LPIS is also known as *Registre Parcellaire Graphique* (RPG). This database is available with an open license [2] and is updated yearly (in general with a delay of 2 years) on the basis of the farmer's Common Agricultural Policy (CAP) [6]. For comparison purposes, each parcel of database used in the main document was intersected with a corresponding LPIS parcel. Some parcels were not defined in the LPIS file, which explains why the number of parcels available for the LPIS analysis is slightly smaller than the number of parcels obtained when using the customer database.

Examples of parcel delinations obtained with LPIS and the proprietary parcellation system are depicted in Figure S20. The parcel frontiers obtained using LPIS are generally less accurate than those resulting from the proprietary system motivating the use of a buffer around the different parcels and robust zonal statistics.

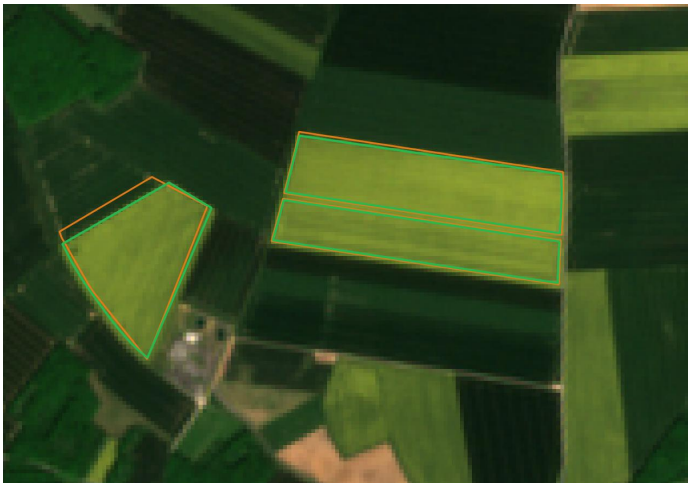


Figure S20. Example of parcel boundaries (rapeseed crop, growing season 2017/2018). In orange: customer database, in green: LPIS database.

Anomaly detection was run with an outlier ratio of 20% using these two different databases. The numbers of detected anomalies for each category are depicted in Figure S21. No significant difference can be observed when using the customer and LPIS parcels, showing that the proposed detection method is robust to this type of changes (probably because robust zonal statistics are used for anomaly detection).

[2] https://www.data.gouv.fr/fr/datasets/58d8d8a0c751df17537c66be/, online accessed 8 July 2020

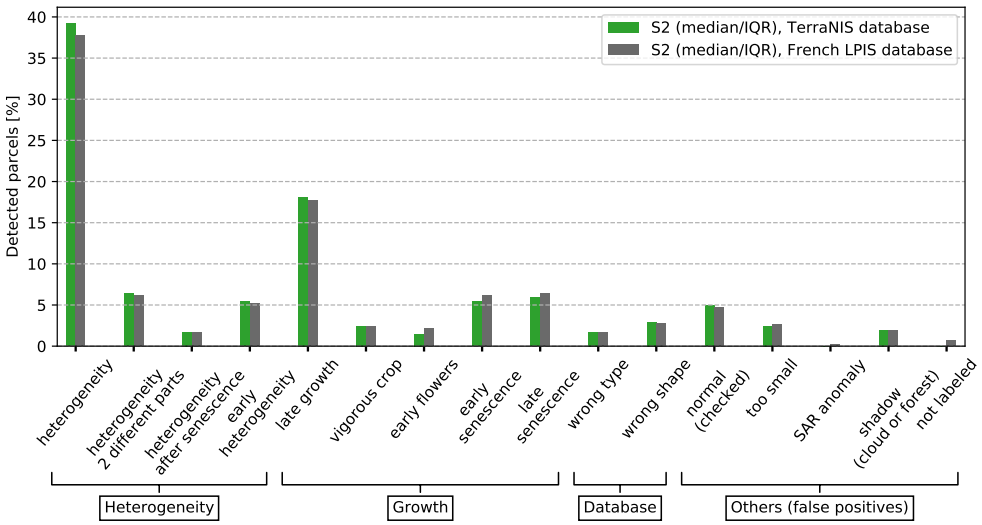


Figure S21. 100×(Number of detected parcels in each category / Number of detected parcels). LPIS and proprietary parcellation databases are compared with the IF algorithm and an outlier ratio equal to 20%.

References

1. Constantinou, V. PyNomaly: Anomaly detection using Local Outlier Probabilities (LoOP). *J. Open Source Softw.* **2018**, *3*, 845. doi:10.21105/joss.00845.
2. Jaakkola, T.; Diekhans, M.; Haussler, D. Using the Fisher kernel method to detect remote protein homologies. ISMB; AAAI Press: Menio Park, CA, USA, 1999; pp. 149–158.
3. Aggarwal, C.C., Linear Models for Outlier Detection. In *Outlier Analysis*; Springer International Publishing: Cham, 2017; chapter 3, p. 93. doi:10.1007/978-3-319-47578-3_3.
4. Liu, F.T.; Ting, K.M.; Zhou, Z. Isolation Forest. IEEE ICDM; , 2008; pp. 413–422. doi:10.1109/ICDM.2008.17.
5. Zhao, Y.; Nasrullah, Z.; Li, Z. PyOD: A Python Toolbox for Scalable Outlier Detection. *J. Mach. Learn. Res.* **2019**, *20*, 1–7.
6. Barbottin, A.; Bouty, C.; Martin, P. Using the French LPIS database to highlight farm area dynamics: The case study of the Niort Plain. *Land Use Policy* **2018**, *73*, 281 – 289. doi:https://doi.org/10.1016/j.landusepol.2018.02.012.