Fig. 1. Positive (top left) and negative (top right) number of catches per month (x axis) and 5 year classes. Catches dataset is affected by strong seasonality. Positive catches (at least 1 *C. imicola*, red points, bottom right) and negative catches (bottom left, green points) maps. *C. imicola’s* presence shows an evident clustered pattern.

Fig. 2. Scatter plot of the difference in prevalence between test and sample (x axis) and corresponding test size (percentage, on y axis) resulting from the brute force procedure (panel a). The red highlighted area containing the 277 samples complying with the chosen constraints are zoomed in panel b. Most of suitable samples come from assigning nine pixel to the test dataset. Panel c shows the total sample size distribution (train plus test), whilst two random example of the (277) selected samples are mapped in panel d (points belonging to train and test are green and red colored respectively)

Fig. 3. Algorithm performance comparison in terms of AUC, Accuracy, Se and Sp (panel a). Spatial prediction (neglecting the timing of catches) using Xlgboost tree model (highest AUC value of 0.984) is shown in panel b for absence (on the left) and presence (on the right) location in the test dataset. Red and green dots represents presence and absence prediction respectively.

Fig. 4. Xlgboost class probability, 1 km resolution, for April the 1st (left) and October the 15th (right), year 2018.

Fig. 5. Distribution of the 277 model performance in terms of AUC, Se, Sp (panel a) and the corresponding variable importance ranking summarized as median (panel b, color coding variable group, independently from the temporal lag). Night temperature group (yellow) was clearly the most important, followed by NDVI group (dark green), whilst TRMM group resulted the least important.