Title: Toward coherent space-time mapping of seagrass cover from satellite data: example of a Mediterranean lagoon
Author(s): Guillaume Cyril Henri Goodwin et al.
MS No.: egusphere-2022-1501
MS type: Research article
Special Issue: Monitoring coastal wetlands and the seashore with a multi-sensor approach

**General comments**

This study is investigating ways to improve the performance of remote sensing methods to predict the presence seagrass consistently over time. The addition of bathymetry, sediment size and coordinates data are indeed very relevant and deserving of such study.

The introduction section is well documented and written.

However, I find the methods and overall presentation of the results quite confusing and hard to interpret. Furthermore, very little attention is given to the ecological relevance of the findings and their possible application in other areas with different seagrass species and cover. The discussion section needs to include how the finding of the study fit in the seagrass remote sensing literature detailed in the introduction and how change detection can be improved.

I have suggested in the comment below various ways I think the authors can significantly improve the quality of this publication. I would strongly recommend major revisions before publication.

**Specific comments**

| | |
|---|---|
| Lines 105-110 | Please provide some more background information about the field survey data. It is not clear how those maps which will be used as training data were produced in the first place (remote sensing ?). |
| Lines 114,115 | Please provide more information about how and why the specific patch of seagrass near Chioggia was digitized. |
| Line 125 | Please provide explanation for choosing the specific 0.75m threshold. |
| Figure 3 | This figure is very complex to interpret. Consider revising how this information is displayed to only show curated results. |
| Description of random forest section | Could you confirm that the input data for the random forest models are individual pixels. Please provide more information about training/validation/test data split ratios. Also give ratio of pixel classified as 0 and 1 (balanced ?) |
| Definition of model performance section | I am confused and possibly concerned about the method used to compare model performances. The main issue for me is the variability in size of the testing datasets. I'm assuming the full lagoon datasets have more |

| | |
|---|---|
| | pixels than the inlet ones therefore making it more prone to errors. This is quite complex and maybe it would be more useful to report the average F1 score for similar testing sets (eg. Full lagoon vs inlet). |
| Figure 6 | Consider updating with the comments from above |
| Results – Effect of added feature | Comparing model performance only based on metrics can be misleading. Indeed a lot of factors can bias those metrics such as sample size and class imbalanced. It would be very useful for the reader to have some actual visual representation (maps) of the prediction vs ground truth. I am suspecting that the Southern zone and Northern zone differences are due to training pixel class imbalance.<br>To me all the comparisons between the models trained on specific year and zone just highlight the effect of training bias. I suggest instead to conserve the training scenarios based on the different features combination (spectral, location, environment) and pool the 2002 and 2017 training data together. Adding cross validation to your random forest would ensure the results are not influenced by a particular year or area. Consider also doing more fine tuning of the number of trees and nodes. I understand you chose 100 trees to make it easier for the vote count to show as a probability percentage. |
| Figure 7 | I don't think this figure is really that informative and I would suggest to be moved to the appendix. Consider here to summarize better the data from the figure. Try to better identify the clear patterns. What I would be looking for in this section is how does the different feature combination of the training scenarios influences the probabilities of a pixel to be classified as seagrass or bare. More importantly try to identify where those potential misclassifications occur (e.g. near the edges of the meadows), what could be causing it, and how adding features such as location, depth and grain size might improve the predictions. |
| Figure 8 | Again here I found the figure quite hard to read due to the large number of points and them clumping. I would suggest here a boxplot showing the range difference of F1 score before and after correction for the categories represented by the various marker type and colours. |

| | |
|---|---|
| Effect of time-based correction | As before judging the effect of the time-based correction solely on F1 score might be misleading. What I would be looking for here is what is the influence of the time-based correction on the seagrass area predicted. Does it create more cohesive meadows by reducing the salt and pepper effect of pixel-based predictions? Does it allow large areas that were not classified previously to get predicted properly now? What are the ecological implications of such differences in predictions |
| Figure 10 | I find this kind of figure much more informative that the ones in the previous result section. This clearly shows the value of the additional features in the RF models combined with the time-based correction. I like also the addition of the effect if the correction on the predicted seagrass area.<br>This is the most compelling figure so far and I would have like to see more of that sooner in the paper.<br>I know understand what the digitised patch was for. It would help the reader to explain the purpose of this patch earlier in the method. |
| Line 339-340 | Please expand on that point. This is a very interesting a relevant observation. Show how the improvements to the model are beneficial for detecting change in seagrass area in greater detail. This is the most important implication and needs to be highlighted more. |

Technical comments

| | |
|---|---|
| Lines 32,35 | Inconsistent references format |
| Line 36 | Typo to be deleted (-0cm) |