Classification accuracy and compatibility across devices of a new Rapid-E+ flow cytometer

Branko Sikoparija[1], Predrag Matavulj[2], Isidora Simovic[1], Predrag Radisic[1], Sanja Brdar[1], Vladan Minic[1], Danijela Tesendic[3], Evgeny Kadantsev[4], Julia Palamarchuk[4] and Mikhail Sofiev[4]

[1]BioSense Institute Research Institute for Information Technologies in Biosystems, University of Novi Sad, Novi Sad, 2100, Serbia
[2]Institute for Data Science, University of Applied Sciences North Western Switzerland, Windish, 5210, Switzerland
[3]Department of mathematics and informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, 21000, Serbia
[4]Finnish Meteorological Institute, Helsinki, Erik Palmenin Aukio 1, FI-00560, Finland

Correspondence to: Mikhail Sofiev (Mikhail.Sofiev@fmi.fi)

Anonymous Referee #1 (Citation: https://doi.org/10.5194/egusphere-2024-187-RC1)

This manuscript "Classification accuracy and compatibility across devices of a new Rapid-E+ flow cytometer" describes the evaluation of a new instrument, the Rapid-E+, upgraded from a previous model made by Plair SA, and its ability to monitor pollen compared alongside a manual Hirst-type sampler. The necessary training of a classification algorithm to distinguish pollen types is detailed and lab evaluation is followed up by field evaluation, and cross-comparison with instruments at other sites to assess method generalisability. The study is thorough and comprehensive, looking into the detail of the different modalities of data obtained for different pollen types across different instruments. The manuscript is of rigorous scientific quality and reports findings that are useful in this field to further the advancement of automated pollen monitoring. It is written and presented concisely and generally clearly, with ample supporting information in the Appendices. There are only some minor technical points that I would address before continuing to publication.

Reply: The authors would like to thank Referee #1 for constructive and positive suggestions on how to improve the manuscript further. Below we answer the questions and indicate the changes we have made to the revised manuscript.

Abstract

Line 22: I would use the term 'instrument' instead of 'monitor'.

Reply: Corrected as suggested throughout text.

Introduction

Line 29: "Buters et al. 2022"

Reply: Corrected as suggested.

Line 30: "monitoring instruments"

Reply: Corrected as suggested.

Materials and Methods

Line 49-50: Not sure in this sentence exactly how the Rapid-E+ compares to the Rapid-E. Perhaps alter to "In particular the Rapid-E+ samples at a faster flow rate of 5 l min-1 (compared to 2.8 l min-1 for the Rapid-E), and records all particles passing through a 447 nm scattering laser into 4 size bins (>0.3 µm, >0.5 µm, >1 µm, >5 µm) unlike the Rapid-E which…?" (does the Rapid-E not have different size bins?)

Reply: The statement is expanded to compare differences and now reads:
"In particular, Rapid-E+ samples at a faster flow rate of 5 l min^{-1} (compared to 2.8 l min^{-1} for the Rapid-E). Also, regardless the operation mode, Rapid-E+ records concentration of all particles passing through a 447 nm scattering laser (classified into 4 size bins: >0.3 µm, >0.5 µm, >1 µm, >5 µm), while Rapid-E only records concentration of particles above operation mode determined size limit."

Line 55-56: "also allows for adjusting the gain of the fluorescence spectrum and lifetime detectors"
Reply: Corrected as suggested.

Line 72: "Three Rapid-E+ air flow cytometers were involved in this study."
Reply: Corrected as suggested.

Line 72: "…in Novi Sad, Serbia, …"
Reply: Corrected as suggested.

Line 73: "the Novi Sad laboratory" is very nondescript. Details about the organisation that runs the Novi Sad laboratory may be helpful, and the environment?
Reply: As suggested, we have specified that device worked indoors during creation of the training dataset and then set to work outside.

Line78: "The test period allowed for the exploration of measurement performance of the automatic bioaerosol monitoring instrument in a variety of conditions characteristic of the Pannonian Plain in [where?]. This region contains a large diversity of pollen and fungal spores…" This sentence was quite long so I suggest splitting it into two, e.g. where I have done so.
Reply: Corrected as suggested

Line 82: "the period of seasonal allergies" – perhaps a little more description specifically as to what these seasonal allergies are in this place?
Reply: The sentence is extended and now reads:
"In the study region, the period of seasonal pollen allergies (i.e. tree pollen season from January to April and grass pollen season from April to September) is extended by the weed pollen season from July to the end of October when large quantity of ragweed pollen is recorded in the air (Sikoparija et al., 2018)"

Line 83: "when large quantities of ragweed pollen are recorded in the air"
Reply: Corrected as suggested.

Line 85: "the main features of diurnal variations"
Reply: Corrected as suggested.

Line 89: "Reference pollen for training was collected locally."
Reply: Corrected as suggested

Line 98: "to ensure identity" - could you explain this better?
Reply: This part was indeed confusing, so we removed it from the text.

Line 102: "exposed to pollen using the Swisens Atomizer"
Reply: Corrected as suggested.

Line 103: "expose pollen to the Novi Sad and Osijek devices.
Reply: Corrected as suggested.

Line 106: “validating”
Reply: Corrected as suggested.

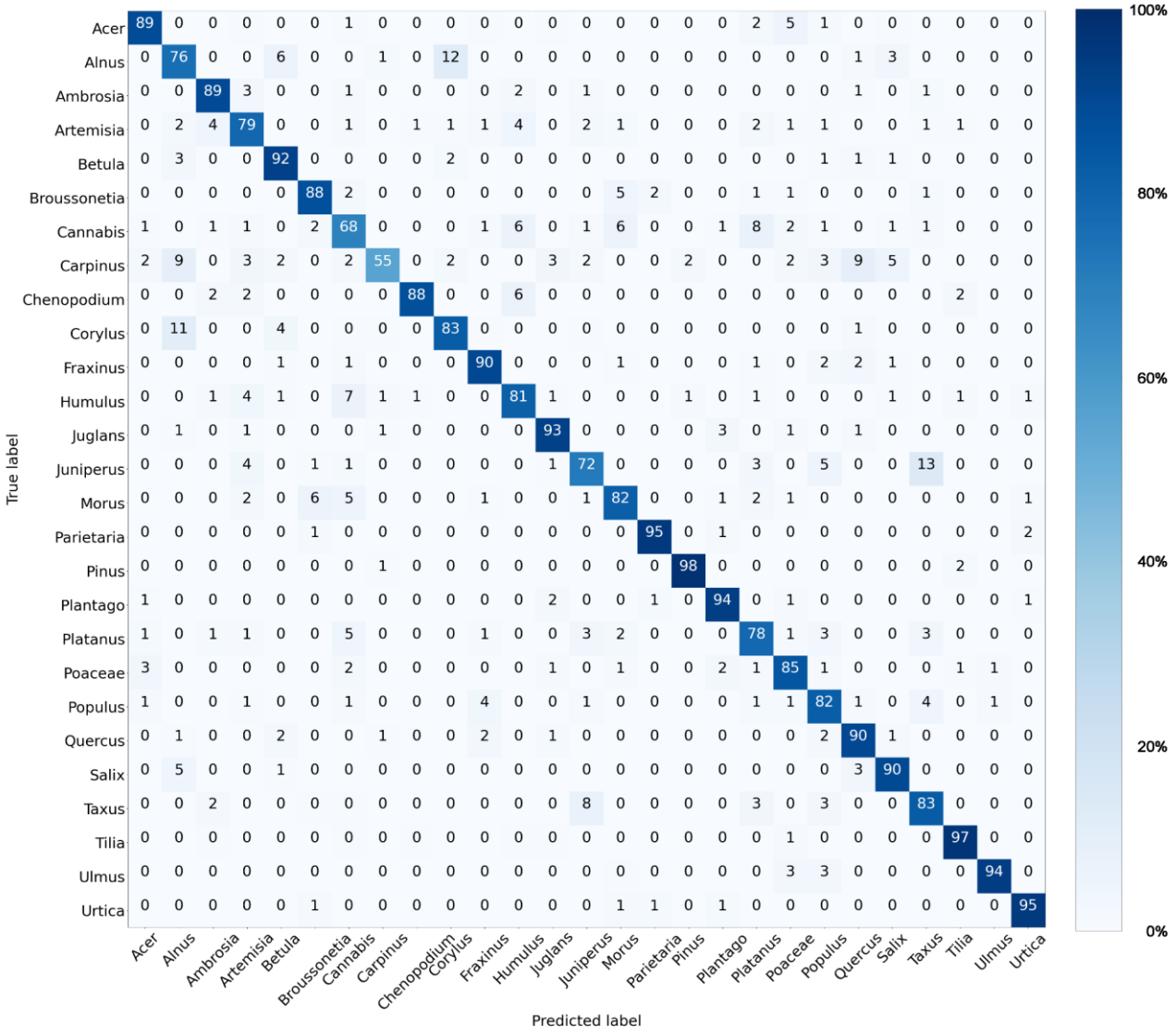
Line 109: Could say “colocated” instead of side-by-side.
Reply: Corrected as suggested.

Results and discussion
Line 201: Are these precision, recall and F1 scores averaged across scores for each pollen classification? If so, just mention they are averaged to avoid confusion, if not, I am unsure how the score differs from the discrimination of pollen from “other”.
Reply: Yes, that is correct. The F1 scores were calculated for each class and then averaged. It is now indicated in the text, as suggested.

Line 207: By ‘the classification algorithm with high accuracy’ do you mean the one that achieved F1 score of 0.86 as opposed to 0.84? Or simply that the algorithm managed to distinguish these pollen types with high accuracy, regardless as to which? Perhaps it may be better to write something like one of the following, depending on which you meant to avoid confusion…
“It is interesting to note that the latter classification algorithm (with merged classes) distinguished Urtica and Parietaria from Brousonetia despite these pollen grains being morphologically similar.”
Or
“It is interesting to note that the classification algorithm distinguished Urtica and Parietaria from Brousonetia with high accuracy, despite these pollen grains being morphologically similar.”
Reply: Yes, it is correct. And we appreciate the suggestion for improving clarity. The sentence now reads:
“It is interesting to note that the classification algorithm distinguished Urtica and Parietaria from Brousonetia with high accuracy, despite these pollen grains are morphologically similar.”

Fig. 2: The numbers and names are a bit small and blurry, would be good to make the characters a little larger if possible.
Reply: The figures were created in sufficient resolution, and we believe importing them into Word may have affected their quality. We expect that in the published version, after typesetting, the original files will be used, so they won't be blurry. Since the confusion matrices present 27 classes, increasing the font size is not feasible. Therefore, we suggest arranging the panels of Figure 2 in a vertical orientation, which could result in a 100% increase in the size of the panel and thus improve the font size as well.

(A)


True label
Predicted label
Acer
Alnus
Ambrosia
Artemisia
Betula
Broussonetia
Cannabis
Carpinus
Chenopodium
Corylus
Fraxinus
Humulus
Juglans
Juniperus
Morus
Parietaria
Pinus
Plantago
Platanus
Poaceae
Populus
Quercus
Salix
Taxus
Tilia
Ulmus
Urtica
100%
80%
60%
40%
20%
0%

(B)

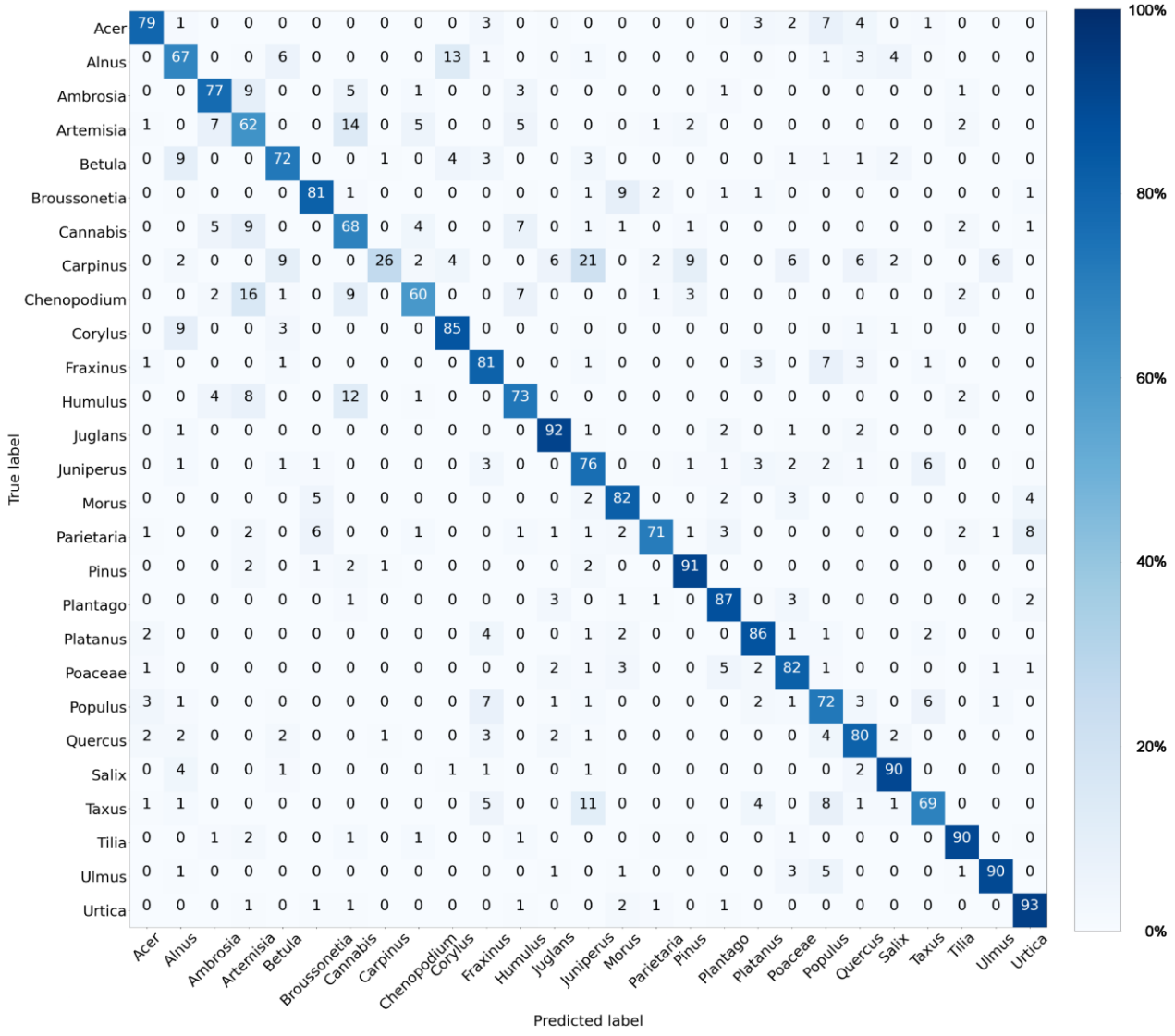


Figure 2: Confusion matrices depicting pollen classification performance on test dataset measured in (A) "pollen mode" and (B)

Line 226: what are the exact dates referred to here?

Reply: The dates for the indicated period, 3-7 May 2023, were added.

Line 235: Best to define PSLs in brackets for good measure as it is mentioned for the first time in this manuscript.

Reply: The "(Polystyrene Particles)" was added after PSLs when mentioned for the first time as suggested.

Line 241: At a glance, this sentence was a little confusing, I would correct it to something like: "Automatic detections of total pollen, as well as Juglans, Morus and Ambrosia, have a statistically significant positive correlation with…"

Reply: Corrected as suggested.

Line 243: “for most pollen classes” or “for most of the pollen classes”
Reply: Corrected as suggested.

Line 245: Perhaps rephrase as, for example, “Pollen grains that occur simultaneously in the air had a clear tendency to be confused amongst each other, which was expected…”
Reply: We kept the original sentence here.

Line 261: “As demonstrated for the Rapid-E, this problem also exists for the Rapid-E+.”
Reply: This sentence is changed following the suggestion from other participant of the public discussion, and section now reads:
“As a result, classification performance falls when a model trained on a reference dataset from one device is tested on a reference dataset from another one, which was demonstrated for Rapid-E (Matavulj et al., 2021). The same problem exists in Rapid-E+ (Fig. 4). The algorithm created on the training dataset collected with the Novi Sad device failed to identify the same reference pollen collected with both Osijek and FMI devices (average F1 score = 0.01 in both cases)”

Line 278: I would probably start a new sentence and replace the second i.e. before ‘different timing…’ with something else. This sentence is a bit confusing and long. Is it saying that since some pollen classes were comparable across devices, the differences observed across others shouldn’t be due to doing lab work at different times and different methods of pollen exposure to the instrument? Or are you saying each lab followed the same procedures so it shouldn’t be an issue?
Reply: This section is shortened and now reads:
“When analysing the results of the cleaning reference data for the same pollen measured with different devices, we noticed a significant difference for most pollen classes, except for Platanus, Salix and Betula. Different timing of the lab work and different methods of exposing the device to pollen cannot explain observed differences but it is rather attributed to differences in device sensitivity to the scattering and/or fluorescence signals.”

Fig. 5 writing font too small and am unsure what I am looking at in 5D, can labels be added to the x, y and colour axes?
Reply: We have increased the font size used in Figure 5. Regarding panel D, it presents an image from the scattering light as described in the sub-chapter 2.2: "In addition, the intensity of light, scattering from a 637 nm laser, is recorded as an image using a 4x4 pixel detector". We have expanded the caption of Figure 5 to give more details.

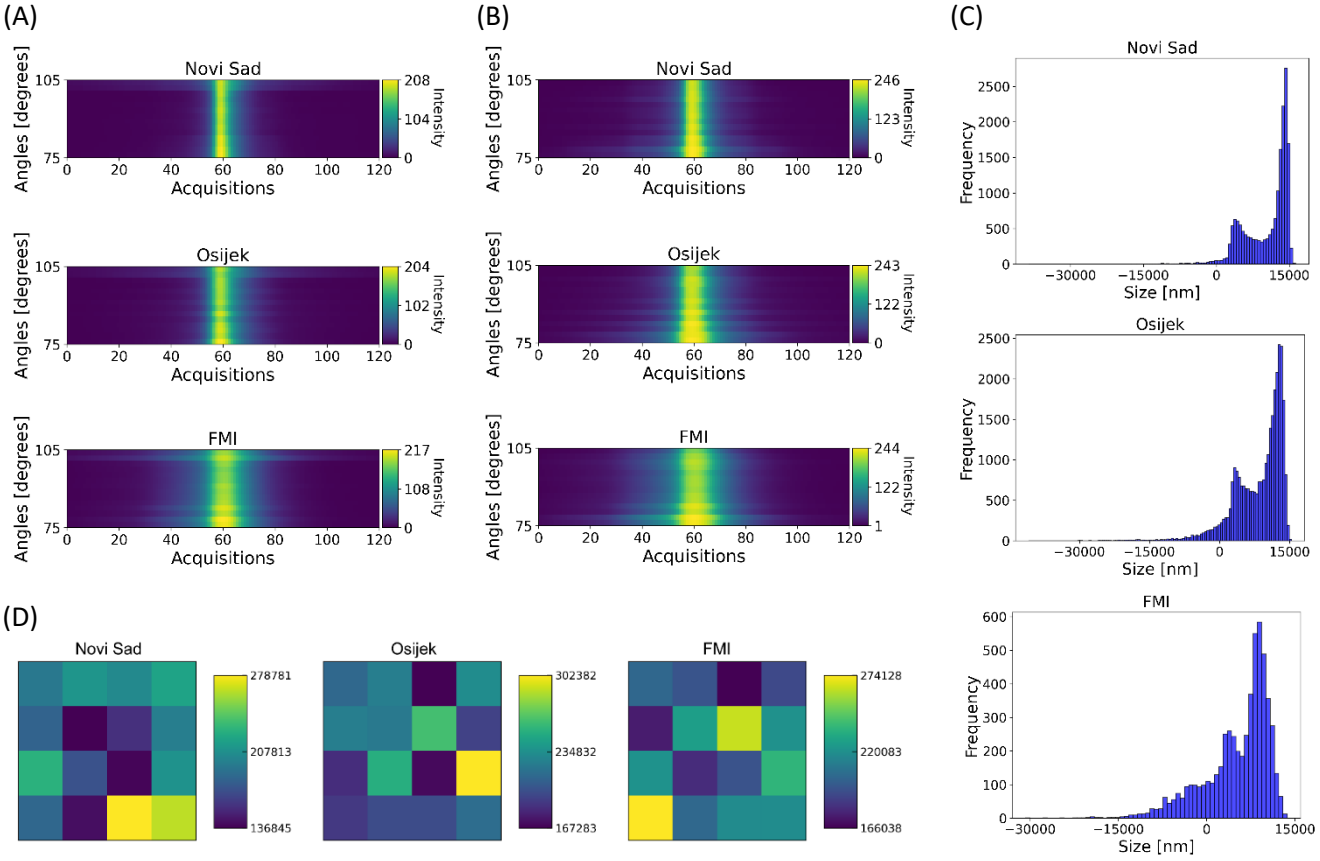
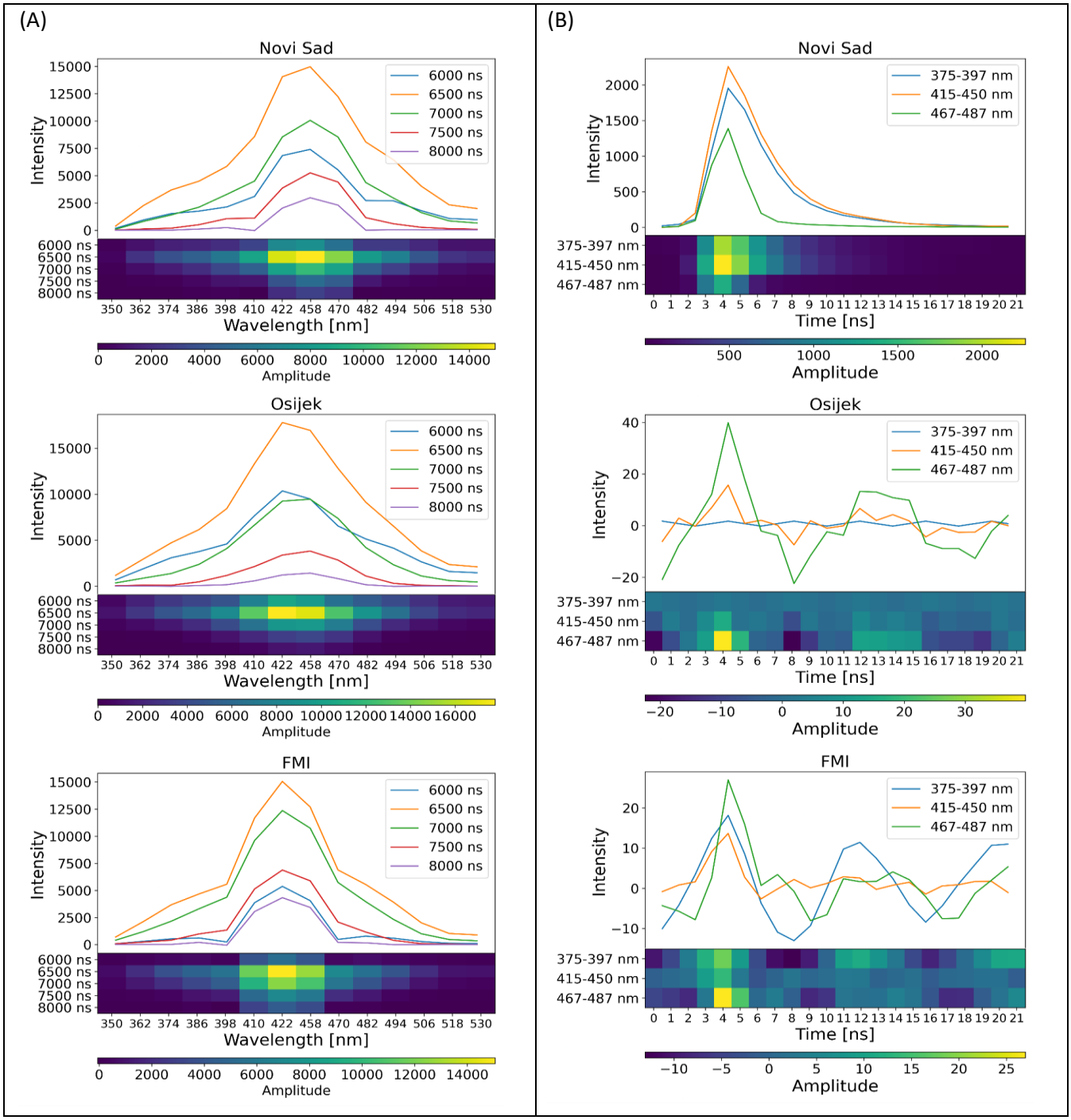


Figure 5: Comparison of reference Betula pollen measurements in "pollen mode" on Novi Sad, Osijek and FMI Rapid-E+ devices after preprocessing: (A) average 447 nm laser perpendicular polarisation scatter, (B) average 447 nm laser parallel polarisation scatter, (C) histogram of size distribution (D) average unitless intensity of 637 nm laser scattered light, recorded as an image using a 4x4 pixel detector.

Fig. 6 again writing font too small.

Reply: We have increased the font size used in Figure 6.



References

Matavulj, P., Brdar, S., Racković, M., Sikoparija, B. and Athanasiadis, I. N.: Domain adaptation with unlabeled data for model transferability between airborne particle identifiers. in: Proceedings of the 17th International Conference on Machine Learning and Data Mining (MLDM 2021), New York, USA, https://doi.org/10.5281/zenodo.5574164, 2021.

Sikoparija, B., Marko, O., Panic, M., Jakovetic, D., and Radisic, P.: How to prepare a pollen calendar for forecasting daily pollen concentrations of Ambrosia, Betula and Poaceae?, Aerobiologia, 34, 203-217, https://doi.org/10.1007/s10453-018-9507-9, 2018.