We thank the reviewers for their constructive comments. Their suggestions were taken into account and helped improve the article's quality. In response to the comments from the three reviewers, several changes have been made to the manuscript. First, the study's objective has been rephrased to better clarify that this work constitutes a first methodological step in the development of a pipeline for analyzing airborne pollen. Important clarifications were also provided regarding the data cleaning step and the model's ability to distinguish pollen grains from non-pollen debris using the characteristic fluorescent signature of sporopollenin. Regarding the model, details were added regarding the application of the data balancing fonction, the strict separation of training and validation sets, and the use of Breiman's random forest model. Finally, the limitations associated with model transfer between instruments were discussed in greater detail, and several terminological and typographical corrections suggested by the reviewers were incorporated throughout the manuscript.

Please find below our detailed answers to each of the concerns raised.

Note that the line numbers correspond to those of the revised manuscript.

RC1: Anonymous Referee #1, 01 Feb 2026

The study is oriented towards identifying airborne pollen and the authors imply that flow cytometry is an efficient alternative for microscopical identification. It is particularly valuable to see that analysis of flow cytometry measurements seems to enable discrimination between different species of the same plant genus which is usually not possible in routine microscopical analysis. The manuscript is well written, and it clearly describes possibilities of standard flow cytometers for identification pollen.

However, in my opinion the manuscript lacks tests and discussion on the applicability of proposed method (flow cytometry measurement and the developed random forest classification model) for analysing aerosol samples. The authors emphasized importance of adapting the models to real environment samples (lines 259-264). But in my opinion for Atmospheric Measurement Techniques more than just theoretical discussion is needed when linking to atmospheric measurements. Without tests on atmospheric samples, it is just a speculation that proposed approach has a "potential for large-scale urban pollen monitoring".

We thank the reviewer for this important comment. We acknowledge that the manuscript may overstate the goal of the model developed. We have revised the text to articulate more clearly that this work focuses on developing and validating a robust reference classification framework based on flow cytometry measurements of known pollen taxa. Such controlled conditions are a prerequisite before the method can be applied to atmospheric samples, which are inherently more complex and heterogenous.

L65-67 : We modified to : « Our study aims to develop a classification model capable of identifying pollen from urban environments at species and genus levels. Here we present the necessary first step in the development of a broader methodological pipeline for the analysis of airborne pollen. »

There are several aspects that should be addressed/discussed:

1.How the identification algorithm handles large diversity of aerosols (i.e. fungal spores, dust clusters, starch) which could be in the same size class as pollen. So far, the studies showed that

very good classification algorithms in lab settings (as seen in confusion matrix) in real environment tend to fail.

We thank the reviewer for asking this question. We note also that Referee #2 had a similar concern, which further highlights that we need to be clearer on this issue. Prior to model training, we performed a data cleaning step (Figure A2 and Lines 127–131) to separate pollen grains from "debris" classified as non-pollen (other) in our training dataset. To distinguish these two groups, we first used excitation violet laser fluorescence parameters (detectors PB450 and Violet610), while verifying that the distinction held true for the same groups across the other variables. This excitation/emission range is characteristic of sporopollenin which contains fluorophores that in turn are specific to pollen grains [1]. The model is therefore able to identify all particles sharing the fluorescence signature of pollen; everything else is categorised as "other". Even if some non-pollen aerosols (e.g. fungal spores, dust, starch) may emit autofluorescence in similar wavelengths, the probability that any such particle would simultaneously reproduce the sporopollenin-specific fluorescence signature and all multi-parametric characteristics used by the model (size, granularity, multi-channel fluorescence) remains very low. The robustness of our approach relies precisely on this combined multi-parametric classification.

[1] Pöhlker, C., Huffman, J. A., Förster, J.-D., and Pöschl, U.: Autofluorescence of atmospheric bioaerosols: spectral fingerprints and taxonomic trends of pollen, Atmos. Meas. Tech., 6, 3369–3392, https://doi.org/10.5194/amt-6-3369-2013, 2013.

In response to the Reviewer's comment, we made the following changes to the manuscript.

L146-147 : We added : « This excitation/emission range is characteristic of sporopollenin which contains the fluorophores specific to pollen grains (Pöhlker et al., 2013). »

L136 we also clarify what we consider as debris : « […] debris, that is non-pollen particles, […] »

The authors claim using flow cytometers could enable near-real-time pollen identification. How the aerosol sample is to be processed to be delivered to the flow cytometer (here bear in mind that sample could contain large particles that might require filtering out to avoid clogging).

We acknowledge that the term "near real-time" may have been an overstatement; we have revised the manuscript to more accurately state that our approach is faster compared to traditional methods, in the lab.

L318 ; 329 : We replaced « near-real-time » by « faster »

Regarding the filtration step, the samples contain a large number of particles, and filtration serves two complementary purposes: first, to retain only particles within the size range of interest (5–100 µm, corresponding to pollen grains), and second, to prevent clogging of the flow cytometer.

In addition to size-based filtering, we took care to minimize pollen grain aggregation by resuspending samples in PBS and stirring them regularly during data acquisition with the cytometer, as pollen grains, particularly the larger ones, sediment rapidly (within less than 30 seconds). This protocol is described in lines 92–94 of the manuscript and in more detail in supplementary material.

Accordingly, L94-98 was modified to « Pollen was extracted from the floral units using a filtration system that retained only particles between 5 and 100 µm in diameter, a size range that includes pollen grains but also particles of similar size. Filtration also prevents clogging of the flow cytometer because, as is generally recommended, particles should not exceed one-third to one-fifth of the width of the flow cell, which limits particle size to approximately 100 µm on the CytoFLEX instrument we used. »

We have also added these clarifications to the detailed protocol in supplementary material.

2. Also related to previous item please discuss to what extent the flow cytometry approach could be challenging for quantification. For example, some devices can handle up to 200 microliters of sample for a single analysis which then requires either subsampling from larger volumes in which aerosols are collected or sample should be concentrated.

Instruments such as the CYTOFlex can accommodate volumes up to 5 mL and allow consecutive tube addition.

In our study, samples were centrifuged and resuspended in PBS to limit aggregation. We typically started with approximately 200 µL of resuspended sample and further diluted it when an excessively high particle concentration was observed during sample analysis on the flow cytometer. When necessary, multiple tubes were run per sample, and the resulting counts were summed to obtain the final value for each sample.

These precisions have been added to the detailed protocol in supplementary material.

The approach to rely pollen identification exclusively on flow cytometry measurements that most cytometers routinely used in healthcare is very important. But the use of the same classification algorithm on different devices (even the same model) appeared to be challenging (as authors also clearly noted in lines 273-281). If not possible to test the model on different device measuring same parameters, the authors should at least discuss the measurement uncertainty for each parameter and refer to other studies that observed differences in flow cytometry parameters between different devices.

We agree that absolute signal intensities in flow cytometry are device-dependent and that model transfer across platforms without correction could fail. Our intended operational scenario is deployment of harmonized analyzers of the same model, under bead-based daily QC and targeted detector settings, procedures that are now standard practice in clinical research labs. Under such procedures, studies have demonstrated that comparable readouts can be achieved between instruments, including within a given model/brand and, with specific workflows, even across brands with different optical trains [1].

[1] Solly F, Rigollet L, Baseggio L, et al. Comparable flow cytometry data can be obtained with two types of instruments, Canto II, and Navios. A GEIL study. Cytometry Part A. 2013;83(12):1066-1072.

In response to the comment, we modified L297-307 to add "However, deployment within harmonized analyzers is feasible under standard bead-based daily QC protocols (CS&T/Application Settings for conventional analyzers; SpectroFlo QC beads for spectral systems), which have been shown to control inter-instrument MFI drift within single-digit

percentages (Cornel et al., 2020; Omana-Zapata et al., 2019; Solly et al., 2013b). Channel-wise normalization during data processing further reduces residual variability, and a lightweight domain-adaptation step, based on acquiring a small reference pollen set on the target instrument, can re-anchor feature distributions prior to inference. FSC and SSC parameters remain more sensitive to flow-rate and optical alignment and should therefore be monitored carefully. Where direct comparison between instruments is required, ERF/MESF calibration from NIST's Flow Cytometry Standards Consortium allows comparing fluorescence results between different instruments (Wang and Hoffman, 2017)."

In line 133 authors indicated the training dataset for *Thuja* genus was impossible to clean from debris. Is presence of debris confirmed by microscope? If not, how can you be sure it is not a part of the normal pollen variability? The pollen from Thuja (and many other Cupressaceae) tends to break in wet environment resulting in separation of exine from the resto of pollen grain. Could it be that those separated exines are the "debris" you see in the data.

We confirmed by microscopy that intact pollen grains and debris, mainly cone debris, were indeed present in our samples. We observed by microscopy that the majority of pollen grains were not broken. Pollen grains tend to break apart when they have been stored for too long in a humid environment (more than approximately 6 months), which was not the case here. However, regardless of the cause, in our case, we were unable to distinguish the pollen grains from other debris present in our Thuja samples and were therefore unable to include them in the model training.

L273 : We added « we had visually confirmed the presence of intact pollen grains »

In Table A1, authors reference is missing for an accurate scientific name (e.g. *Ambrosia artemisiifolia* L.). Genus should be written in cursive font and also should include author references

Thanks for the comment; it has been updated in the revised manuscript, see Table A1.

RC2: Anonymous Referee #2, 02 Feb 2026

Overview:

In this study, Tardiff et al. sought to distinguish tree, grass & weed pollen in an urban area (Montréal) using flow cytometry data & a machine learning algorithm (specifically random forest). They identified granularity & fluorescence parameters from violet (Violet610_A) & blue (PB450_A) lasers to be the most distinctive features for discerning a pollen species, though species accuracy (F1 = 0.76) in the random forest-model was lower than genus accuracy (F1 = 0.90). I generally found the manuscript to be well written, with few spelling, grammatical issues. This study presents a strong step toward scalable pollen classification using non-imaging cytometry. However, some concerns and requests for clarification are pointed out below:

We thank the reviewer for the positive assessment of our manuscript and for the constructive comments and suggestions.

Major comments:

Line 94-95: For this study, I understand that the purity of the pollen can be critical in analyzing the data obtained from the flow cytometer because the proportion of non-pollen particles that can be similar in size or fluorescence intensity can influence the classification. However, there is no data to validate the pollen's purity using other methods. For example, would it be possible to compare the pollen purity by an image analysis of the extracted pollen subsample under a light/fluorescence microscope?

We thank the reviewer for asking this question. We note also that Referee #1 had a similar concern, which further highlights that we need to be clearer on this issue. To ensure pollen purity, we performed prior to model training a data cleaning step (Figure A2 and Lines 127–131) to separate pollen grains from "debris" classified as non-pollen (other) in our training dataset. To distinguish these two groups, we first used excitation violet laser fluorescence parameters (detectors PB450 and Violet610), while verifying that the distinction held true for the same groups across the other variables. This excitation/emission range is characteristic of sporopollenin which contains fluorophores that in turn are specific to pollen grains [1]. The model is therefore able to identify all particles sharing the fluorescence signature of pollen; everything else is categorised as "other". Even if some non-pollen aerosols (e.g. fungal spores, dust, starch) may emit autofluorescence in similar wavelengths, the probability that any such particle would simultaneously reproduce the sporopollenin-specific fluorescence signature and all multi-parametric characteristics used by the model (size, granularity, multi-channel fluorescence) remains very low. The robustness of our approach relies precisely on this combined multi-parametric classification.

[1] Pöhlker, C., Huffman, J. A., Förster, J.-D., and Pöschl, U.: Autofluorescence of atmospheric bioaerosols: spectral fingerprints and taxonomic trends of pollen, Atmos. Meas. Tech., 6, 3369–3392, https://doi.org/10.5194/amt-6-3369-2013, 2013.

In response to the Reviewer's comment, we made the following changes to the manuscript:

L146-147 : We added : « This excitation/emission range is characteristic of sporopollenin which contains the fluorophores specific to pollen grains (Pöhlker et al., 2013a). »

L136 we also clarify what we consider as debris : « […] debris, that is non-pollen particles, […]»

Line 143-144 = Since the data is unbalanced, the authors chose to use synthetic minority over-sampling to normalise the data. However, I was under the impression that oversampling can cause model overfitting & an inaccurate representation of the smallest minority classes (~300 – 35000 is very unbalanced). Were steps taken to avoid/ensure that didn't happen? It may be at least worth a mention in the discussion.

We thank the reviewer for this comment. Only four taxa (Acer saccharum, Gramineae spp, Juglans cinerea, Picea abies) were oversampled. We chose a threshold of 1,000 pollen grains per species to avoid excessive oversampling of species for which we had insufficient grains. We have added a new table in appendices with this information (Table A2).

We made the following changes to the manuscript to address this issue:

L. 157: We replaced « (min=306; max=35307) » by « (Table A2) »

L161-162: We added « Only four taxa were oversampled (Acer saccharum, Gramineae spp, Juglans cinerea, and Picea abies). »

Line 145-146 = Reads as synthetic minority over-sampling is done on the samples, then the dataset is split into 70% training & 30% validation. If so, would this not lead to overfitting & inflated precision metrics, since the training data is used in the validation set? Perhaps I've misinterpreted what's written, or this doesn't matter. If so, clarification may be needed in the order of steps taken.

We thank the reviewer for this important comment. SMOTE balancing function was indeed applied before splitting the dataset into training and validation sets; however, for our objective, this does not introduce overfitting or inflated accuracy metrics. The validation set, although balanced, was never used in any way during model training. The Random Forest classifier was trained exclusively on the 70% training portion. The 30% validation set, whether balanced or not, was used exclusively to determine whether the model was capable of correctly identifying species it had never encountered during training. It is important to note that the purpose of SMOTE was to provide the classifier with a balanced training set in order to prevent it from being biased toward the majority class. Whether the validation set is balanced or imbalanced has no bearing on this objective, as long as those samples, as was the case here, were not seen during training.

We acknowledge that this ordering of steps may appear unconventional, and we have clarified this in the revised manuscript:

L161-163 : We added « Only four taxa were oversampled (Acer saccharum, Gramineae spp, Juglans cinerea, and Picea abies). The purpose of balancing data was to provide the classifier with a balanced training set to prevent it from being biased toward the majority class. »

L164-165 : We added « The validation set was not used for model training. The Random Forest classifier was trained exclusively on the 70% training portion. »

Line 160 = While F_1-scores are useful for measuring model performance in a single metric, the likely strong class imbalance and use of oversampling would suggest that including metrics such as PR AUC in addition to F_1-scores would provide a more complete summary of the model's performance (Saito & Rehmsmeier, 2015 https://doi.org/10.1371/journal.pone.0118432).

Given that our dataset is balanced and that oversampling was performed only for four species, we believe that the F1 score remains relevant.

Minor comments:

Lines 103-106: The authors state that the pollen's fluorescence depends on the fluorescent proteins on its surface (my understanding is that these are not proteins). If so, please provide a reference.

Thank you for your comment; that is indeed an error. It is true that sporopollenin contains fluorescent phenolic compounds, but it is not a protein; we have changed this sentence in the revised manuscript:

L109-110 We changed « ...which excites the fluorescent proteins on the surface of the pollen grain's outer wall. » by « ... which excites fluorescent phenolic compounds present in the sporopollenin of the pollen grain's outer wall. »

Line 148 = At the beginning of 2.4. It is stated that four supervised classification algorithms were tested, and the random forest performed best. There are many types of random forest classifiers, such as random forest by randomisation, which deals well with unbalanced/noisy data. Which random forest classifiers (Breiman?) were tested? It may be worth mentioning why the particular random forest was chosen over other random forest classifiers.

We used indeed Breiman's original Random Forest model, and since our training dataset is balanced, there was no need to use a different Random Forest model - as a reminder, very few species were oversampled, and the SMOTE function was primarily used for undersampling.

Accordingly, on L154, we have added the reference for random forest model: (Breiman, 2001) Breiman, L.: Random forests. Machine learning, 2001.

RC3: Anonymous Referee #3, 24 Feb 2026

General comment

This preprint describes an innovative method for measuring and identifying pollen grains using a standard laboratory flow cytometer and machine learning tools. The authors made considerable efforts to collect and prepare a comprehensive pollen reference database, which they then used to test their method. The results are promising, showing that liquid flow cytometry coupled with machine learning algorithms could become an effective alternative to manual pollen counting and identification. I only have a few questions related to the content and some technical corrections. In my opinion, no major changes are needed.

We would like to thank the reviewer for the very favorable assessment of our manuscript, for the constructive comments and suggestions, and for all the minor technical corrections; they were very helpful.

Specific comments

- l.88-91 “Their inclusion enabled the model to learn to discriminate tree pollen from other common airborne particle types, as real-world environmental samples typically comprise a heterogeneous mix of tree, grass, and weed pollen, along with various non-pollen particulates.” Are non-pollen particles a majority in your samples? What proportion did you observe?

We thank the reviewer for the comment. The proportion of debris varies considerably from one sample to another, as it depends on how easy (or difficult) it is to separate the pollen from the rest of the floral unit, as well as on the filtration stage. The objective during model training is to include a limited number of non-pollen particles so that the model has an option for defining an “other” category, but not to achieve an exhaustive representation of the debris that might be found in the samples; we therefore do not consider it relevant to include the proportions in the manuscript.

When balancing the data, we kept only 1,000 for the species-level model and 10,000 for the genus-level model, like the other categories.

- l.104-106 “Depending on their peptide composition, these proteins absorb light at a certain wavelength and emit light radiation at a different wavelength in return producing a characteristic fluorescence signature that varies among species.” Is the fluorescence signal expected to differ from airflow cytometry due to the suspension in PBS?

The PBS we used is Dulbecco’s phosphate-buffered saline (without calcium and without magnesium), as described by Dunker et al. and Aloisi et al. [1][2]. This is a standard solution used in flow cytometry; it is a neutral, isotonic buffer solution that has no effect on fluorescence.

[1] Aloisi, Iris; Cai, Giampiero; Tumiatti, Vincenzo; Minarini, Anna; Del Duca, Stefano (2015): Natural polyamines and synthetic analogs modify the growth and the morphology of Pyrus communis pollen tubes affecting ROS levels and causing cell death. Plant Sci 239, pp. 92–105. DOI: https://doi.org/10.1016/j.plantsci.2015. 07.008.

[2] Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, et al. Pollen analysis using multispectral imaging flow cytometry and deep learning. New Phytol. 2021;229(1):593–606. https://doi.org/10. 1111/nph.16882

To address this issue, on L98-101, we added « The filtrate was then suspended in Dulbecco's phosphate-buffered saline (PBS) solution, a standard neutral isotonic buffer commonly used in flow cytometry to minimize aggregation (Aloisi et al., 2015; Dunker et al., 2021b) (see detailed protocol in the supplementary material). »

- l.106-108 "For each laser, avalanche photodiode (APD) detectors measure the intensity of light emitted at different wavelengths using ten filters: 450/45, 525/40, 610/20 (violet laser), 525/40, 585/42, 690/50, 780/60 (blue laser), 660/10, 712/25, 780/60 nm (red laser)." It would be nice to have an explanation on these values, e.g., for 450/45 what does 450 and 45 represent? Most people in the aerobiology community are not familiar with these measurements.

We thank the reviewer for the comment; this has been clarified in the revised manuscript: L114-118: «Each filter value, such as 450/45, follows a simple convention: the first value corresponds to the central wavelength (in nanometers, nm), which is the midpoint of the light allowed to pass through the filter ; the second corresponds to the bandwidth, i.e., the width of this "window" of light. Thus, a 450/45 filter transmits light between 427.5 nm and 472.5 nm (i.e., 450 ± 22.5 nm). »

- l.108-109 "In addition to fluorescence, two scatter parameters were recorded to describe particle morphology: grain size and granularity." Give a precise definition of what size and granularity represent in this study as they can be misinterpreted. Is the size value the maximum axis of the pollen grain? What values do we expect for granularity?

We have clarified this in the revised manuscript.

L119-126: We changed for "The forward scatter (FSC) measures light diffracted by the pollen grain at a flat angle, reflecting the approximate diameter of the grain. The sidewards scatter (SSC) measures light diffracted by the pollen grain at a right angle, reflecting its granularity." with "The forward scatter (FSC) detects light scattered at low angles in the forward direction, which correlates with the cross-sectional area of the particle, equivalent to a spherical diameter. For non-spherical particles like pollen (prolate, oblate, tricolporate, etc.), FSC reflects an average optical cross-section as the particle passes through the laser in a random orientation. The sidewards scatter (SSC) detects light scattered at ~90° (orthogonal) to the laser beam, which is sensitive to internal complexity and surface irregularities. In pollen, this captures internal granularity, wall sculpturing, apertures, vacuoles/ pollen sacs. The more complex the structure and texture of the pollen grain, the higher the granularity values will be."

- l.125-126 "This resulted in a total of 25 parameter values per particle." Since the previous sentence is quite long I would summarise here the number of parameters: "This resulted in three values for size, two for granularity and 20 for fluorescence, with a total of 25 parameter values per particle."

We made the suggested changes, see L139-140.

- l.127-129 “For each species, pollen grains were manually separated from debris using scatter density plots (size vs. granularity) and histograms of all fluorescence features.” This is a very interesting approach that needs to be shared in the aerobiology community!

We thank the reviewer for this positive feedback.

- l.140-141 “Among these, the Random Forest algorithm showed the best performance and was therefore selected for subsequent analysis.” Which criterion was used to determine that Random Forest showed the best performance? F1 score? Please state it briefly in the text again.

We used the F1 score; this is mentioned in the revised manuscript on L156.

- l.160-173 It is very interesting to compare the species level with the genus level. In these two paragraphs, you describe the taxa that performed with reduced accuracy. What about those that performed best?

L179-180: We added “Most species perform very well, with 75% of species achieving a F1-score above 0.70.”

- l.212-214 “The models achieved high classification performance (F1=0.76 at the species level and 0.90 at the genus level) highlighting the potential of this approach as a scalable alternative to traditional microscopy for pollen identification.” For clarity, it would be good to state that this performance was achieved under lab conditions.

We have added this in the revised manuscript.

L240-242: added « While these findings are promising, they were obtained using reference pollen grains collected directly on trees; further validation using atmospheric samples will be necessary before implementation in an airborne pollen monitoring network. »

Technical corrections

Technical corrections are highlighted in the attached PDF file.

We thank the reviewer very much for taking the time to provide this feedback; we really appreciate it, and we have made all suggested changes in the revised manuscript.