

Reviewer 1

We thank the reviewer for their careful reading of the manuscript and for the valuable comments and suggestions.

Reviewer: No code or training dataset is made directly available, which significantly hinders the reproducibility of the study and the reuse of the model for operational or research purposes. This is particularly regrettable given that dedicated portals would allow this with minimal effort. For example, the Sylva/AutoPollen data portal or the Zenodo repository could be used for the training datasets, and GitHub could be used for the code. The issue is exacerbated by the lack of information on model training and architecture.

Thank you for this suggestion. To ensure data reproducibility, we decided that the training datasets collected during the campaigns will be made publicly available via an open-access portal. The datasets will be accessible here (the link will be added after our data stewardship unit confirms the correctness of the data). In the Supplementary, we have included a PDF file showing a view of the repository.

The screenshot shows the RODBUK dataset page for "SWISENS POLENO JUPITER TRAINING DATASETS". The page is part of the University of Wrocław's Faculty of Earth Sciences and Environmental Management. It features a draft status and a warning to submit for review. The dataset is described as containing 11 datasets for five tree taxa (Alnus, Betula, Fraxinus, Pinus, Quercus) with detailed event information. It includes a keyword "pollen detection, machine learning, holography images, fluorescence spectra, real time detection" and is licensed under CC0 1.0. The page also has tabs for Files, Metadata, Licenses and terms, and Versions, along with a search bar and filter options.

RODBUK Search ▾ User Guide Support English ▾ Szymon Tomczyk ▾

Uniwersytet Wrocławski
Wydział Nauk o Ziemi i Kształtowanie Środowiska
Faculty of Earth Sciences and Environmental Management
(Uniwersytet Wrocławski)

Uniwersytet Wrocławski > Faculty of Earth Sciences and Environmental Management >

Info – This draft version needs to be submitted for review. When ready for sharing, please [submit it for review](#).

SWISENS POLENO JUPITER TRAINING DATASETS

Draft **Unpublished**

 Tomczyk, Szymon; Werner, Małgorzata; Malkiewicz, Małgorzata, 2026, "SWISENS POLENO JUPITER TRAINING DATASETS", <https://doi.org/10.34616/DMTZAB>, Uniwersytet Wrocławski, DRAFT VERSION ⓘ

[Cite Dataset ▾](#) [Learn about Data Citation Standards](#)

Access Dataset ▾

- [Submit for Review](#)
- [Edit Dataset ▾](#)
- [Link Dataset](#)
- [Contact Owner](#) [Share](#)

Description ⓘ

The data were created during training sessions conducted with the Swisens Poleno Jupiter developed by Swisens AG in Switzerland. The repository contains 11 datasets representing five different tree taxa: five datasets for *Alnus*, two for *Betula*, one for *Fraxinus*, one for *Pinus*, and two for *Quercus*. Each dataset includes information for every recorded event, corresponding to individual particles detected during the training process. For each event, the repository provides two holographic images of the particle, three .npy files, and a .json file containing detailed information about the analysed particle.

Keyword ⓘ pollen detection, machine learning, holography images, fluorescence spectra, real time detection

License/Data Use Agreement  [CC0 1.0](#)

Files **Metadata** **Licenses and terms** **Versions**

Search this dataset... 🔍

Filter by
File Type: All ▾ Access: All ▾

[+ Upload Files](#)

⬆️ Sort ▾

We have changed the data availability section as follows:

The training datasets are available at: (the link will be added)

In response to the reviewer's comment, a brief summary of the key features of the model architecture was added to the Methods section entitled "Model architecture". This issue is now addressed in the manuscript as follows:

"The pollen classification model was based on a customised convolutional neural network architecture specifically optimised for processing greyscale images generated by the SwisensPoleno digital holography module. Unlike standard pre-trained architectures such as VGG16, which are primarily designed for RGB image recognition tasks, the applied network was adapted to the characteristics of monochromatic holographic data, reducing unnecessary model complexity and improving compatibility with the input format. The model processed two orthogonal pollen images separately before combining the extracted features through fully connected layers. Additionally, fluorescence measurements could be incorporated through a dedicated network branch, enabling classifications both with and without fluorescence input (Crouzy et al., 2025; Erb et al., 2025; Sauvageat et al., 2020).

Non-biological particles were removed prior to classification using a deterministic morphological filter described by Sauvageat et al. (2020). Water droplet training datasets were created from operational measurements collected during fog and rain events and manually cleaned to remove artefacts such as aggregates and debris. The model was trained entirely from scratch without the use of pre-trained neural networks. As emphasised by Erb et al. (2025), reliable pollen classification requires training datasets that reflect the full variability of each pollen taxon, including samples collected from different plants, locations, and meteorological conditions. This underlines the importance of the present study. The detailed network architecture is available on GitHub (MeteoSwiss Biometeorology Team, 2025)."

Reviewer: The benchmark model ("old model") shows reasonable correlations with reference measurements, but there are massive scaling issues (Figure 3). This could be related to issues when applying scaling factors used for converting raw events to concentrations, as different versions of the Swisens Poleno software have been known to circulate. I strongly recommend that the authors double-check how scaling is applied when using the benchmark model, as this has a dramatic effect on the metrics for exposure class determination (leading to the old model systematically overestimating classes), the RMSE, and consequently the conclusions of the paper.

We used the first version of the model together with the default scaling factor available at the time of analysis. We acknowledge that scaling factors can be further optimized using local measurements, which may further improve model performance. However, the aim of this study was to compare the models under consistent processing conditions using the resources available at that stage. The results also highlight the importance of local data, both for model training and for future calibration of the scaling factor.

Reviewer: Line 71: “advanced instrument” I suggest removing “advanced” as this adds no information per se.

The term “advanced” has been removed, and the sentence now refers simply to “an instrument”.

Reviewer: Line 75: “detections” remove “s”.

Corrected as suggested; the term now appears as “detection”.

Reviewer: Line 76: remove “advanced”.

The term “advanced” has been removed.

Reviewer: Line 81–82: Which “reference pollen database” are the authors referring to? How can this database be accessed?

We thank the reviewer for this important comment. Our intention was to emphasize that building appropriate datasets for algorithm training is a crucial step; however, we agree that the original wording may have been unclear. The sentence has been revised to:

“The reference pollen database provided with the Swisens Poleno System comprises 14 taxonomic classes and an additional class for water droplets. It was originally developed based on data collected at the site where the detector was built in Switzerland, without adaptation to new locations. In this study, the reference database used in the Swisens Poleno Jupiter system was further developed through extension and refinement in order to improve detection performance.

Reviewer: Line 92: “These tree species” – suggestion to use “genera” instead of “species”.

This was a language inaccuracy and has now been corrected to “genera”.

Reviewer: Line 109–111: Maybe compare hourly correlations you obtain with the results from Chappuis et al. (2020), Aerobiologia, for a different automatic pollen monitor.

We thank the reviewer for this valuable suggestion. We regret that we were not initially aware of this publication. The manuscript has been revised accordingly. The following sentence has been updated to better reflect the existing literature:

“Although frequently mentioned in aerobiology, its influence on hourly pollen variability derived from automatic measurements has rarely been investigated compared to standard local meteorological drivers, with the first attempt in this direction made by Chappuis et al. (2020).

In addition, we have now included this reference in the Discussion section:

“Analysis of hourly pollen percentages revealed distinct diurnal patterns consistent with daily meteorological rhythms, as also reported by Chappuis et al. (2020).”

The reference has been added to the Literature as follows:

Chappuis, C., Tummon, F., Clot, B. et al. (2020). Automatic pollen monitoring: first insights from hourly data. *Aerobiologia*, 36, 159–170. <https://doi.org/10.1007/s10453-019-09619-6>

Reviewer: Line 119: “20 meters” what about surrounding buildings height /height of the urban layer if any.

We have extended the description as follows: “The Swisens Poleno Jupiter detector is installed in Wrocław (southwestern Poland), on the roof of the Department of Climatology and Atmosphere Protection, University of Wrocław, at an elevation of 20 meters above ground level, with no other high-rise structures in the immediate surroundings”.

Reviewer: Line 132: “airborne cytometer” to be replaced by “airflow cytometer”

The term has been corrected, and “airborne cytometer” has been replaced with “airflow cytometer” in the revised manuscript.

Reviewer: Line 157-158: Why were Poaceae, particularly important for respiratory allergies disregarded in the study?

This work is part of a broader effort to expand and improve the automatic detection system. In the first stage, we focused on tree pollen taxa, for which the detection performance is currently more robust. Herbaceous taxa, including *Poaceae*, are being addressed in the second stage, as their reliable detection requires further development and refinement of the classification algorithms.

Although *Poaceae* pollen is known for its strong allergenic potential, its detection efficiency with the Swisens Poleno system is currently lower compared to the tree pollen taxa. This is primarily due to the greater complexity associated with herbaceous taxa,

including the high diversity of grass species and the increased effort required to collect sufficiently representative training datasets.

Therefore, in line with the aim of this study-to analyze meteorological influences and the diurnal variability of pollen concentrations-we prioritized taxa for which the highest data quality and reliability could be ensured. Ongoing work is focused on improving the detection of herbaceous taxa, and a dedicated publication addressing *Poaceae* and related species is planned as part of the second stage.

Reviewer: Line 221: How was the size of the buffer determined?

The 2.5 km buffer was chosen to capture the main local vegetation surrounding the monitoring site. This distance ensures the inclusion of the most relevant tree stands, i.e. Park Szczytnicki and the Odra River corridor. Additionally, it has been shown that local pollen emissions typically extend to distances of approximately 1 km or more (Sofiev et al., 2012; Airborne Pollen Transport), which supports the choice of a buffer large enough to capture this effect.

Reviewer: Line 265: “Cruozy” -> “Crouzy”

The name has been corrected from “Cruozy” to “Crouzy” in the revised manuscript.

Reviewer: Line 324: “Polleno” takes one “l”, check throughout the manuscript as this appears one more time on line 413.

It has been corrected throughout the manuscript.

Reviewer: Figure 3: I suggest to use a capital letter in front of each genus name

Thank you for this helpful comment. The issue has been corrected in Figure 3 and consistently revised in the remaining figures where applicable.

Reviewer: Line 340, 348 and 352: the massive RMSE differences might be related to improper scaling.

Related to the main point number 1.

Reviewer: Figure 5, 7 and the S1 to S5 should be made self-contained by the caption for the reader to understand all the symbols used and different metrics represented.

Thank you for the suggestion. The captions have been revised to be more detailed and self-contained to improve clarity for the reader. The updated versions are provided below.

Figure 5. Performance of automatic ML models compared with Hirst reference data, shown as a performance diagram. The diagram is based on daily data from the 2024–2025 pollen season. The metrics include success ratio (SR), probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI).

Figure 7. Spearman correlation matrix between meteorological variables and pollen concentrations. The correlation matrix is based on daily pollen season data (2024–2025) and corresponding meteorological data for the same period. Abbreviations: hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Figure 1S. The hourly Spearman correlation matrix between meteorological variables and *Alnus* pollen concentrations for the 2024–2025 season. hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Figure 2S. The hourly Spearman correlation matrix between meteorological variables and *Betula* pollen concentrations for the 2024–2025 season. hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Figure 3S. The hourly Spearman correlation matrix between meteorological variables and *Fraxinus* pollen concentrations for the 2024–2025 season. hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Figure 4S. The hourly Spearman correlation matrix between meteorological variables and *Pinus* pollen concentrations for the 2024–2025 season. hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Figure 5S. The hourly Spearman correlation matrix between meteorological variables and *Quercus* pollen concentrations for the 2024–2025 season. hum – relative humidity, rain – precipitation sum, sun – sunshine duration, temp – air temperature at 2 m, wind – wind speed.

Reviewer: Line 501-510: the case of rainfall could be discussed in more details as the result is a bit counter intuitive (rainfall is expected to scavenge airborne pollen).

In the discussion section, we have added the following statement: “However, rainfall was not always statistically significant, highlighting the complex nature of its influence on pollen concentrations, as some studies even report increases in concentrations after precipitation events (Kluska et al., 2020). The effect varies with rainfall intensity and type, as precipitation processes are associated with pollen wash-out. Kluska et al. (2020) showed that under low-intensity rainfall ($<1.0\text{--}2.5\text{ mm}\cdot\text{h}^{-1}$), no decreases and even slight increases in pollen concentrations can occur.”

Reviewer: Table 2: the results on PBL are interesting but again counter intuitive, how could that be explained? One might expect a compression effect.

There is a positive correlation, which suggests a relationship; however, the low correlation values indicate that the process is likely more complex. This is emphasised in the Discussion section. However, the reviewer is correct that further explanation is needed, and therefore this part has been expanded as follows:

“suggesting that other factors, such as local turbulence or canopy-level processes, could play a more important role in its short-term variability. It is reflected in planetary boundary layer (PBL) processes, where studies suggest that katabatic flows can enhance pollen concentrations at night, while daytime convective turbulence promotes dispersion, contributing to variability near the surface. More generally, differences between pollen types arise from interacting factors such as source distribution, meteorological conditions, PBL structure, pollen properties, flowering phenology, and other environmental influences (Andújar-Maqueda et al., 2025).”

Reviewer: Line 544 “Dissucion” typo

The typo “Dissucion” has been corrected to “Discussion.”

Reviewer: Whole discussion: correlation vs. causality issues, for example “crucial” Line 642, “stronger effect” Line 658, “influential” Line 660, “responsive”

We have revised the manuscript to address concerns regarding potential overinterpretation of correlations and causal language. Specifically, we have systematically replaced terms implying causality or strong inference (e.g., “crucial factor,” “stronger effect,” “influential,” “responsive,” and “key factor”) with more neutral, association-based terminology.

Reviewer: Line 674 or “shaping” Line 690”.

Lines 674 and 690 have been revised to replace causal wording (e.g. “shaping”) with neutral, correlation-based phrasing.

Reviewer: Line 717: Pine pollen might not be relevant for allergies.

The reviewer is correct; however, pine pollen can reach high concentrations and is also considered allergenic. Therefore, despite its relatively weak allergenic potency compared to other taxa, it is still relevant to include it in this analysis.

Reviewer 2

We thank the reviewer for their careful reading of the manuscript and for the valuable comments and suggestions.

Reviewer: In line 56 of the manuscript you mention that many monitoring station are currently converting to automatic pollen samplers. It is true, yet these types of samplers remain very expensive. This creates a situation of a few locations in each country with very accurate pollen monitoring, but also large gaps in space with many locations that have no data et al. To create a monitoring network with a good spatial coverage, low-cost sensors should be developed and tested.

We thank the reviewer for this important comment. We fully agree that, despite the ongoing transition toward automatic pollen monitoring systems, their high cost remains a major limitation for the development of dense observational networks. As a result, most countries still rely on a limited number of high-precision stations, which leads to substantial spatial gaps in pollen data coverage. We also agree that the development and validation of low-cost sensors is a promising direction for improving spatial representativeness. This aspect was emphasized more clearly in the revised manuscript.

“However, Hirst-type traps remain the reference against which other pollen and fungal spore detection methods are evaluated (Tummon et al., 2021). At the same time, the development and validation of low-cost sensors are essential for expanding monitoring networks, as they can help identify optimal locations for the subsequent deployment of automatic samplers and improve overall spatial coverage.”

Reviewer: Comment on line 76: I agree that automatic pollen detection has not achieved its full potential and maturity, but I also wonder when further improvements will still make considerable additional contributions (and to what purpose). You collected high-quality reference data for retraining and validating the pollen detection model. For every country/sensor, this will have to be done again.

We thank the reviewer for this insightful comment. We agree that automatic pollen detection requires local calibration and reference datasets for each region and instrument. However, further improvements remain important, particularly through expanding the range of classified particles and taxa. In this context, continuous model development can enhance classification performance by including additional pollen taxa, fungal spores, and other airborne particles such as Saharan dust, thereby increasing the applicability and usefulness of automatic monitoring systems. Data collected from different locations should also be made available across stations to improve detection efficiency in new instruments, which further highlights the relevance of this work. The prepared datasets can be used in a range of studies and applied to various detection systems. Importantly, within this study, we provide a pollen dataset collected during the research campaign, which can support future model development, calibration, and intercomparison studies across different detection systems. The dataset is available at: (the link will be added after our data stewardship unit confirms the correctness of the data). In the Supplementary Materials, we have included a PDF file showing a view of the repository.

RODBUK

Search ▾User GuideSupportEnglish ▾Szymon Tomczyk ▾



Uniwersytet Wrocławski
Wydział Nauk o Ziemi
(Katedra Geologii i Geofizyki)

Faculty of Earth Sciences and Environmental Management
(Uniwersytet Wrocławski)

Uniwersytet Wrocławski > Faculty of Earth Sciences and Environmental Management >

Info – This draft version needs to be submitted for review. When ready for sharing, please submit it for review.

SWISENS POLENO JUPITER TRAINING DATASETS

Draft **Unpublished**



Tomczyk, Szymon; Werner, Małgorzata; Malkiewicz, Małgorzata, 2026, "SWISENS POLENO JUPITER TRAINING DATASETS", <https://doi.org/10.34616/DMTZAB>, Uniwersytet Wrocławski, DRAFT VERSION ⓘ
Cite Dataset ▾[Learn about Data Citation Standards.](#)

Access Dataset ▾
Submit for Review
Edit Dataset ▾
Link Dataset
Contact OwnerShare

Dataset Metrics ⓘ
0 Downloads ⓘ

Description ⓘ

The data were created during training sessions conducted with the Swisens Poleno Jupiter developed by Swisens AG in Switzerland. The repository contains 11 datasets representing five different tree taxa: five datasets for Alnus, two for Betula, one for Fraxinus, one for Pinus, and two for Quercus. Each dataset includes information for every recorded event, corresponding to individual particles detected during the training process. For each event, the repository provides two holographic images of the particle, three .npy files, and a .json file containing detailed information about the analysed particle.

Keyword ⓘ

pollen detection, machine learning, holography images, fluorescence spectra, real time detection

License/Data Use Agreement

[CC0 1.0](#)

FilesMetadataLicenses and termsVersions

Search this dataset... 🔍

Filter by
File Type: All ▾Access: All ▾

+ Upload Files

⬆ Sort ▾

Reviewer: The relevance of the diurnal concentration pattern analysis remained unclear to me until I received the results section. You seem to believe that information on the high-risk times of day will benefit the allergy sufferers, and that it enables you to assess the influence of meteorological parameters. I understand the link with the meteorological parameters, yet I doubt if warning allergy sufferers of high-risk moments during the day for will be implementable. Referring to my first point, the lack of a good spatial network will probably have a larger impact.

We thank the reviewer for this comment. We agree that limited spatial coverage of monitoring networks is currently a major constraint for operational allergy warnings. The aim of the diurnal analysis in this study is primarily to improve the understanding of sub-daily variability and its relationship with meteorological drivers. In this context, time of day may serve as an additional predictor in forecasting approaches, for example in machine learning models that integrate meteorological variables and concentration data. Thus, while direct implementation of within-day warning schemes may be limited at present, such information can still contribute to improving predictive models and future warning systems.

Reviewer: Interesting that the old model overestimated the duration of the season for Quercus and Betula. It is also interesting that the old model shows much shorter periods for 2025 compared to 2024. Is there any explanation for this?

The performance of the old model, which is based solely on holography, may vary over time depending on the presence of false signals, i.e. particles resembling specific pollen taxa, which can occur in the atmosphere and potentially interfere with the classification. This aspect was already included in the manuscript.