Low-cost Monitoring of Pollen Bioaerosols with OPCs and Machine Learning

Sophie A. Mills[1,2], A. Robert MacKenzie[1,2], Francis D. Pope[1,2]
[1]Department of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK
[2]Birmingham Institute of Forest Research, University of Birmingham, B15 2TT, UK

The Aerosol Society
UNIVERSITY OF BIRMINGHAM | BIFOR
CENTA
Natural Environment Research Council

Introduction

Up to 40% of industrialised populations suffer from pollen allergies and pollen can interact with cloud processes, affecting weather and climate.

Current pollen monitoring standard: Hirst-type trap and pollen grains counted manually under the microscope - laborious, usually low (daily) time resolutions and delay in data accessibility.

Technological advancements and automation needed. Recently developed high-performance instruments generally use automated imaging and machine learning techniques but are very expensive.

We demonstrate a method using machine learning to obtain pollen information from optical particle counter (OPC) sensors which provide remotely-accessible, real-time, high-time resolution data. These would be useful supplements to hybrid pollen monitoring networks and cover greater spatial resolution with affordable costs.

Methods

Trained machine learning models (neural network & random forest) to estimate pollen concentrations from OPC particle size data, with and without temperature and RH, supervised on Hirst-measured pollen concentrations.

Utilised explainable AI methods (SHAP) to interpret the models and evaluate model-learned relationships between OPC particle size bins and concentrations of different pollen taxa.

Continuing work includes using fitted models to predict pollen concentration in other environments, to assess method generalisability and investigate pollen dynamics in previously inaccessible contexts.

Results

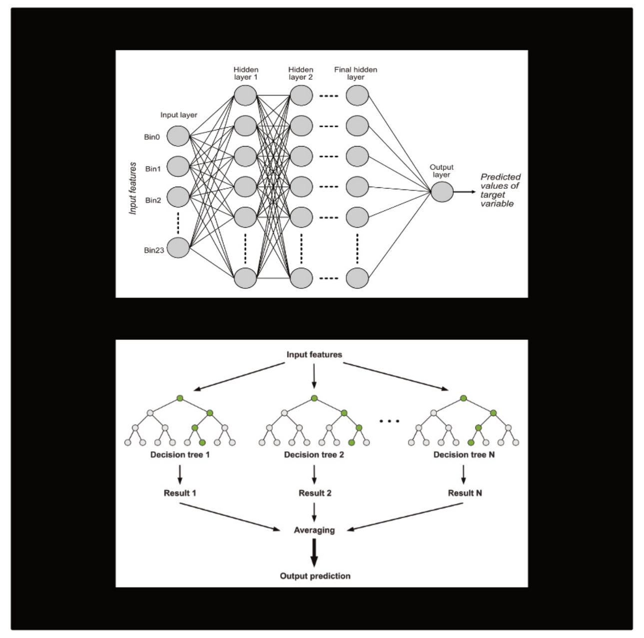
Models have been optimised, by hyperparameter tuning and regularisation techniques, and generally achieve good metrics when evaluated on an unseen test dataset. Grass, oak, birch and pine pollen taxa can be estimated to varying degrees of accuracy and training models on coarser, daily (as opposed to hourly) time resolution generally results in higher R2 scores though less detail.

Models trained with meteorological input features tend to perform better on the test dataset from the same environment as training data. However, the strong influence of RH and temperature on these models may be an issue for generalisability in other environments, such as other climates or indoors. Models based solely on particle size information may be more generalisable.

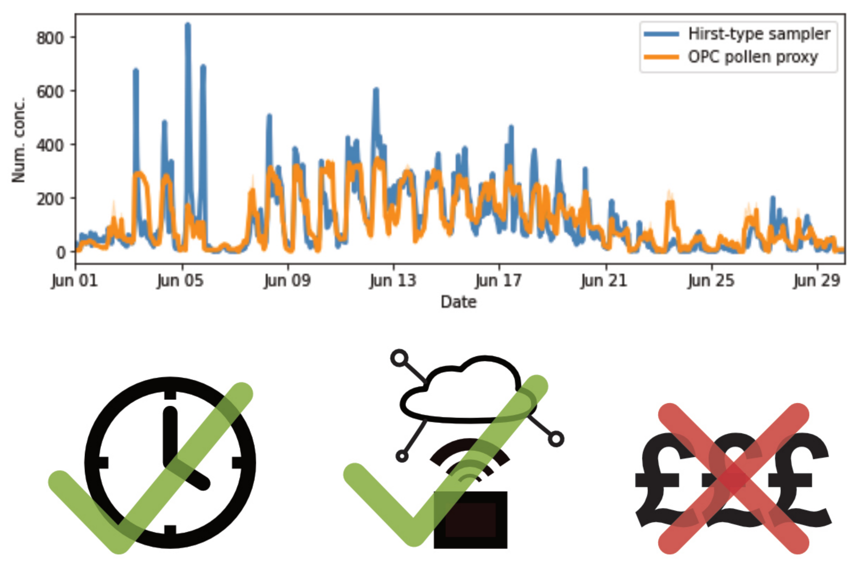
OPC sensors were not good at detecting larger particle sizes, so the models appear to have learned to estimate pollen concentrations based on subpollen-sized particles <10 µm. Total pollen in this context was largely represented by grass (and pine) pollen so the particle size bin correlations appear similar. Oak pollen is distinctly different and strongly positively correlated with particles at 1.7-2.3 µm.

Low-cost OPC sensor

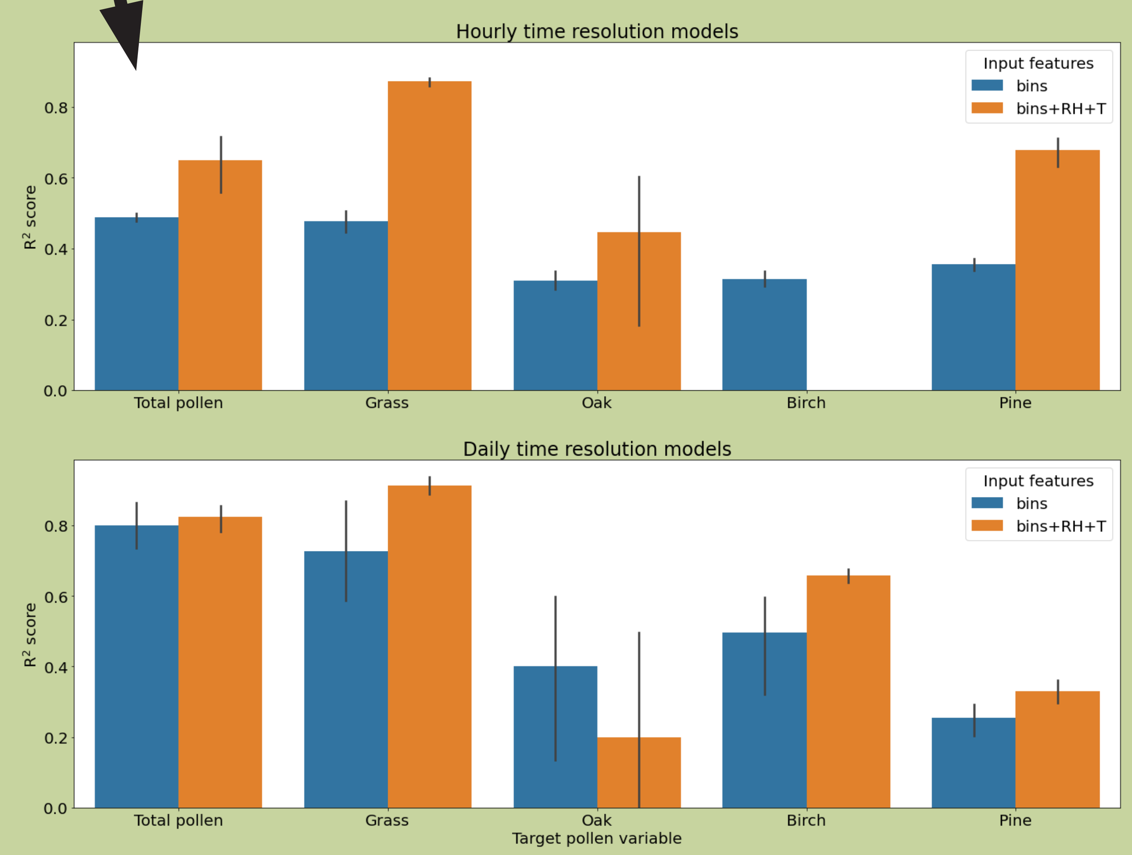
Machine learning methods

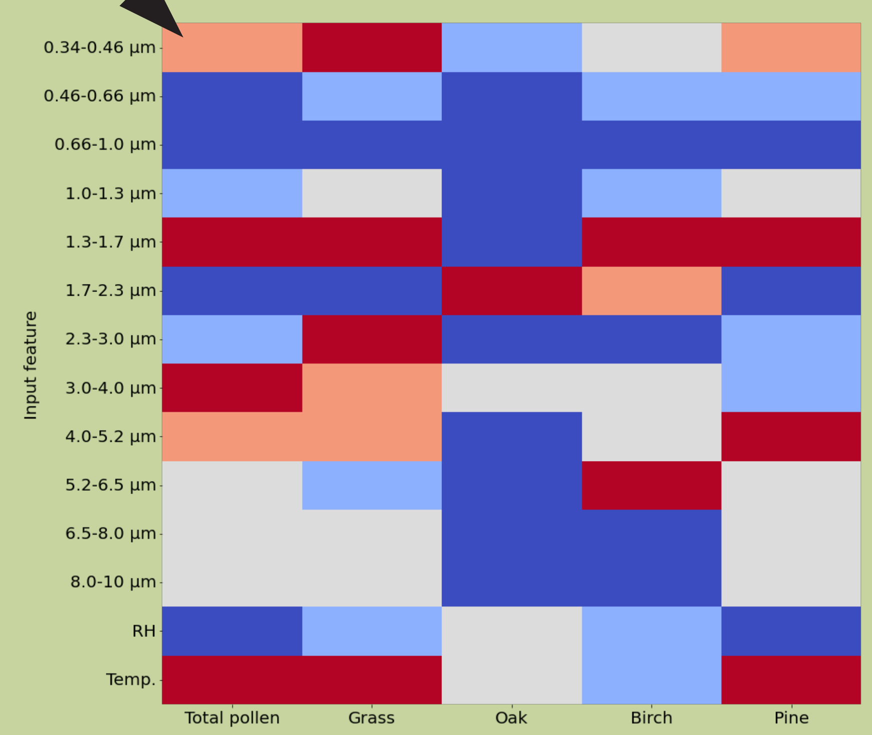


Real-time, remotely accessible, low-cost method for monitoring pollen

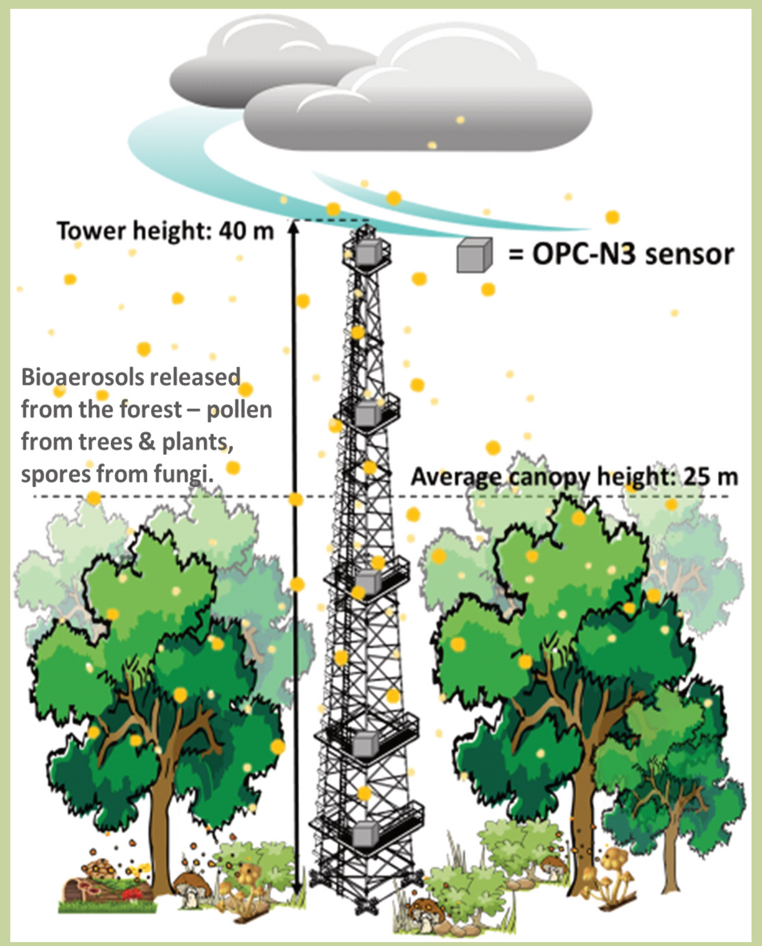


Use the QR code above to see our recently published paper for more details on our ML methods - *"Constructing a pollen proxy from low-cost Optical Particle Counter (OPC) data processed with Neural Networks and Random Forests"* by Mills et al., 2023.





Visually observable (nonlinear) relationships between input features and target variables from plotted SHAP values. Red = positive correlation; blue = negative correlation.

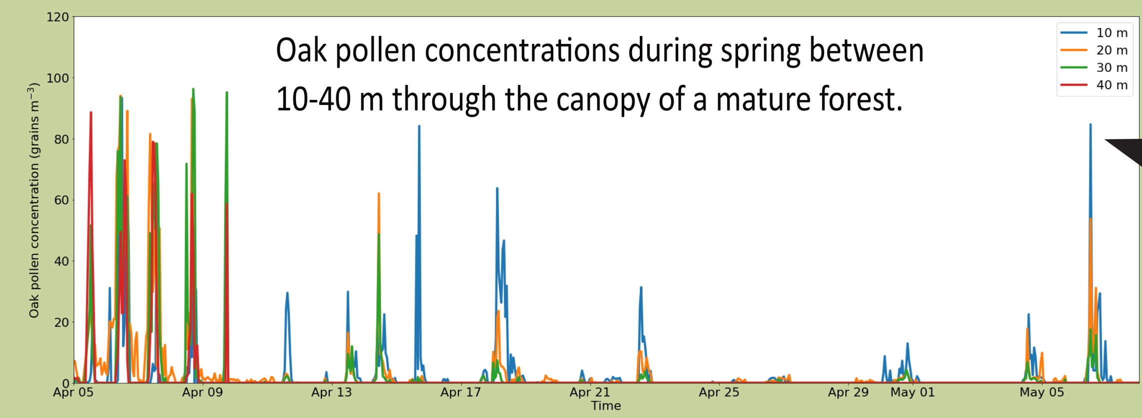


BIFoR FACE

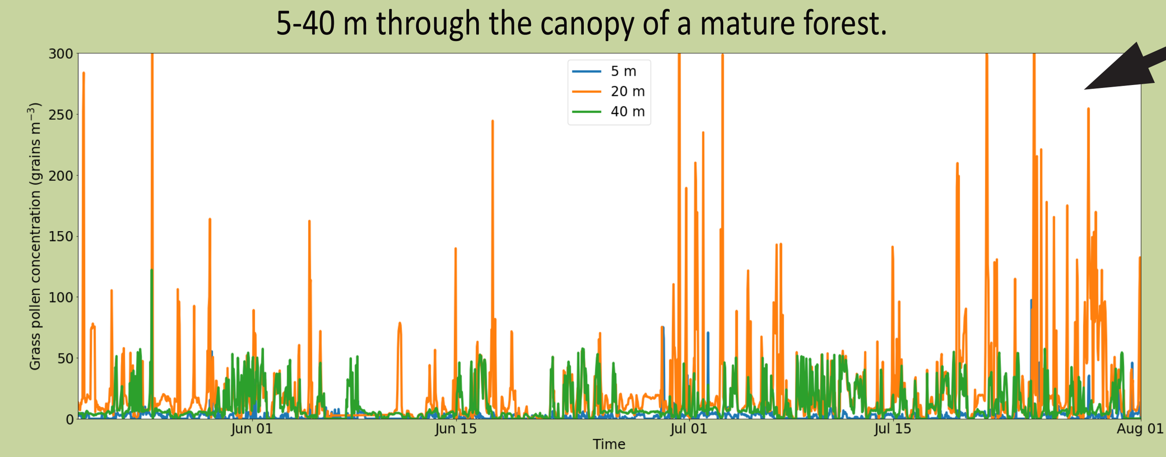
Method applied to OPC data collected at the Birmingham Institute of Forest Research (BIFoR) Free Air CO_2 Enrichment (FACE) facility to investigate pollen dynamics within this mature forest site.

BIFoR FACE is a large-scale, long-term experiment designed to address the impact of climate and environmental change on woodlands. Arrays within the forest are fumigated with elevated and ambient levels of CO_2 and we have OPC data collected from within them. The effect of elevated CO_2 treatment on pollen concentrations remains inconclusive from our data so far.

Further OPC data have been collected from sensors at different heights up the flux tower in the forest (see left). The time series (see right) show oak (above) and grass (below) estimated pollen concentrations, during their respective pollen seasons.



Grass pollen concentrations during summer between 5-40 m through the canopy of a mature forest.



BIFoR FACE - Results

Oak pollen: Sensors above the canopy at 30 and 40 m (green & red) initially measure higher concentrations of oak pollen before diminishing, but as the season progresses the sensors lower in the canopy at 10 and 20 m (blue & orange) measure higher concentrations relatively. This is congruent with observations that catkins higher in the canopy, exposed to more sunlight, tend to mature and release pollen earliest.

Grass pollen: Concentrations appear lowest beneath the canopy at 5 m (blue) and highest at 20 m mid-canopy level (orange). Grass pollen is unlikely to have come from the forest floor but has either travelled from outside the forest or the algorithm may be confused by other particles of similar size pattern from the canopy.

This method of estimating pollen concentrations still has limitations in terms of accuracy, however we emphasise the novel wealth of information that can be obtained with networks of low-cost sensors in dynamic environments such as this. We anticipate further advancements and accumulation of training data could significantly improve accuracy.

Contact email:
sam919@student.bham.ac.uk