'Finding a needle in a haystack' - monitoring pollen with low-cost sensors 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

UNIVERSITY OF BIRMINGHAM | BIFOR | CENTA
UKRI Natural Environment Research Council

Introduction

Why pollen?

- Airborne pollen plays a vital role in the reproductive cycle and spread of plants, exacerbates health conditions e.g. allergic rhinitis and asthma, and has potential to affect cloud processes and climate.
- It is yet unclear how changing climate and rising CO_2 will affect pollen production and its role in the atmosphere.

Why low-cost sensors?

- Current pollen monitoring methods have many limitations e.g. time consuming, high cost, low spatiotemporal-resolution.
- The low-cost optical particle counter (OPC) sensors we implement have advantages of high time resolution, real-time data and deployment flexibility.

Overview

With extensive datasets captured across the BIFoR FACE facility we can investigate pollen activity as it is released from the forest with temporal and spatial resolution that has not been possible before.

Acknowledgements

Special thanks to Andrew Tanner who helped functionalise the sensors, the BIFoR team who facilitated their deployment onsite at BIFoR, and Jose Manzano and Fiona Tummon who facilitated their deployment at the EUMETNET AutoPollen - ADOPT COST Action Intercomparison Campaign in Munich.

Questions? Please feel free to contact:
sam919@student.bham.ac.uk

The challenge - 'finding a needle in a haystack'

- Our OPC sensors simply record particle number information across 24 sized bins ranging bewteen 0.35-40 µm. From this data we need to isolate a signal for pollen.
- We have been exploring methods to isolate a potential pollen signal and evaluate these methods by comparing OPC particle bin data to parallel data from a reference Hirst volumetric pollen sampler, obtained from the EUMETNET AutoPollen - ADOPT COST Action (CA18226) Intercomparison Campaign 2021 in Munich, Germany.

Ref

'Pollen Proxy' method:

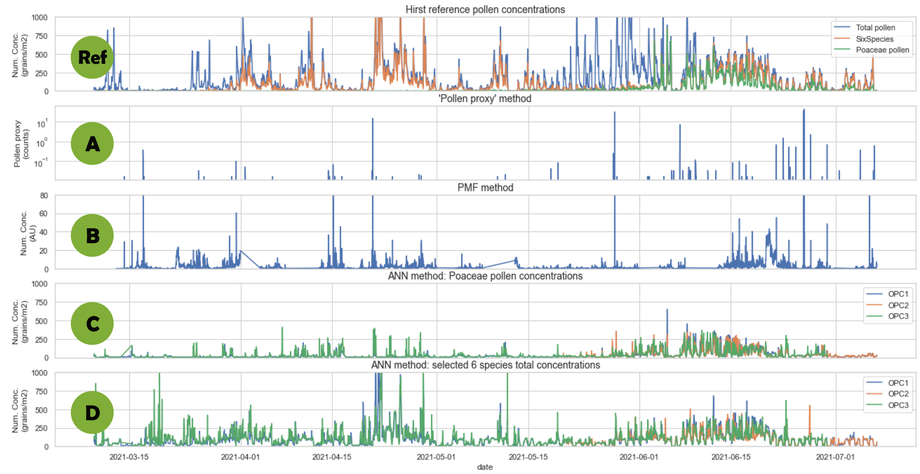
A

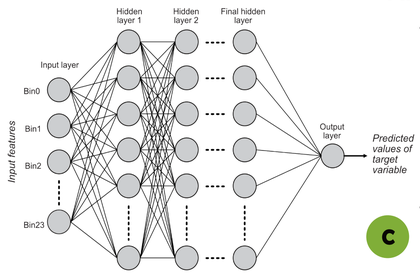
- Sum of all particles in bins between 10-40 µm (i.e. pollen size range).
- Neglible correlation was found with pollen reference.

Positive Matrix Factorisation (PMF) method:

B

- Source apportionment method whereby a PMF model is used to obtain the contribution of various aerosol sources over a time series, often used to identify specific air pollutants.
- Factors (i.e. sources) were identified with some correlation to certain pollen species.





Simplified diagram of ANN model.

Artificial Neural Network (ANN) method:

- Neural network model constructed taking input features of sized bin particle counts (and humidity) and trained against target variable - e.g. Poaceae (grass) pollen or a collection of 6 select species (including grass, oak and birch) - to produce predicted concentrations for the target pollen species based on OPC data.

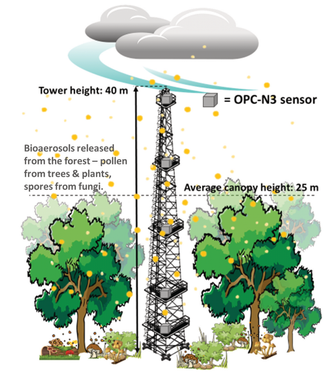
C

- Poaceae pollen model predictions achieved R^2 value of 0.61 against Poaceae reference and 0.25 against total pollen.

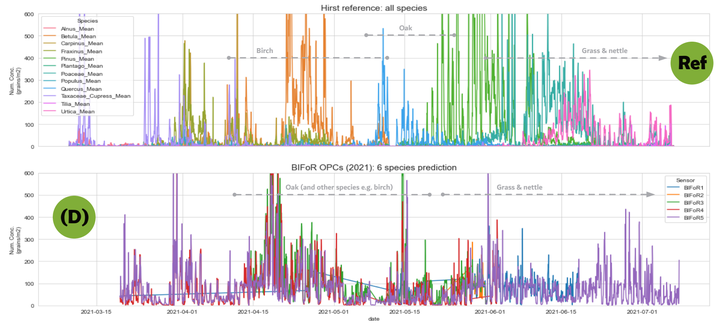
D

- 'Six species' model achieved R^2 value of 0.58 against the collective 6 species and 0.43 against total pollen reference concentrations.

Pollen activity at BIFoR FACE



- OPC datasets have been captured at the BIFoR FACE facility for varying heights through the canopy up to 40 m (see diagram) and in both elevated and ambient CO_2 arrays during multiple pollen seasons.
- We can use the model trained on the Munich Intercomparison Campaign dataset to predict pollen concentrations from the OPCs at the BIFoR FACE facility.



- From comparison between the reference pollen concentrations in Munich during 2021 and predicted concentrations for the 'six species' model applied to the BIFoR OPC data over the same period, a picture can be constructed of the pollen seasons in the forest.
- Being predominantly an oak forest, activity observed from April-May is most likely due to oak pollen release (as well as other species such as birch) which appears to have started a few weeks earlier in the UK than Germany. Meanwhile, activity occurring from late May to July would be due to grass and nettle pollen.

Continuing work...

- Can the ANN model be improved to make even more accurate predictions?
- BIFoR OPCs intercomparison - are there significant differences between pollen concentrations at different heights or in different arrays?
- Can we observe a possible effect of CO_2 levels on pollen production and what implications are there for the future?