Posted on 22 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10500275.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Applying machine learning and deep learning to forecast allergic pollen using environmental, land surface and NEXRAD radar parameters

Gebreab Zewdie[1], David Lary[2], Xun Liu[2], Daji Wu[3], and Estelle Levetin[3]

[1]University of Texas at Dallas
[2]The University of Texas at Dallas
[3]The University of Tulsa

November 22, 2022

Abstract

Airborne allergic pollen is a well known trigger for several cases of public health issues affecting millions of people. The effect is highly prevalent in the temperature region, especially in the North American region and Europe. For example, about 50 million Americans are affected by pollen caused allergy and similarly, studies show quite a significant population of Europe is affected. Contrary to the higher abundance of pollen in rural areas, pollen allergy is severe in urban areas than rural environments. Of all sources, it is the Ambrosia pollen that affects most due to its abundant production, strong allergic potency and its high prevalence near urban areas. Hence estimating the concentration of allergic pollen in the ambient atmosphere and notifying the public is crucial for people with allergies and health professionals who care for them. In this workshop, we present estimation of allergic pollen (particularly Ambrosia pollen) using advanced machine learning methods and input parameters from a suite of sources ranging from land surface to global reanalysis models and NEXRAD weather radar measurements at location of Tulsa, Oklahoma. We will present results of the machine learning model tested using an independent dataset and characterization of each atmospheric and land surface parameters' importance for the machine learning estimation.

Applying machine learning and deep learning to forecast allergic pollen using environmental, land surface and NEXRAD radar parameters

Gebreab K. Zewdie, David J. Lary, Xun Liu, Daji Wu and Estelle Levetin, The University of Texas at Dallas, AGU Fall 2018 Meeting, Washington DC

Objectives

- The objective of this research is to use machine learning/deep learning to forecast allergic pollen using atmospheric weather and land surface (ECMWF and MERRA) and NEXRAD radar parameters.

Pollen measurements

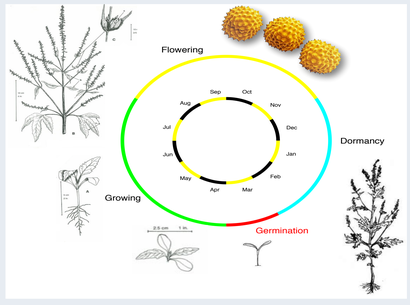


Figure 1: Life Cycle of Ambrosia.

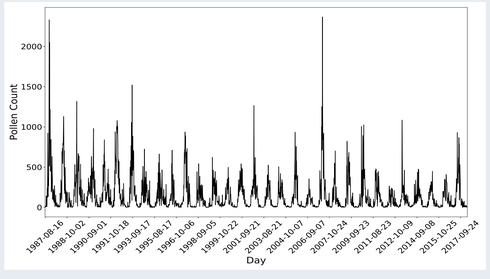


Figure 2: Pollen measurements made from 1987 to 2017.

Introduction: predictor variables

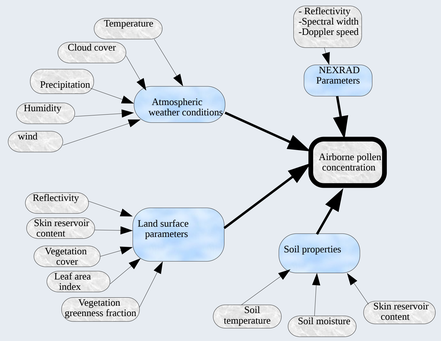
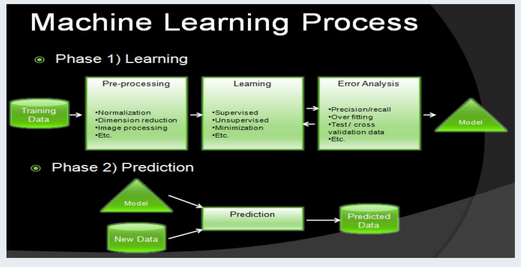


Figure 3: Atmospheric and land surface parameters controlling the distribution pollen

- We use a suite of meteorological and land surface parameters from the ECMWF, MERRA and NEXRAD radar to forecast allergic pollen
- This parameters control the production and dispersion of allergenic pollen
- We use deep neural networks, random forest and extreme gradient boosting.

Machine Learning

- Machine learning employs mathematical and statistical approaches that 'learn' by example from a training data set.
- It is suitable for problems in which we do not have the functional relation between the output parameter(s) we would like to estimate and the set of predictors.
- We develop the model using the training set (the data set from 1987 - 2011)



Neural Network/Random Forest

Neural networks are learning algorithms whose performance is inspired in analogy the flow of information in the human brain.

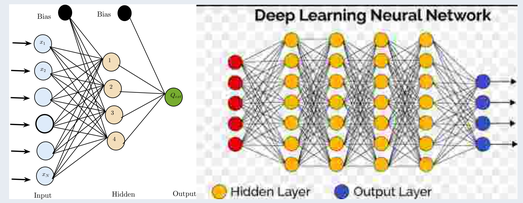


Figure 4: The neural network architecture.

$$Q = b^2 + \sum_i^4 w_i^2 f\left(b_i^1 + \sum_{j=1}^N w_{ij} x_j\right) \qquad (1)$$

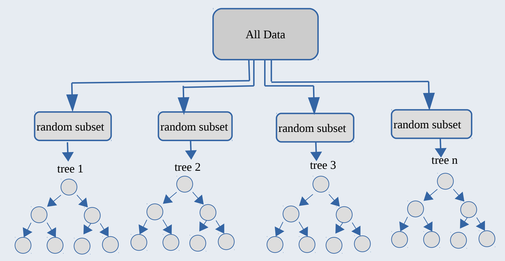


Figure 5: The Random Forest random sampling method

Results: Variable Importance estimated

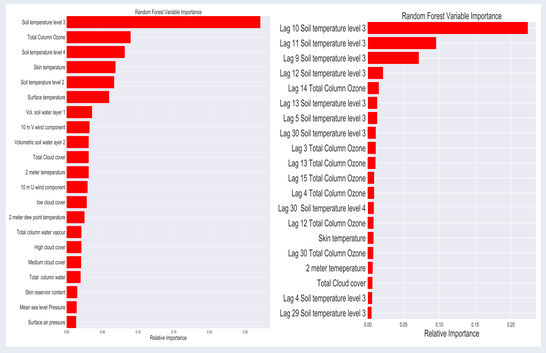


Figure 6: Variable importance estimated using the random forest machine learning method.

Results: forecasted pollen

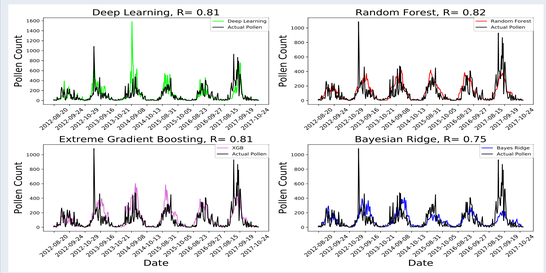


Figure 7: Forecasted pollen using independent test data for the 2012-2017 pollen season.

Results: estimated error

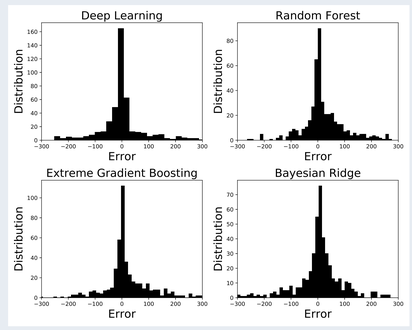


Figure 8: Error estimated for the four machine learning methods shown in Figure 7.

Pollen estimated using NEXRAD

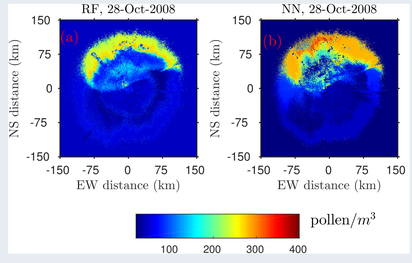


Figure 9: Pollen estimated over a large spatial area using NEXRAD radar data.