

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

Health Impacts of Pollen

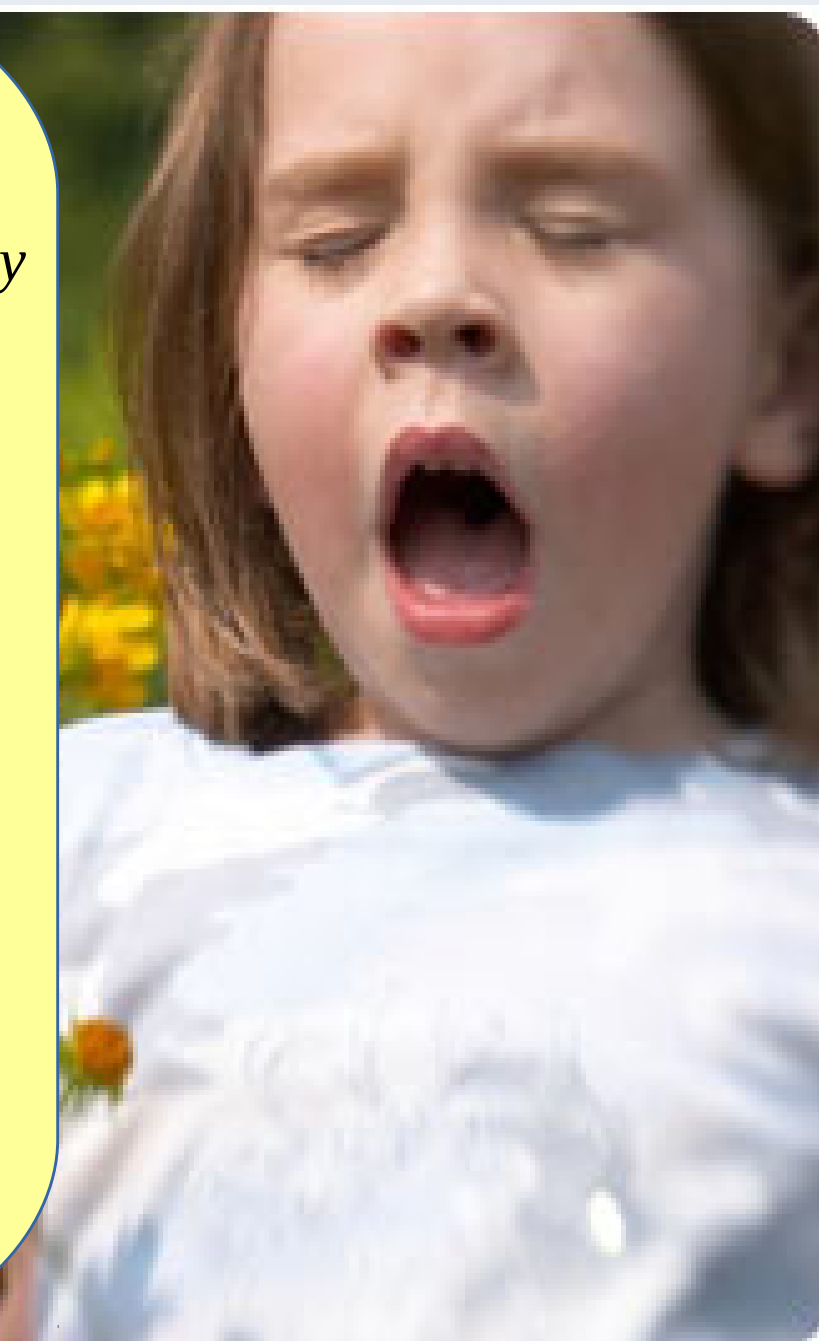
≈50 million Americans have pollen allergy
In 2005 in the US alone:
3.5 million work days lost.
2 million school days lost.

Health care costs increase from 6 billion in 2000 to 11 billion in 2006.

Increase in Asthma attack.

Large number of people visit emergency room.

Leads to death in highly sensitive people during high pollen season



- 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

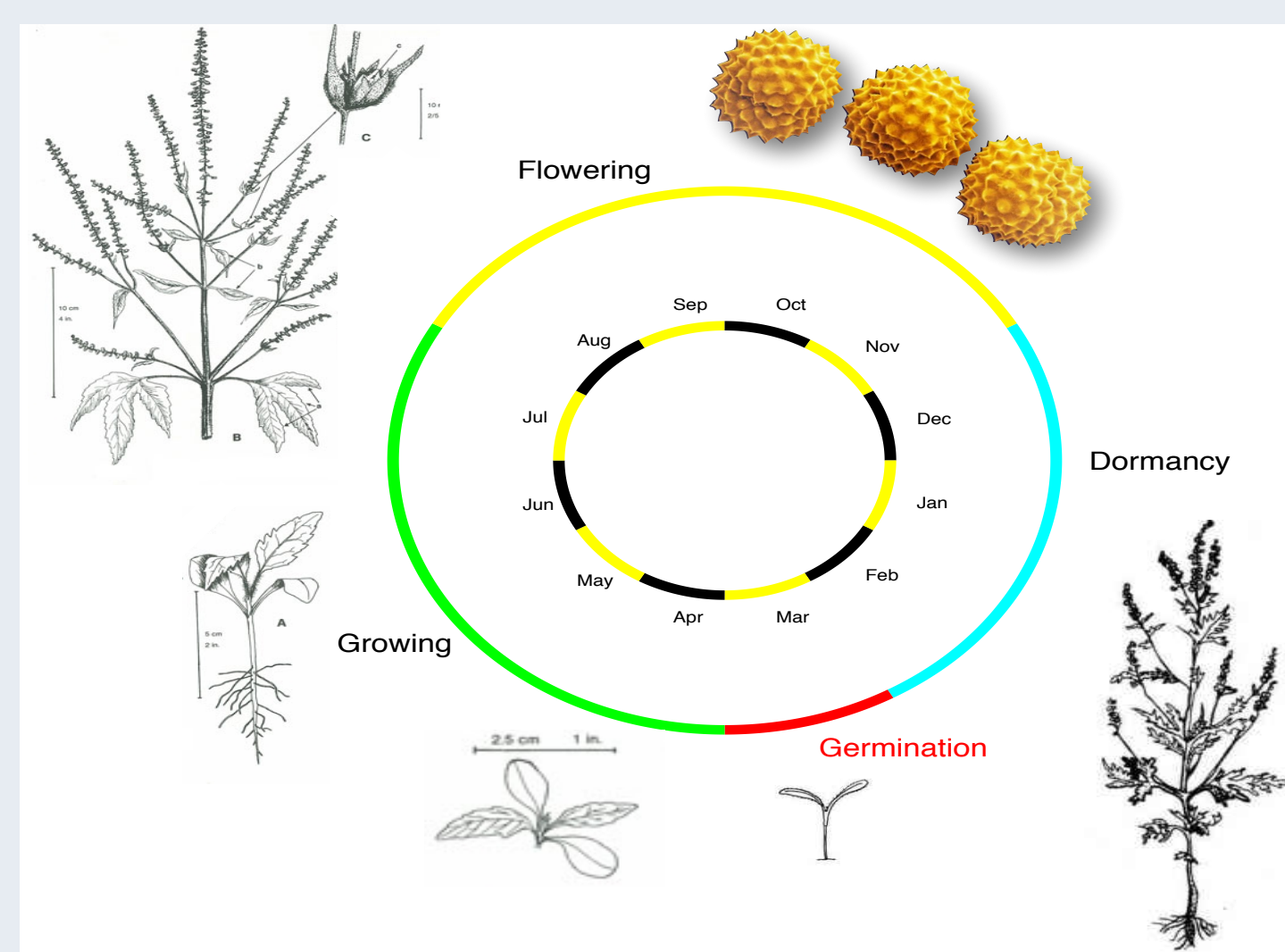


Figure 1: Life Cycle of Ambrosia.

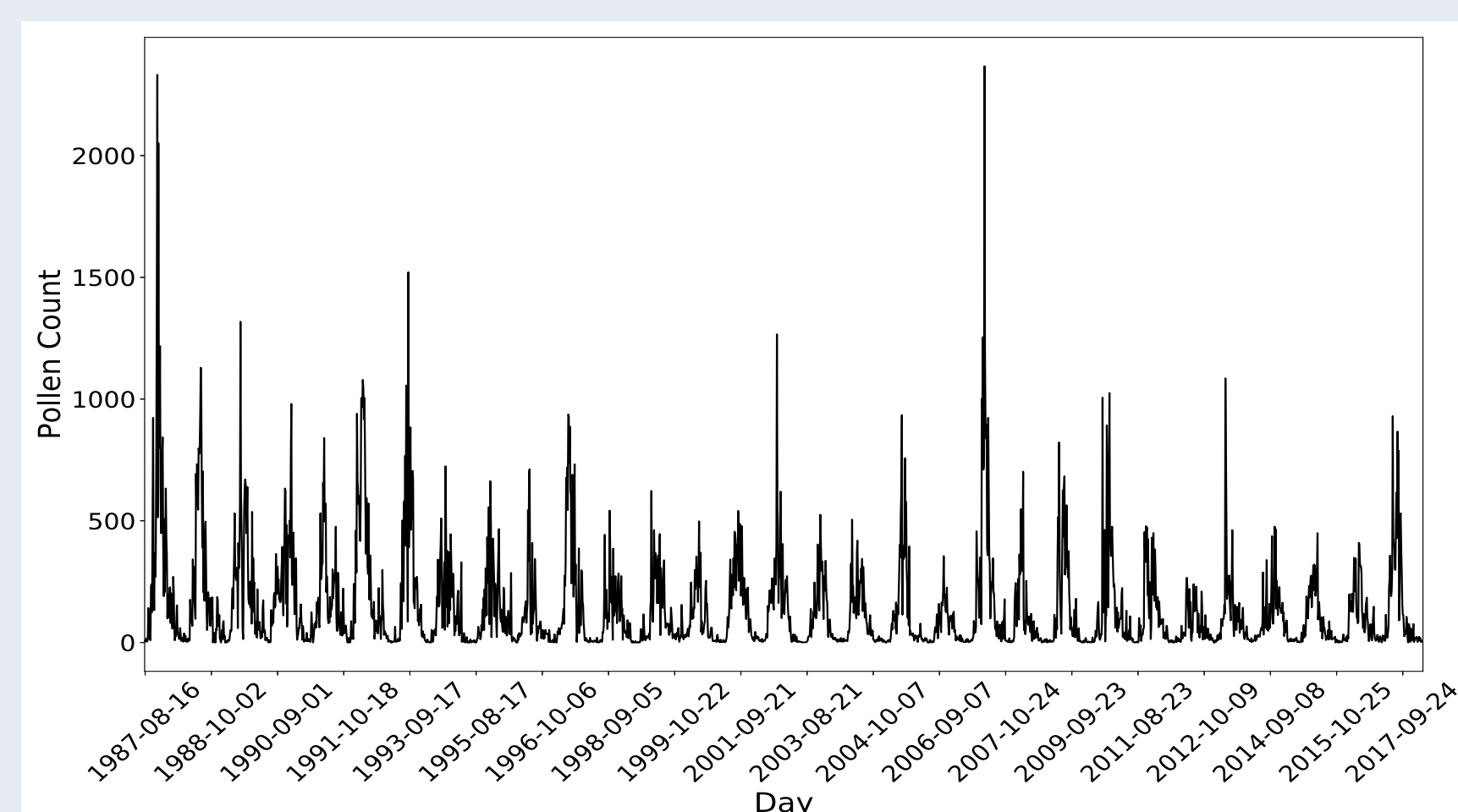


Figure 2: Pollen measurements made from 1987 to 2017.

Introduction: predictor variables

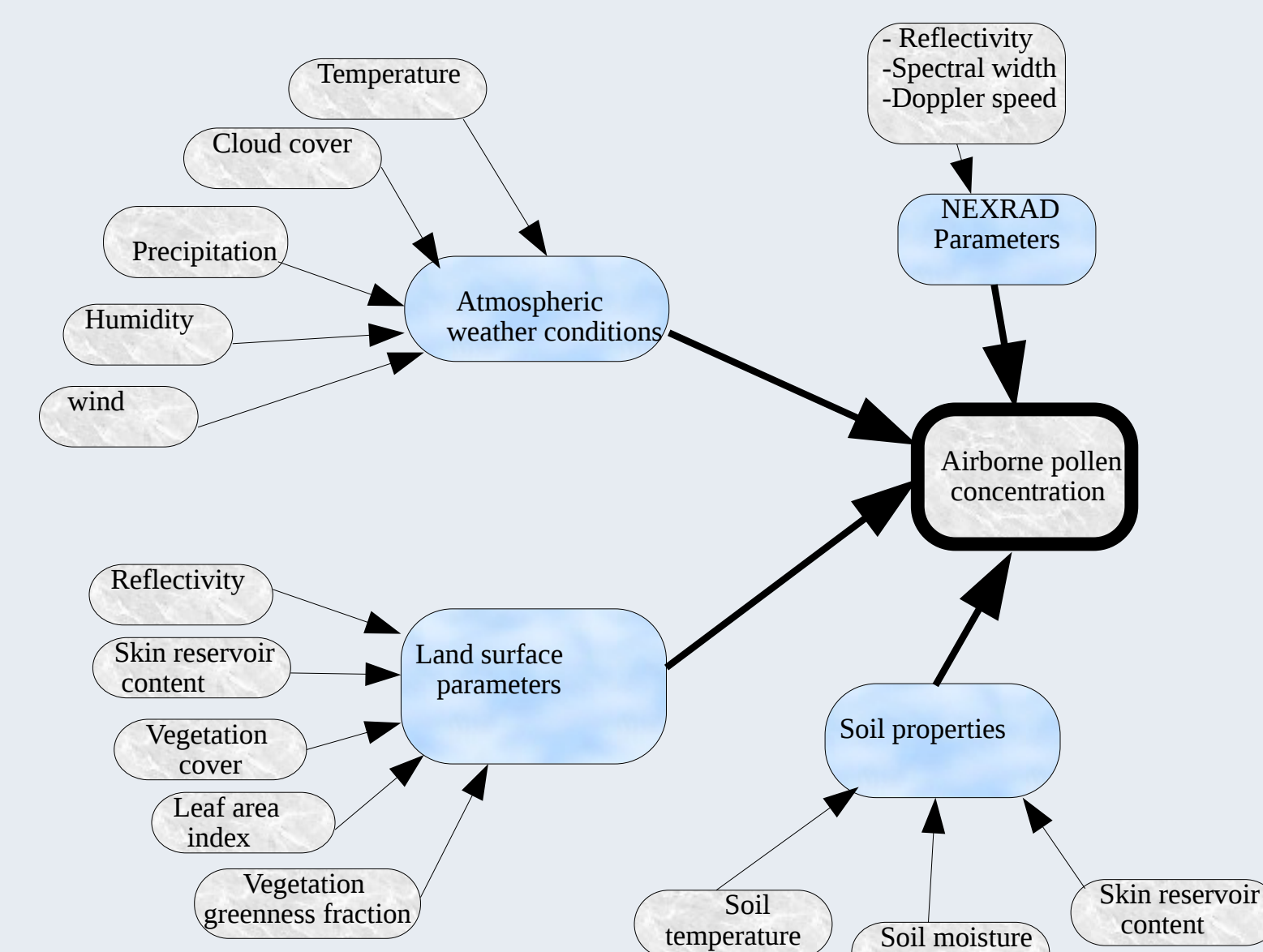
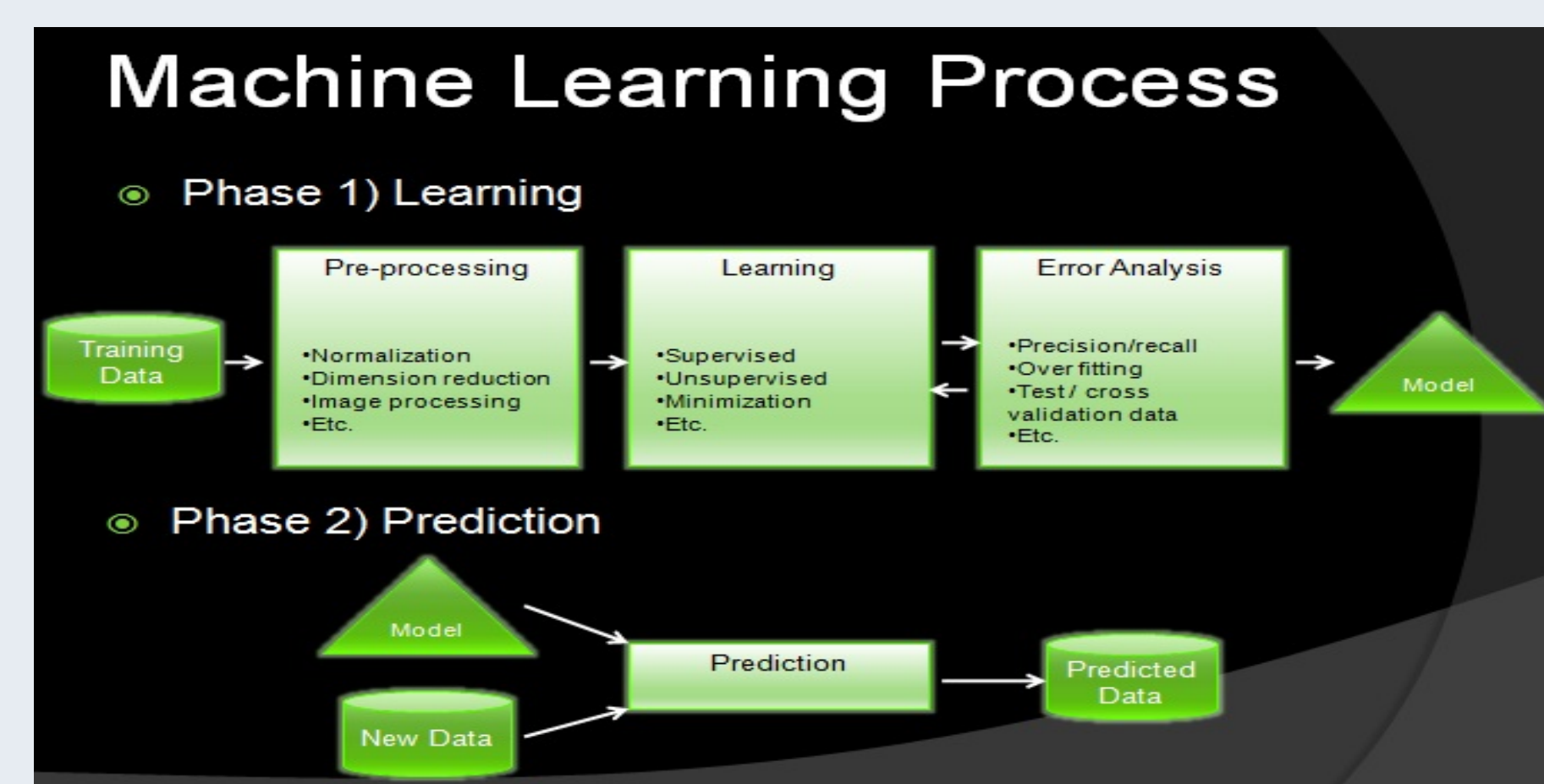


Figure 3: Atmospheric and land surface parameters controlling the distribution pollen

- 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.

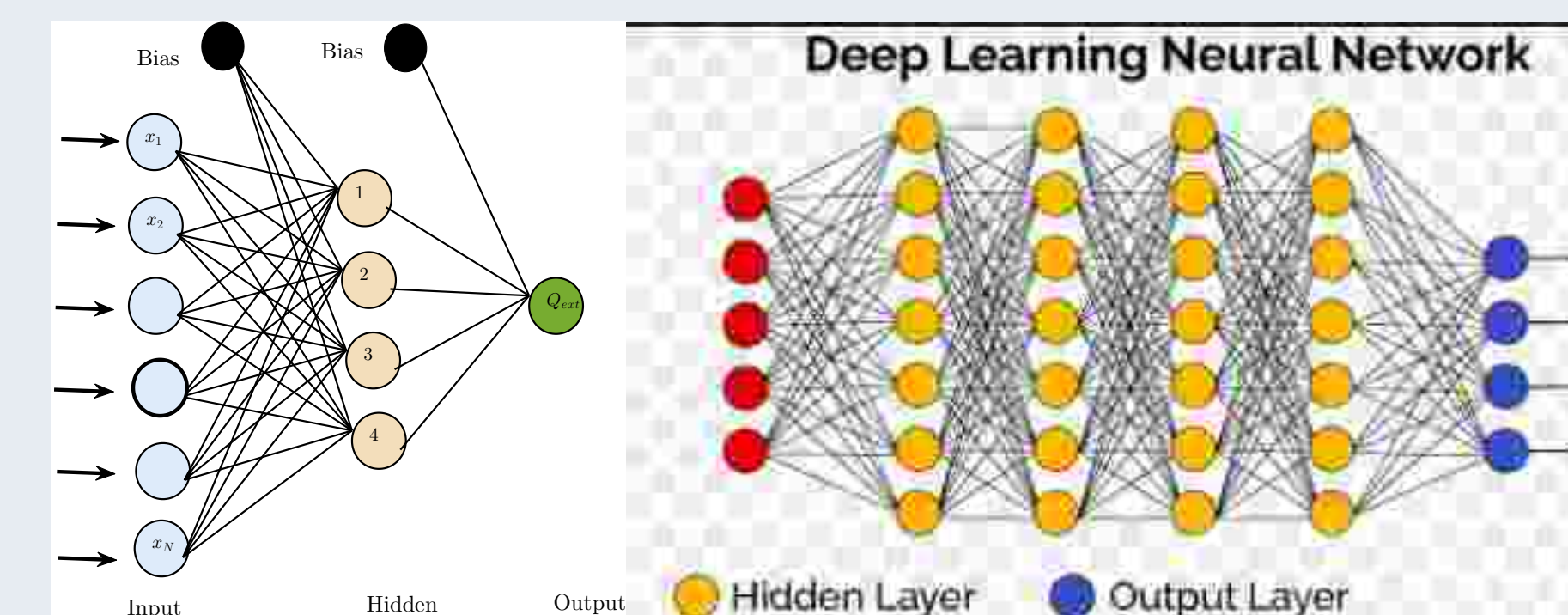


Figure 4: The neural network architecture.

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

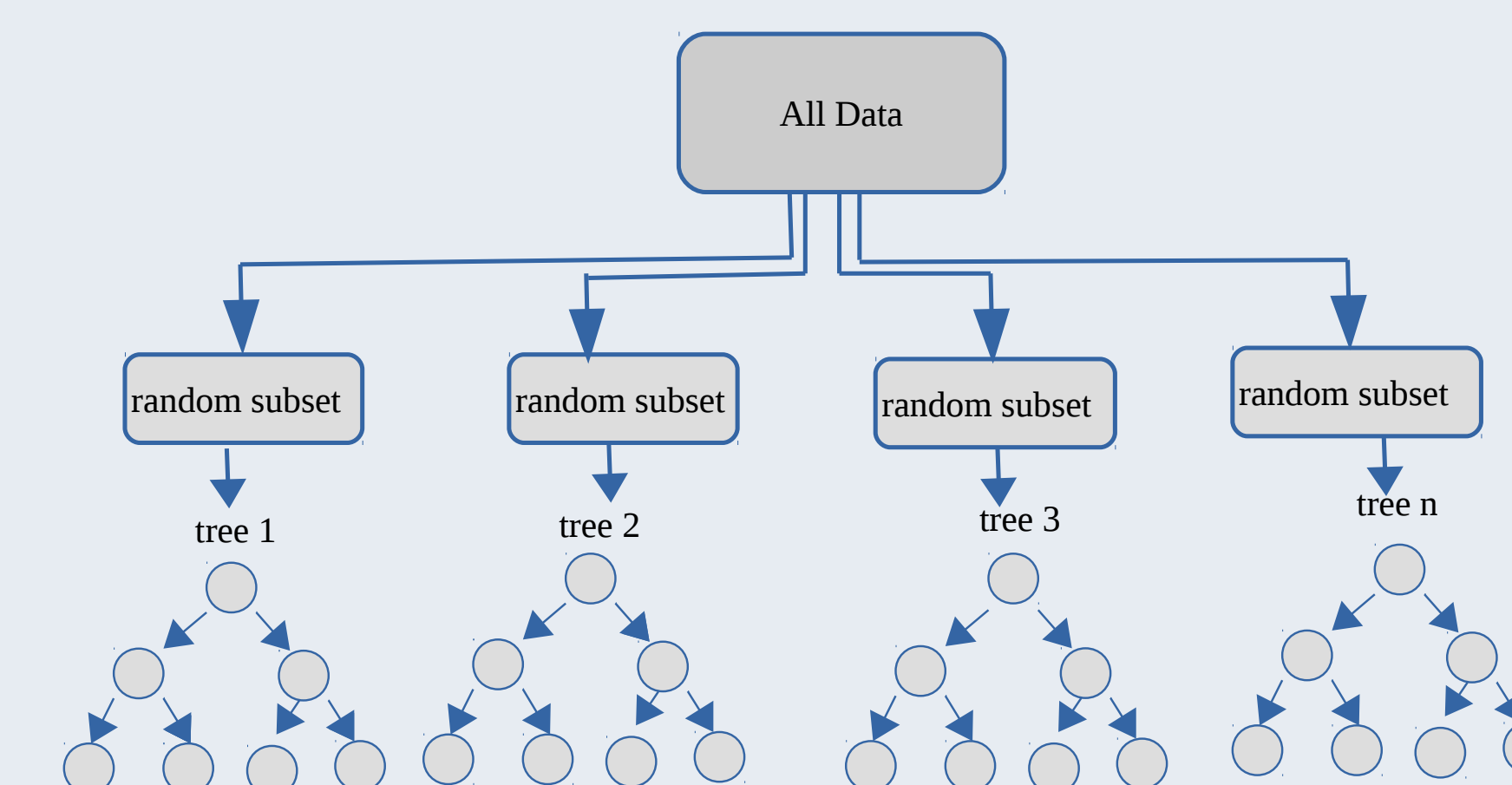


Figure 5: The Random Forest random sampling method

Results: Variable Importance estimated

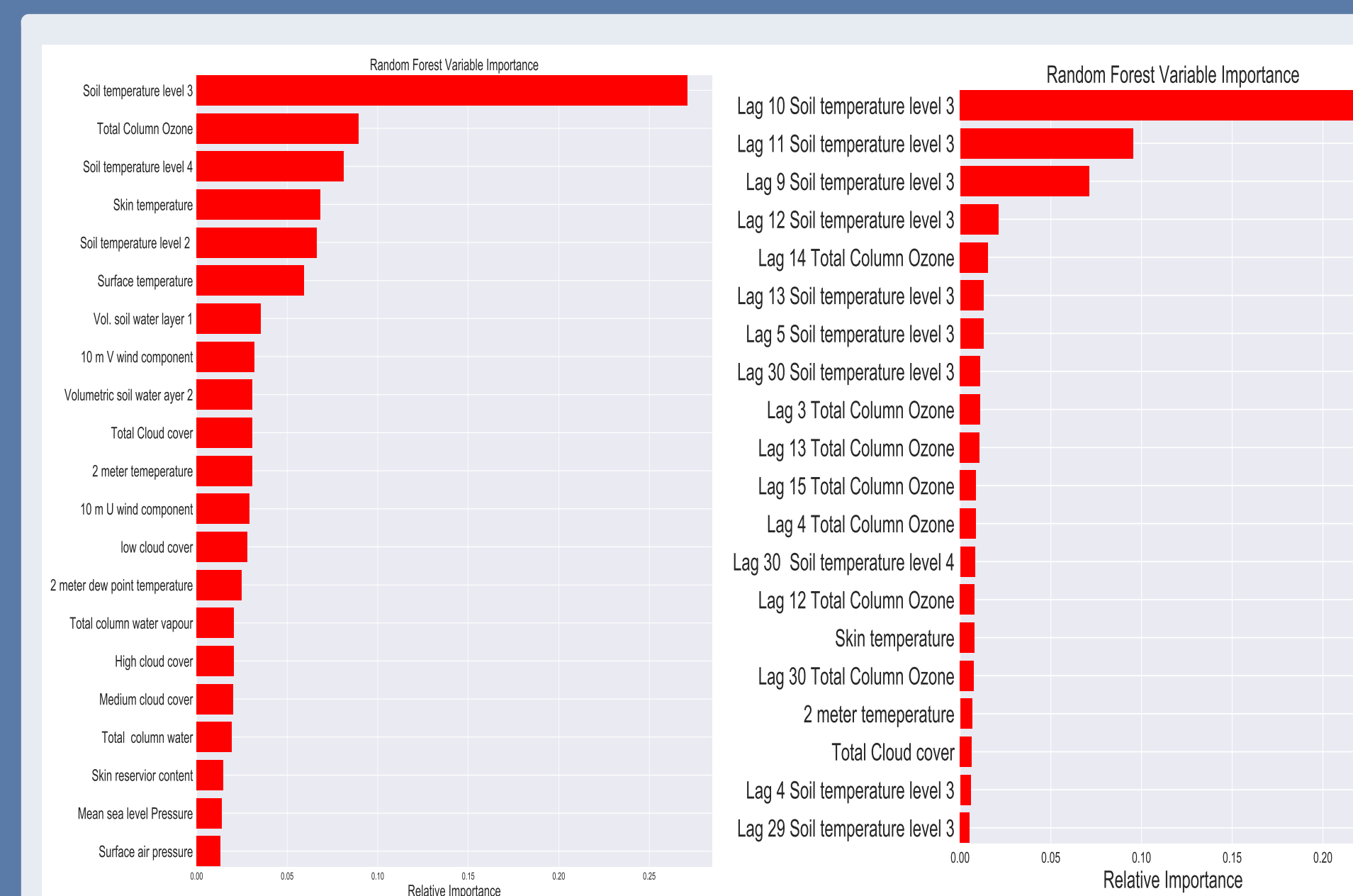


Figure 6: Variable importance estimated using the random forest machine learning method.

Results: forecasted pollen

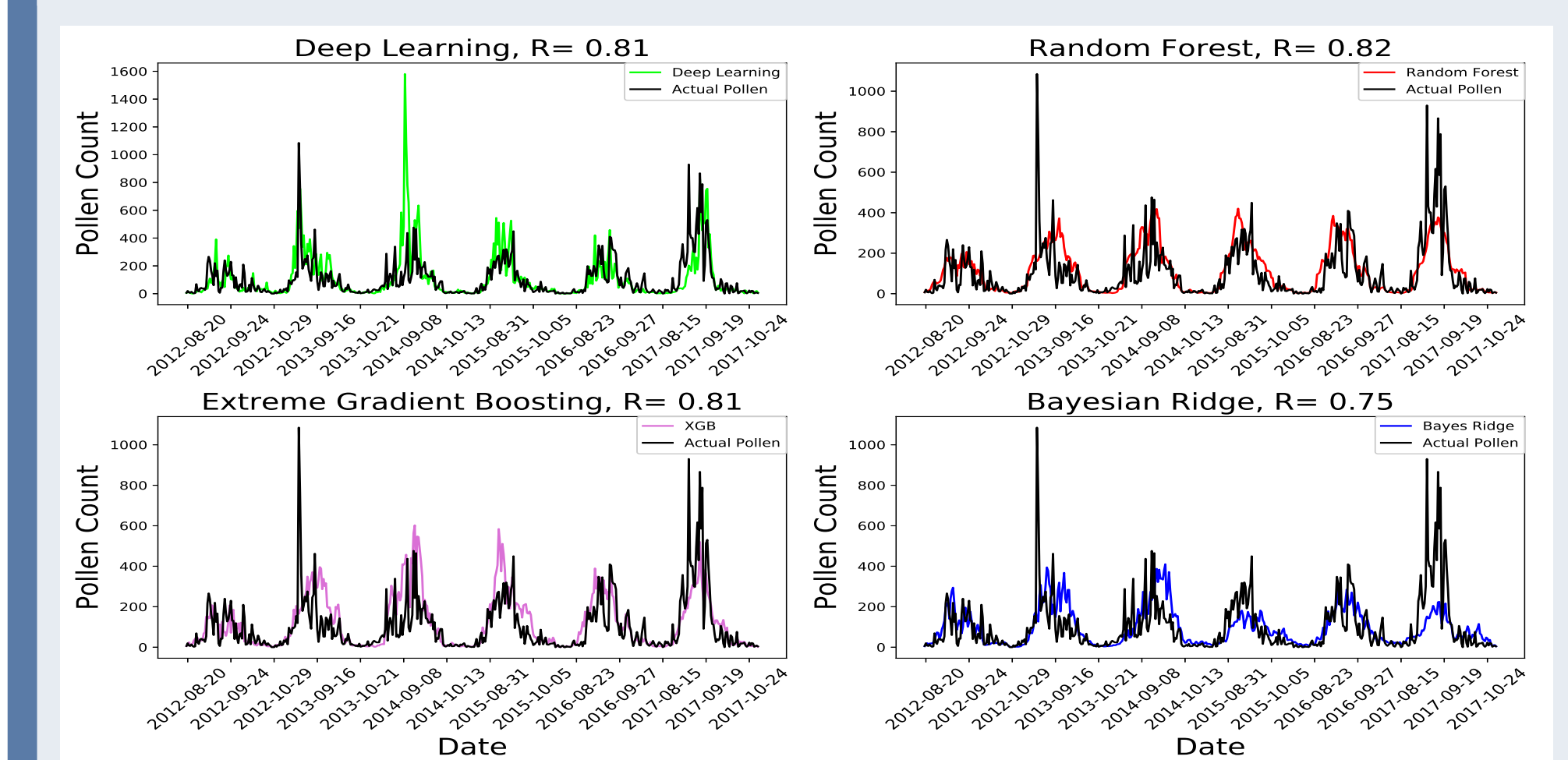


Figure 7: Forecasted pollen using independent test data for the 2012-2017 pollen season.

Results: estimated error

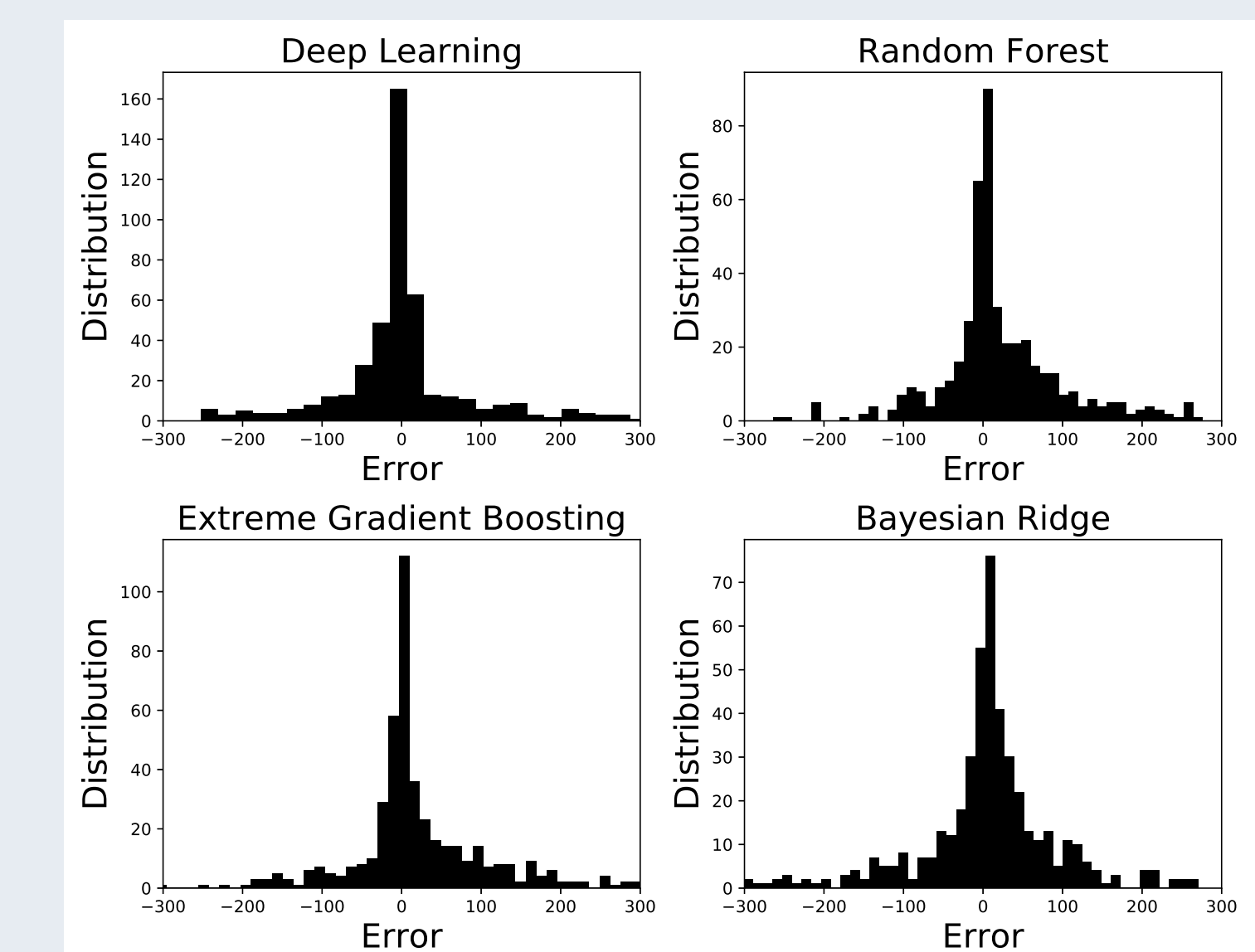


Figure 8: Error estimated for the four machine learning methods shown in Figure 7.

Pollen estimated using NEXRAD

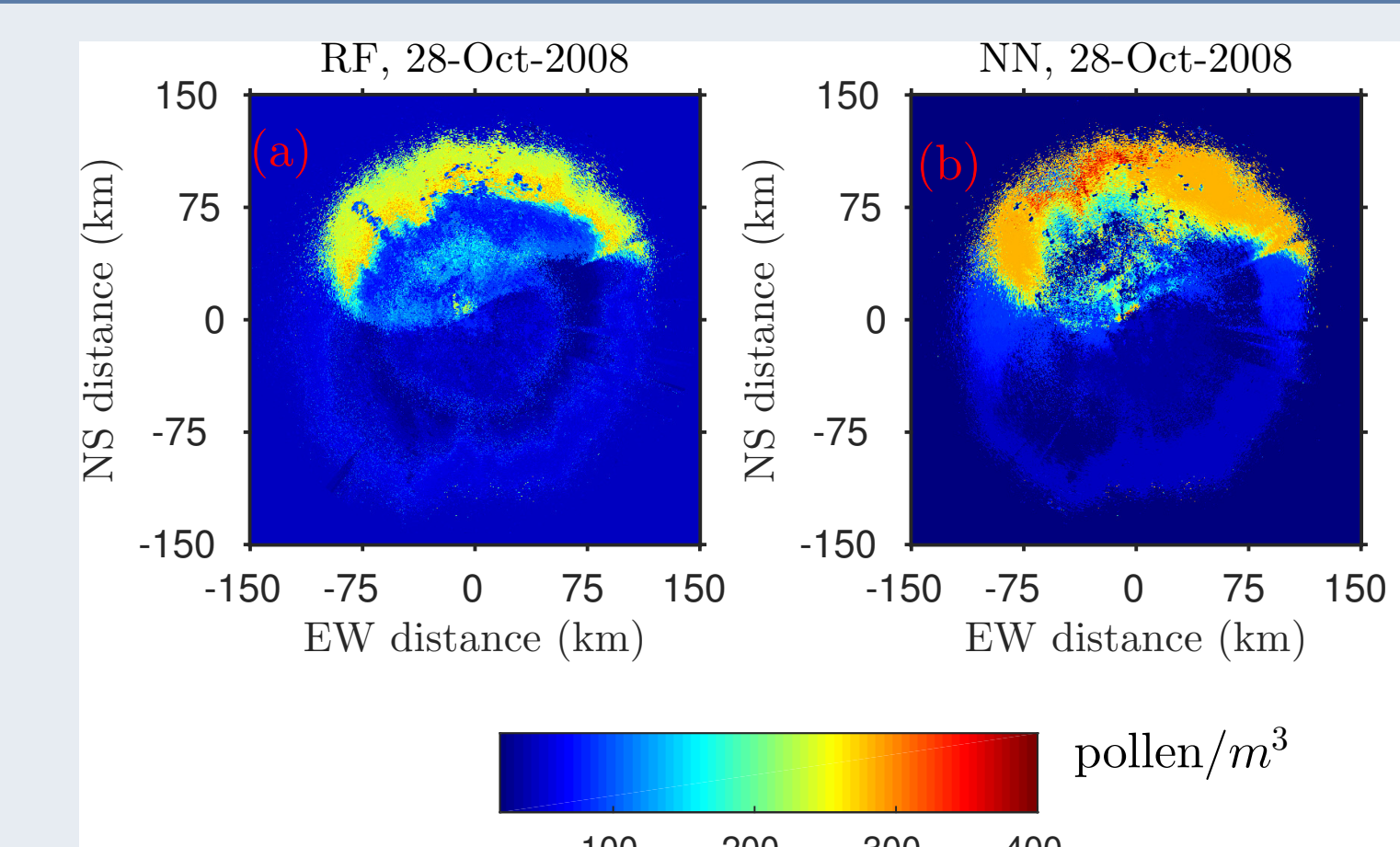


Figure 9: Pollen estimated over a large spatial area using NEXRAD radar data.