# Machine learning methods for atmospheric molecular level applications

Vitus Besel[1*], Milica Todorović[2], Theo Kurtén[3], Patrick Rinke[4] and Hanna Vehkamäki[1]
[1] Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, 00560 Helsinki, Finland
[2] Department of Mechanical and Materials Engineering, University of Turku, FI-20014 Turku, Finland
[3] Institute for Atmospheric and Earth System Research/Chemistry, University of Helsinki, 00560 Helsinki, Finland
[4] Department of Applied Physics, Aalto University, FI-00076 Helsinki, Finland

## 1. Motivation

Atmospheric aerosol particles have an important influence on the climate and air quality. Most atmospheric aerosol particles are formed by the clustering of gas-phase molecules in the atmosphere. Highly oxygenated molecules (HOMs) play a significant role in this aerosol particle formation process, but they are difficult to study due to the large number of HOM species in the atmosphere (> $10^7$).

## 2. Research

We investigate HOM saturation vapor pressures ($p_{Sat}$) on a big data scale. The $p_{Sat}$ is calculated with quantum chemistry in the context of a large scale calculation workflow (Merlin). We calculate subsets of available HOMs and train a Gaussian Process Regression machine learning model within an Active Learning loop to predict $p_{Sat}$.

## 3. Research goals

- Create HOM database
- Reliably predict $p_{Sat}$ of previously unseen HOM
- Evaluate Active Learning Loop picking strategies