

## MACHINE LEARNING CLASSIFIER FOR DETERMINING AEROSOL PARTICLE FORMATION DAYS

M. A. Zaidan<sup>1,\*</sup>, V. Haapasilta<sup>1</sup>, R. Relan<sup>2</sup>, H. Junninen<sup>3</sup>, P. P. Aalto<sup>3</sup>, F. F. Canova<sup>1</sup>,  
L. Laurson<sup>1</sup> and A. S. Foster<sup>1,4</sup>

<sup>1</sup>Aalto Science Institute and COMP Centre of Excellence, Department of Applied Physics, Aalto University, P.O. Box 11100, 00076 Aalto, Espoo, Finland

<sup>2</sup>Department ELEC, Vrije Universiteit Brussel, Brussels B-1050, Belgium

<sup>3</sup>Department of Physics, P.O. Box 64, 00014, University of Helsinki, Helsinki, Finland

<sup>4</sup>Division of Electrical Engineering and Computer Science, Kanazawa University, Kanazawa 920-1192, Japan

\*email: martha.a.zaidan@aalto.fi

Atmospheric aerosol particles are minute molecular agglomerations suspended in the air. These small particles take part in a multitude of physical processes, some of which have very consequential societal impact. Atmospheric scientists are thus interested in understanding how various processes modify the properties of the aerosol particles and especially how, why and when these particles form. However, due to the vast physical and chemical complexity of the atmosphere, the paths leading to atmospheric aerosol particle formation are not yet completely understood. One straightforward approach to overcome this issue is to gather more data.

Towards this end, several Stations for Measuring the forest Ecosystem-Atmosphere Relationships (SMEAR) have been established in Finland. These stations measure numerous variables, such as gases, meteorology, radiation, aerosol particles and soil fluxes at different altitudes. At SMEAR II station in Hyytiälä forest alone [1], there are more than 200 observables in total since 1996 (i.e. millions in sample size), producing big data. Manual labor with such big data is challenging and thus robust automated procedures for data analysis are called for [2].

Here, we present a Machine Learning based neural network classifier to automatically determine atmospheric particle formation days - one of the most interesting observables. The method has a classification accuracy of 83% for determining event/non-event days using only aerosol particle concentration data from Differential Mobility Particle Sizer instrument. The proposed method seems promising and suggests that Machine Learning based methods might eventually solve this problem to higher accuracy. The presented method is to be developed further by applying probabilistic models, such as Bayesian-based classifier, and by incorporating more measured variables into the models.

[1] P. Hari and M. Kulmala, *Boreal Env. Res.* 10 (2005) 315-322

[2] M. Kulmala, T. Petäjä, T. Nieminen, M. Sipilä, H. E. Manninen, K. Lehtipalo, M. Dal Maso, P. P. Aalto, H. Junninen, P. Paasonen, I. Riipinen, K. E. J. Lehtinen, A. Laaksonen and V.-M. Kerminen, *Nature Protocols* 7 (2012) 1651-1667.