

EGU2020-8103

<https://doi.org/10.5194/egusphere-egu2020-8103>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## Using machine learning method to classify polar stratospheric cloud types from Envisat MIPAS observations

**Rocco Sedona**<sup>1,4</sup>, Lars Hoffmann<sup>1</sup>, Reinhold Spang<sup>2</sup>, Gabriele Cavallaro<sup>1</sup>, Sabine Griessbach<sup>1</sup>, Michael Höpfner<sup>3</sup>, Matthias Book<sup>4</sup>, and Morris Riedel<sup>4</sup>

<sup>1</sup>Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany (r.sedona@fz-juelich.de)

<sup>2</sup>Institut für Energie- und Klimaforschung (IEK-7), Forschungszentrum Jülich, Jülich, Germany

<sup>3</sup>Institut für Meteorologie und Klimaforschung, Karlsruher Institut für Technologie, Karlsruhe, Germany

<sup>4</sup>University of Iceland, Reykjavik, Iceland

Polar stratospheric clouds (PSC) play a key role in polar ozone depletion in the stratosphere. Improved observations and continuous monitoring of PSCs can help to validate and enhance chemistry-climate models that are used to predict the evolution of the polar ozone hole. Here we present the results of our study in which we explored the potential of applying machine learning (ML) methods to classify PSC observations of infrared limb sounders. Two datasets have been considered. The first dataset is a collection of infrared spectra captured in Northern Hemisphere winter 2006/2007 and Southern Hemisphere winter 2009 by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument onboard ESA's Envisat satellite. The second dataset is the cloud scenario database (CSDB) of simulated MIPAS spectra. We first performed an initial analysis to assess the basic characteristics of these datasets and to decide which features to extract from them. More than 10,000 Brightness temperature differences (BTDs) features have been generated and fed as input to the ML methods instead of directly using the infrared spectra. Next, we assessed the use of ML methods for the reduction of the dimensionality of this large feature space using principal component analysis (PCA) and kernel principal component analysis (KPCA) as well as the classification with the random forest (RF) and support vector machine (SVM) techniques. All methods were found to be suitable to retrieve information on the composition of PSCs. Of these, RF seems to be the most promising method, being less prone to overfitting and producing results that agree well with established results based on conventional classification methods.