EGU2020-10472
https://doi.org/10.5194/egusphere-egu2020-10472
EGU General Assembly 2020

Using Machine Learning for processing Big Data of Copernicus Satellite Sensors at the Example of the TROPOMI / Sentinel-5 Precursor and Sentinel-4 Cloud Product

Fabian Romahn[1], Athina Argyrouli[1,2], Ronny Lutz[1], Diego Loyola[1], and Victor Molina Garcia[1]
[1]German Aerospace Center (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, Germany
[2]Technical University of Munich, Department of Civil, Geo and Environmental Engineering Chair of Remote Sensing Technology, Munich, Germany

The satellites of the Copernicus program show the increasing relevance of properly handling the huge amount of Earth observation data, nowadays common in remote sensing. This is further challenging if the processed data has to be provided in near real time (NRT), like the cloud product from TROPOMI / Sentinel-5 Precursor (S5P) or the upcoming Sentinel-4 (S4) mission.

In order to solve the inverse problems that arise in the retrieval of cloud products, as well as in similar remote sensing problems, usually complex radiative transfer models (RTMs) are used. These are very accurate, however also computationally very expensive and therefore often not feasible in combination with NRT requirements. With the recent significant breakthroughs in machine learning, easier application through better software and more powerful hardware, the methods of this field have become very interesting as a way to improve the classical remote sensing algorithms.

In this presentation we show how artificial neural networks (ANNs) can be used to replace the original RTM in the ROCINN (Retrieval Of Cloud Information using Neural Networks) algorithm with sufficient accuracy while increasing the computational performance at the same time by several orders of magnitude.

We developed a general procedure which consists of smart sampling, generation and scaling of the training data, as well as training, validation and finally deployment of the ANN into the operational processor. In order to minimize manual work, the procedure is highly automated and uses latest technologies such as TensorFlow. It is applicable for any kind of RTMs and thus can be used for many retrieval algorithms like it is already done for ROCINN in S5P and will be soon for ROCINN in the context of S4. Regarding the final performance of the generated ANN, there are several critical parameters which have a high impact (e.g. the structure of the ANN). These will be evaluated in detail. Furthermore, we also show general limitations of ANNs in comparison with RTMs, how this can lead to unexpected side effects and ways to cope with these issues.

With the example of ROCINN, as part of the operational S5P and upcoming S4 cloud product, we show the great potential of machine learning techniques in improving the performance of classical

retrieval algorithms and thus increasing their capability to deal with much larger data quantities. However, we also highlight the importance of a proper configuration and possible limitations.