

EGU2020-21593

<https://doi.org/10.5194/egusphere-egu2020-21593>

EGU General Assembly 2020

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



A Machine Learning Approach to Cloud Masking in Sentinel-3 SLSTR Data

Samuel Jackson¹, Jeyarajan Thiyagalingam^{1,2}, and Caroline Cox¹

¹Rutherford Appleton Laboratory, Science and Technology Facilities Council, Harwell Campus, Didcot, OX11 0QX

²Oxford e-Research Centre, Engineering Science, University of Oxford, 7 Keble Rd, Oxford, UK, OX1 3QG

Clouds appear ubiquitously in the Earth's atmosphere, and thus present a persistent problem for the accurate retrieval of remotely sensed information. The task of identifying which pixels are cloud, and which are not, is what we refer as the cloud masking problem. The task of cloud masking essentially boils down to assigning a binary label, representing either "cloud" or "clear", to each pixel.

Although this problem appears trivial, it is often complicated by a diverse number of issues that affect the imagery obtained from remote sensing instruments. For instance, snow, sea ice, dust, smoke, and sun glint can easily challenge the robustness and consistency of any cloud masking algorithm. The cloud masking problem is also further complicated by geographic and seasonal variation in acquired scenes.

In this work, we present a machine learning approach to handle the problem of cloud masking for the Sea and Land Surface Temperature Radiometer (SLSTR) on board the Sentinel-3 satellites. Our model uses Gradient Boosting Decision Trees (GBDTs), to perform pixel-wise segmentation of satellite images. The model is trained using a hand labelled dataset of ~12,000 individual pixels covering both the spatial and temporal domains of the SLSTR instrument and utilises the combined channels of the dual-view swaths. Pixel level annotations, while lacking spatial context, have the advantage of being cheaper to obtain compared to fully labelled images, a major problem in applying machine learning to remote sensing imagery.

We validate the performance of our mask using cross validation and compare its performance with two baseline models provided in the SLSTR level 1 product. We show up to 10% improvement in binary classification accuracy compared with the baseline methods. Additionally, we show that our model has the ability to distinguish between different classes of cloud to reasonable accuracy.