



Contribution ID: 62

Type: **Oral presentation**

## ESoWC: A Machine Learning Pipeline for Climate Science

*Wednesday, 16 October 2019 16:00 (20 minutes)*

**Thomas Lees** [1], Gabriel Tseng [2], Simon Dadson [1], Steven Reece [1]

[1] University of Oxford

[2] Okra Solar, Phnom Penh

As part of the ECMWF Summer of Weather Code we have developed a machine learning pipeline for working with climate and hydrological data. We had three goals for the project. Firstly, to bring machine learning capabilities to scientists working in hydrology, weather and climate. This meant we wanted to produce an extensible and reproducible workflow that can be adapted for different input and output datasets. Secondly, we wanted the pipeline to use open source datasets in order to allow for full reproducibility. The ECMWF and Copernicus Climate Data Store provided access to all of the ERA5 data. We augmented this with satellite derived variables from other providers. Finally, we had a strong focus on good software engineering practices, including code reviews, unit testing and continuous integration. This allowed us to iterate quickly and to develop a code-base which can be extended and tested to ensure that core functionality will still be provided. Here we present our pipeline, outline some of our key learnings and show some exemplary results.

**Primary authors:** Mr LEES, Thomas (University of Oxford); TSENG (WILL PRESENT REMOTELY), Gabriel (Okra Solar); Prof. DADSON, Simon (University of Oxford); Dr REECE, Steven (University of Oxford)

**Presenters:** Mr LEES, Thomas (University of Oxford); TSENG (WILL PRESENT REMOTELY), Gabriel (Okra Solar)

**Track Classification:** Workshop: Building reproducible workflows for earth sciences