
opernicus
Europe's eyes on Earth
Predicting and
Understanding
Drought
ECMWF
#ESoWC2019

Project Overview

Robust, useful pipeline

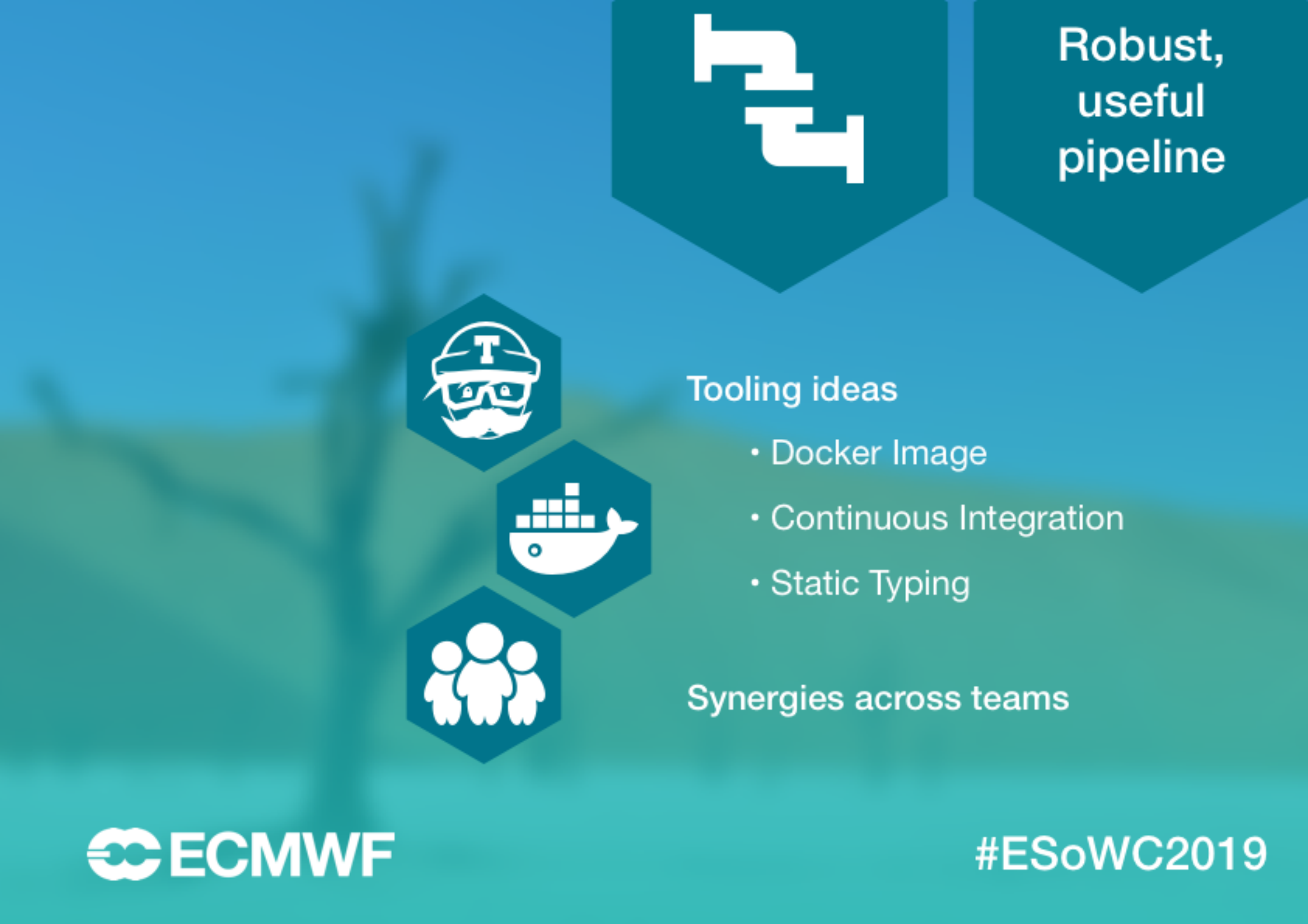
Predictive System for Agricultural Drought

Interpretable Machine Learning

Communicate Results

ECMWF

#ESoWC2019


Robust, useful pipeline
Tooling ideas
• Docker Image
• Continuous Integration
• Static Typing
Synergies across teams
ECMWF
#ESoWC2019

Robust, useful pipeline

Export / download

Inputs: Parameters in the pipeline defining what needs to be downloaded.
Outputs: Raw data, saved in an easily accessible location (e.g. locally)

Engineering

Inputs: Raw data exported above
Outputs: Arrays, split into test and train sets, which can be directly fed into a machine learning model

Model Training

This step may manipulate the data further (depending on the model being used), but should result in predictions for the test set and a saved, trained model.

Analysis

Take the trained model and the predictions, and explain them.
In addition, there should be tools here to explore how good the model's predictions are.

Predictive System

Agricultural Drought

Target: Vegetation Health Index (VHI) at end of Season

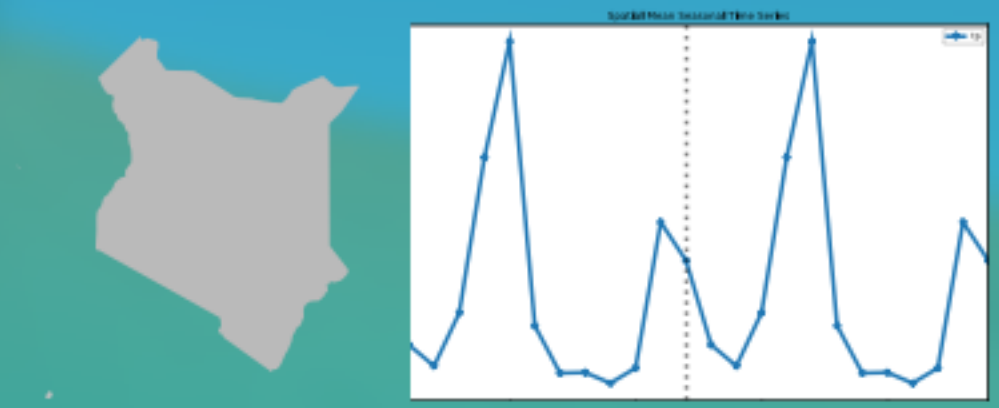
Inputs: Meteorology (Temp, Precip), Hydrology (Soil Moisture), Climate Vars (Nino3.4), Static Variables (Orography)

Challenges: Encoding spatial-temporal information, how to utilise climate variables, masking, Vegetation health observed from satellites?

ECMWF

#ESoWC2019

Predictive System



ECMWF

#ESoWC2019

Predictive System

Aim 1: Forecast

From preceding conditions can we forecast a value ahead of time? This requires us to learn the relationship between the previous conditions and the current conditions.

Goal: Predictive Accuracy

Challenge: Are we reproducing / competing with physically based models?

Predictive System

Aim 2: Identify Correlations

Can we use the high dimensional fields (SST, SLP) to identify connections between remote regions in space and time.

Goal: Identify climate drivers of variability from data

Challenge: Crazy high dimensions (atmospheric levels [z], geographic regions [x,y], lagged in time, multiple variables - 5D space)

Predictive System

Decision Trees

Gradient Boosted Trees: XGBoost
Random Forest: Scikit-learn

Neural Networks

RNNs: PyTorch
Linear Models: PyTorch
Segmentation Models: PyTorch

ECMWF

#ESoWC2019

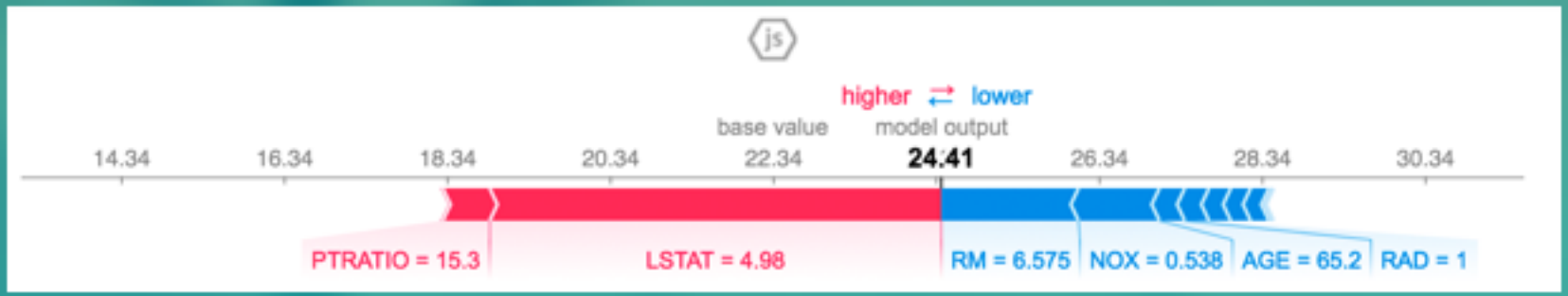
Interpret Models

SHAP Values

A method for assigning payouts to players (features) depending on their contribution to the total payout (predictive accuracy).

Package: SHAP (https://github.com/slundberg/shap)

Visualisations:



ECMWF

#ESoWC2019

Communicate Results

Flexible Pipeline

An experimental pipeline for easily exchanging models, definitions and input features.

Extensible: Python classes and funtions

Well Documented: Example notebooks, test suite

ECMWF

#ESoWC2019

Communicate Results

Blog Posts

Blog posts throughout the process outlining our thinking and our implementations.

More polished blog posts to follow towards the end of the ESoWC Project

ECMWF

#ESoWC2019

Communicate Results

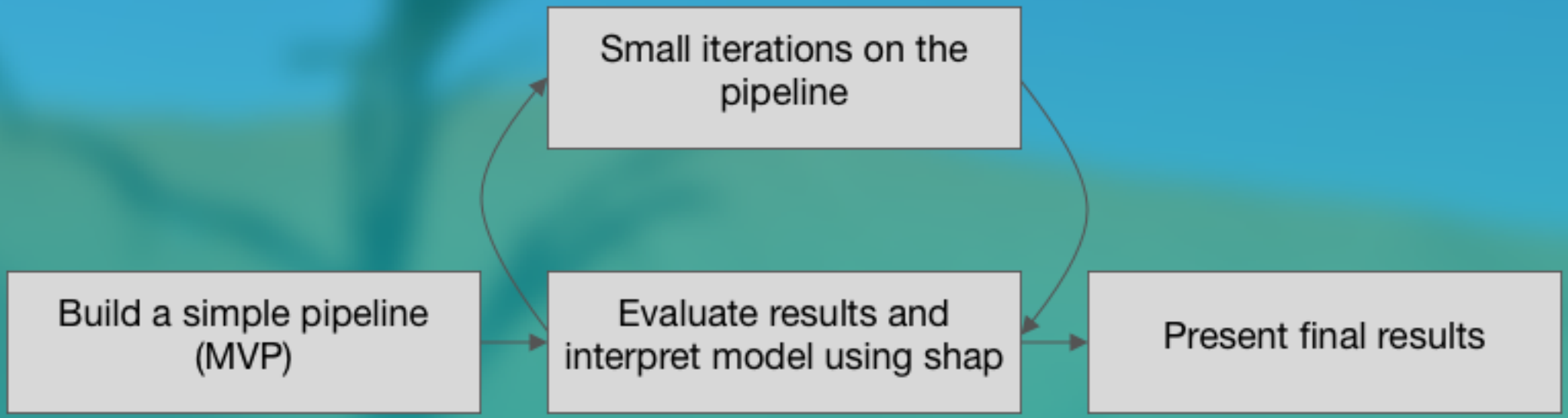
Academic Papers

- Overview of the pipeline and functionality (JOSS, GMD).
- Overview of predictive accuracy and the algorithmic approach.
- Overview of scientific insights (e.g. SSTs in Western Pacific Warm Pool)

ECMWF

#ESoWC2019

Development Plan



Experiments

- choice of thresholds
- choice of definition
- choice of variable

Drought metrics vs. Drought impacts

Identify Teleconnections in SST/SLP data

Does soil moisture offer predictability for rainfall?

Can we quantify the human element in drought risk?

Can we quantify the human element in drought risk?

Combine ML Vegtation with SEAS5 Precipitation

Questions

Communication: we have been using slack to communicate between ourselves, and have found it very effective. Would an ESoWC-ML slack be helpful? (If not,)
Would mentors like to be added to our internal slack?
In general, how should inter-team communication happen?
Questions about how open this should be; private repo for scripts?

Plan: Is our project plan aligned with what ECMWF has in mind?

Data: Access to SEAS5 forecast data
Are we limited by the same API requests? - `limit is 10,000`
For the disaster database how is 'drought' defined? Is it meteorological, agricutural, hydrological? Or is it solely focused on the impacts of an event (like EMDAT)?

Infrastructure: Using google cloud research credits for compute-heavy tasks - individual requests or grouped under ECMWF?

Predictive System

Hydrological Drought

Target: Soil Moisture Time Series

Inputs: Meteorology (Temp, Precip), Vegetation Health (NDVI, VHI), Climate Vars (Nino3.4), Static Variables (Orography, Soil Type)

Challenges: Encoding spatial-temporal information, how to utilise climate variables, masking, is the data correct?

ECMWF

#ESoWC2019

Predictive System

Meteorological Drought

Target: Standardised Precipitation Index (SPI)

Inputs: Preceding Meteorology (Temp, Precip), Hydrology (Soil Moisture), Climate Vars (Nino3.4), Static Variables (Orography, Soil Type)

Challenges: Encoding spatial-temporal information, how to utilise climate variables, Performance vs. SEAS5

ECMWF

#ESoWC2019

Predictive System

Hydrological Drought 2

Target: Streamflow

Inputs: Preceding Meteorology (Temp, Precip), Hydrology (Soil Moisture), Climate Vars (Nino3.4), Static Variables (Orography, Soil Type)

Challenges: Encoding spatial-temporal information, how to utilise climate variables, Performance vs. SEAS5