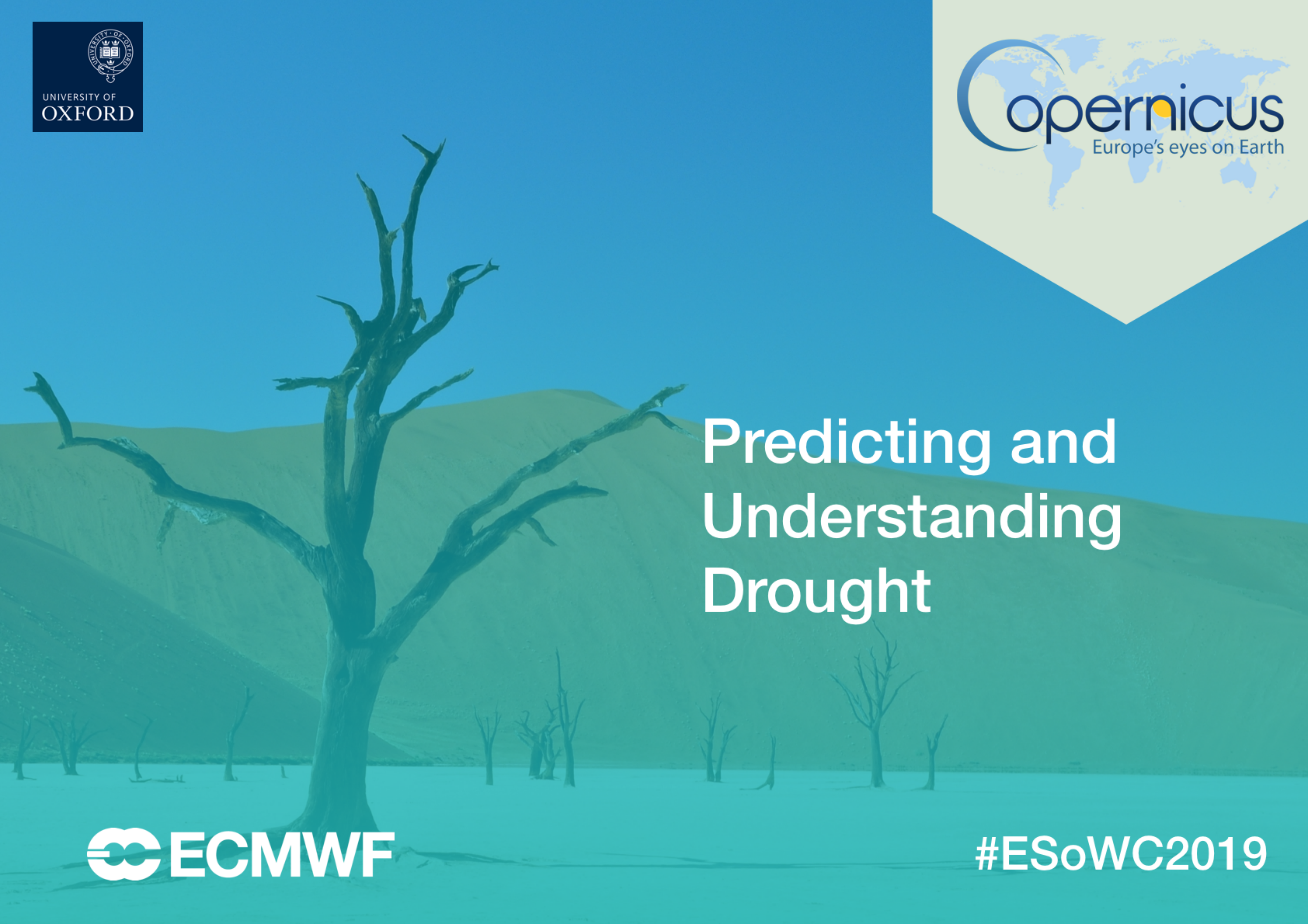


UNIVERSITY OF OXFORD

Copernicus
Europe's eyes on Earth

Predicting and Understanding Drought

ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

To use **ECMWF/Copernicus open datasets** to evaluate machine learning (ML) techniques to **better predict one specific kind of an extreme weather event**, e.g. drought or hurricanes; provide templates for future ML work

ECMWF

#ESoWC2019

Project Overview

Robust, useful pipeline

Predictive System for Agricultural Drought

Interpretable Machine Learning

Communicate Results

UNIVERSITY OF OXFORD

Robust, useful pipeline

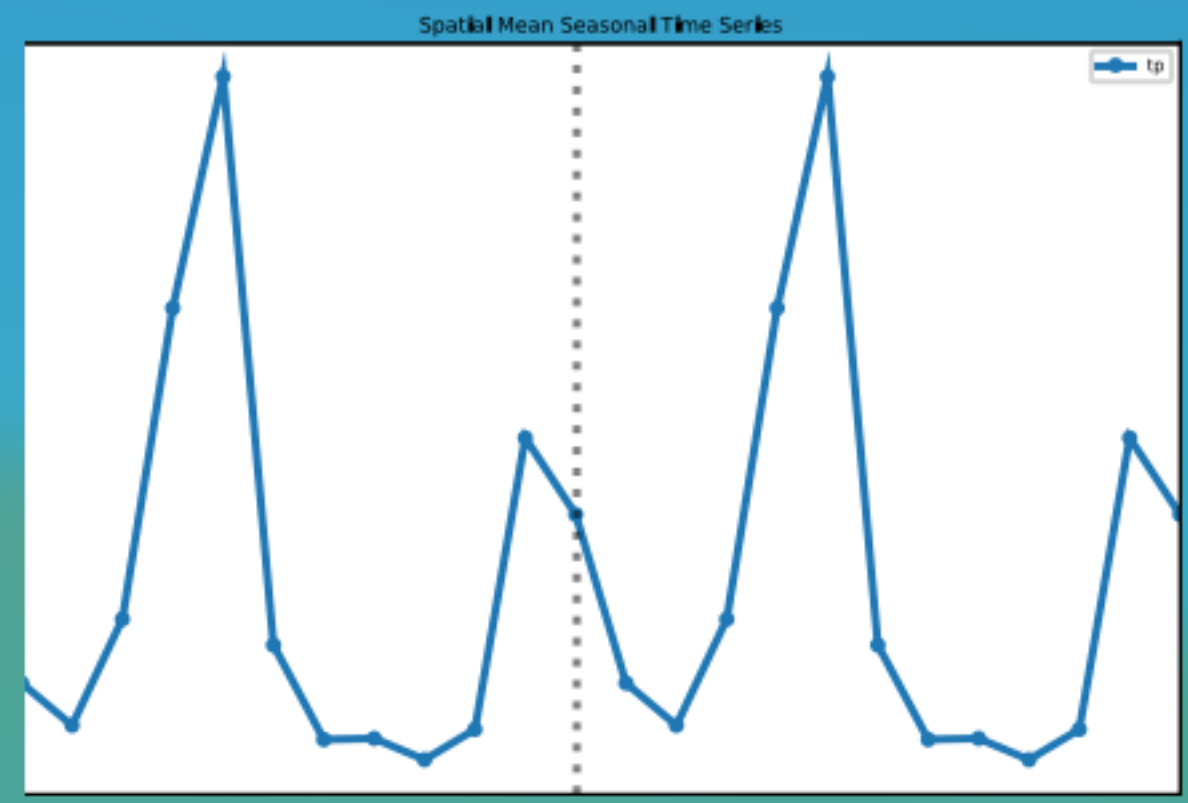
Export / download → Engineering → Model Training → Analysis

ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

Predictive System



ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

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

UNIVERSITY OF OXFORD

Predictive System

Table 7: Indicators monitored by the drought early warning system

Type of indicator	Examples of indicators monitored	Types of impact
Biophysical	Rainfall data Vegetation condition State of water sources	Environmental
Production	Livestock body condition Milk production Livestock migration Livestock mortality Crop production	Livestock production Crop production
Access	Terms of trade (meat/maize) Milk consumption Distances to water	Markets Access to food and water
Utilisation	MUAC (Mid-Upper Arm Circumference) Coping strategies	Nutrition Coping strategies

ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

Predictive System

Table 2: *Classification scheme used for computing the Combined Drought Indicator. Note that the delta symbol (Δ) is used as a prefix to indicate anomalies, and "m-1" is used as a suffix to indicate the month previous to the current one.*

LEVEL	*COLOUR*	*CLASSIFICATION CONDITION*
Watch		SPI-3 < -1 or SPI-1 < -2
Warning		SMA > 1 and (SPI-3 < -1 or SPI-1 < -2)
Alert		ΔFAPAR < -1 and (SPI-3 < -1 or SPI-1< -2)
Partial recovery		(ΔFAPAR < -1 and (SPI-3 $_{m-1}$ < -1 and SPI-3 > -1)) or (ΔFAPAR < -1 and (SPI-1 $_{m-1}$ < -2 and SPI-1 >-2))
Full recovery		(SPI-3 $_{m-1}$ < -1 and SPI-3 > -1) or (SPI-1 $_{m-1}$ < -2 and SPI-1 > -2))

ECMWF

#ESoWC2019

Predictive System

Decision Trees

Gradient Boosted Trees: XGBoost
Random Forest: Scikit-learn

Neural Networks

RNNs: PyTorch
Linear Models: PyTorch
Segmentation Models: PyTorch

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)

UNIVERSITY OF OXFORD

Predictive System

Aim 1: Forecast

Operationalise

$$f(X_{train}) =$$

$$f(X_{SEAS5}) =$$

where

precip temp soil M

$X =$ t-1, t-2, ..., t-n

ECMWF

#ESoWC2019

Predictive System

Aim 2: Identify Correlations

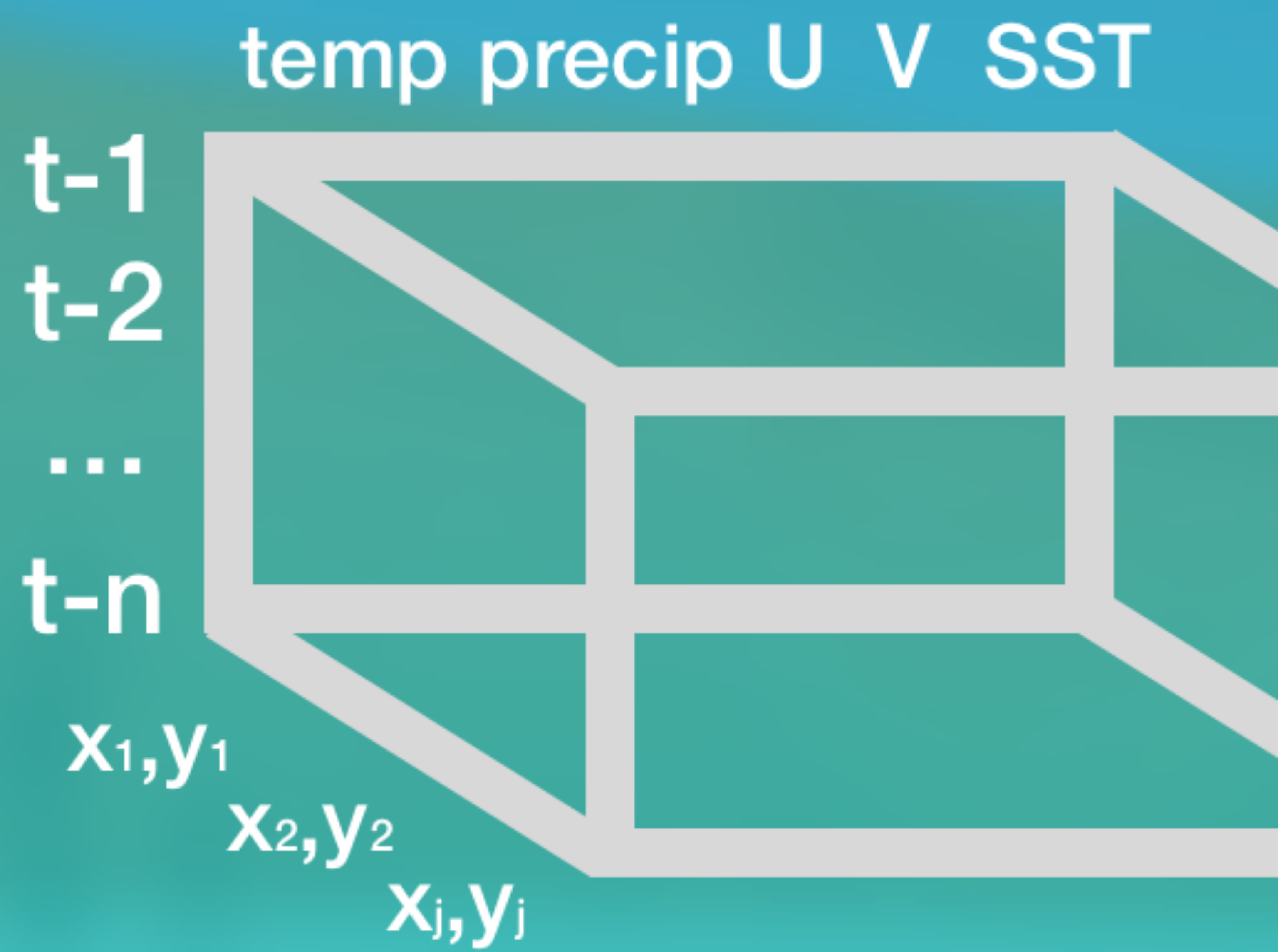
Operationalise

$$f(X_{train}) =$$

$$f(X_{now}) =$$

where

$X =$



ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

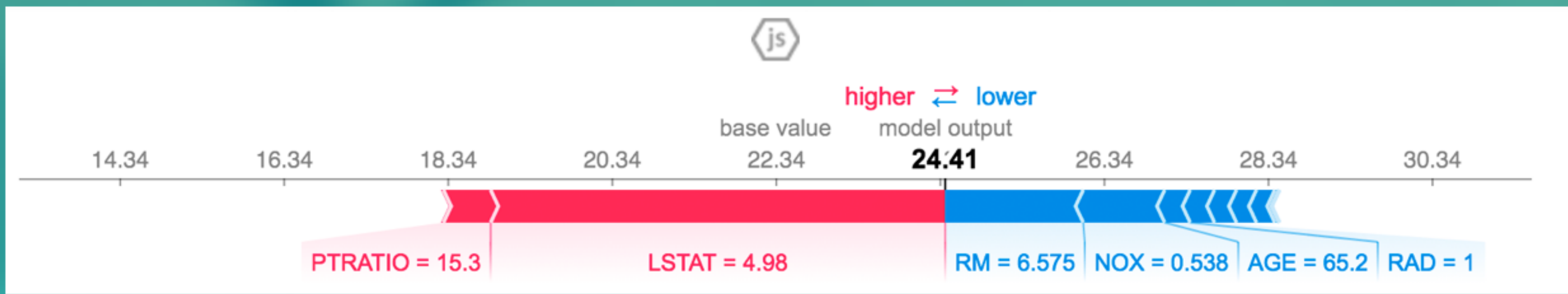
Interpret Models

SHAP Values

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

$$\text{importance of } j = f\,(\text{with } j) - f\,(\text{without } j)$$

Visualisations:



sum of all SHAP values = difference of prediction - base
(24.41 - 22.34)

ECMWF

#ESoWC2019

Interpret Models

Where do we find 'predictive skill'?

How do we compare with physically-based forecasts?

Which of our features give us the best improvement in skill?

#ESoWC2019

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?

Combine ML Vegtation with SEAS5 Precipitation

UNIVERSITY OF OXFORD

Input variables

(VERY) initial results

	lst_night_1	lst_day_1	precip_1	sm_1	lst_night_2	lst_day_2	precip_2	sm_2	lst_night_3	lst_day_3	precip_3	sm_3	target
0	-0.622899	-0.683058	-0.754043	-1.138327	-0.208489	-0.178908	-0.752656	-0.747571	0.136853	0.498644	-0.705800	-1.085976	0.5559
1	-0.622899	-0.644524	-0.751266	-1.138327	-0.161005	-0.156430	-0.753820	-0.747571	0.244772	0.456900	-0.721717	-1.085976	0.5136
2	-0.471812	-0.689481	-0.750568	-1.138327	-0.156688	-0.204597	-0.754299	-0.747571	0.326790	0.514700	-0.726781	-1.085976	0.4928
3	-0.527930	-0.503234	-0.752355	-0.981666	-0.195539	-0.124319	-0.752928	-0.979183	0.240455	0.594979	-0.711250	-1.115977	0.4576
4	-0.510663	-0.509656	-0.751587	-0.981666	-0.078986	-0.121108	-0.753940	-0.979183	0.210238	0.646357	-0.710568	-1.115977	0.4357

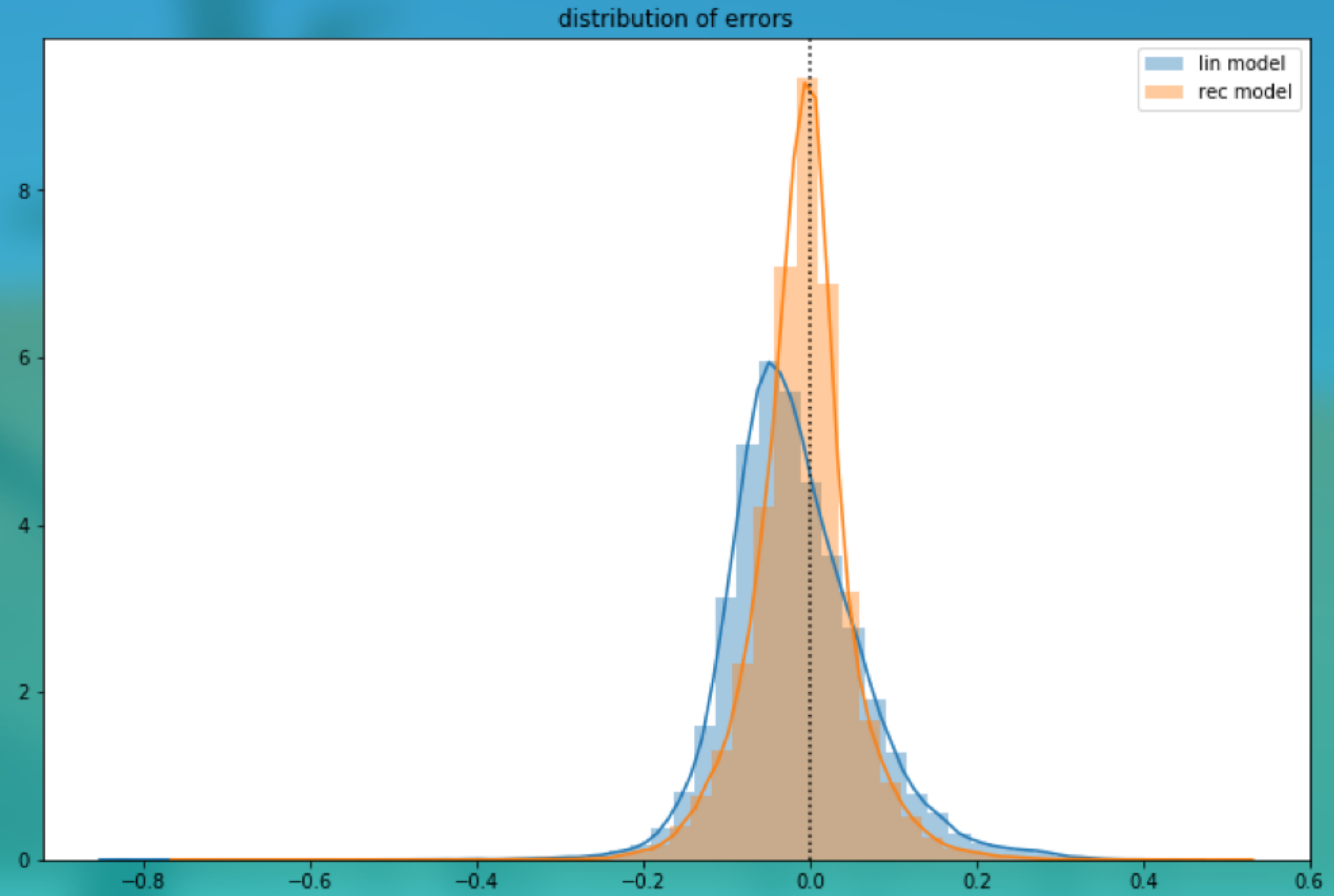
Model Comparison

(VERY) initial results

Model	RMSE	RMSE (no veg)
Linear Regression	0.040	0.084
Feedforward neural network	0.038	0.070
Recurrent neural network	0.035	0.060

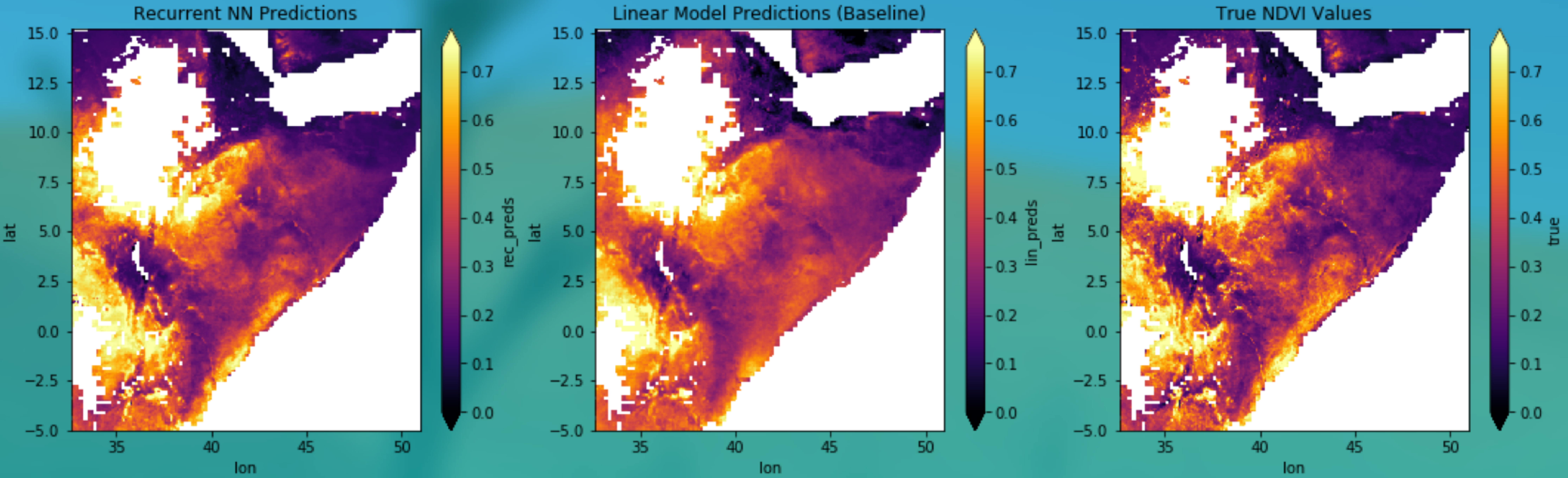
Model Comparison

(VERY) initial results



Model Comparison

(VERY) initial results

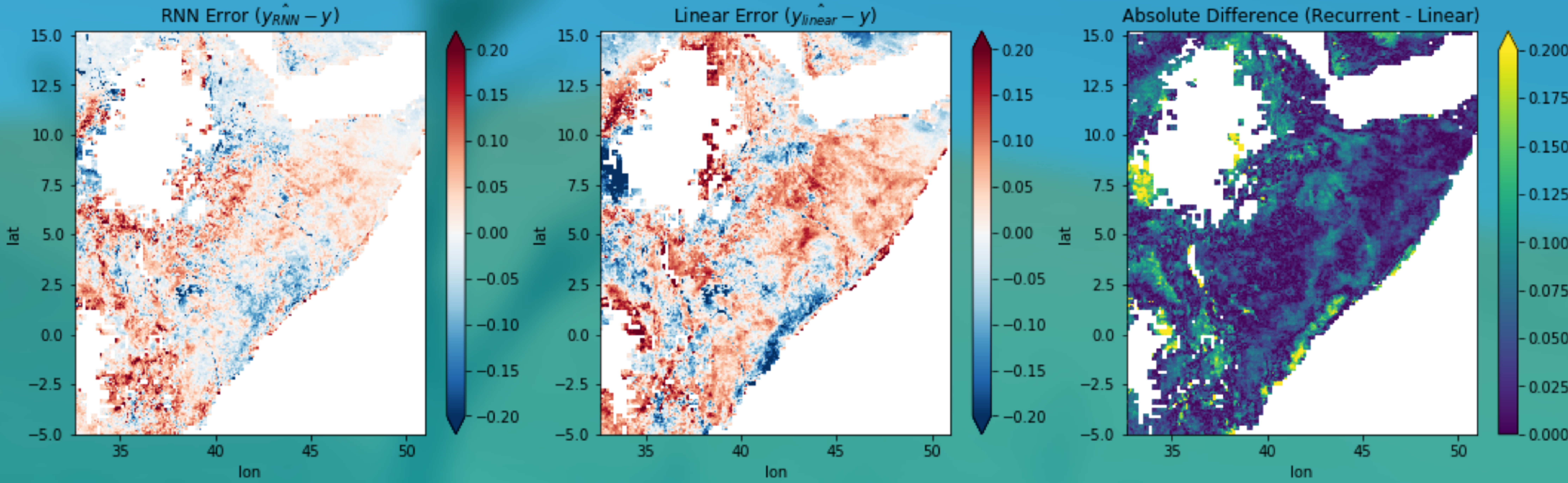


ECMWF

#ESoWC2019

Model Comparison

(VERY) initial results

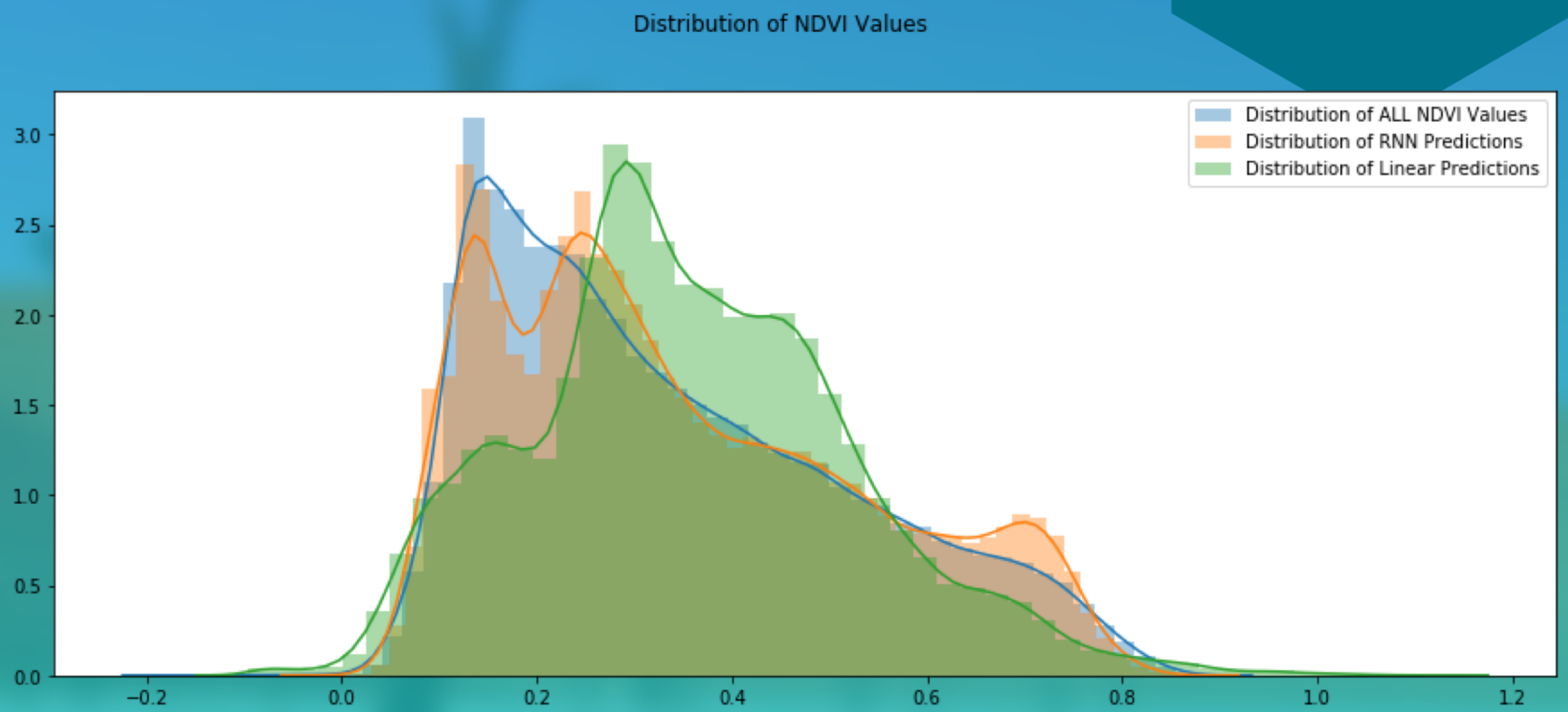


ECMWF

#ESoWC2019

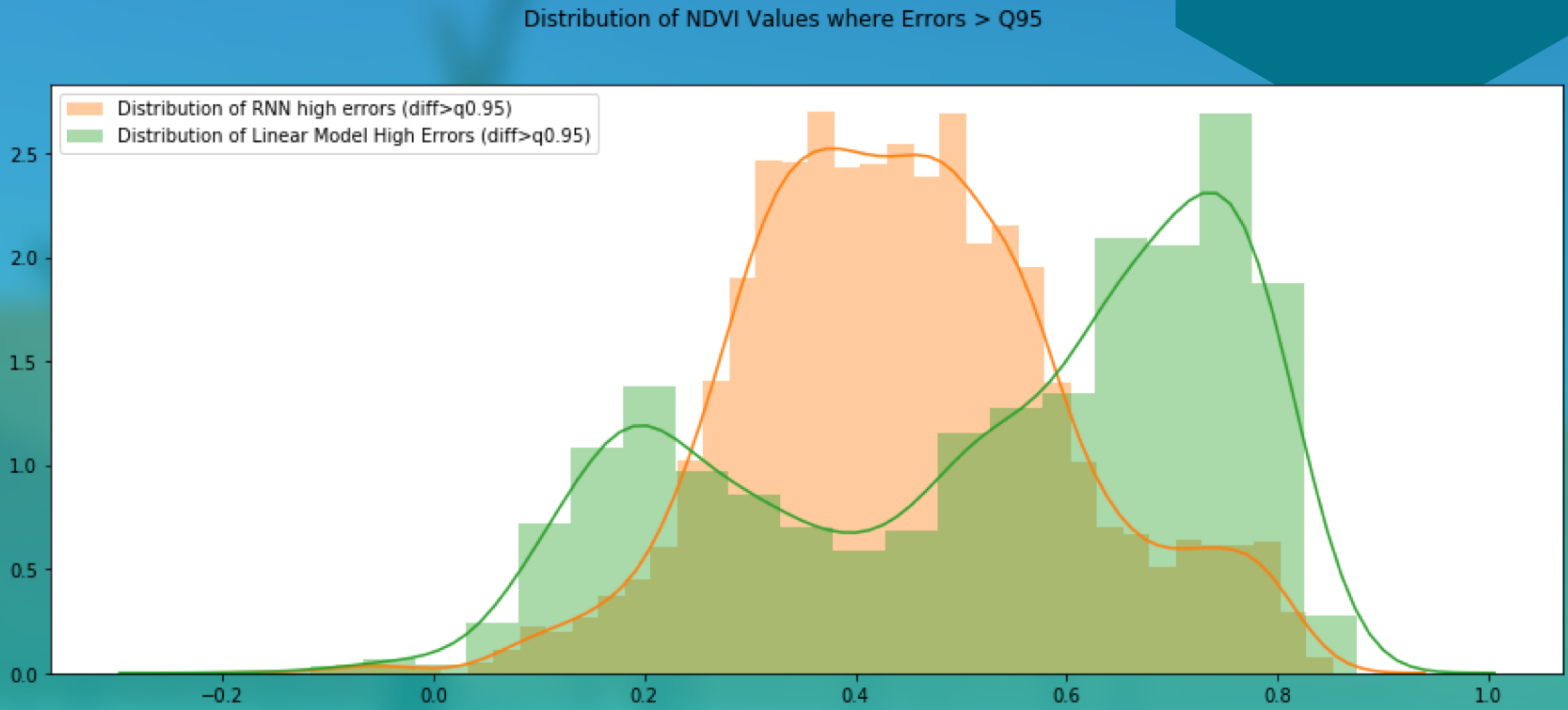
Model Comparison

(VERY) initial results



Model Comparison

(VERY) initial results



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

#ESoWC2019

UNIVERSITY OF OXFORD

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

UNIVERSITY OF OXFORD

https://github.com/esowc/ml_drought

@tommylees112

ECMWF

#ESoWC2019

Appendix

UNIVERSITY OF OXFORD

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

UNIVERSITY OF OXFORD

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

UNIVERSITY OF OXFORD

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

ECMWF

#ESoWC2019

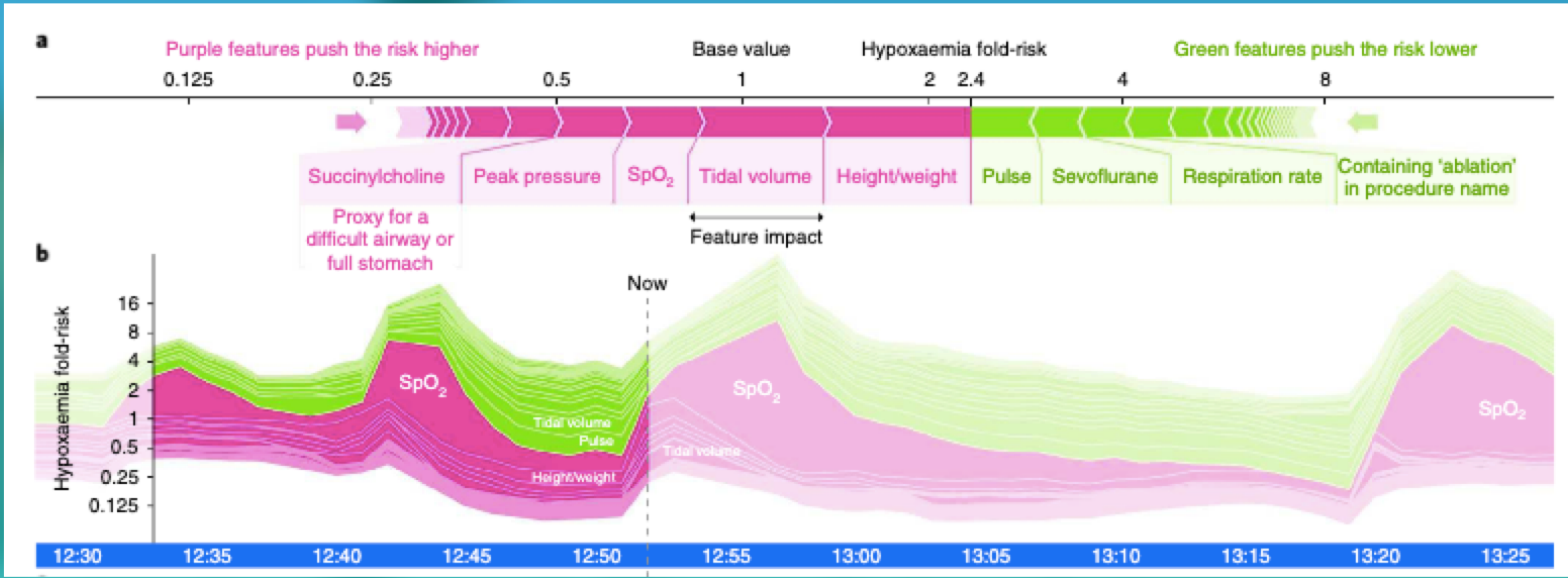
Interpret Models

SHAP Values

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

Interpret Models

SHAP Values



UNIVERSITY OF OXFORD

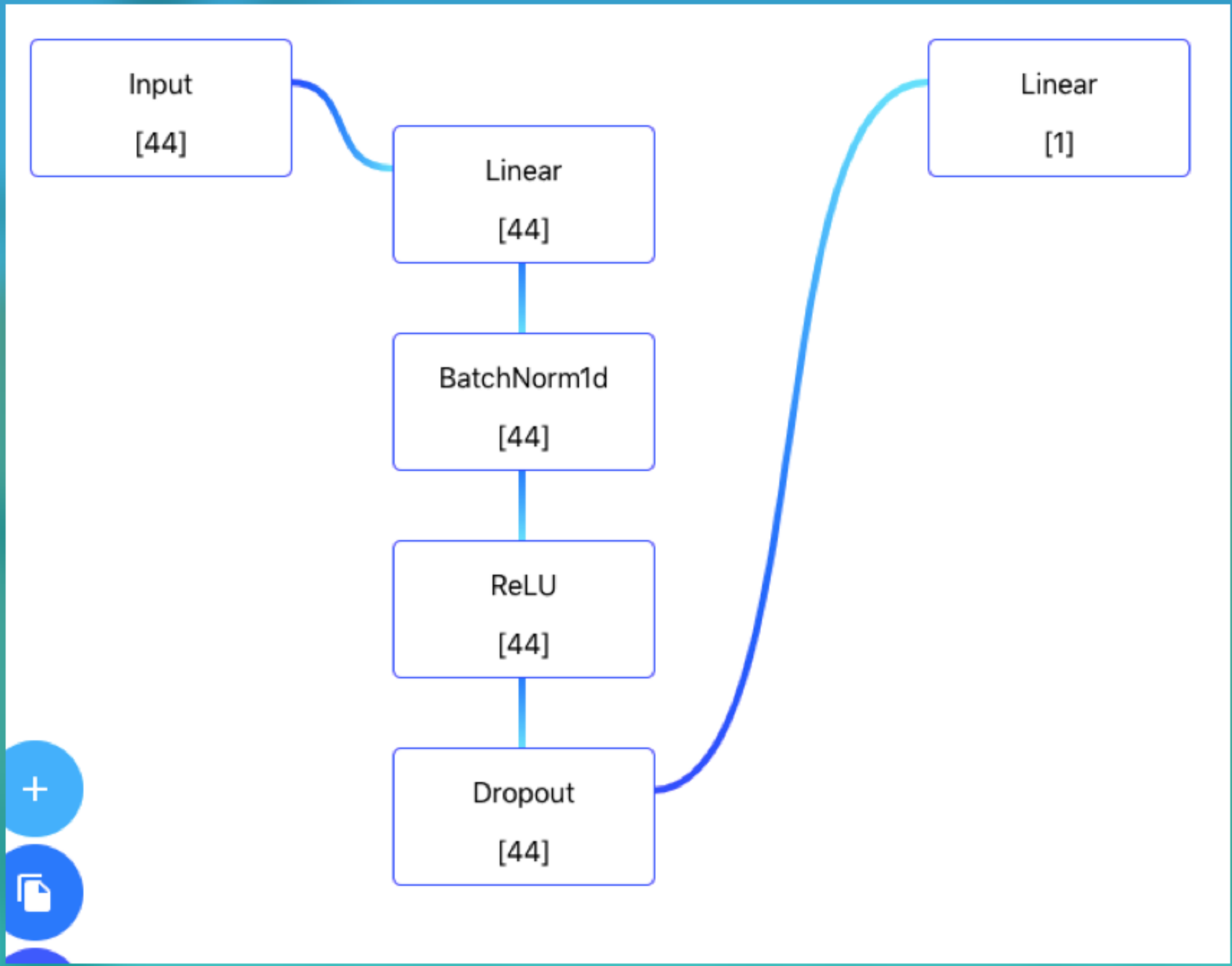
ECMWF

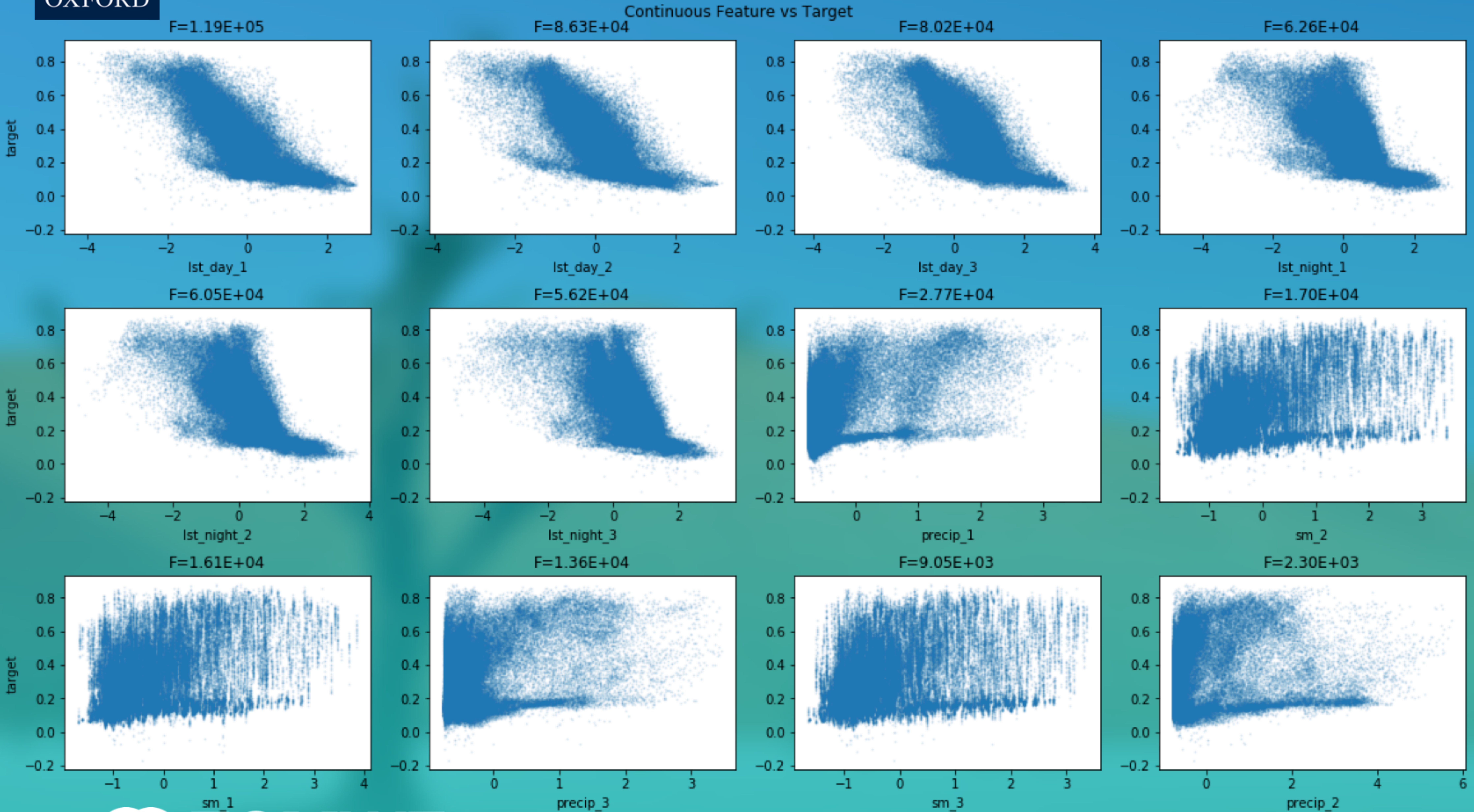
#ESoWC2019

UNIVERSITY OF OXFORD

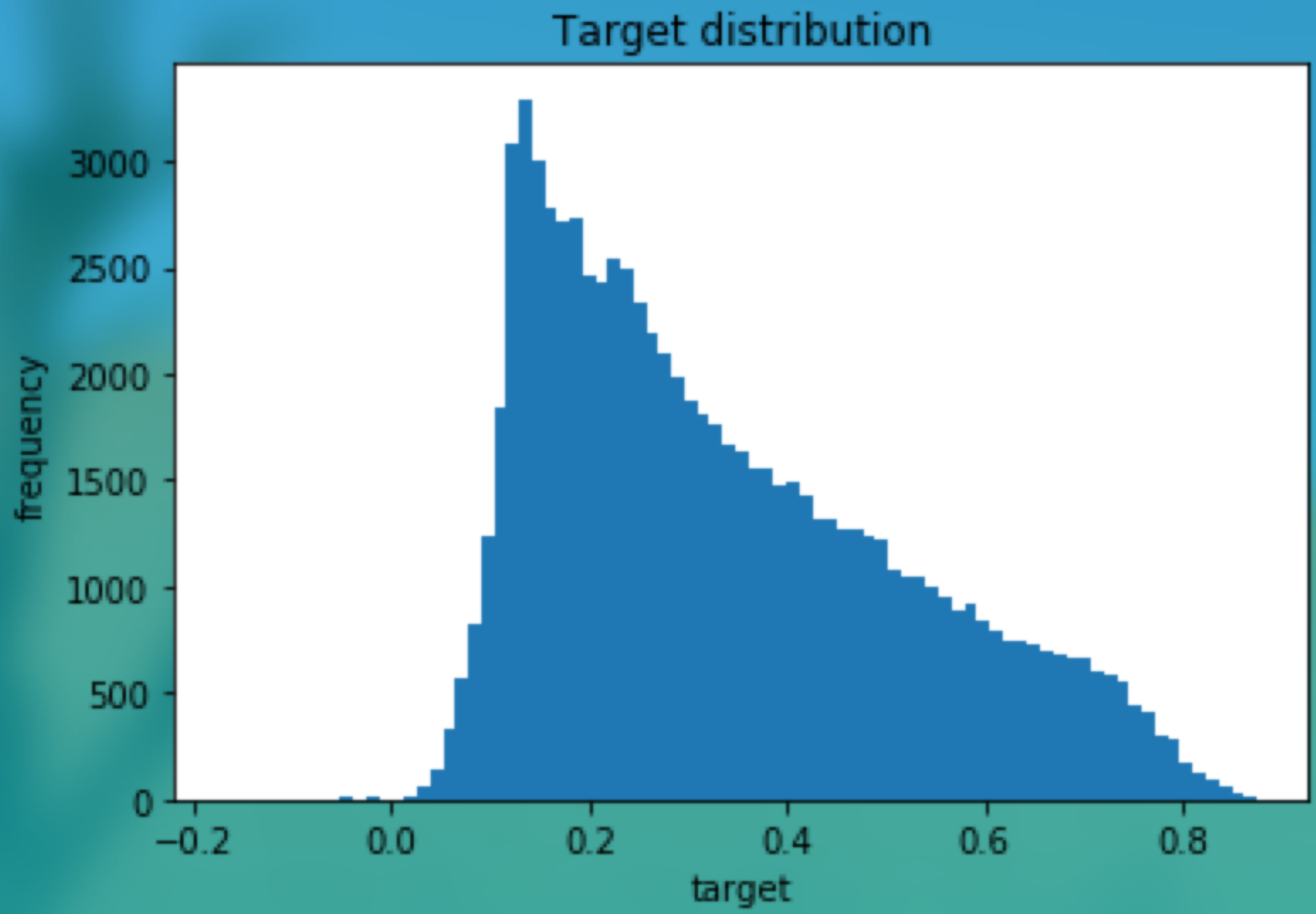
(VERY) initial results

The feedforward model




Continuous Feature vs Target
F=1.19E+05
F=8.63E+04
F=8.02E+04
F=6.26E+04
target
lst_day_1
lst_day_2
lst_day_3
lst_night_1
F=6.05E+04
F=5.62E+04
F=2.77E+04
F=1.70E+04
target
lst_night_2
lst_night_3
precip_1
sm_2
F=1.61E+04
F=1.36E+04
F=9.05E+03
F=2.30E+03
target
sm_1
precip_3
sm_3
precip_2

#ESoWC2019



ECMWF

#ESoWC2019

UNIVERSITY OF OXFORD

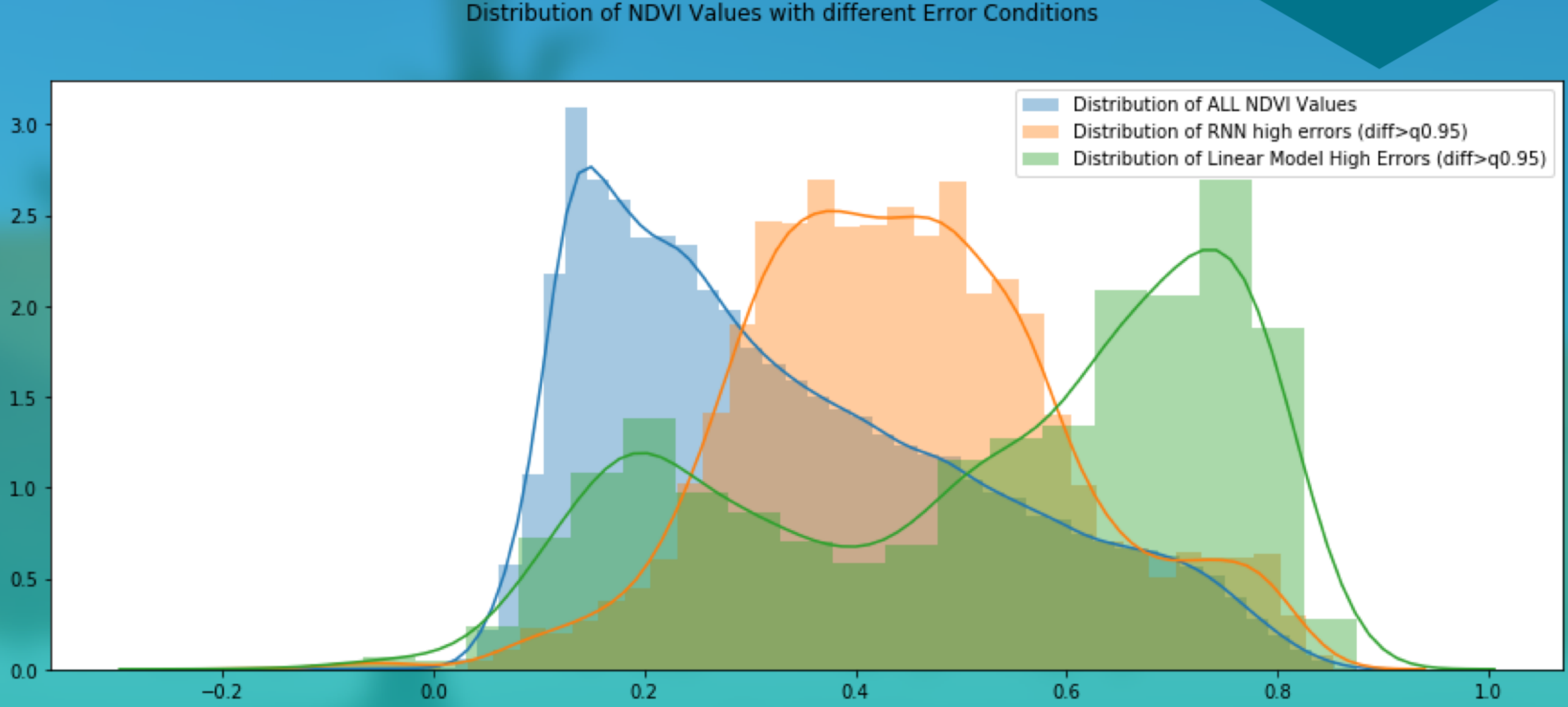
```python
def process(self, pred_month=6, target='ndvi_anomaly'):
    """Preprocesses the raw data, and saves it. Specifically, the following
    preprocessing happens:
    1. `gb_year` and `gb_month`, which are the dates relative to
        `pred_month`, are added.
    2. `ndvi_anomaly` is calculated
    3. NaN values (and missing data) is removed from the dataframe.
    4. Normalizes all values to have mean 0 and std 1

    Parameters
    ----------
    pred_month: int
        The month for which the target value should be predicted. This value will be
        predicted using the preceding 11 months of data
    target: str
        The target variable being predicted

    A processed CSV and a .json object containing the values used to normalize
    each variable are saved.
    """
```

ECMWF

#ESoWC2019

Model Comparison

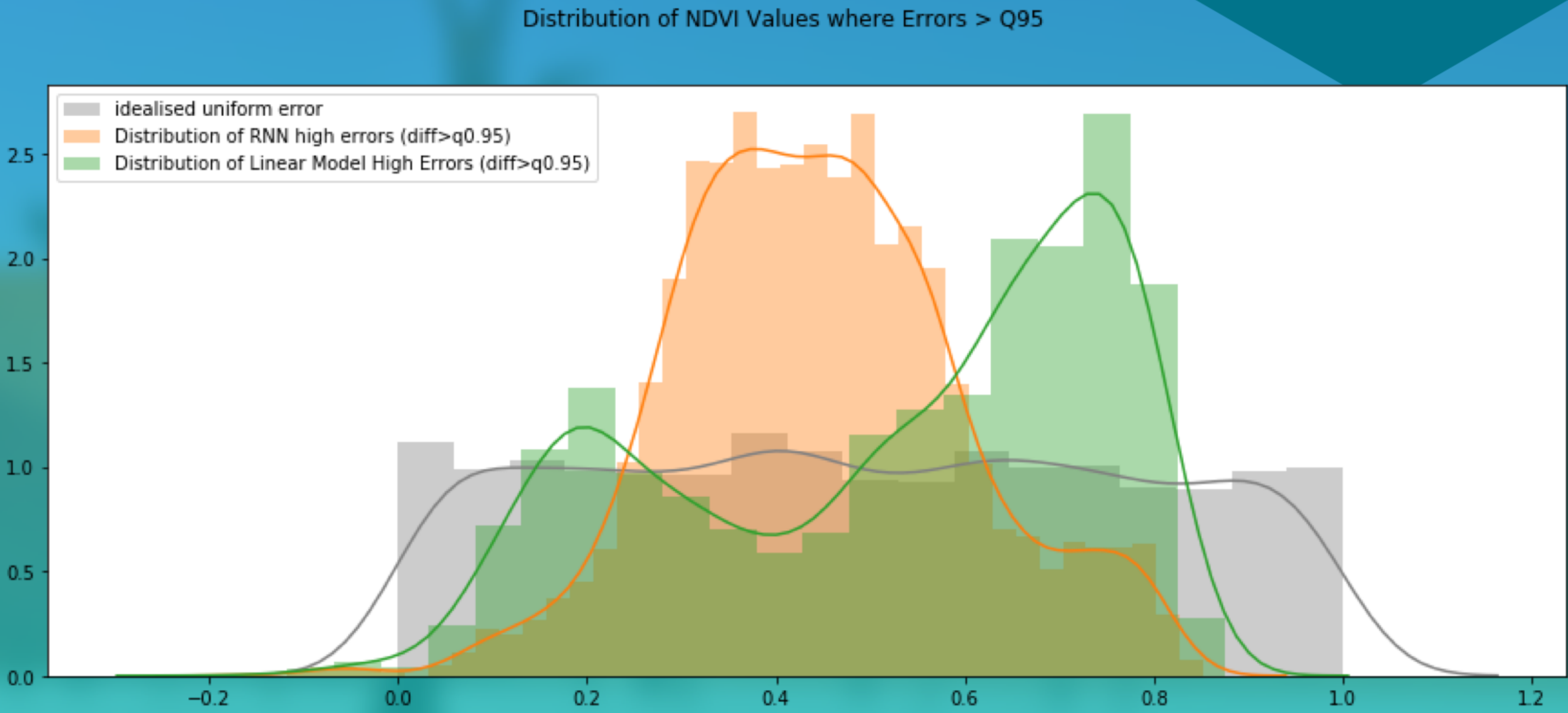
(VERY) initial results



#ESoWC2019

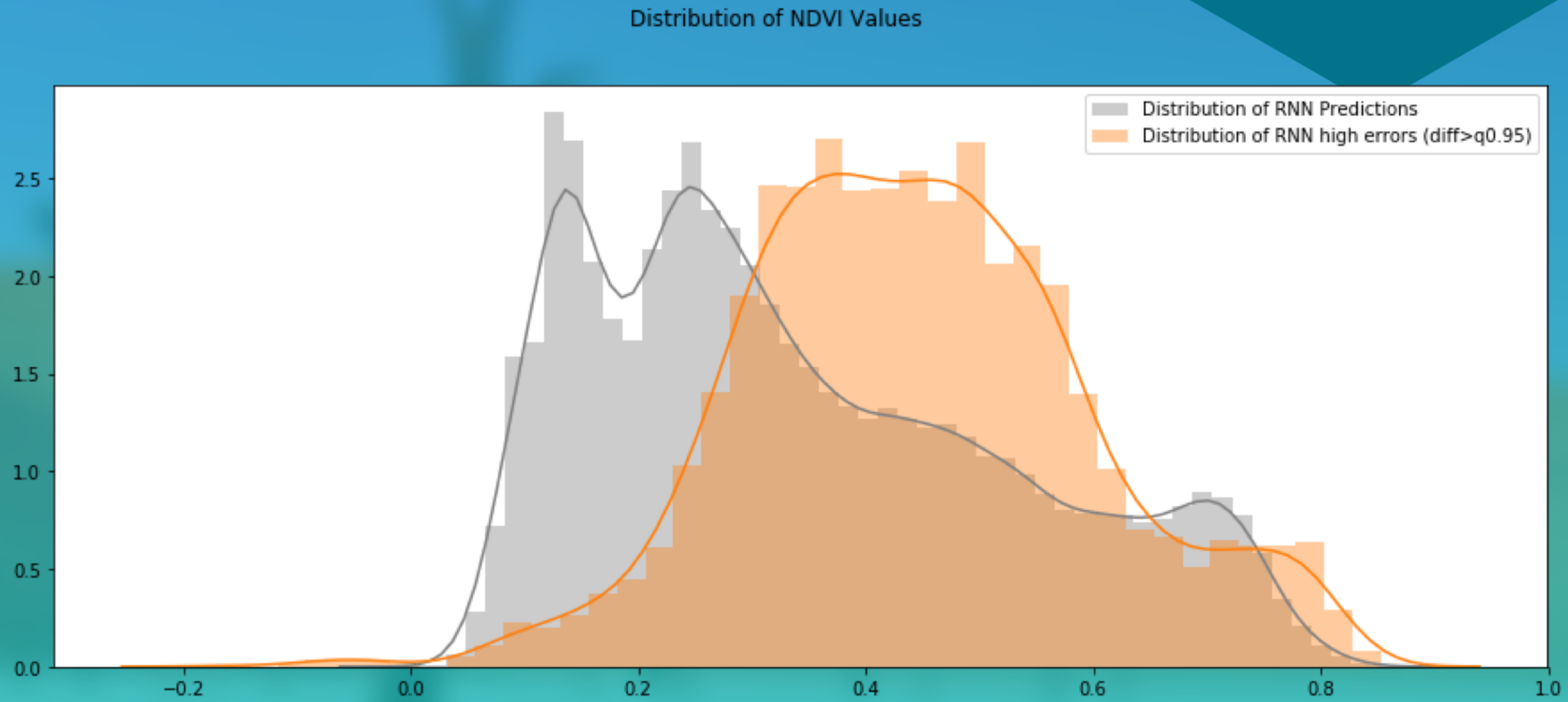
Model Comparison

(VERY) initial results



Model Comparison

(VERY) initial results

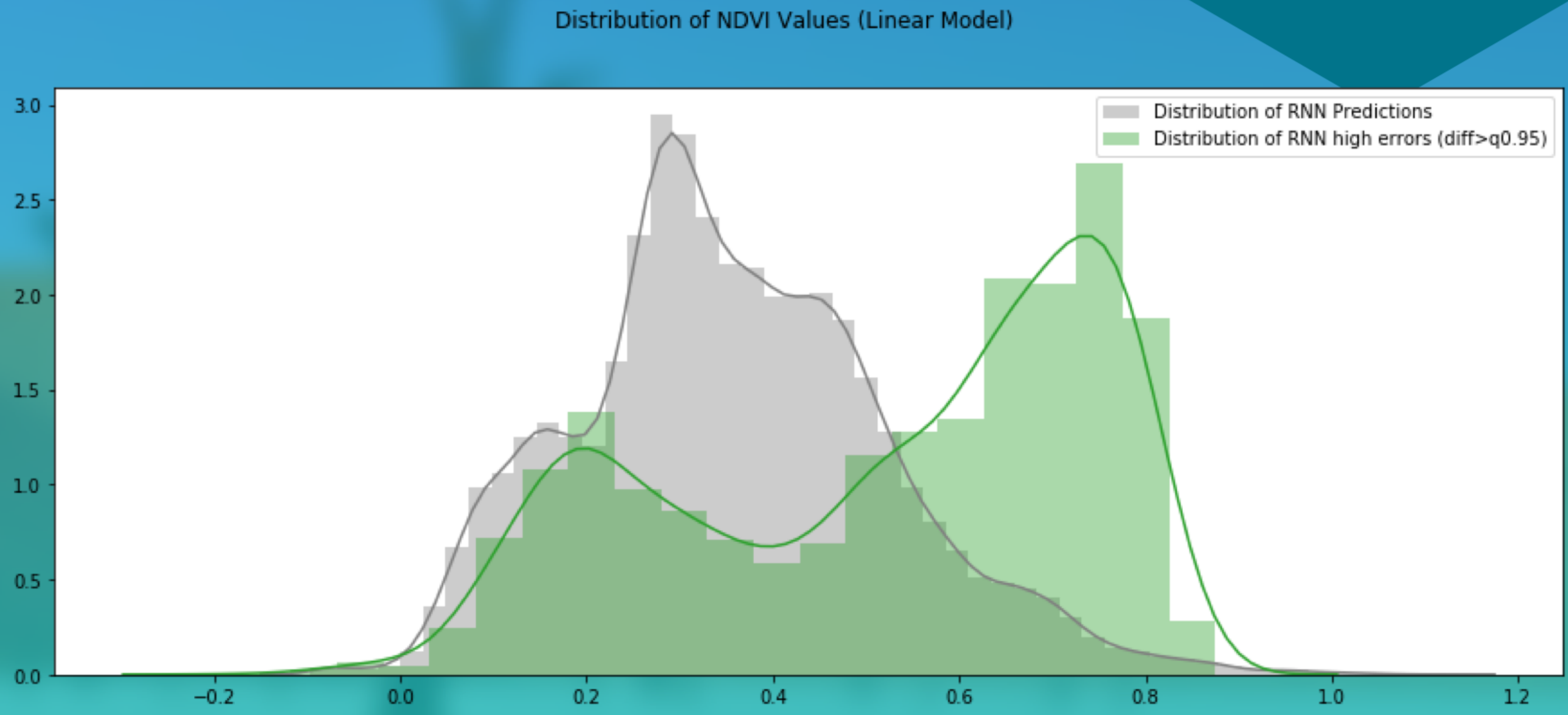


ECMWF

#ESoWC2019

Model Comparison

(VERY) initial results



#ESoWC2019