Developing Machine-Learning Based Emulators for the JULES Land Surface Model

Dr Rob Parker - Research Fellow

National Centre for Earth Observation

Co-authors: Cristina Ruiz Villena, Jasdeep Anand, Tristan Quaife, Natalie Douglas, Ewan Pinnington, Emily Black, Phillip Kershaw, Jack Leland, Andrew Hartley, Andy Wiltshire, Eleanor Burke, Camilla Mathison

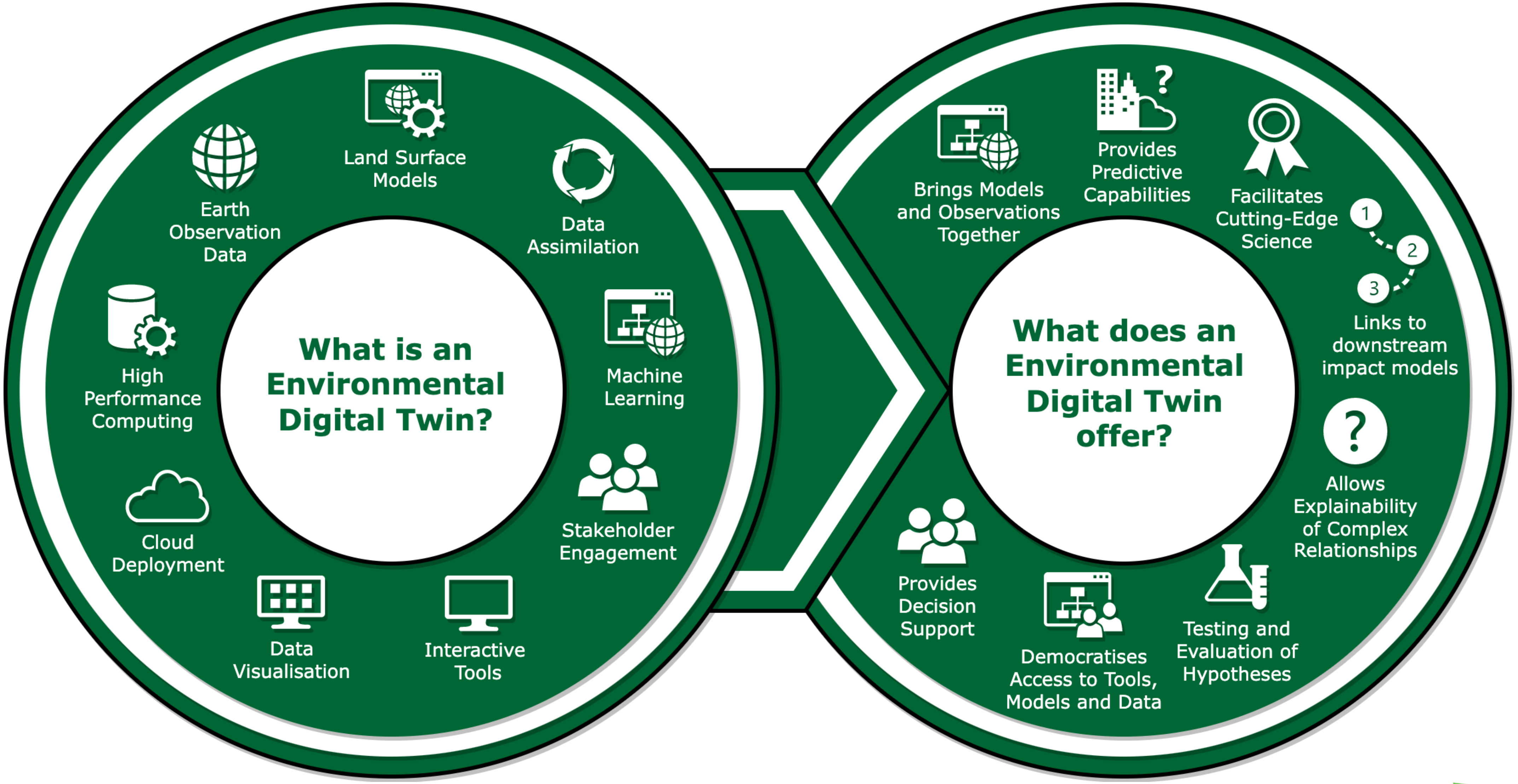
What is a Digital Twin?

- A digital representation of a physical system…
- With some predictive capability (i.e. a model)…
- That is data-driven (e.g. Earth Observation, in-situ, citizen data, etc)…
- Capable of providing decision support to stakeholders

- Lots of different components and considerations that span a whole host of scientific, logistical, technical and IT areas
- Potentially hugely complex and ambitious
- Can extend beyond *just* environmental science (e.g. economics, social sciences, public health, etc)

Rob Parker | rjp23@le.ac.uk

What is a Digital Twin and what does it offer?



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

esa

UKRI Natural Environment Research Council

Rob Parker | rjp23@le.ac.uk

Benefits of emulating the JULES land surface model

Emulator can accurately reproduce JULES simulations but also:

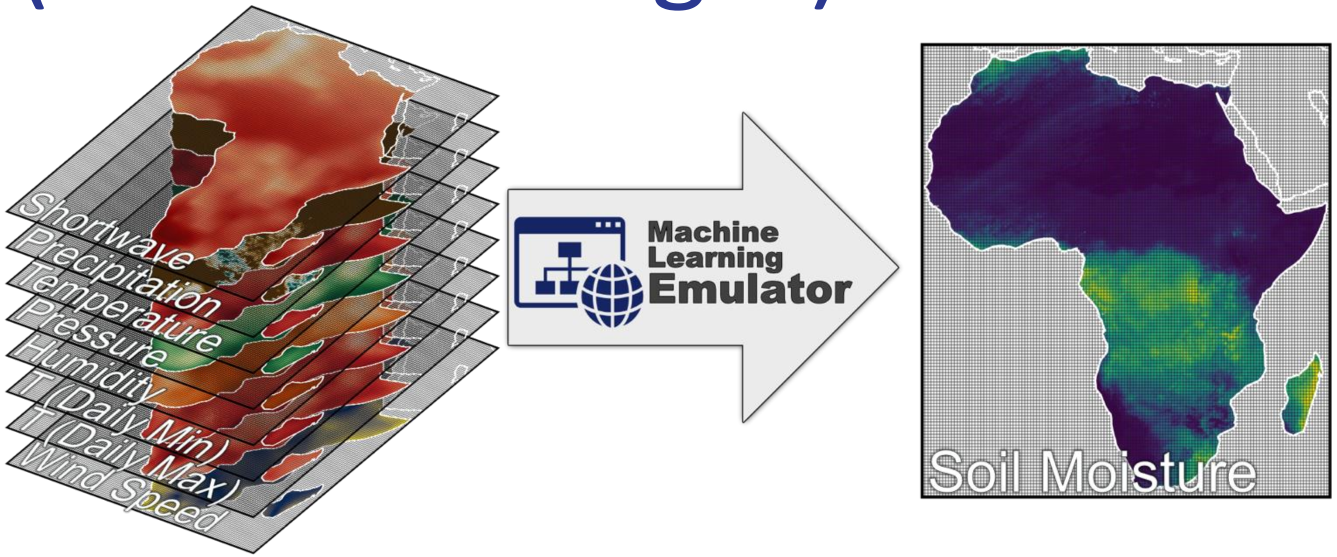
- is extremely fast (years per millisecond)
  - can run huge ensembles, sample uncertainties, etc
- is extremely simple/lightweight (deployed in cloud/notebook/etc)
  - makes JULES far more accessible to non-expert users
  - can be embedded into climate services
- allows explainability of model (Explainable AI methods)
- can be driven by other data (e.g. EO data)
  - constrained by the "physics" within JULES
  - but means we can potentially out-perform JULES by combining JULES and EO data
  - can run at whatever resolution we have available input data for

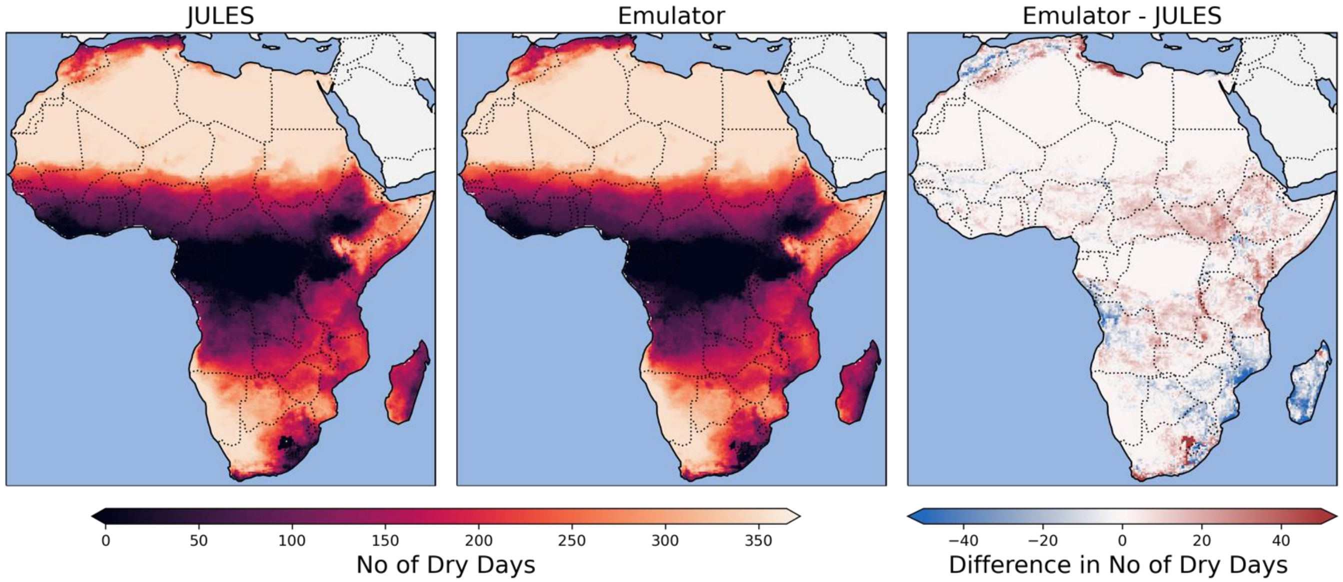
Two NCEO projects related to this work:

- ESA Digital Twin Earth - Drought - Soil moisture over Africa
- ESA IMITATE - Carbon Cycle - GPP over Europe

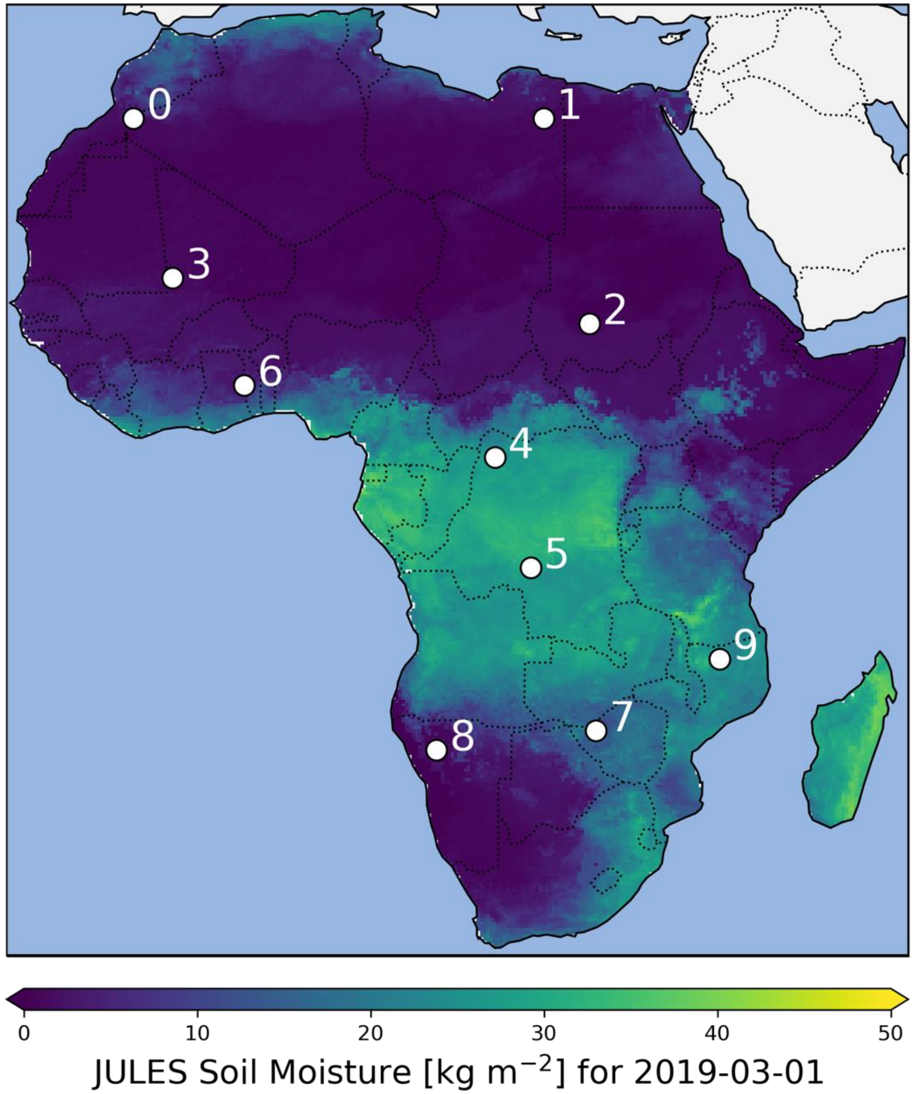
ESA Digital Twin Earth (African Drought)

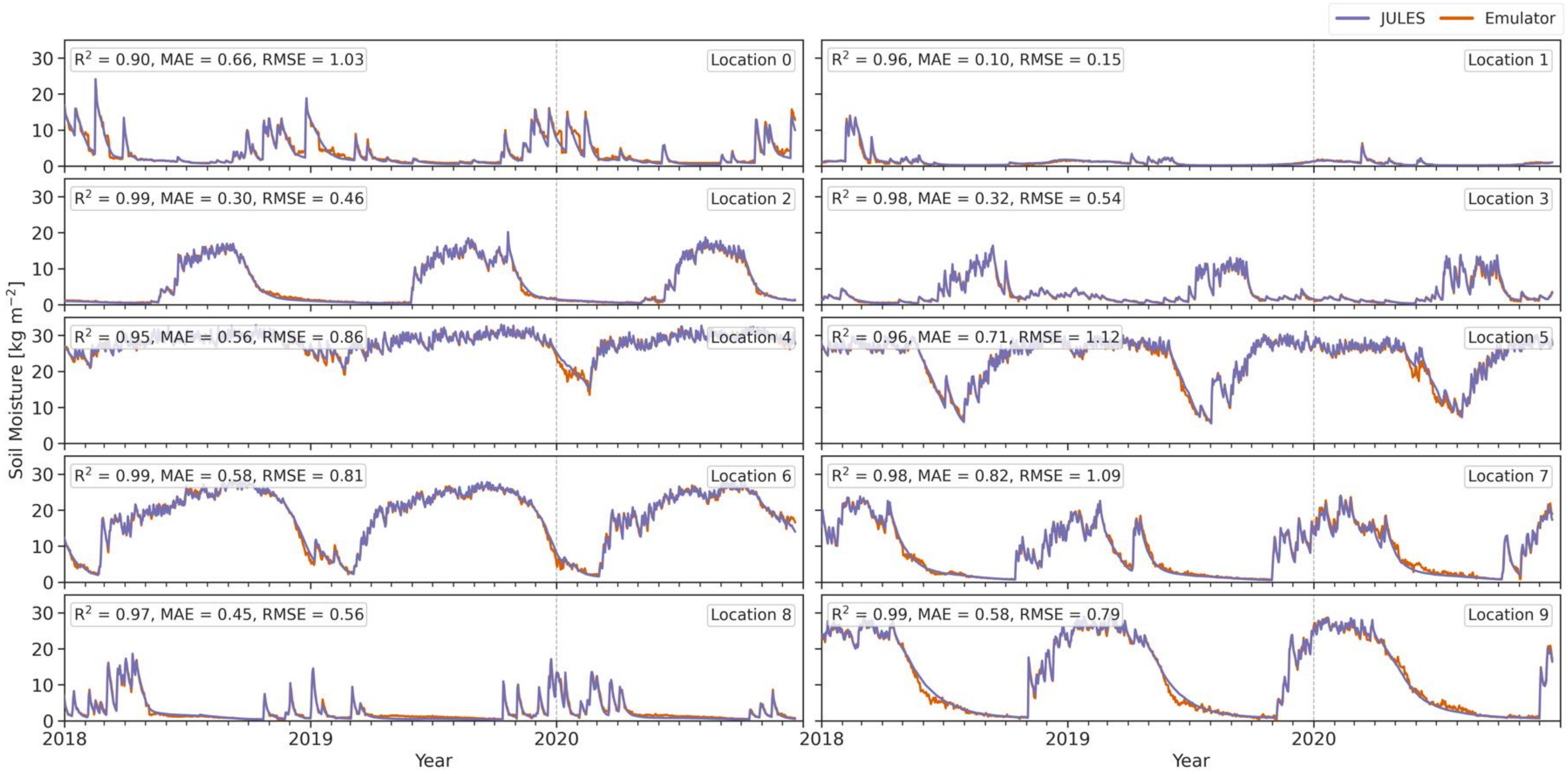
- We've used machine learning to emulate the complex, computationally expensive models in a very fast and light-weight way
- Produce drought metrics - currently **wet season length**, **start date of wet season** and **number of dry days**
- Widgets for these are deployed within our **Interactive Data Portal**
- Emulator is **extremely fast** and **runs in the web-browser**, allowing users to ask their own questions based around soil-moisture response to climate





Evaluation of Emulator

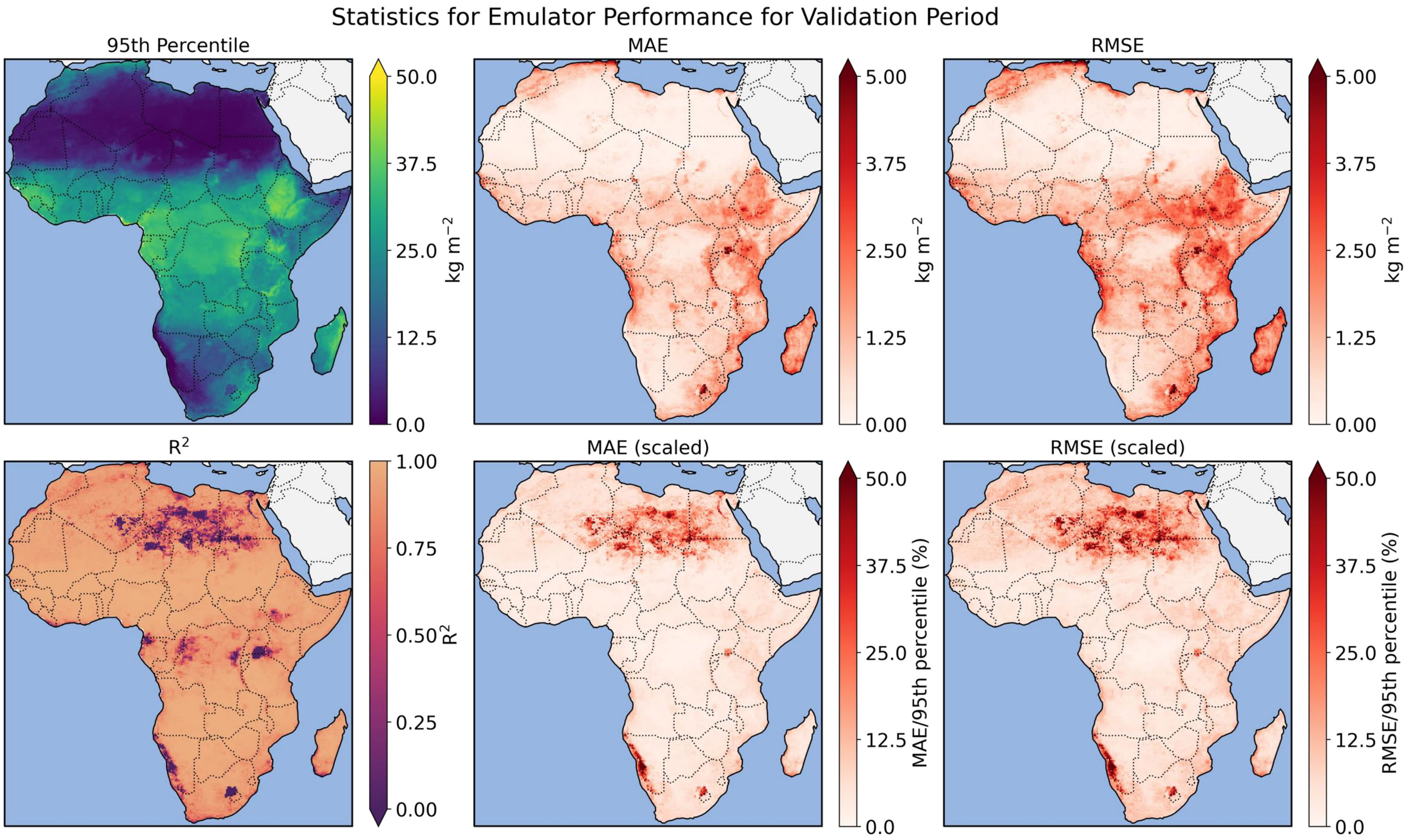




- ❑ Emulator performs exceptionally well and reproduces results of JULES model

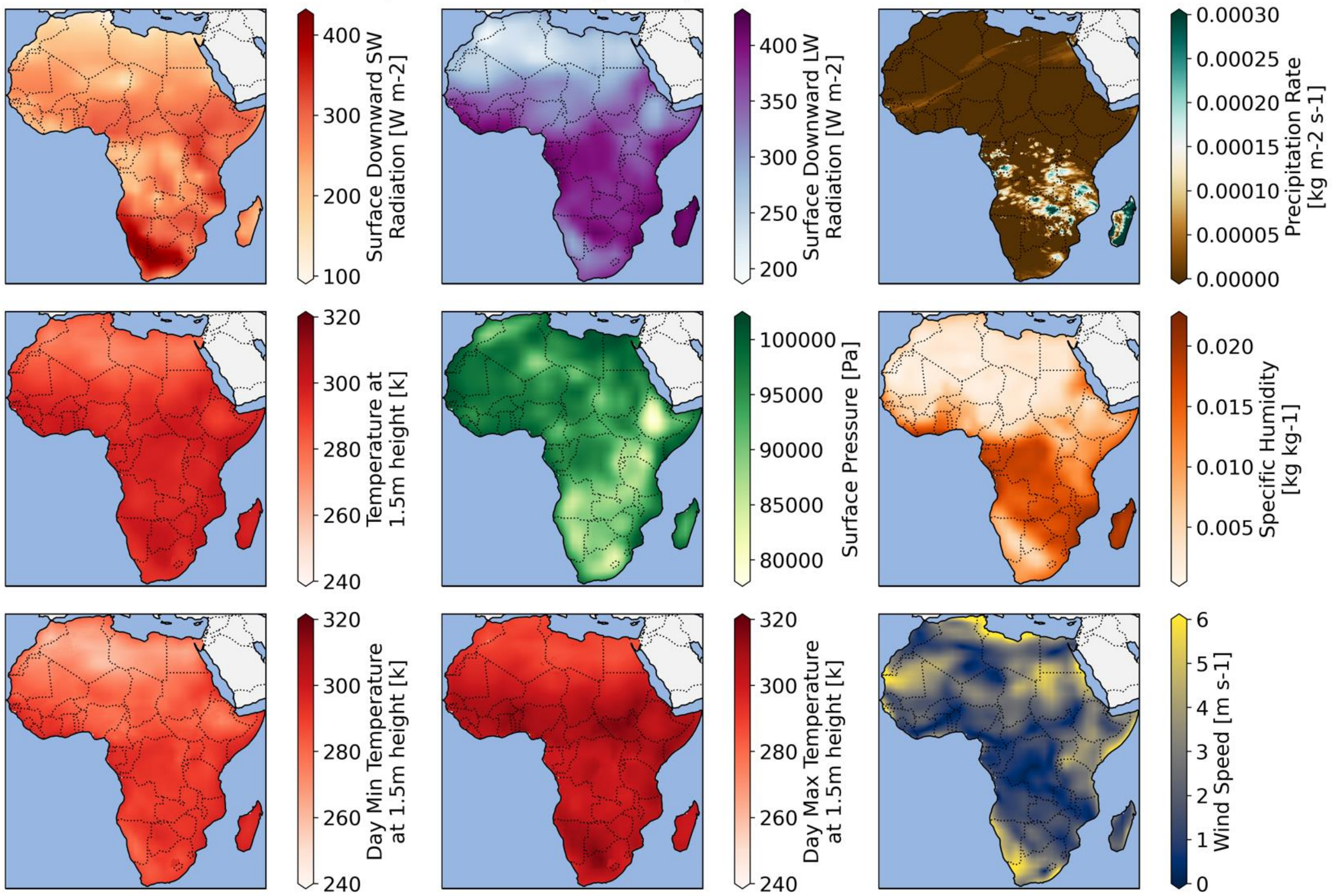
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

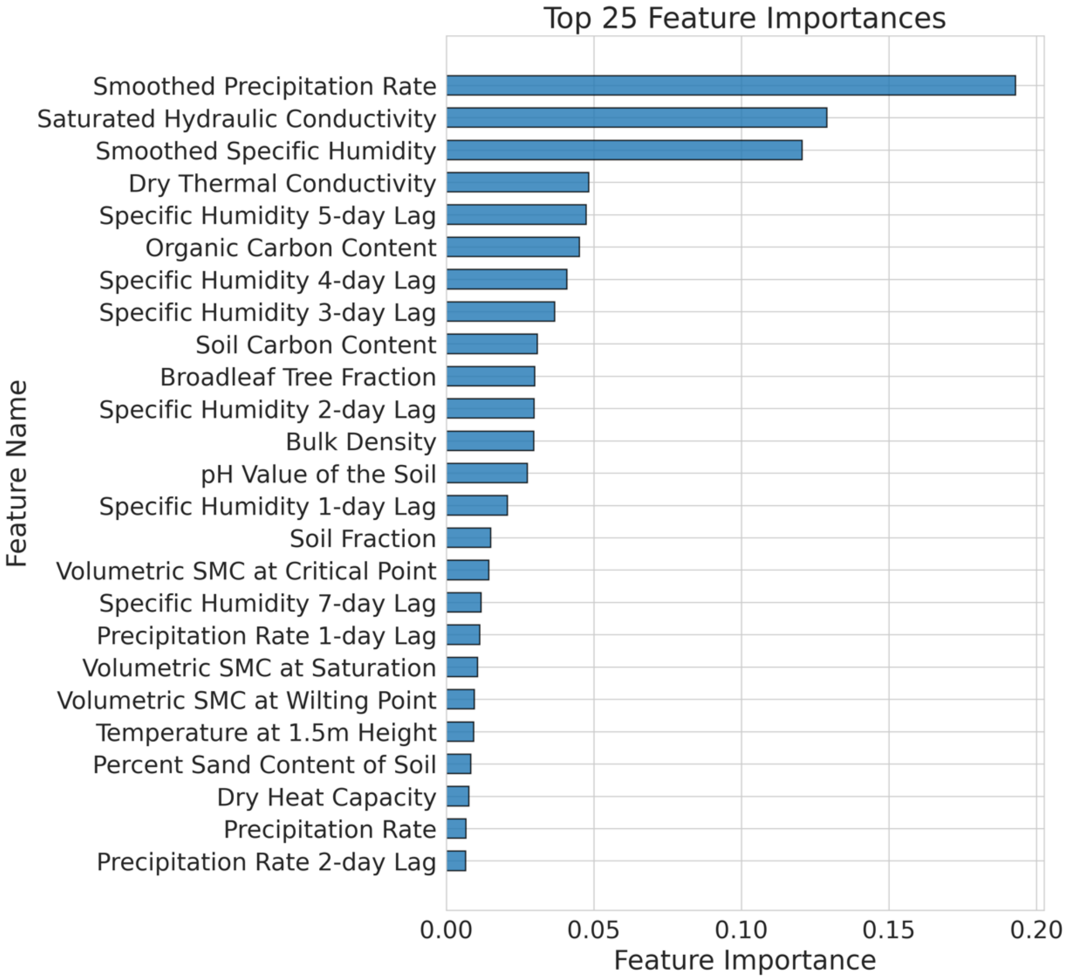
esa UKRI Natural Environment Research Council


Statistics for Emulator Performance for Validation Period
95th Percentile
50.0
37.5
25.0
12.5
0.0
kg m⁻²
MAE
5.00
3.75
2.50
1.25
0.00
kg m⁻²
RMSE
5.00
3.75
2.50
1.25
0.00
kg m⁻²
R2
1.00
0.75
0.50
0.25
0.00
R2
MAE (scaled)
50.0
37.5
25.0
12.5
0.0
MAE/95th percentile (%)
RMSE (scaled)
50.0
37.5
25.0
12.5
0.0
RMSE/95th percentile (%)

Rob Parker | rjp23@le.ac.uk

Explainability and Feature Importance



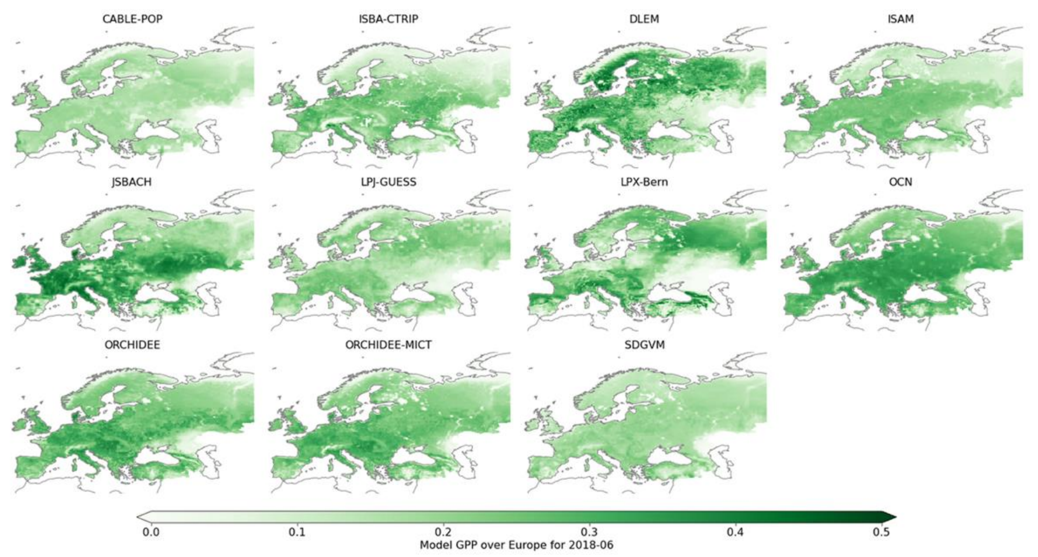


National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL
esa
UKRI
Natural Environment Research Council

Rob Parker | rjp23@le.ac.uk

ESA IMITATE (Carbon Cycle)

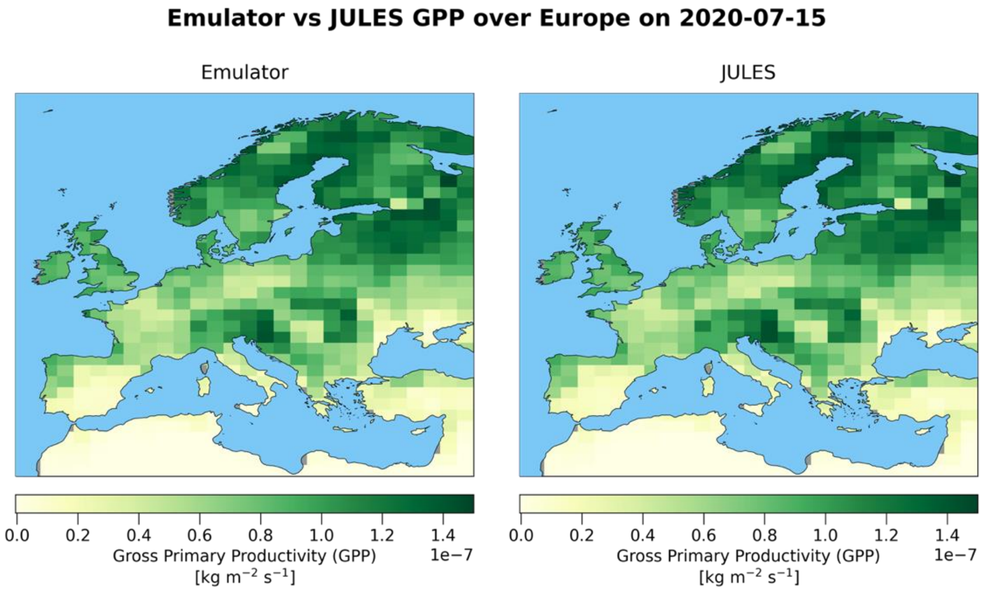
The carbon cycle over Europe is still **highly** uncertain and neither observations nor models alone are capable of addressing these issues.



We are developing machine-learning model **emulators** to replicate simulations from complex land surface model.

Emulators allow greater **understanding** of the model behaviour and let us explore the different relationships between the drivers and carbon fluxes.

We can then use emulator **with** Earth Observation data to derive **new** datasets that are explicitly tied to observations and can make use of their uncertainties.



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

esa

UKRI Natural Environment Research Council

Rob Parker | rjp23@le.ac.uk

JULES Data

All GPP-related inputs

- Gridbox emulator
- Per-PFT emulator

Features used in emulator training:

- Minimum Temperature
- Maximum Temperature
- Soil moisture (**all layers**)
- Soil Moisture Availability Factor
- FAPAR
- SW Downwelling Radiation

1) Can we emulate JULES GPP using all available (relevant) JULES variables?

Observable inputs

- Per-PFT emulator

Features used in emulator training:

- Minimum Temperature
- Maximum Temperature
- Soil moisture (**top layer**)
- Diffuse VIS albedo
- Diffuse NIR albedo

2) Can we emulate JULES GPP using only JULES variables that have an EO equivalent?

EO Data

EO inputs

- Per-PFT emulator

Features used in emulator training:

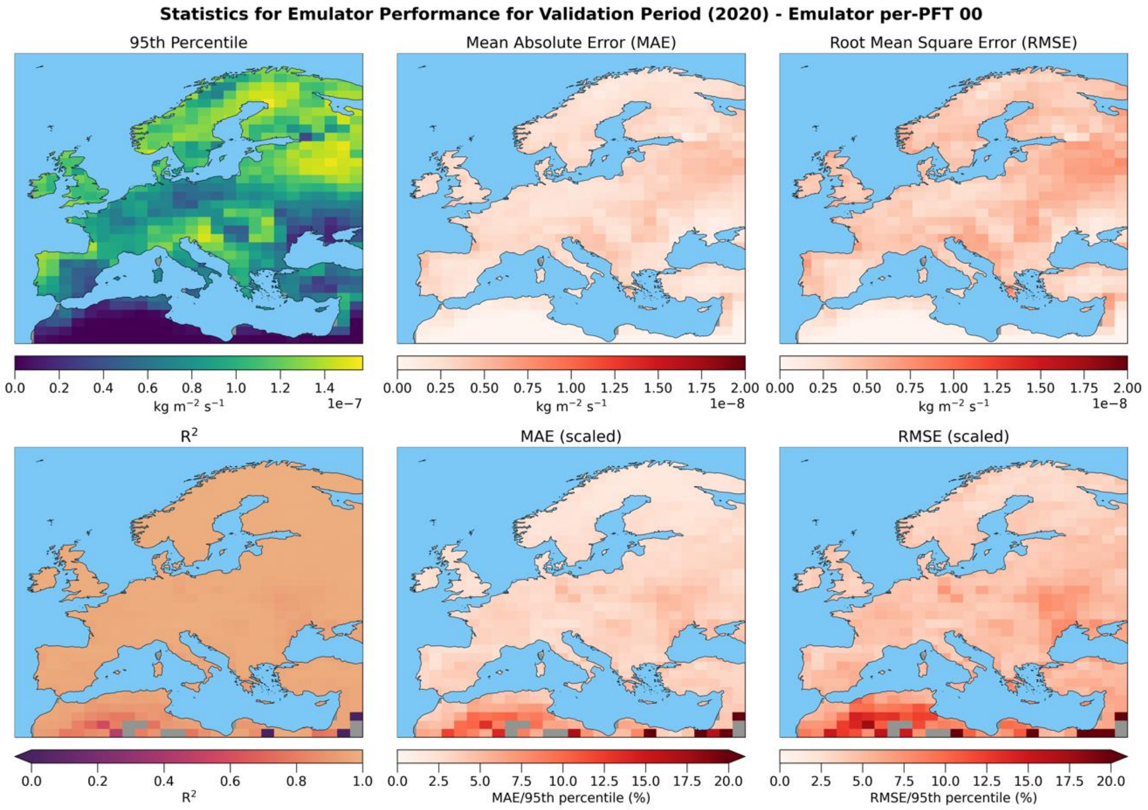
- Minimum LST
- Maximum LST
- Soil moisture (surface)
- White-sky VIS albedo
- White-sky NIR albedo

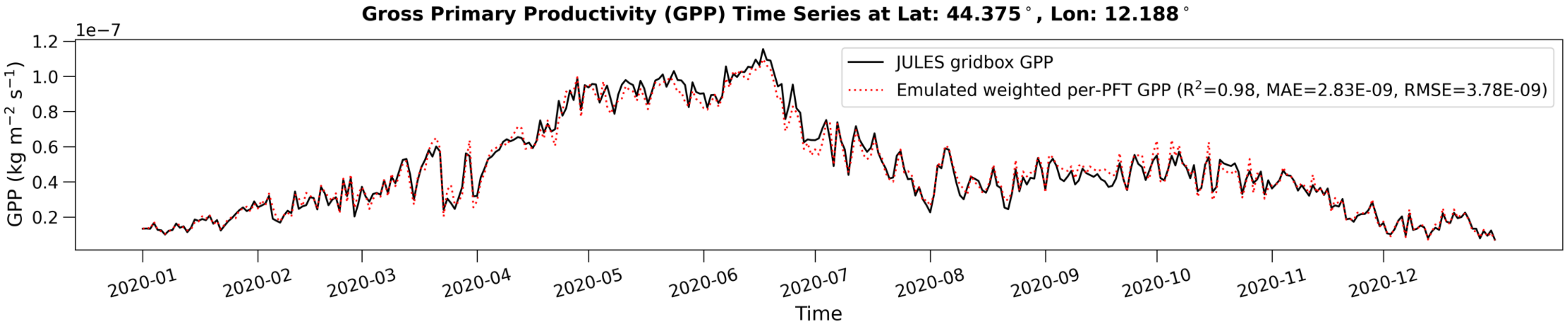
3) Can we use the EO data directly to produce an EO-based GPP data product constrained by JULES process representation?

1) Can we emulate JULES GPP using all available (relevant) JULES variables?

Features

- Minimum Temperature
- Maximum Temperature
- Soil moisture (**all layers**)
- Soil Moisture Availability Factor
- FAPAR
- SW Downwelling Radiation





National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

esa

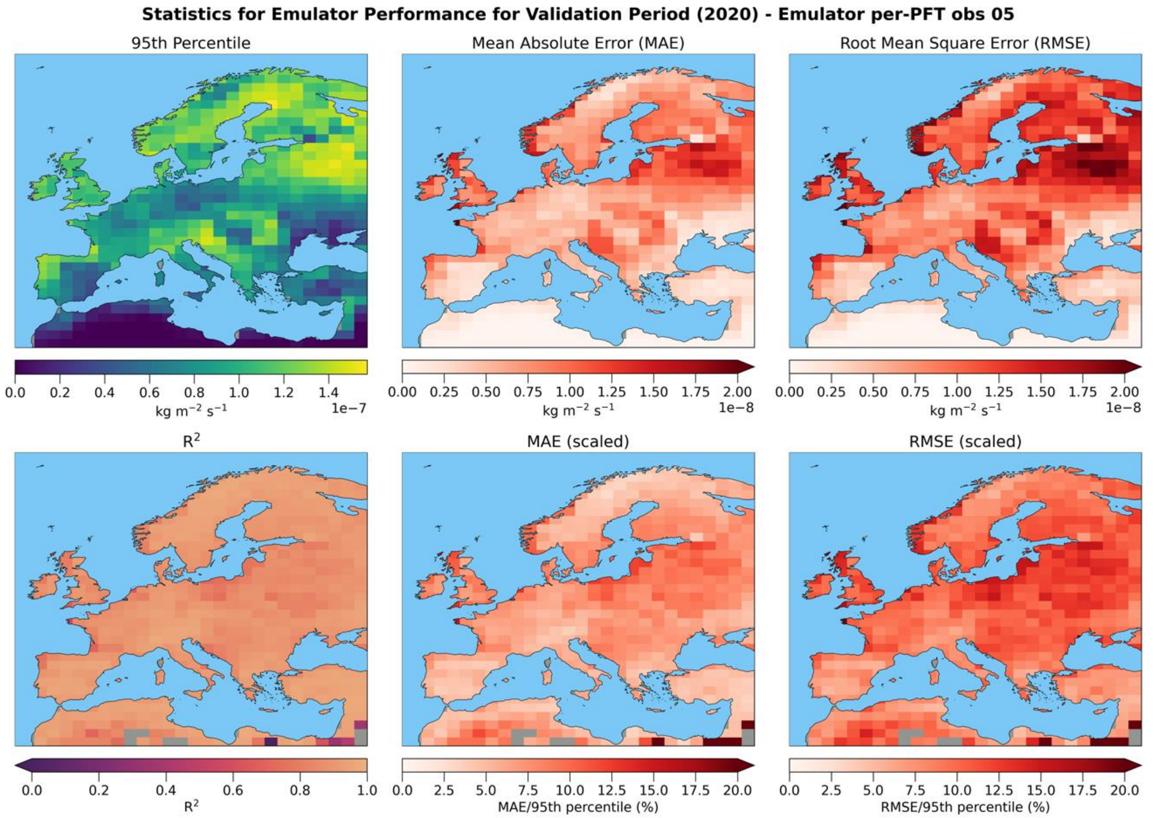
Natural Environment Research Council

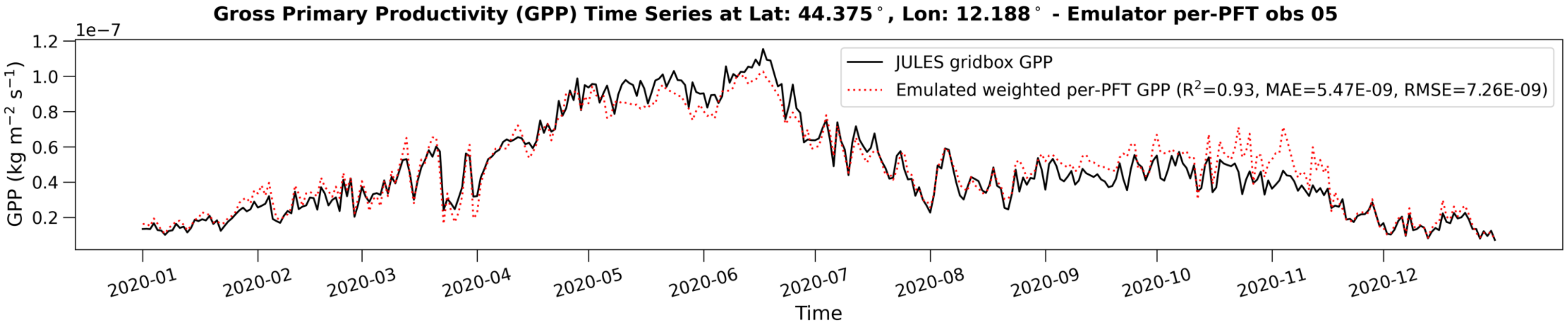
work in progress

2) Can we emulate JULES GPP using only JULES variables that have an EO equivalent?

Features

- Minimum Temperature
- Maximum Temperature
- Soil moisture (**top layer**)
- Diffuse VIS albedo
- Diffuse NIR albedo
- Soil Moisture Availability Factor





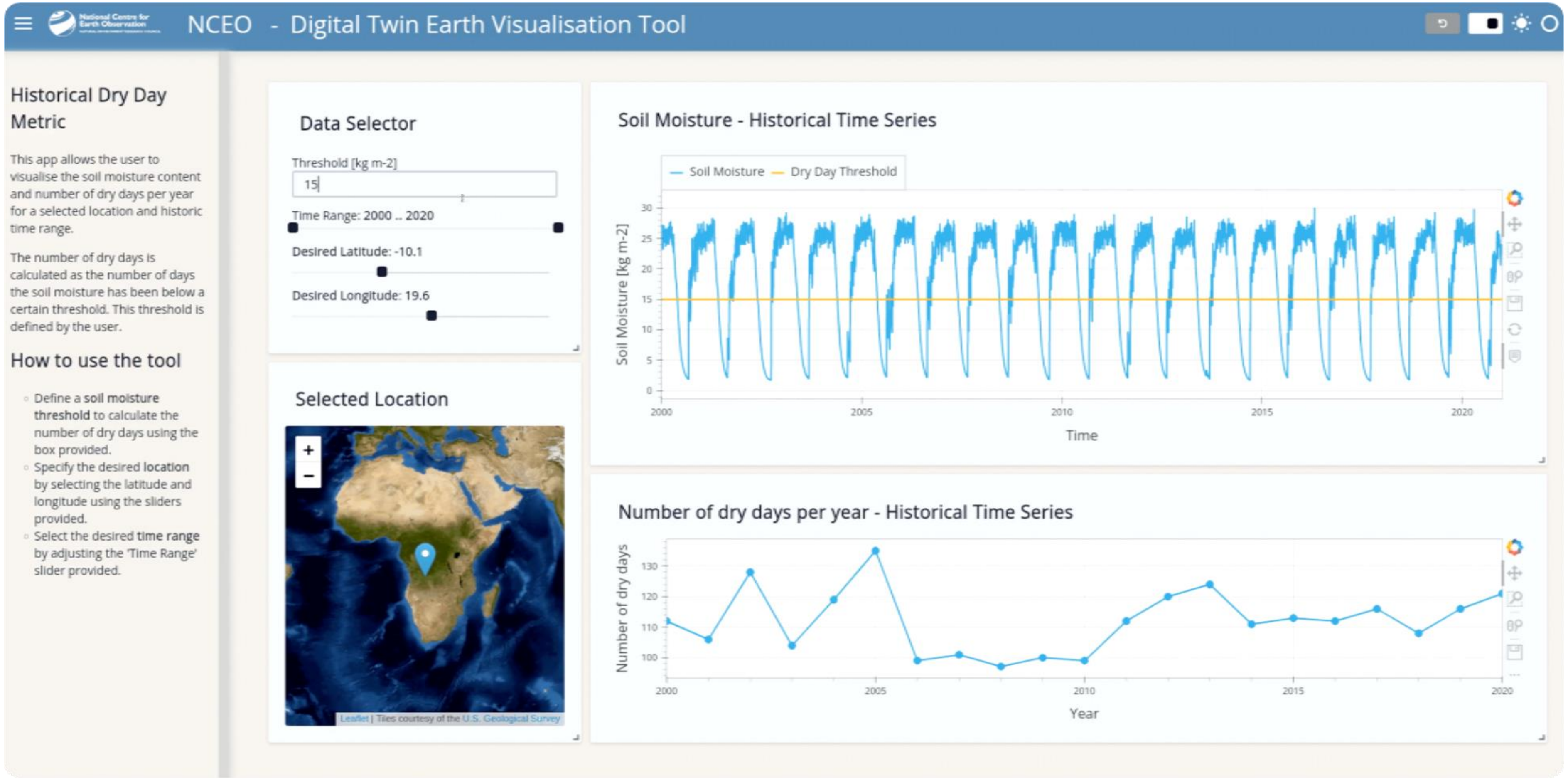
Summary and Conclusions

- We've successfully developed machine-learning based emulators for several different JULES applications (drought and GPP)
- These emulators are (very!) fast, easy to use, etc and open up a range of applications and potentially interesting science
- We're very interested in emulating other bits of JULES
  - Working with UKCEH/MO about a proof-of-concept for JULES-Inferno fires
  - Working with UKCEH/MO on methane emissions and wetland inundation
- If there are JULES simulations where we can easily map the inputs to the output, we can probably build an emulator for it
  - Much more work to do: deploying applications, Explainable AI, model-data fusion by driving with EO data, uncertainty propagation, extending beyond JULES to other land surface models, etc

Extra Slides

NERC Digital Twin Case Study

- In the ESA DTEP project we developed a ML-based emulator for JULES soil moisture over Africa
- This project builds on that and further develops interactive tools for stakeholder engagement
- NCEO (Leicester, Reading, CEDA) with Met Office and STFC-RAL as project partners



next steps

3) Can we use the EO data directly to produce an EO-based GPP data product constrained by JULES process representation?

	Soil Moisture	Land Surface Temperature	Albedo	Land Cover
Product	ESA CCI soil moisture v6.1 COMBINED	ESA CCI LST 3-hourly	MODIS MCD43C3 CMG Albedo	ESA CCI Global Land Cover Maps v2.0.7
JASMIN path	`/neodc/esacci/soil_moisture/data/daily_files/COMBINED/v06.1`	`/neodc/esacci/land_surface_temperature/data/MULTISENSOR_IRMGP/L3S/0.05/v1.00/daily`	N/A https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/61/MCD43C3/	`/neodc/esacci/land_cover/data/land_cover_maps/v2.0.7`
Units	[m^3 m^{-3}]	[K]	[-]	[-]
Time range	1978-11-01 to 2020-12-31	2009-2020	2000-2022	1992 - 2015
Spatial resolution	0.25 degrees	0.05 degrees	0.05 degrees	300 metres