Machine-Learning Emulators of Land Surface Model 'JULES' for African Hydrological Digital Twin Applications

Dr Cristina Ruiz Villena & Dr Rob Parker
National Centre for Earth Observation

Co-authors: Tristan Quaife, Natalie Douglas, Andy Wiltshire, Ross Maidment, Jasdeep Anand

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

EOCIS
EARTH OBSERVATION CLIMATE INFORMATION SERVICE

esa

UNIVERSITY OF LEICESTER

ABOUT DIGITAL TWINS (DTs)

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

Digital Twins can help us make environmental decisions

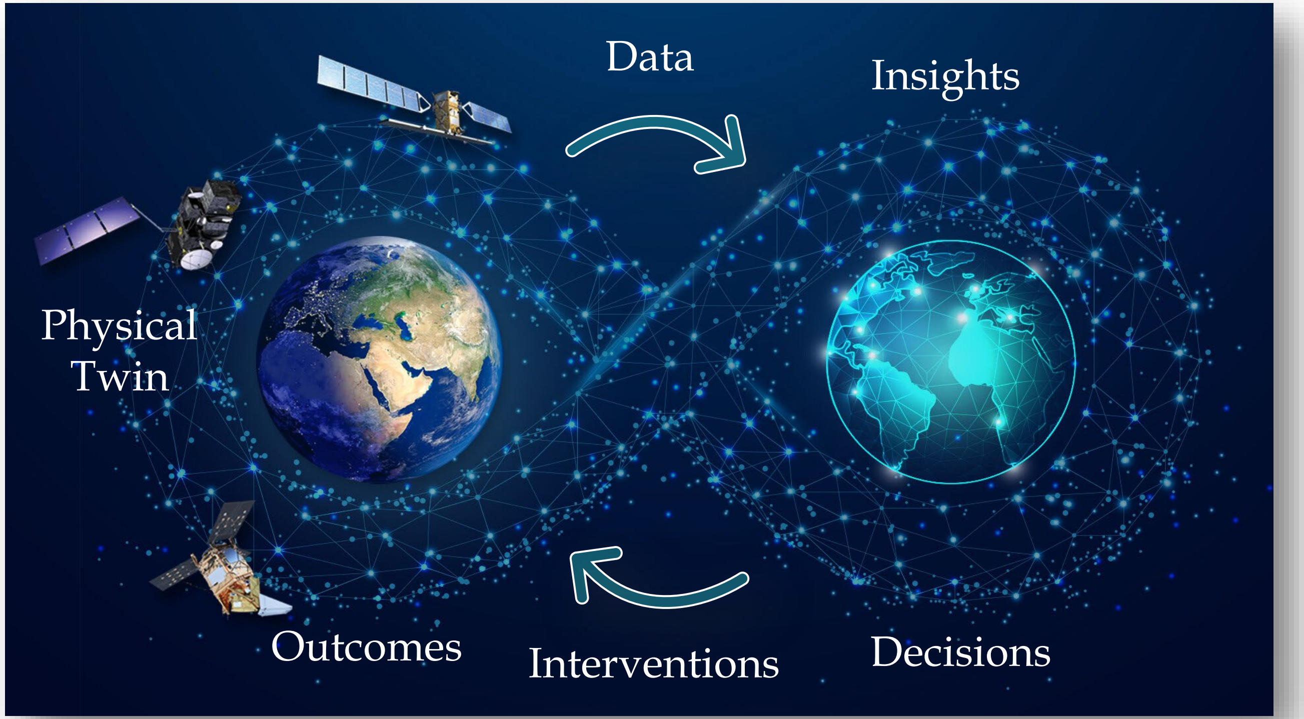


Image credit: adapted from ESA

- ✓ High Performance Computing
- ✓ Cloud Services
- ✓ Machine Learning
- ✓ Data driven
- ✓ What-if scenarios

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UKRI Natural Environment Research Council

Cristina Ruiz Villena | crv2@le.ac.uk

Our approach

1. MODEL

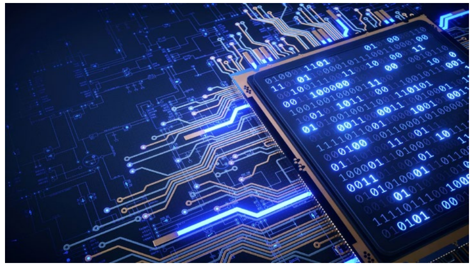
JULES
Joint UK Land Environment Simulator

- **Community** model coordinated by **UK Met Office** and **UKCEH**.
- **Land surface** component of the UK Earth System Model (**UKESM**).
- Major part of UK contribution to **global model intercomparison** projects (e.g. CMIP6), thus informs the IPCC.

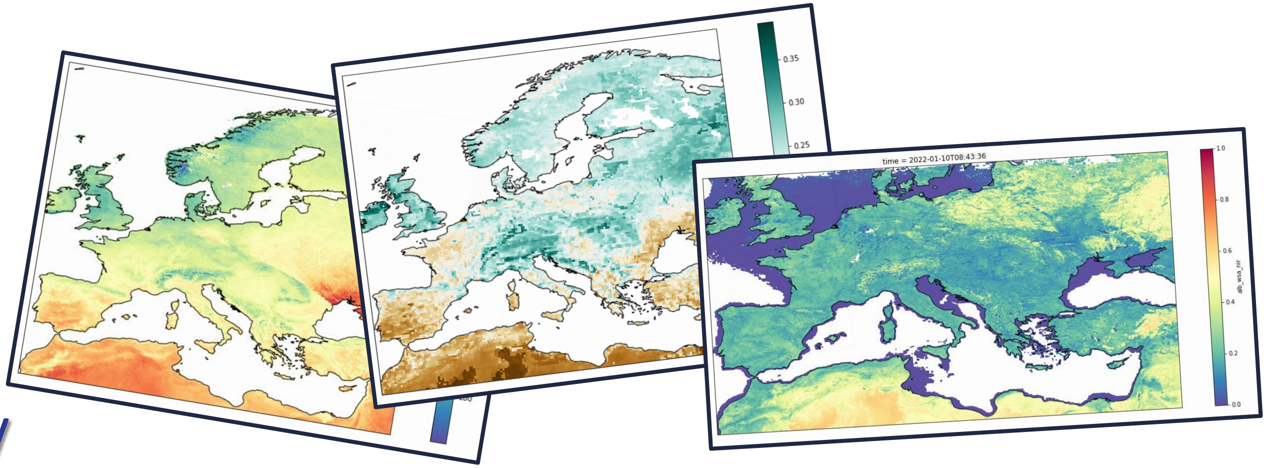
Selected processes

→ Soil moisture
→ Gross Primary Productivity (GPP)
→ Wetlands
→ Wildfires

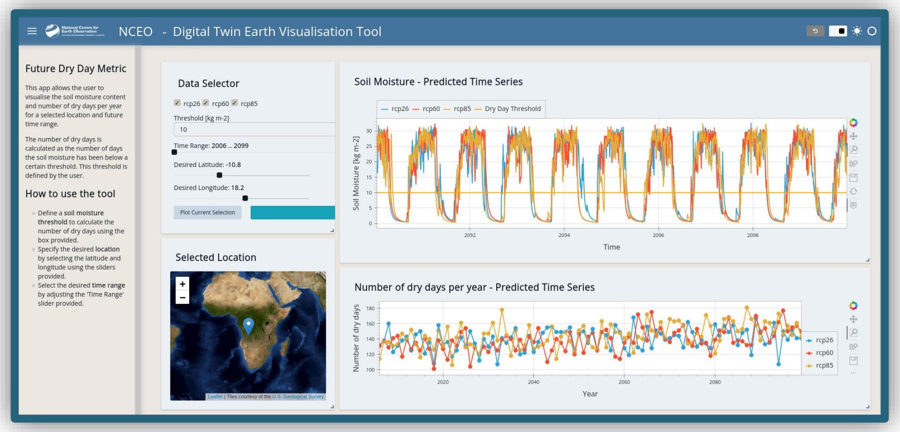
2. MACHINE LEARNING



3. EO DATA



4. INTERACTIVE DASHBOARD



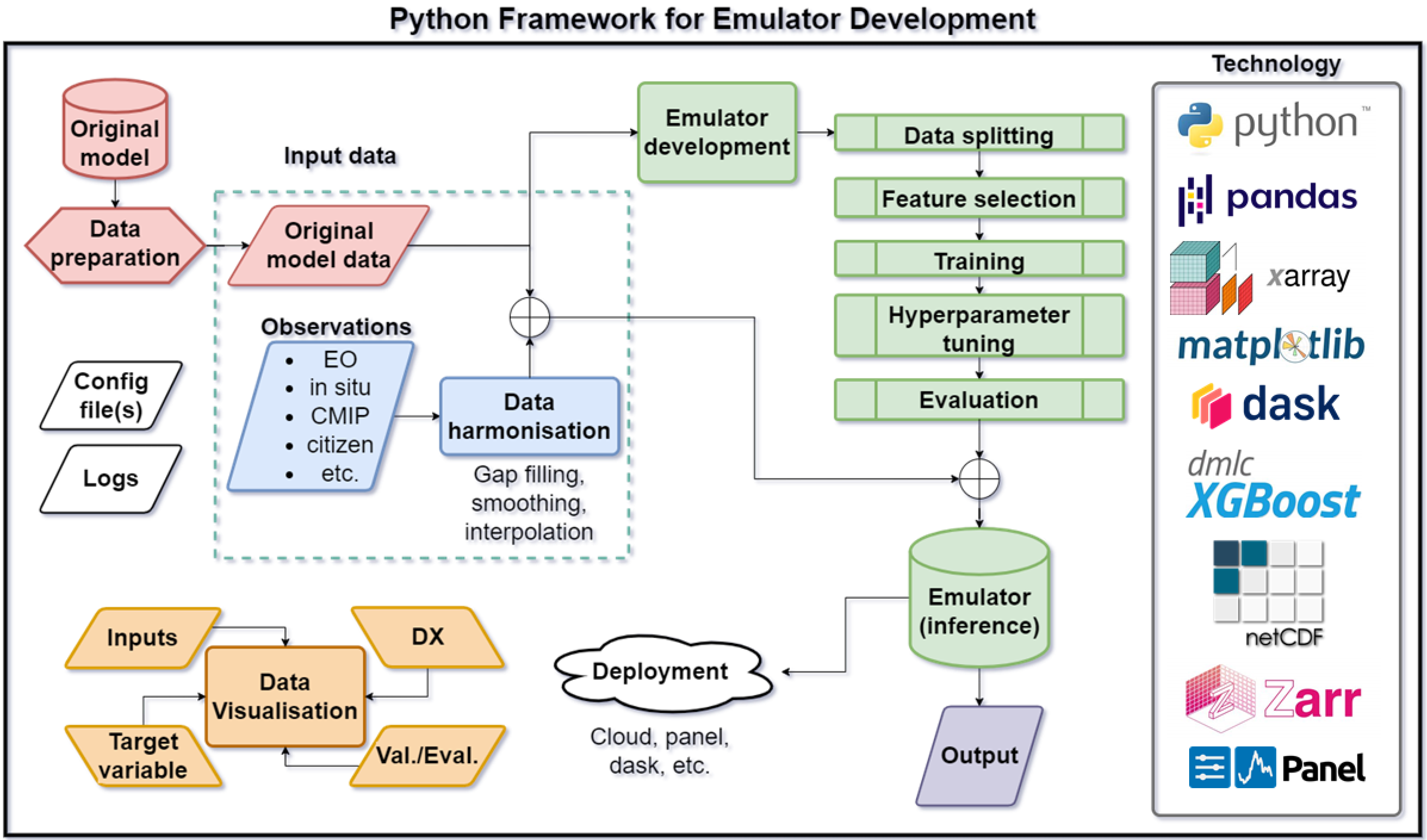
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UKRI Natural Environment Research Council

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

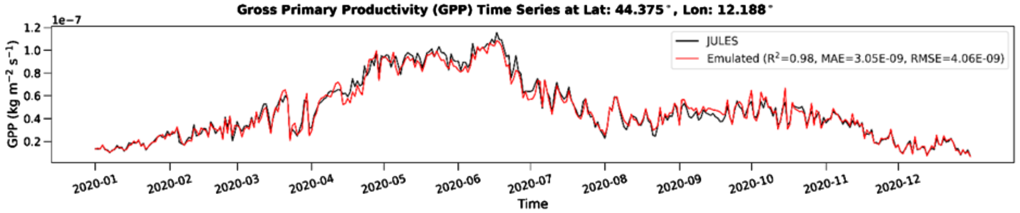


Python Framework for Emulator Development
Original model
Data preparation
Input data
Original model data
Observations
EO
in situ
CMIP
citizen
etc.
Data harmonisation
Gap filling, smoothing, interpolation
Config file(s)
Logs
Emulator development
Data splitting
Feature selection
Training
Hyperparameter tuning
Evaluation
Emulator (inference)
Output
Deployment
Cloud, panel, dask, etc.
Inputs
DX
Data Visualisation
Target variable
Val./Eval.
Technology
python
pandas
xarray
matplotlib
dask
dmlc XGBoost
netCDF
Zarr
Panel

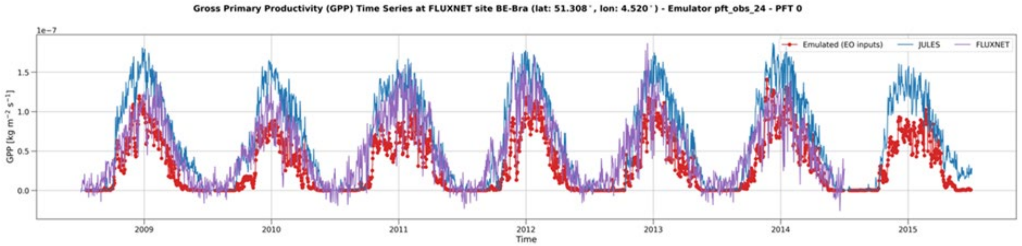
Example: ESA IMITATE Project

- We developed **machine-learning emulators** of **GPP** from **JULES** and generated a new GPP product by using **EO data** as inputs.
- Excellent performance emulating JULES.
- Good agreement with other satellite GPP products and some FLUXNET sites.

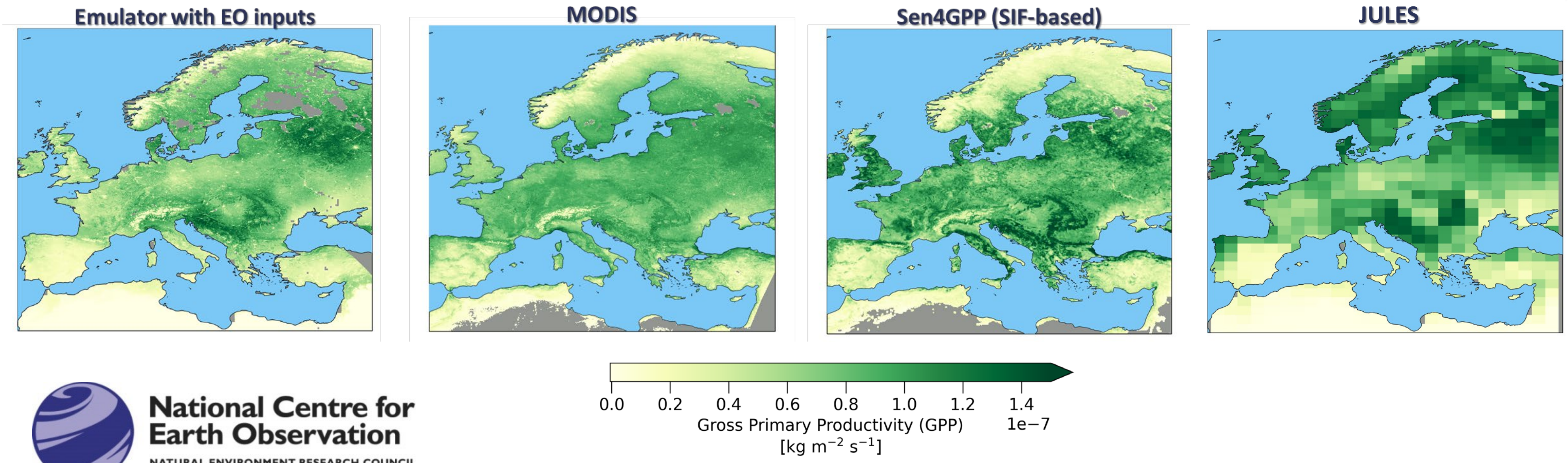
Our emulator replicates JULES really well



With EO inputs: vs JULES vs FLUXNET



GPP monthly average 06-2019



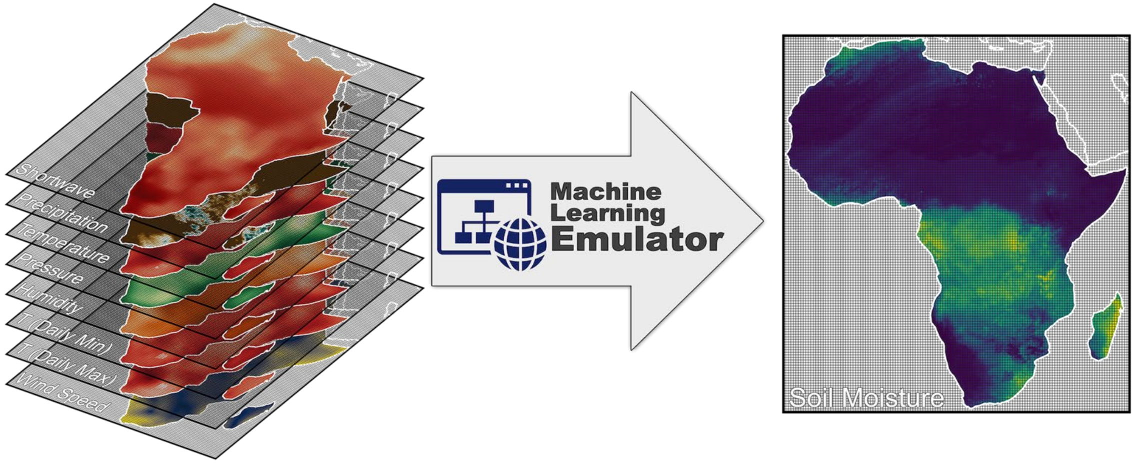
Lessons learned:
- Lots of gap-filling needed.
- Distribution of input variables needs to be similar to JULES.
- Soil moisture at deeper layers is key.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Natural Environment Research Council

Case study: soil moisture in Africa

- ✓ We are building an **emulator** for **soil moisture in Africa**, based on previous work.
- ✓ We will **optimise** and refine it using our **framework**.

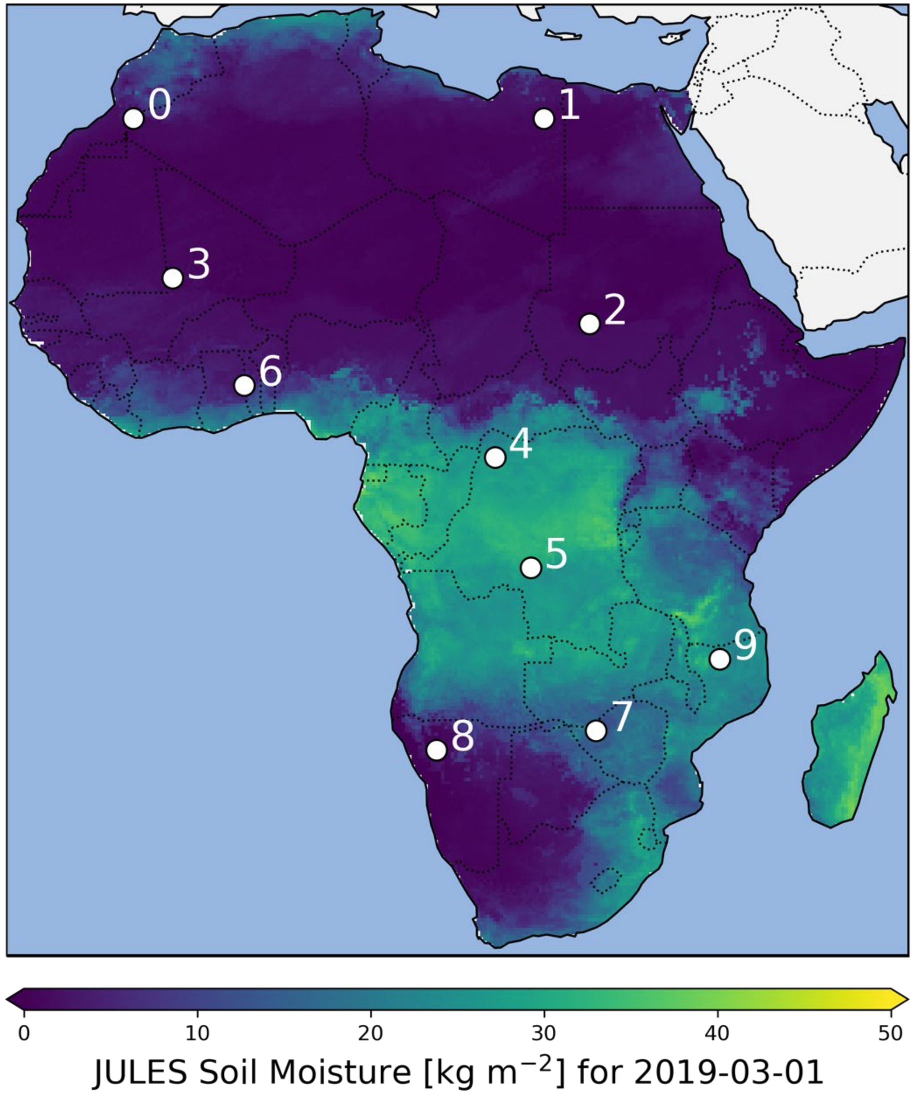


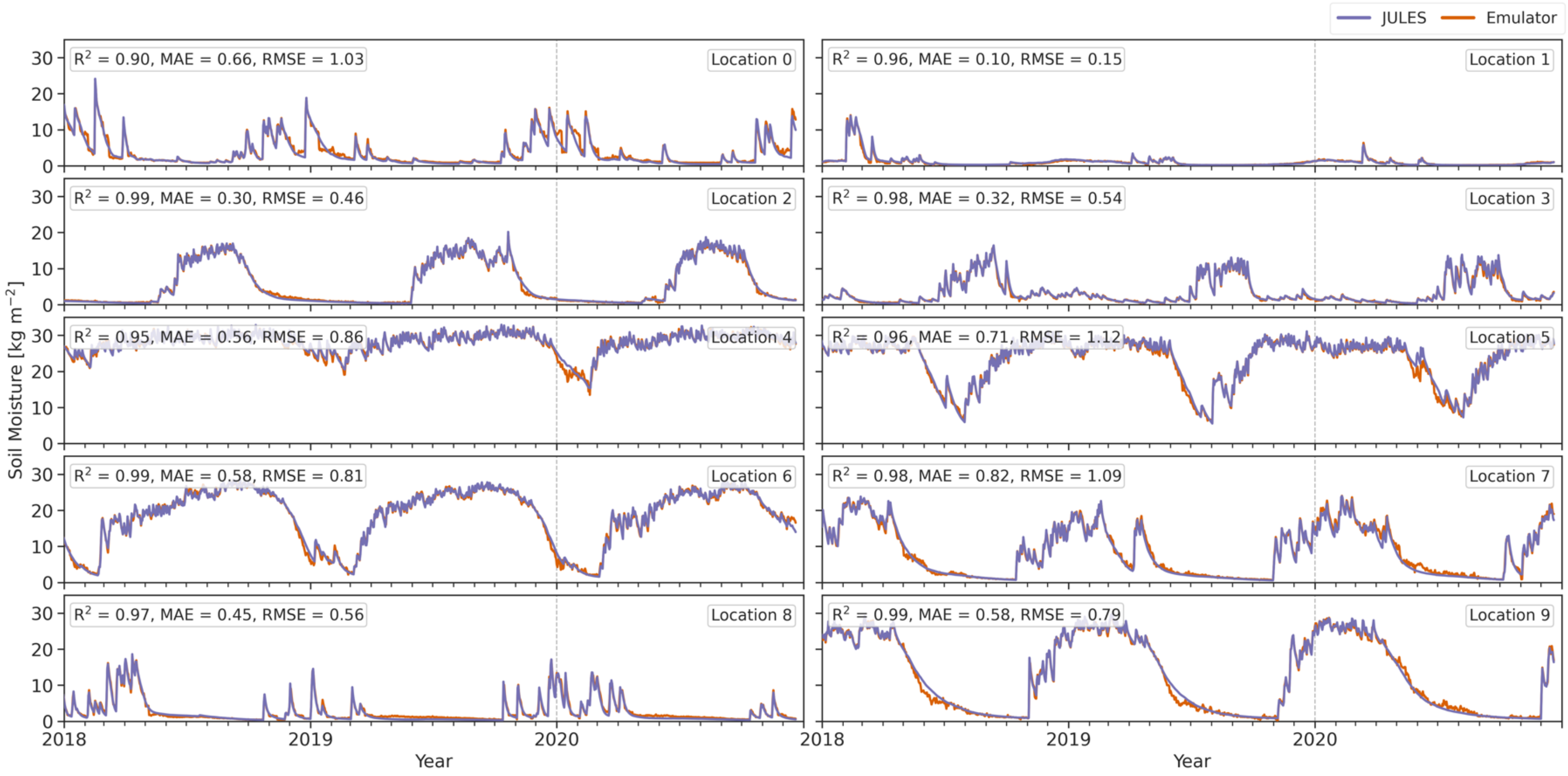
INPUT FEATURES

- Downward solar radiation
- Precipitation
- Daily mean temperature
- Surface pressure
- Specific humidity
- Daily minimum temperature
- Daily maximum temperature
- CO_2 concentration
- [Soil Hydrological variables]
- [Vegetation land cover fractions]
- Soil bulk density
- [Topographic variables]
- Wind speed
- 1 – 7 day lagged precipitation
- 1 – 7 day lagged specific humidity
- 20-day smoothed mean temperature
- 20-day smoothed specific humidity
- 20-day smoothed precipitation

Cristina Ruiz Villena | crv2@le.ac.uk | U. Leicester

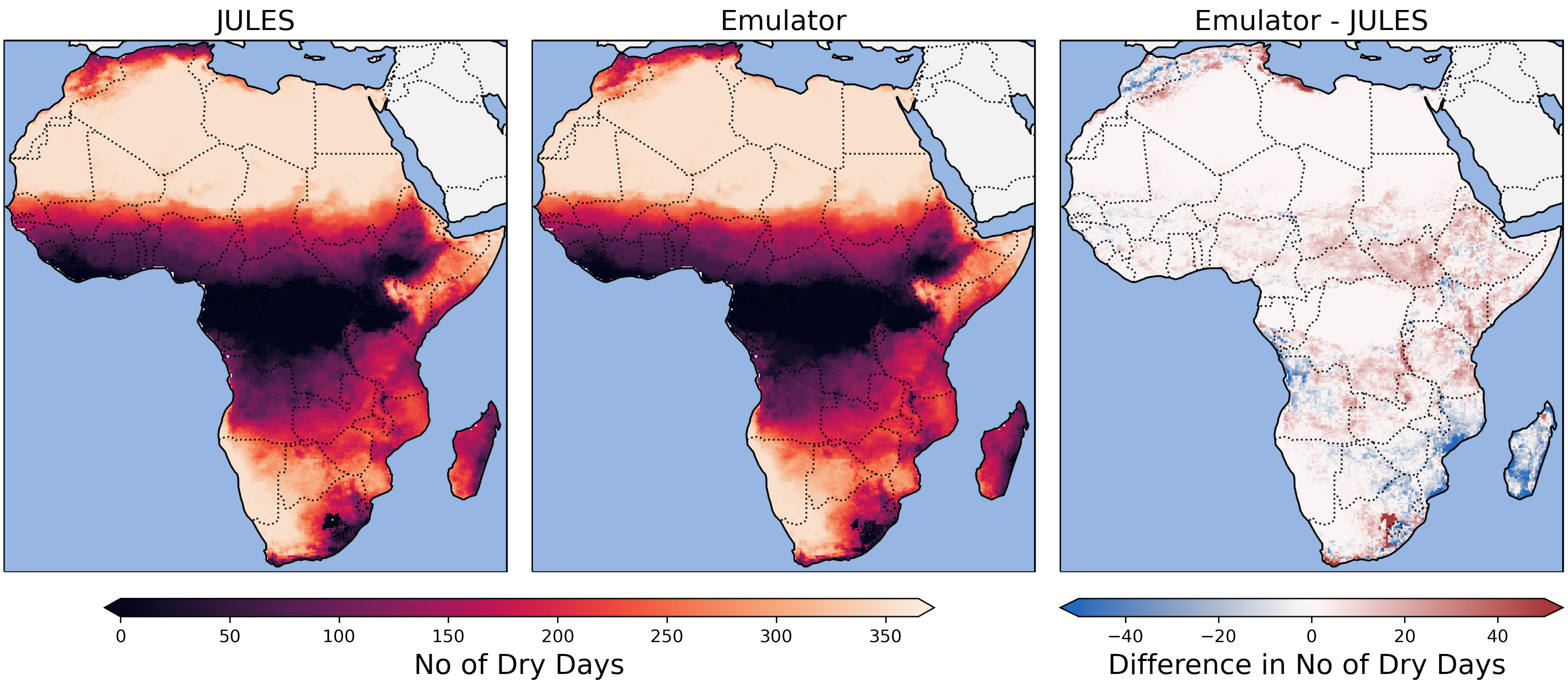
Evaluation of Emulator



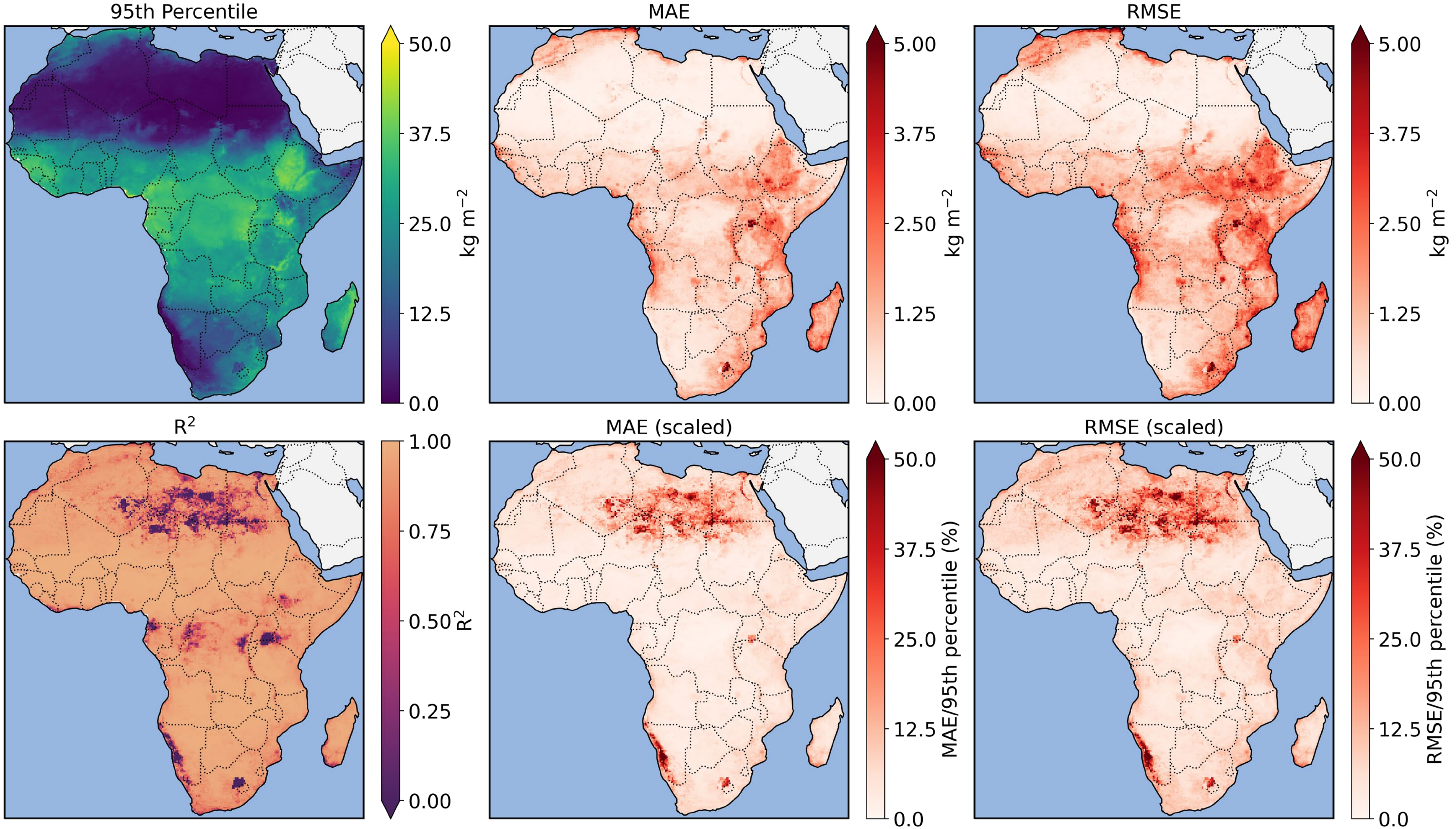


- ❑ Emulator performs exceptionally well and reproduces results of JULES model

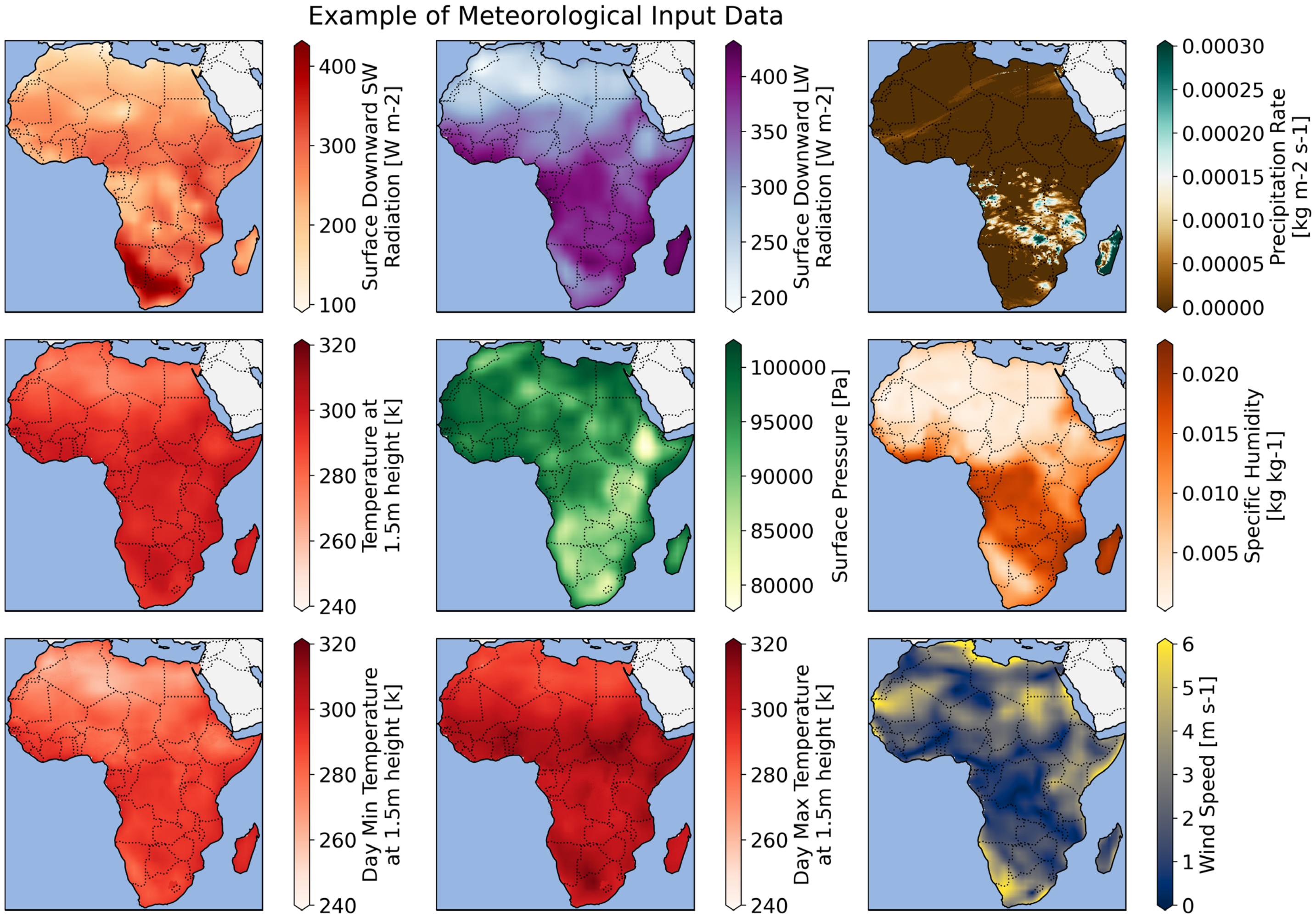
Number of Dry Days – Emulator vs JULES

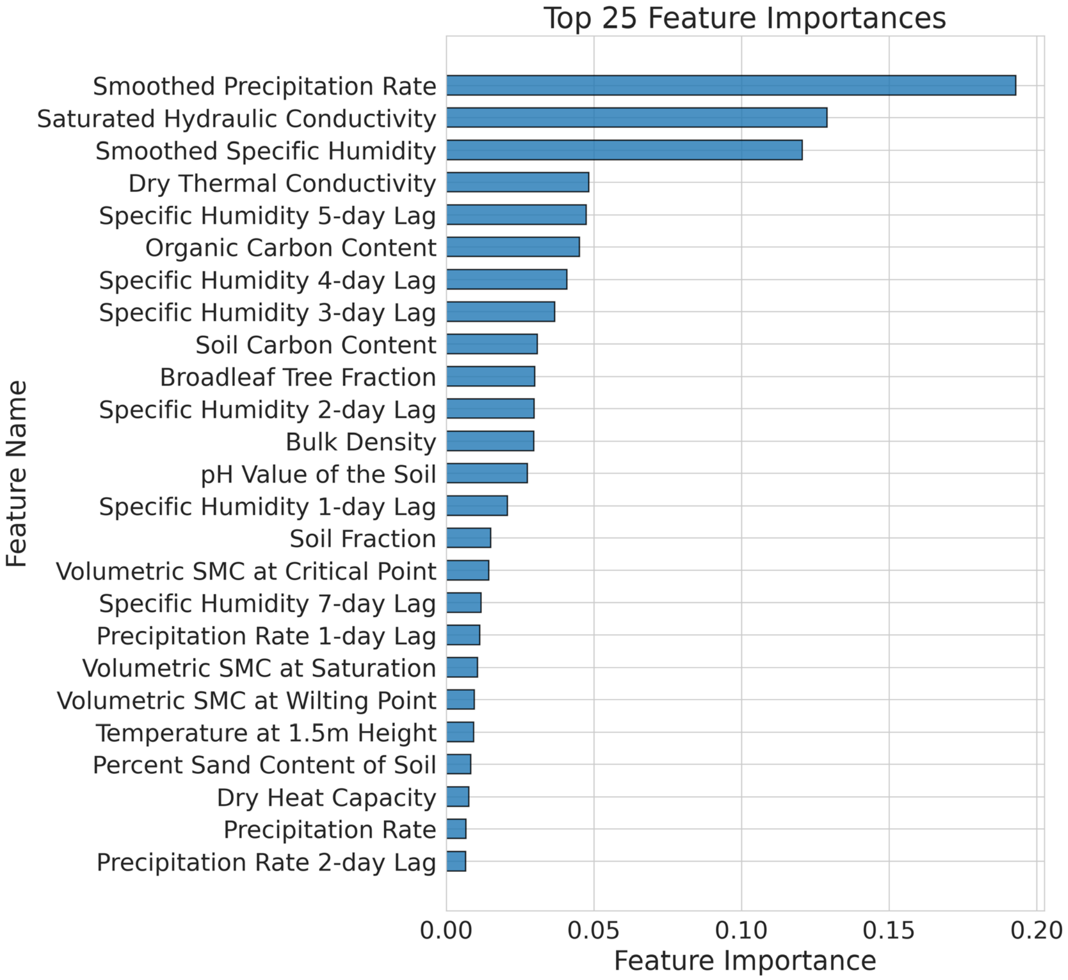


Statistics for Emulator Performance for Validation Period

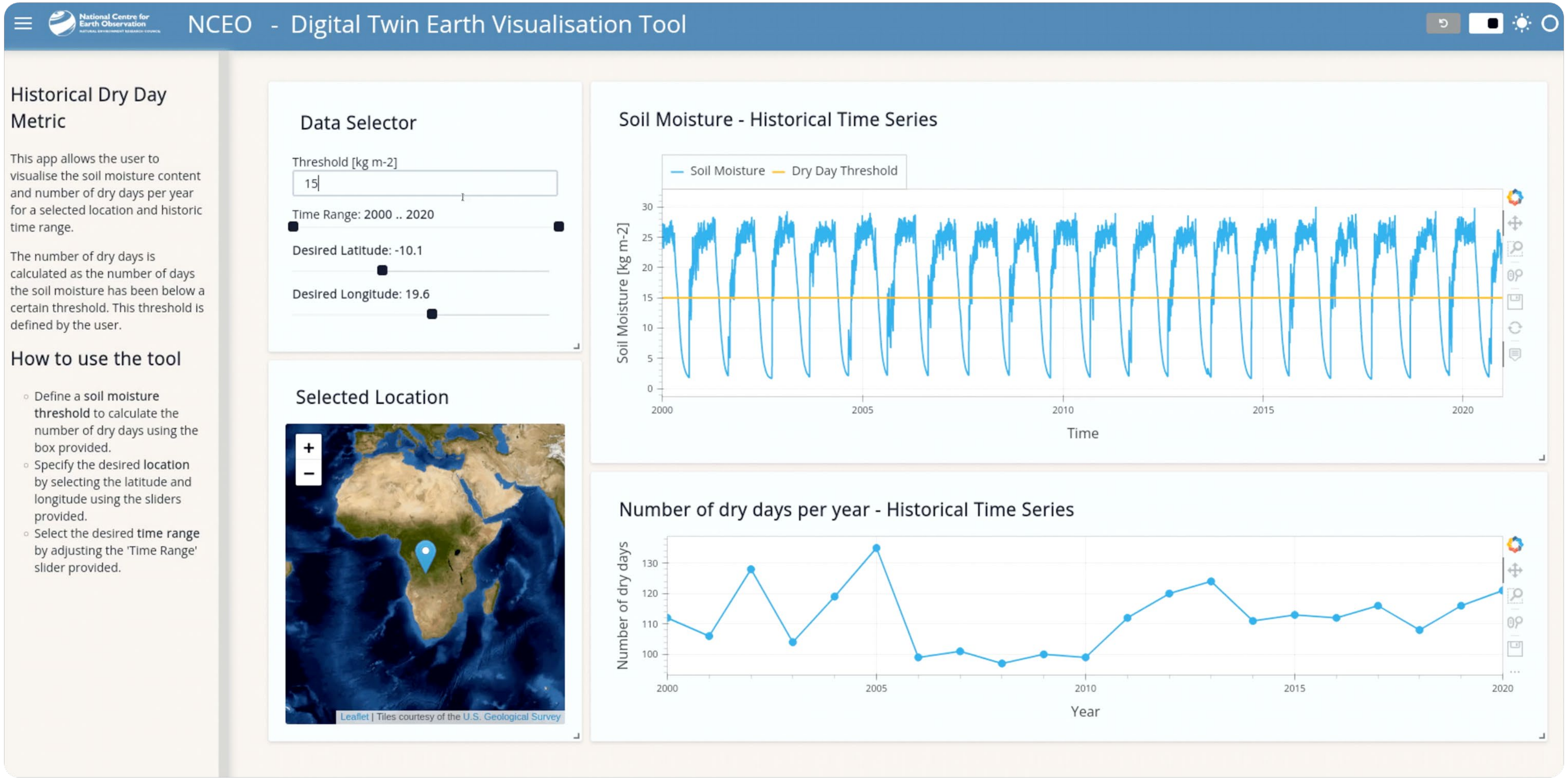


Explainability and Feature Importance





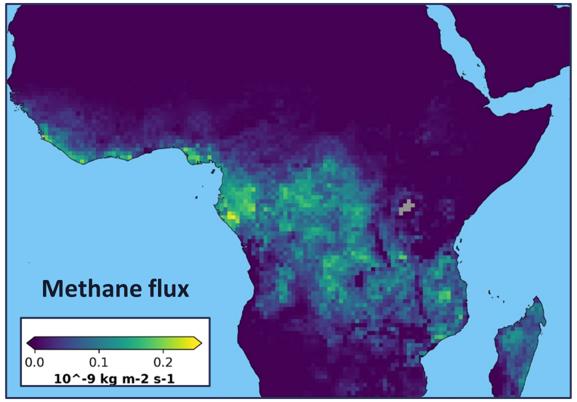
Example of how these emulators can be used

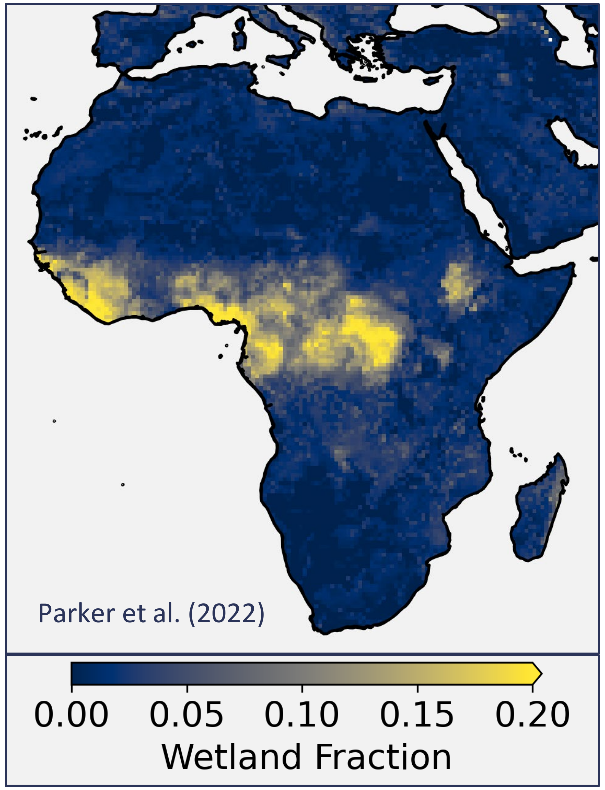


Cristina Ruiz Villena | crv2@le.ac.uk | University of Leicester

Our African Digital Twin Applications

Tropical Wetland Methane





Drought

Heatwaves

Wildfires

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Cristina Ruiz Villena | crv2@le.ac.uk | U. Leicester

UKRI Natural Environment Research Council

Summary

- **Environmental Digital Twins (DTs) are a new paradigm for decision support.**
  - Machine-learning emulators are building blocks.
- **We developed several machine-learning emulators of JULES.**
  - They are very good at emulating JULES.
  - They can be combined with EO data for model-data fusion.
- We are working on emulators for **hydrological applications in Africa** (soil moisture, wetland methane, etc.) with a lot of potential.

Thank you!
Contact:
Email:
crv2@leicester.ac.uk
Twitter:
@DrCristinaRuiz

Future Work

- We will use emulators as the **building blocks** for future **Digital Twins** (e.g. wildfire or drought).
- We will develop climate services around these tools to provide **decision support.**