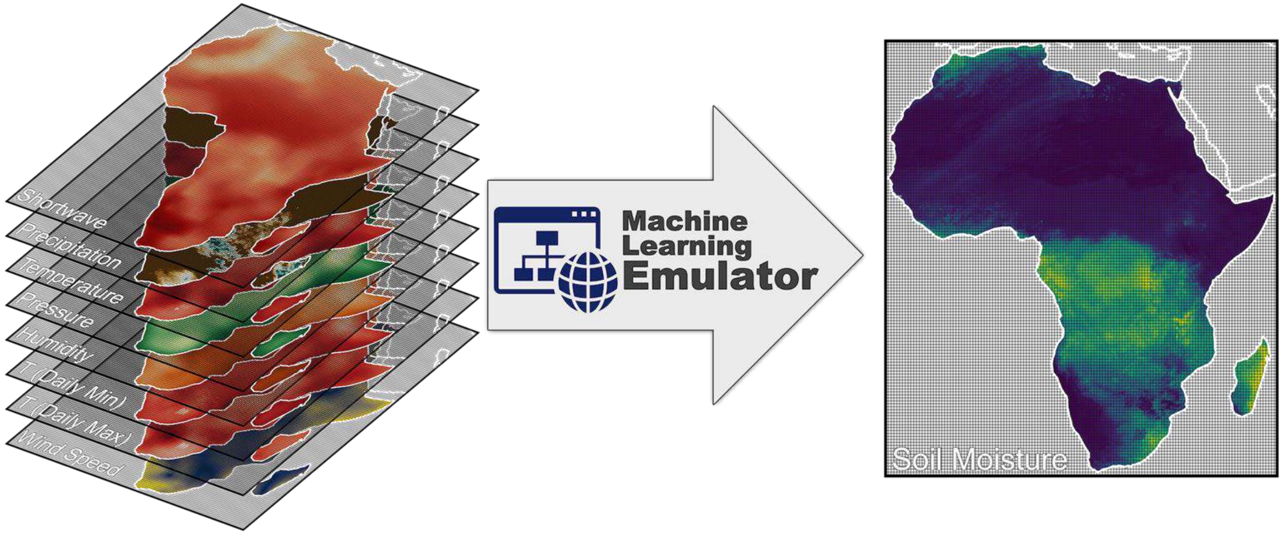
A Machine Learning emulator of JULES soil moisture: Applications to African drought

Jasdeep Anand and Rob Parker
National Centre for Earth Observation
University of Leicester

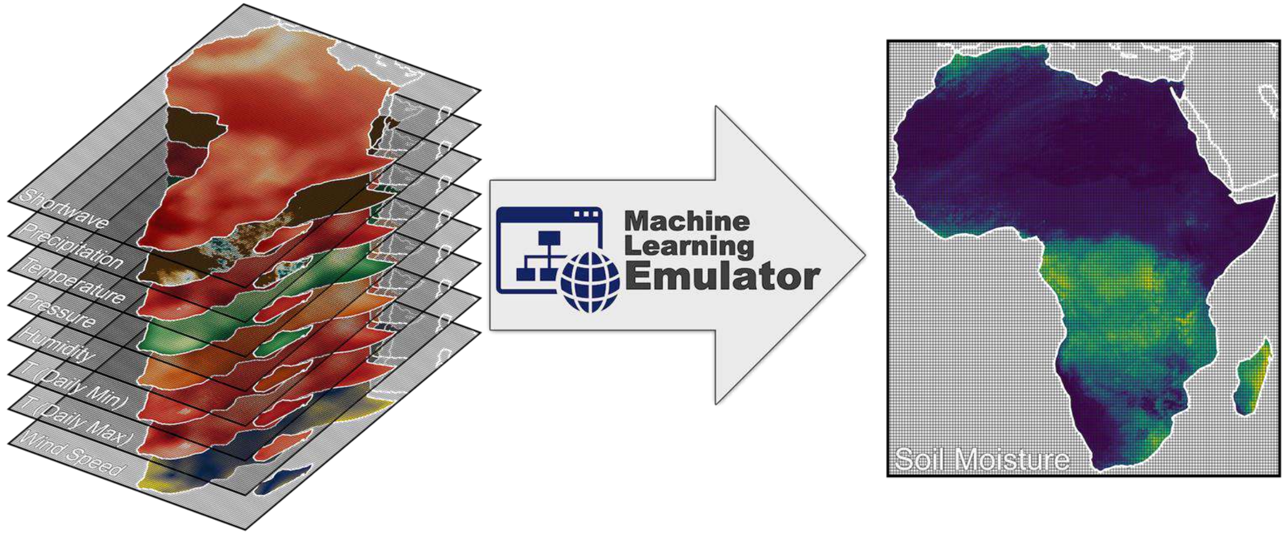
Machine Learning Emulation

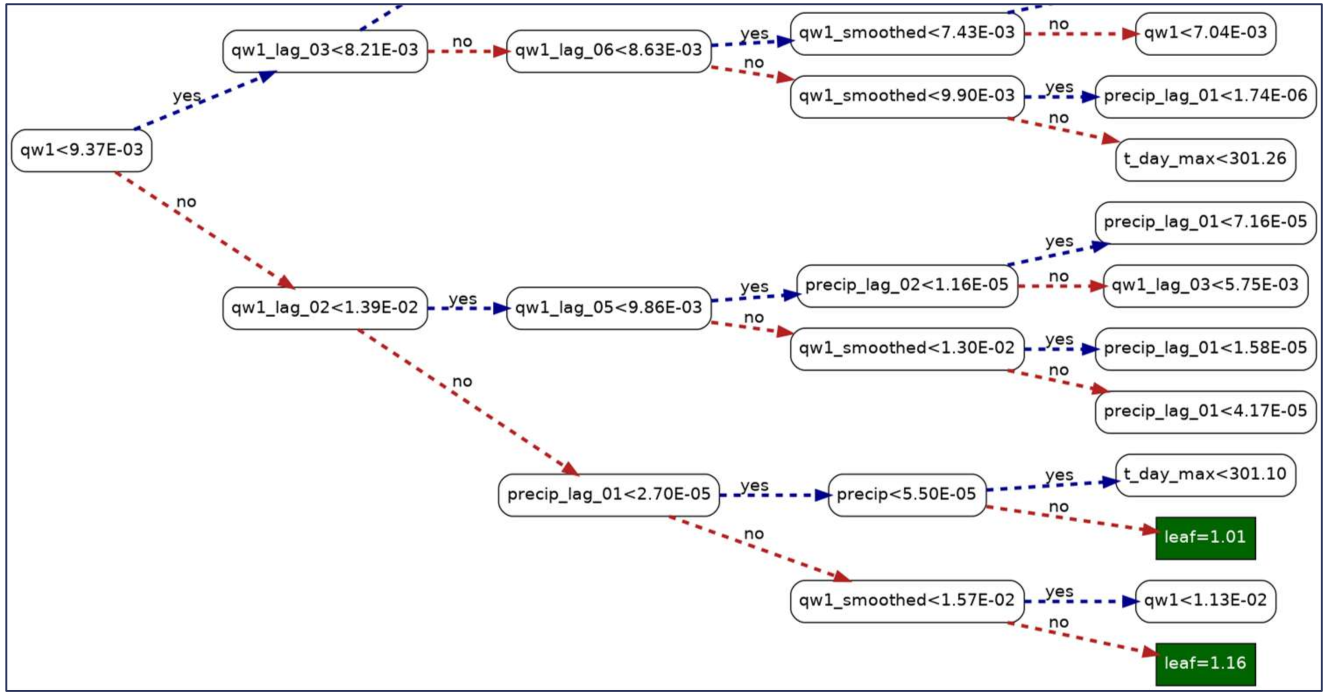
- **Key Question**: Can we emulate JULES soil moisture with a simple/fast machine-learning based algorithm?
- **Requirements and Objectives**:
  - Successfully develop emulator for African soil moisture
    - Needs to be fast and light-weight
    - Capable of running in a Jupyter Notebook and/or in the cloud
  - Assess emulator performance vs JULES simulations
  - Apply emulator to ISIMIP-driven climate scenarios to allow exploration of climate responses



Machine Learning Emulation

- We tested a **range** of machine learning methods:
  - Support Vector Machines, Penalised Regression, Decision Tree Methods
  - **All** worked (surprisingly) well but XGBoost has best performance
- The **Extreme Gradient Boosting** (XGBoost) algorithm
  - Output value of a single tree is determined through succession of **specific value tests** applied to input variables.
  - Gradient descent is used to determine which set of possible trees would **minimise the prediction error**
  - Nonlinear functions can be represented as a large **ensemble of trees** to cover a range of different scenarios.





Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Training data selection

- To train and test the emulator, JULES simulations of African soil moisture were made between **1999-2020** using **TAMSAT (precipitation)** and **CRU-NCEP (temperature)** as input data

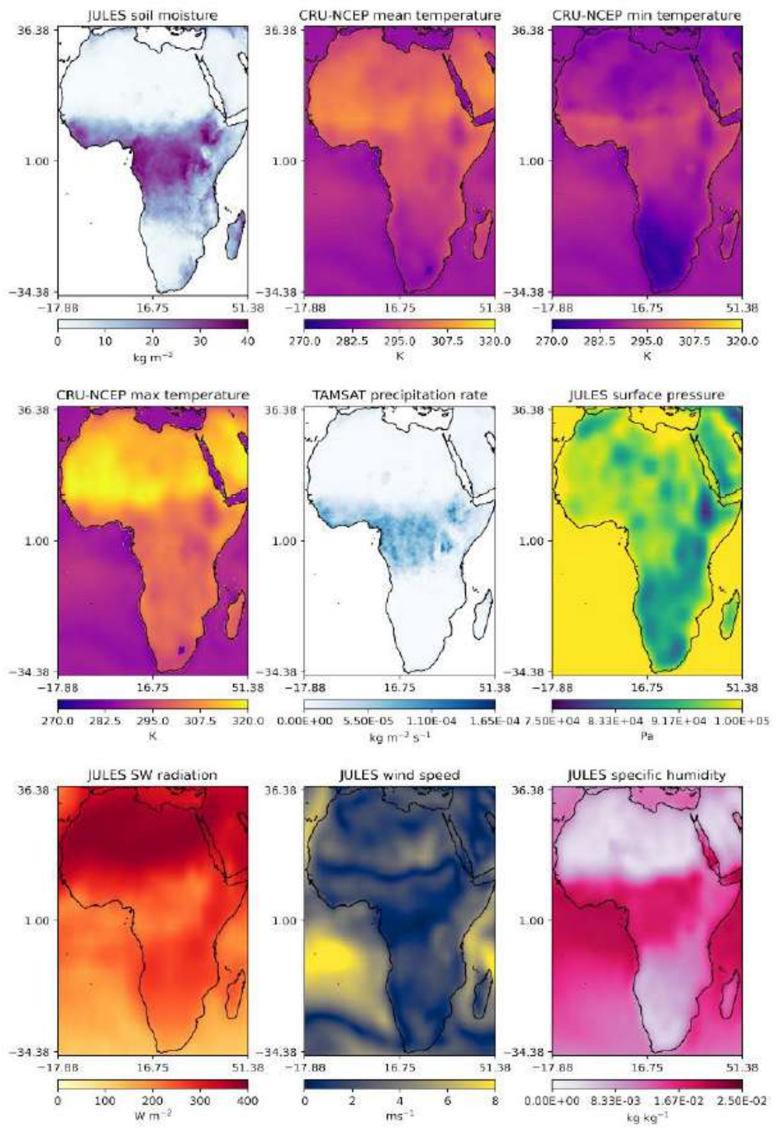
- **2000-2019** used as **training** data, while **2020** reserved for **validation** only

- Initially, **26 variables** were used as training data, taken from the following daily mean variables from the JULES output:

  - Temperature (also daily minimum & maximum)
  - Downward shortwave solar radiation
  - Precipitation
  - Specific humidity
  - Surface pressure
  - Wind speed

- To introduce memory of soil moisture dependence on seasonality and prior weather conditions, we also included:

  - 20-day smoothed temperature, specific humidity, and precipitation
  - 1-7 day lagged specific humidity and precipitation



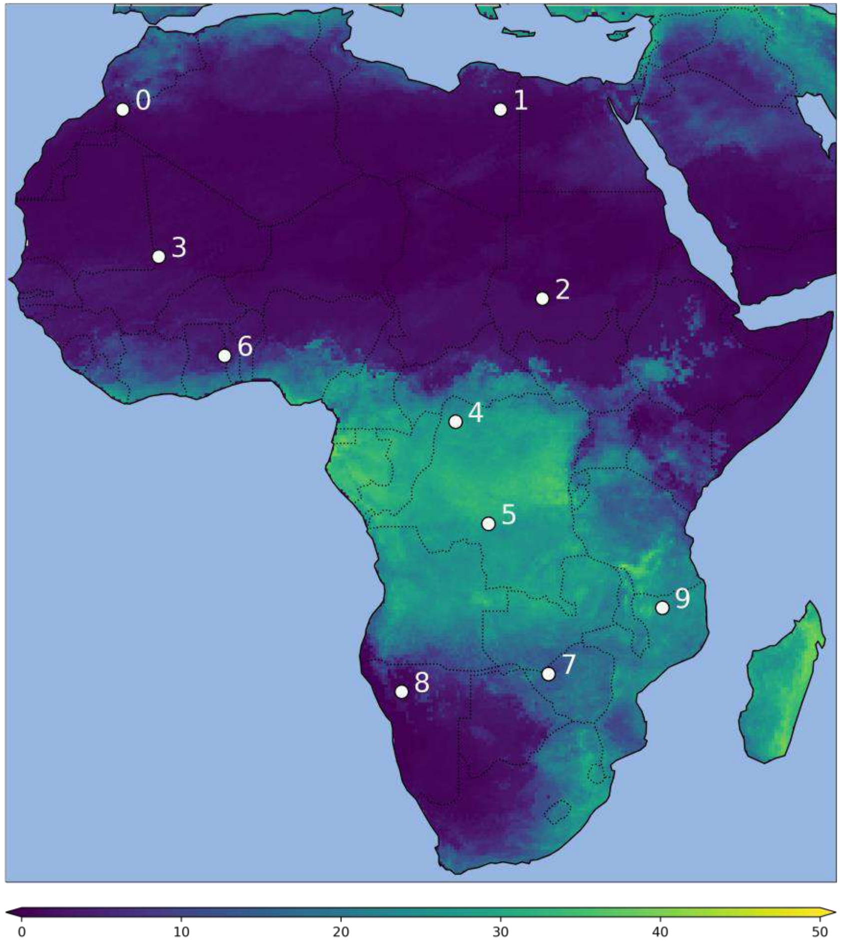
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

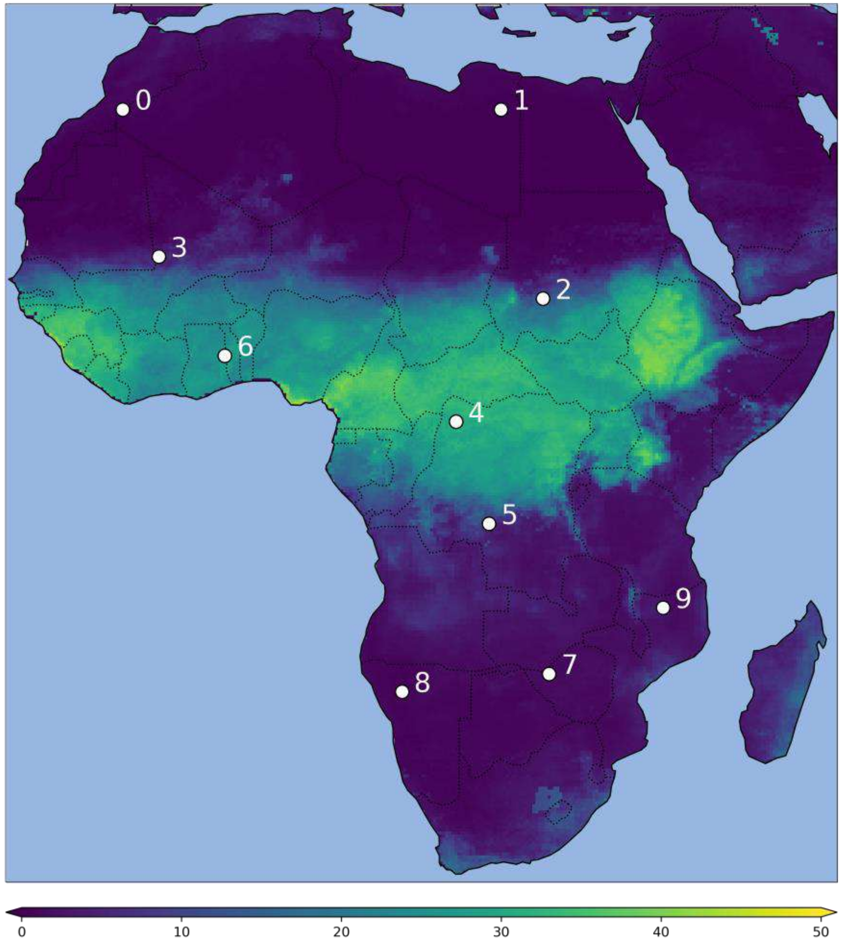
UKRI Natural Environment Research Council

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Training data needs to be representative of all possible African climates – sample data from 10 grid cells



JULES Soil Moisture [kg/m^2] for 2019-03-01



JULES Soil Moisture [kg/m^2] for 2019-08-01

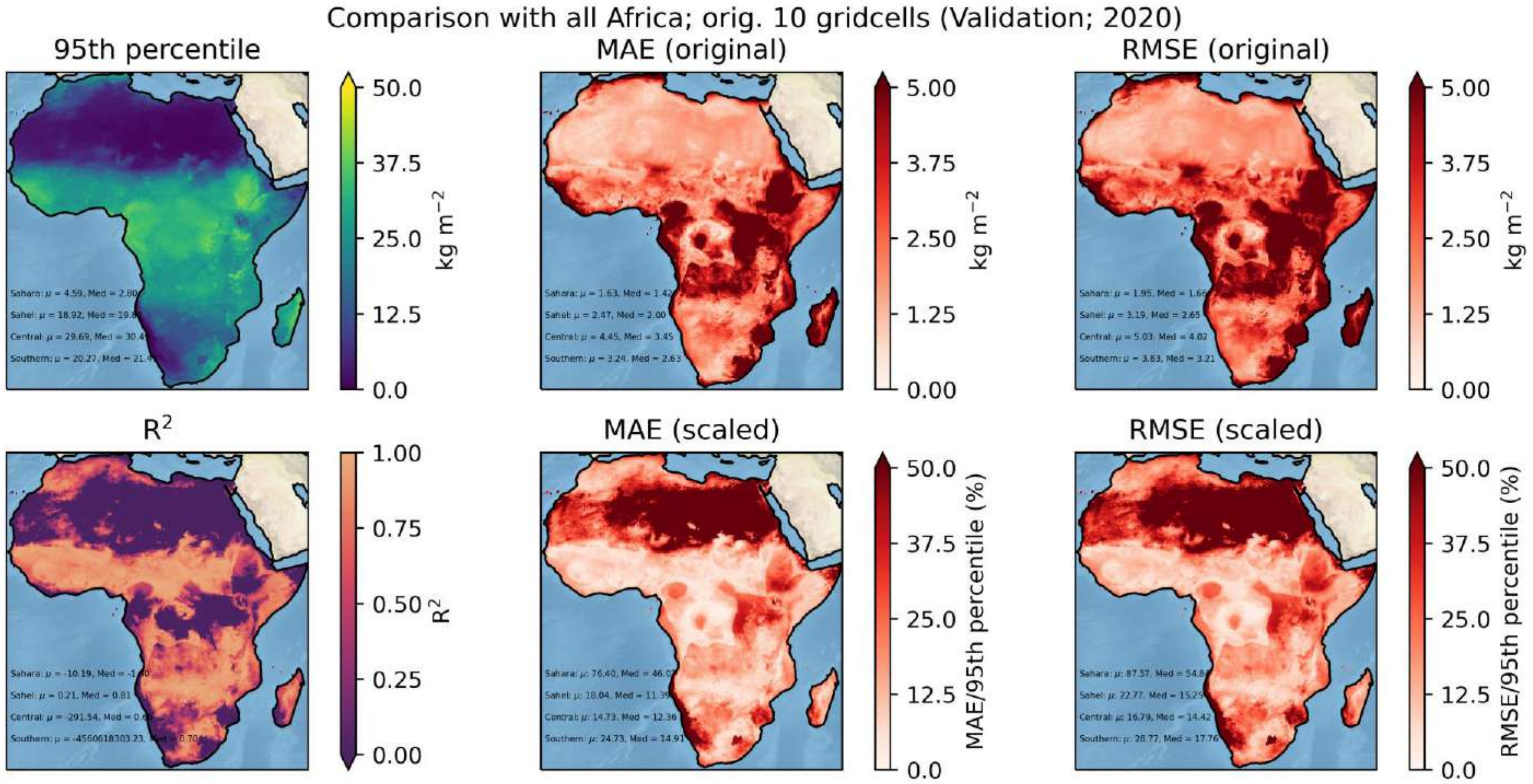
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

UKRI Natural Environment Research Council

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

*Initial results trained with **only 10 grid points** demonstrated good overall agreement, but poor performance over the Sahara and Southern Africa*



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

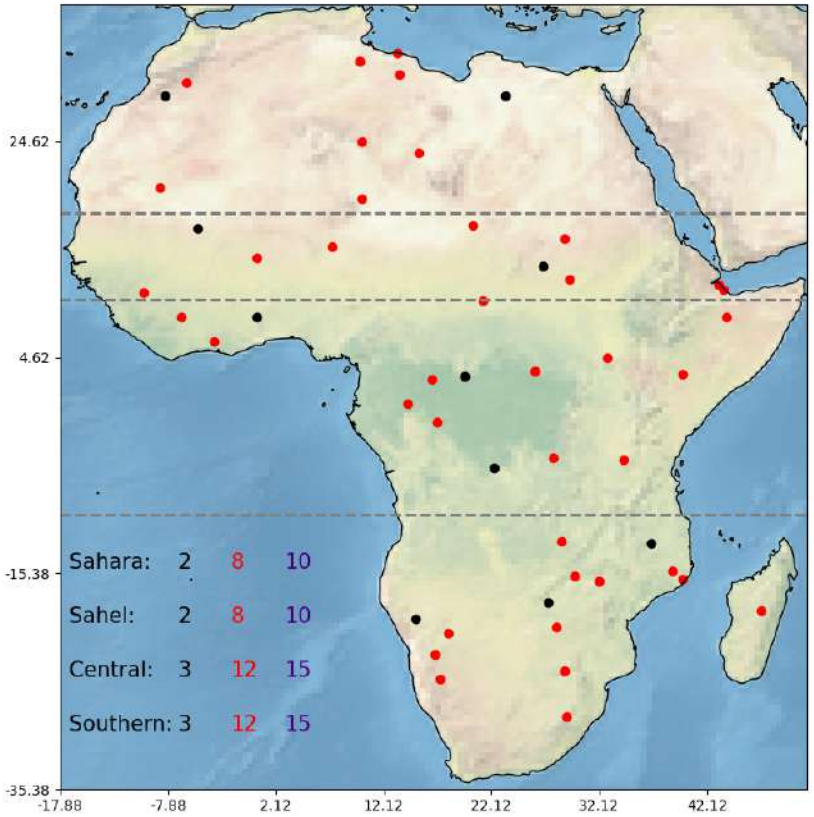
Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

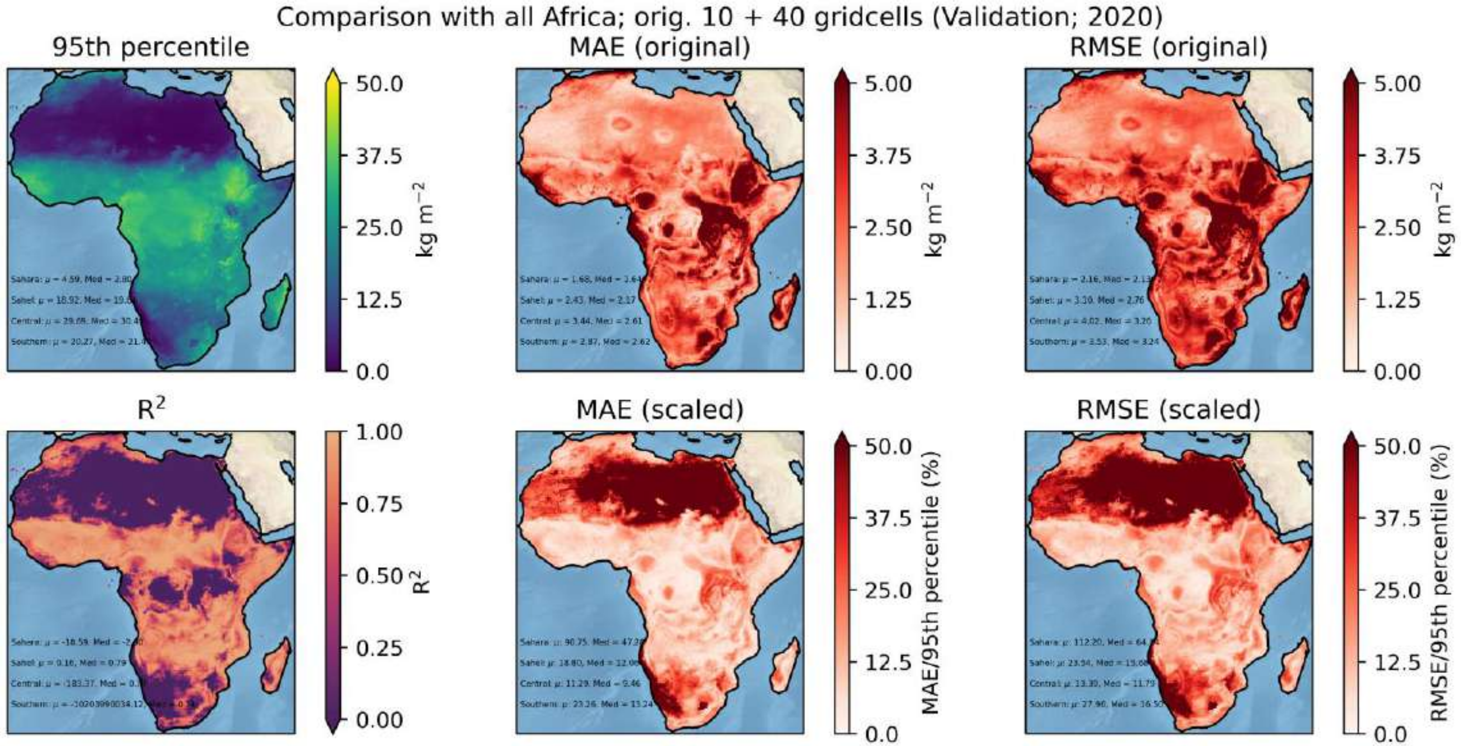
UKRI Natural Environment Research Council

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Adding more grid cells to training dataset helps to improve emulator performance…

Tested 10, 50, 200 and 1000 grid cells in training dataset



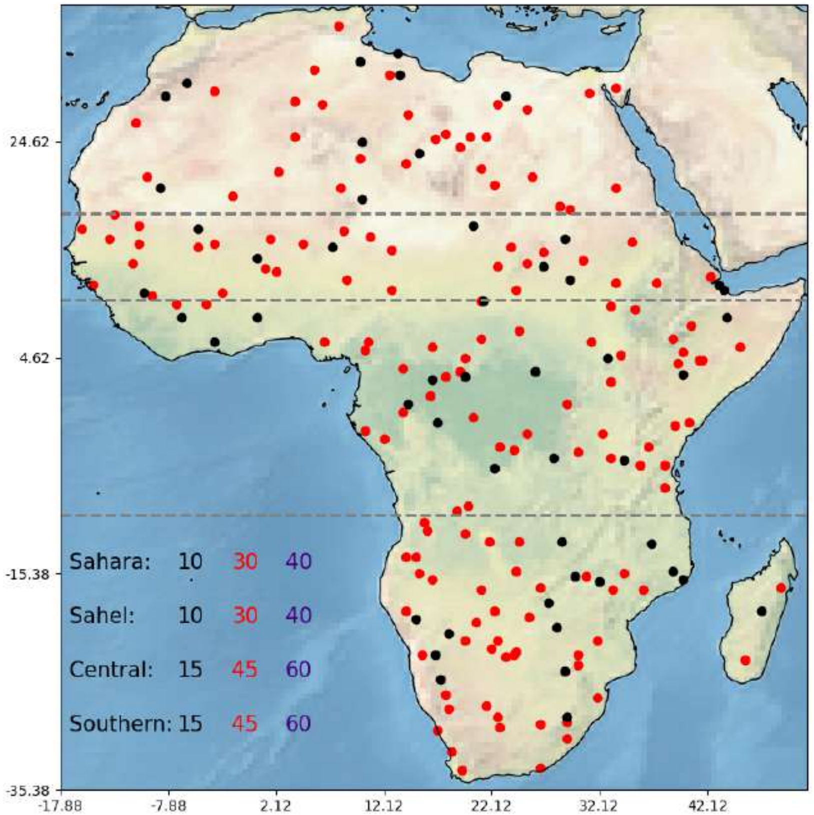


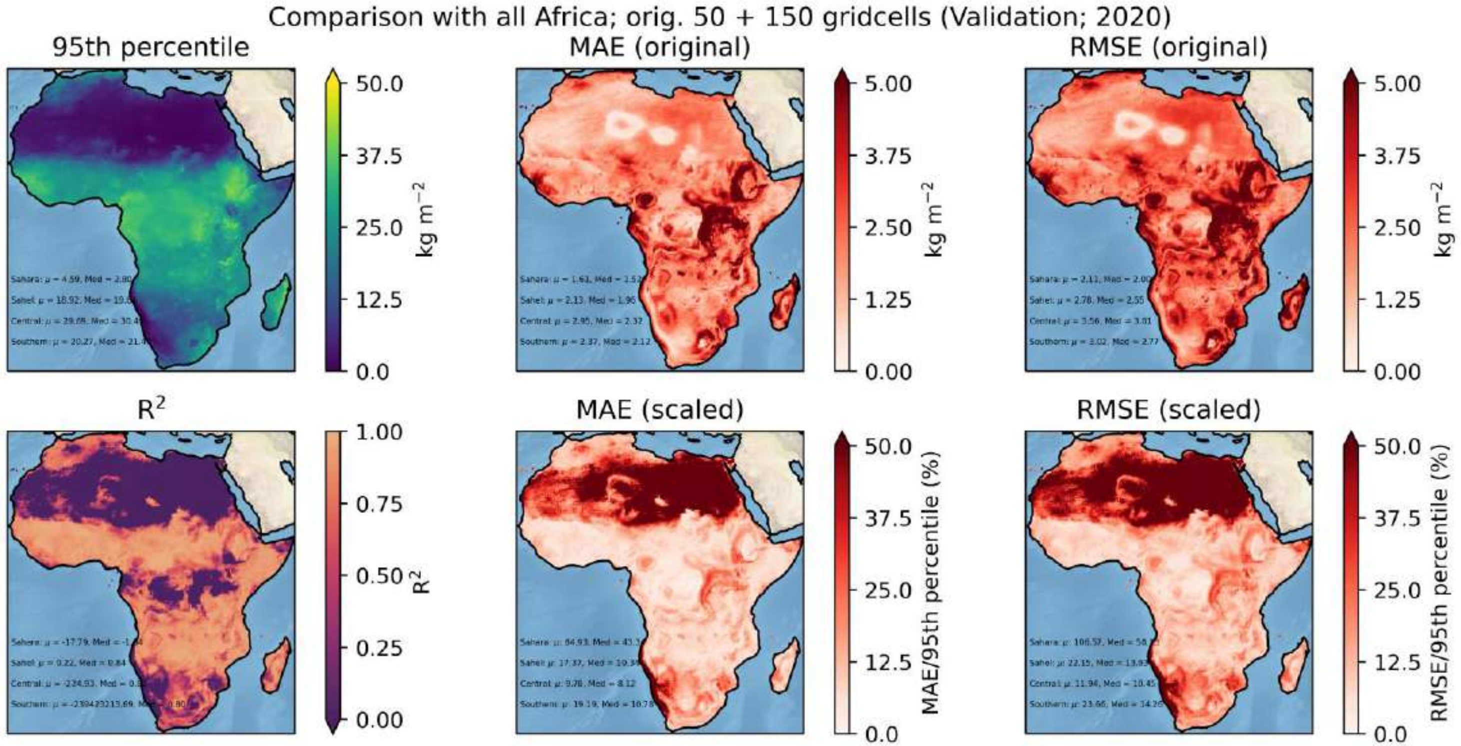
50 training grid cells

Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Adding more grid cells helps to improve emulator performance… to an extent



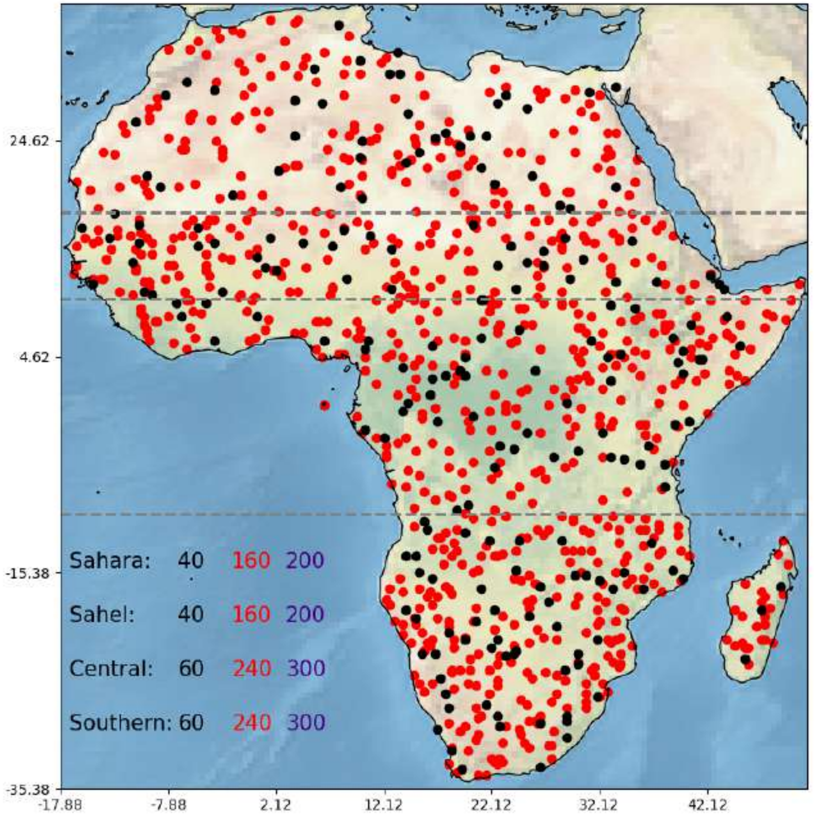


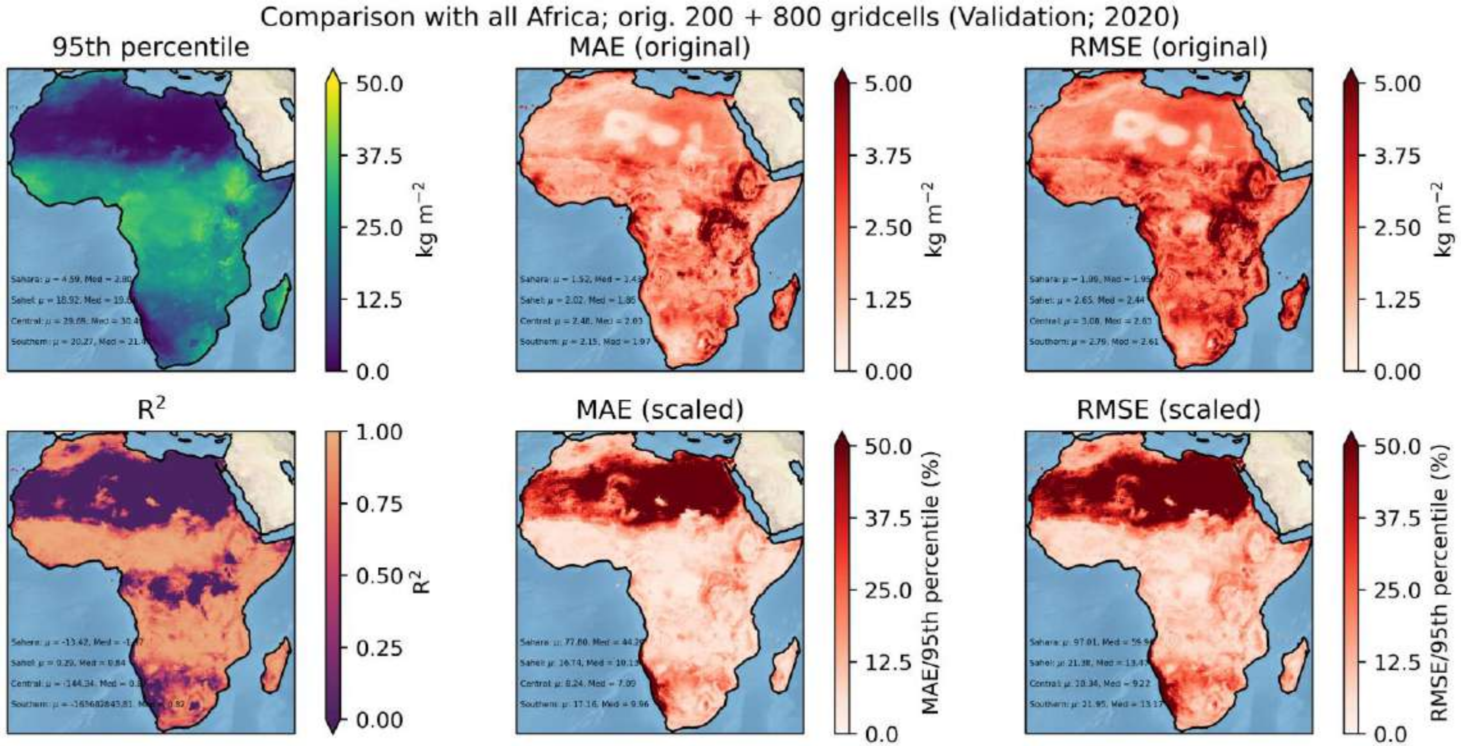
200 training grid cells

Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Adding more grid cells helps to improve model performance… to an extent





1000 training grid cells

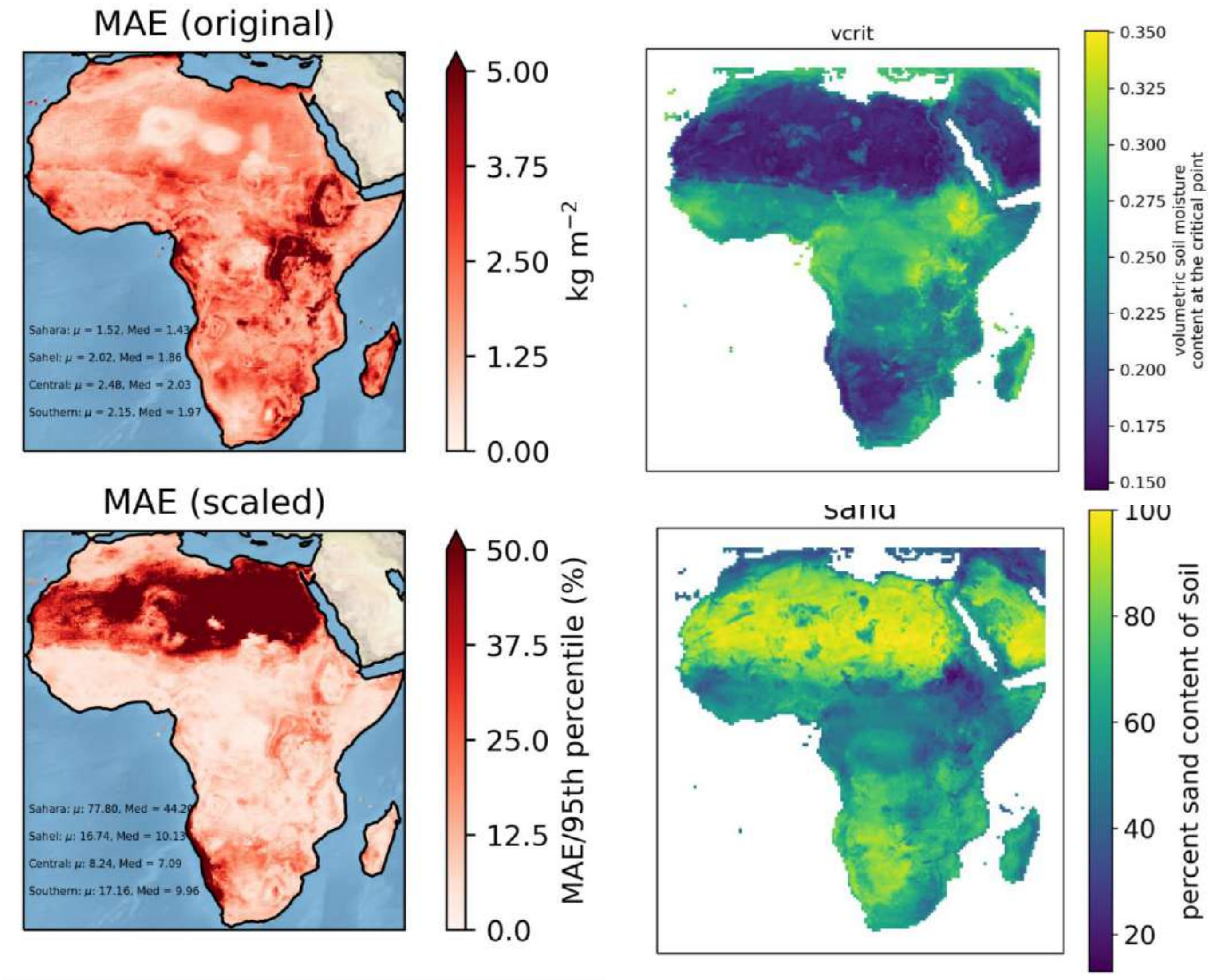
Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Update training dataset with ancillary information

- Despite improved accuracy, the overall model performance is still hampered by poor agreement over the Sahara and Namibian deserts, and savannah in Southern Africa – **Does the emulator need contextual information too?**

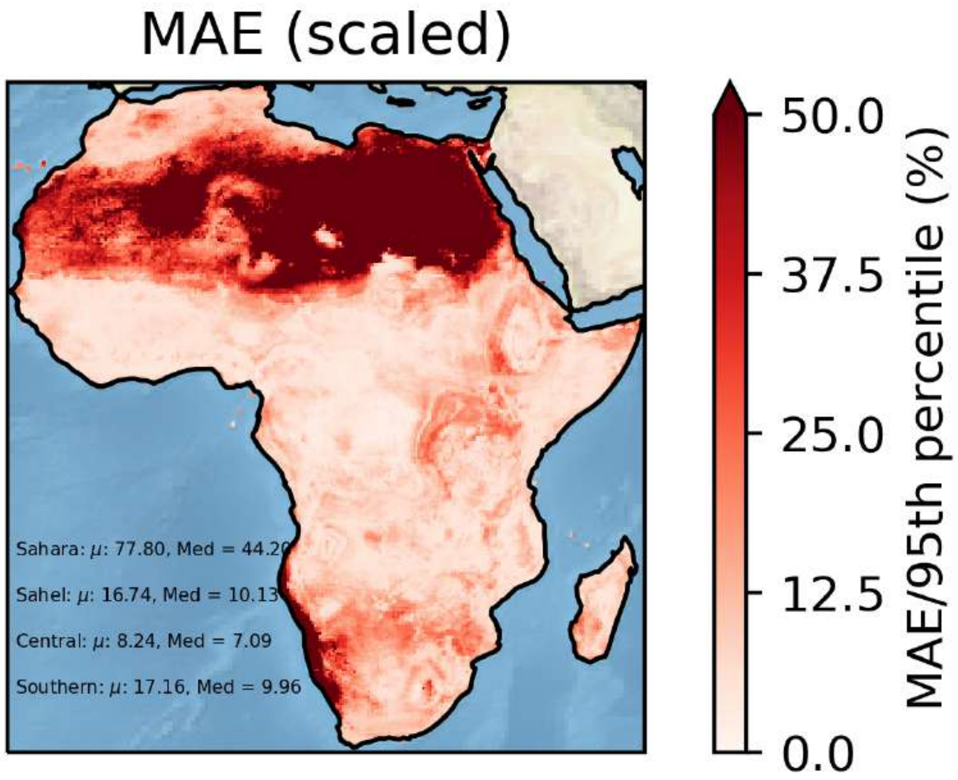
- Our biases have **similar patterns** to some of the ancillary data...

- Solution: Update training dataset with time-invariant JULES **ancillary variables** for each grid cell:
  - Soil type coverage
  - Soil bulk density
  - Vegetation type coverage
  - Soil carbon storage
  - Topography
  - Soil hydrological information

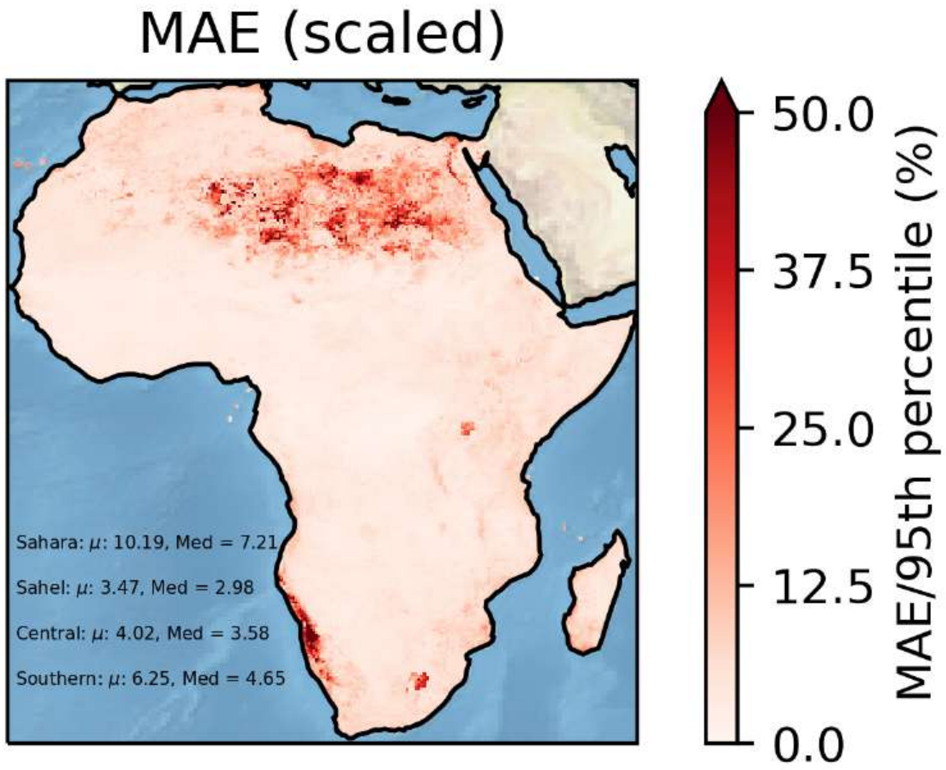


Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

Adding ancillary information to training dataset greatly improves emulator performance



No ancillary data



With ancillary data

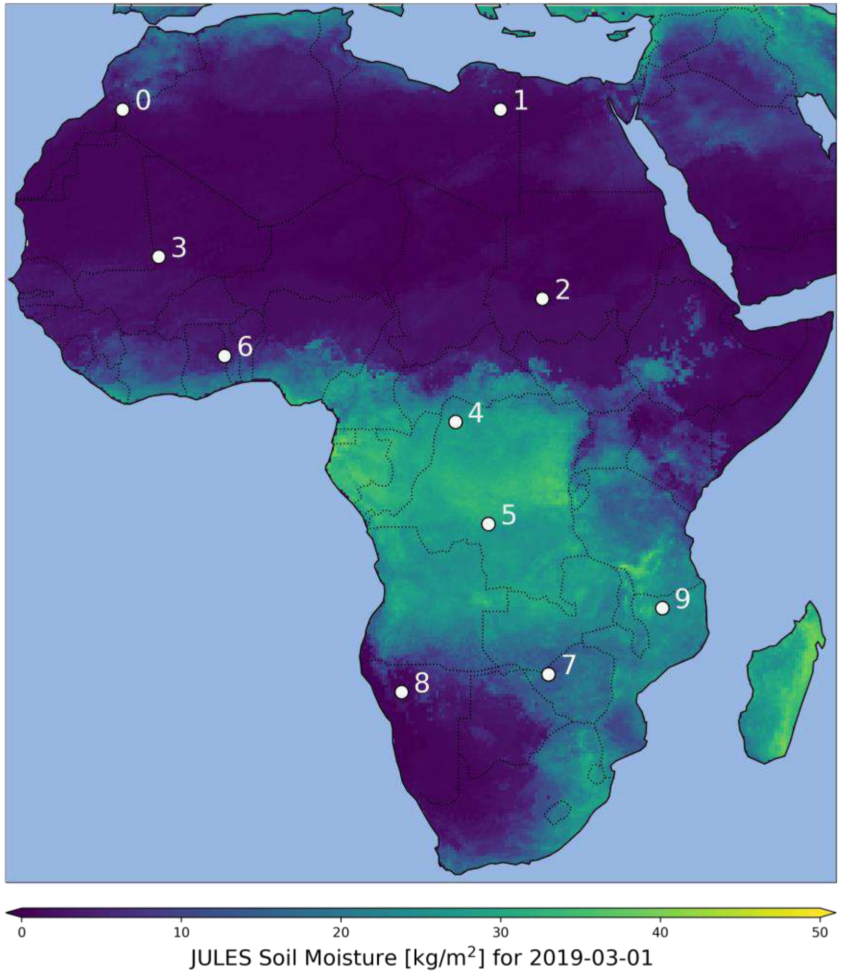
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

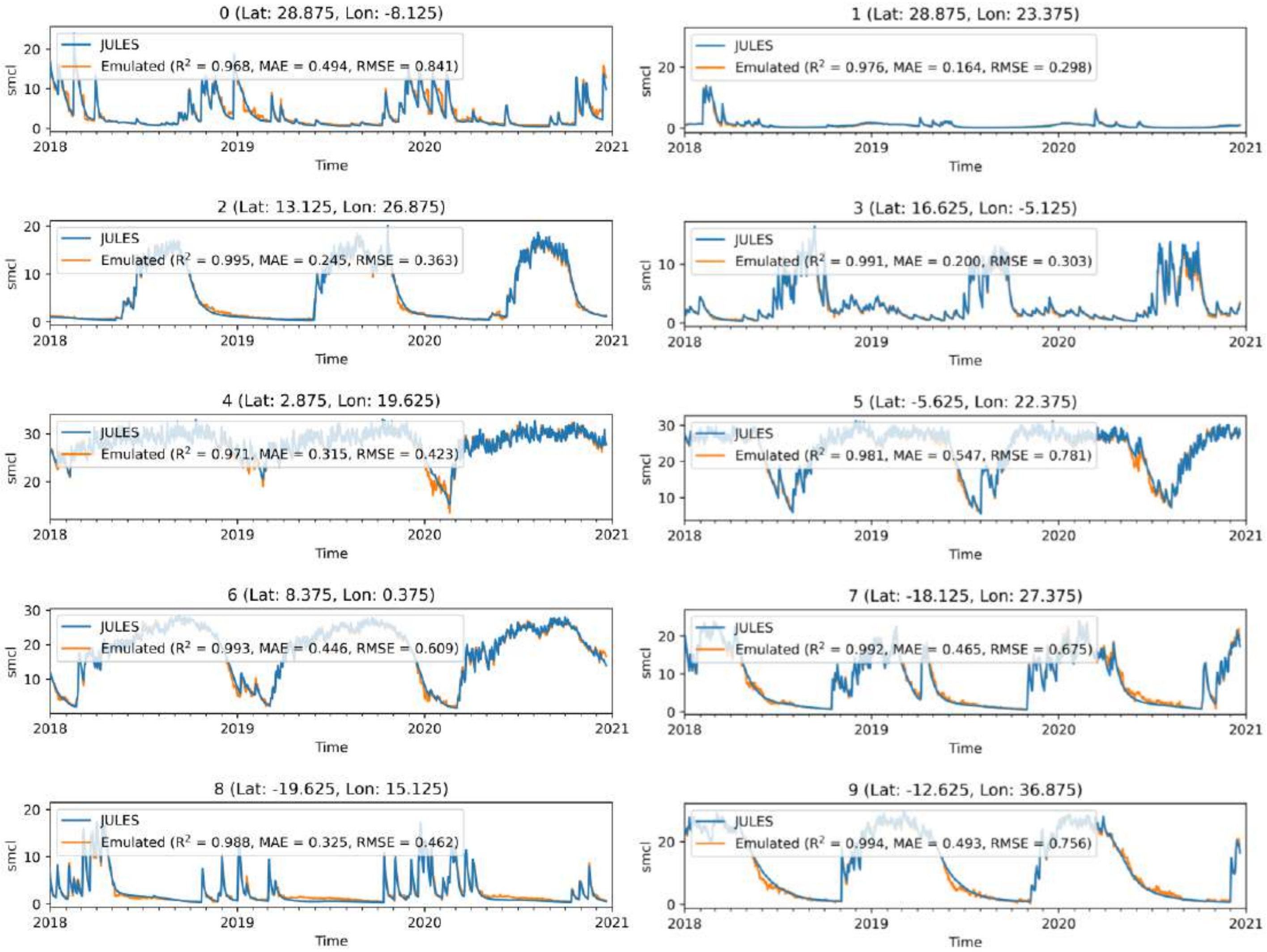
Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

UKRI Natural Environment Research Council

Emulating JULES simulations driven by TAMSAT/CRU-NCEP data

The emulator also shows excellent temporal agreement with the original JULES soil moisture

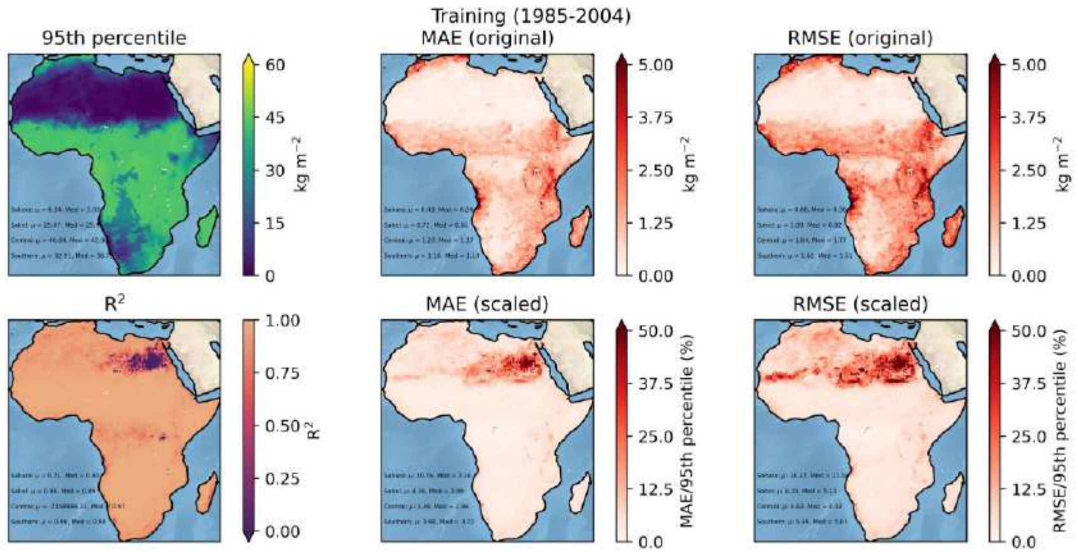




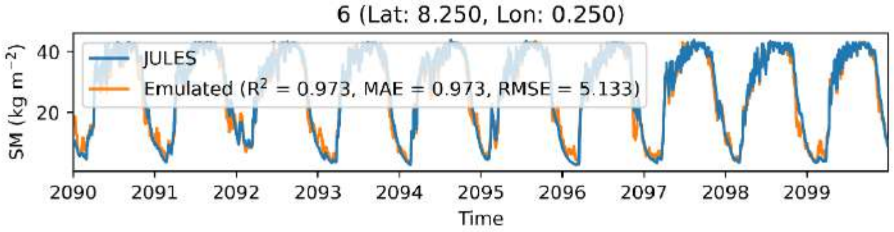
Emulating JULES forecasts using ISIMIP climate scenario data

Emulator retrained using ISIMIP historical data (1985-2005), then applied to predict RCP scenarios from 2006-2099

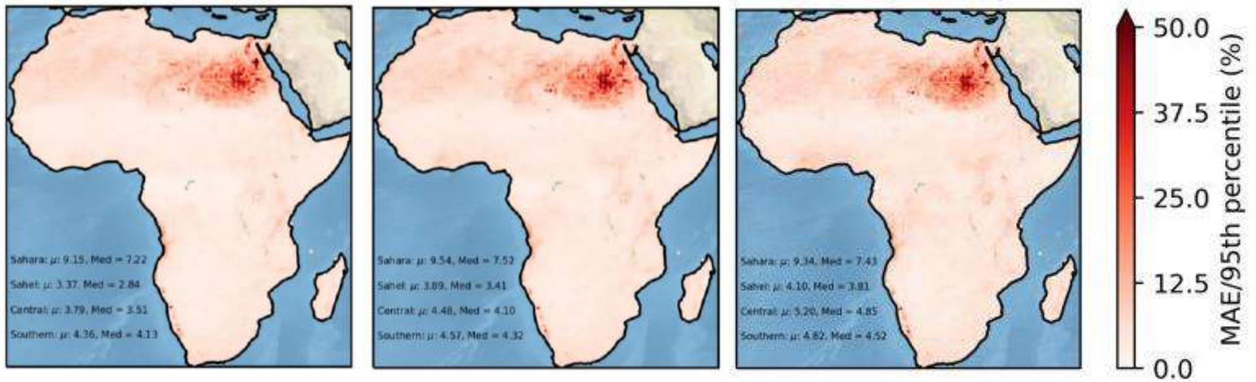
- ❑ JULES ISIMIP setup different than historical setup (resolution, ancils, etc) so model retrained, including [CO_2] as a feature
- ❑ Emulator used to predict soil moisture for ISIMIP-based RCP **climate scenarios**: RCP2.6 (< 2°C rise), RCP6.0 and RCP8.5 (BAU)
- ❑ For all of Africa, the **error is typically < a few %** after training on 1000 locations (~2% of total grid cells)
- ❑ Emulator reproduces 2006–2099 period simulated by JULES **very well**



HADGEM2-ES, RCP2.6, 2090-2099



MAE from emulating ISIMIP simulations



RCP 2.6 **RCP 6.0** **RCP 8.5**

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Jasdeep Anand | jsa13@le.ac.uk | University of Leicester

Natural Environment Research Council

Summary and Conclusions

- **Successfully** developed machine-learning based emulator for JULES soil moisture
- Initial testing of various methods **all** gave good results but settled on XGBoost algorithm
- Emulator is **VERY** fast/light-weight (runs in a few ms per grid cell per year)
- Performance metrics (MAE and RMSE) used to assess performance over whole of Africa
- Emulator performance improved iteratively by:
  - Adding in more training **locations** (better sampling of feature-space)
  - Adding in more training **features** (e.g. soil ancil data)
- Final model performs **extremely well**
  - Error **typically < a few %** compared to original JULES simulation
- Model retrained on JULES ISIMIP historical simulations
  - Required as ISIMIP suite has different settings (different ancils, different resolution, etc)
  - Performance of ISIMIP simulations **very good** across multiple RCPs
  - Emulator can be used to **quickly explore** JULES response to different climate scenarios
- Lots of potential for expanding
  - Extend to other JULES processes
    - Rob/Tristan have proposal under consideration to do this for GPP
  - Extend to other land surface models and cross-compare
  - Do some interesting science with it!

Rob Parker | rjp23@le.ac.uk | University of Leicester