

# JLMP and Evaluation Tools

Chris Jones and Tris Quaife

# JULES Benchmarking and Evaluation

Geosci. Model Dev., 4, 255–269, 2011  
www.geosci-model-dev.net/4/255/2011/  
doi:10.5194/gmd-4-255-2011  
© Author(s) 2011. CC Attribution 3.0 License.



## A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale

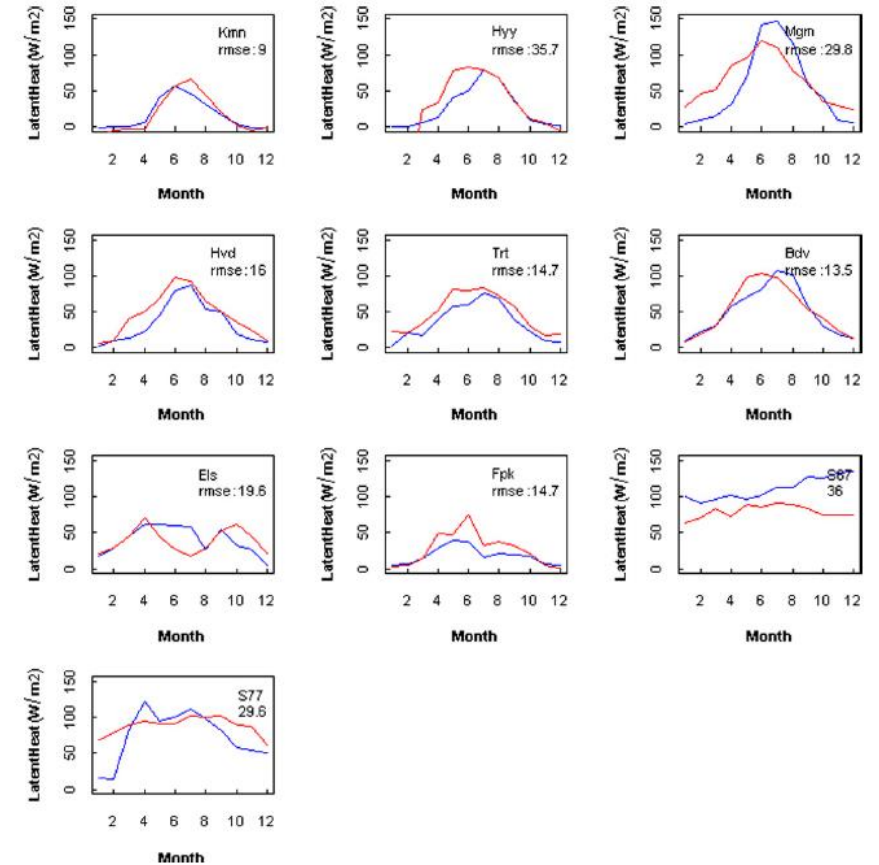
E. Blyth<sup>1</sup>, D. B. Clark<sup>1</sup>, R. Ellis<sup>1</sup>, C. Huntingford<sup>1</sup>, S. Los<sup>2</sup>, M. Pryor<sup>3</sup>, M. Best<sup>3</sup>, and S. Sitch<sup>4</sup>

<sup>1</sup>Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK

<sup>2</sup>Department of Geography, Swansea University, Singleton Park, Swansea, SA2 8PP, UK

<sup>3</sup>Hadley Centre for climate prediction and research, Met Office, Joint Centre for Hydro-Meteorological Research, Wallingford OX10 8BB, UK

<sup>4</sup>School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

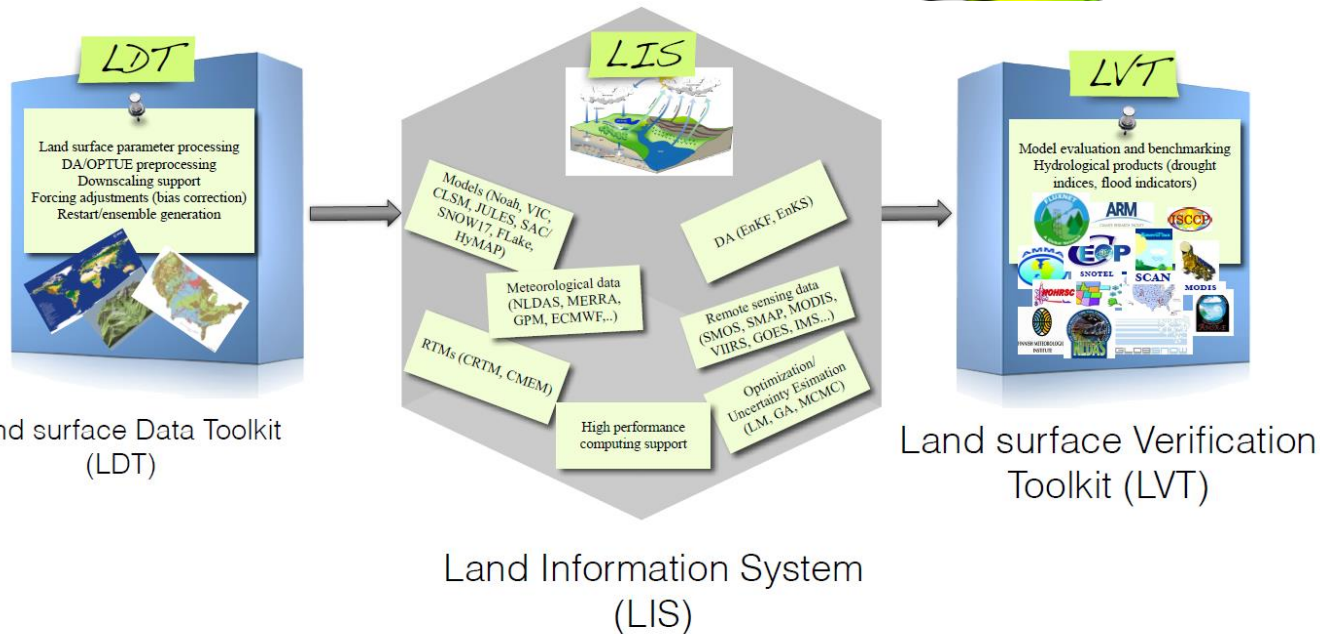


JULES benchmarking setup a few years ago but fell into disrepair.

# JULES Benchmarking and Evaluations

- There is an ongoing need to test our current model configurations and new developments to ensure JULES is suitable for application
  - Physical understanding, process understanding, projections
- JLMP aims to support the community through the availability of suitable tools (including configurations) on shared platforms.
- Evaluation includes comparison with Fluxnet sites, but can include emergent constraints – such as the seasonal cycle
- JULES configurations are applied across time and space scales – evaluation important diurnal-decadal and point to global
- Benchmarking is the comparison of a standard set of metrics between and existing configuration and new development. Do the benchmarks improve?

# Many tools out there in use by the community.



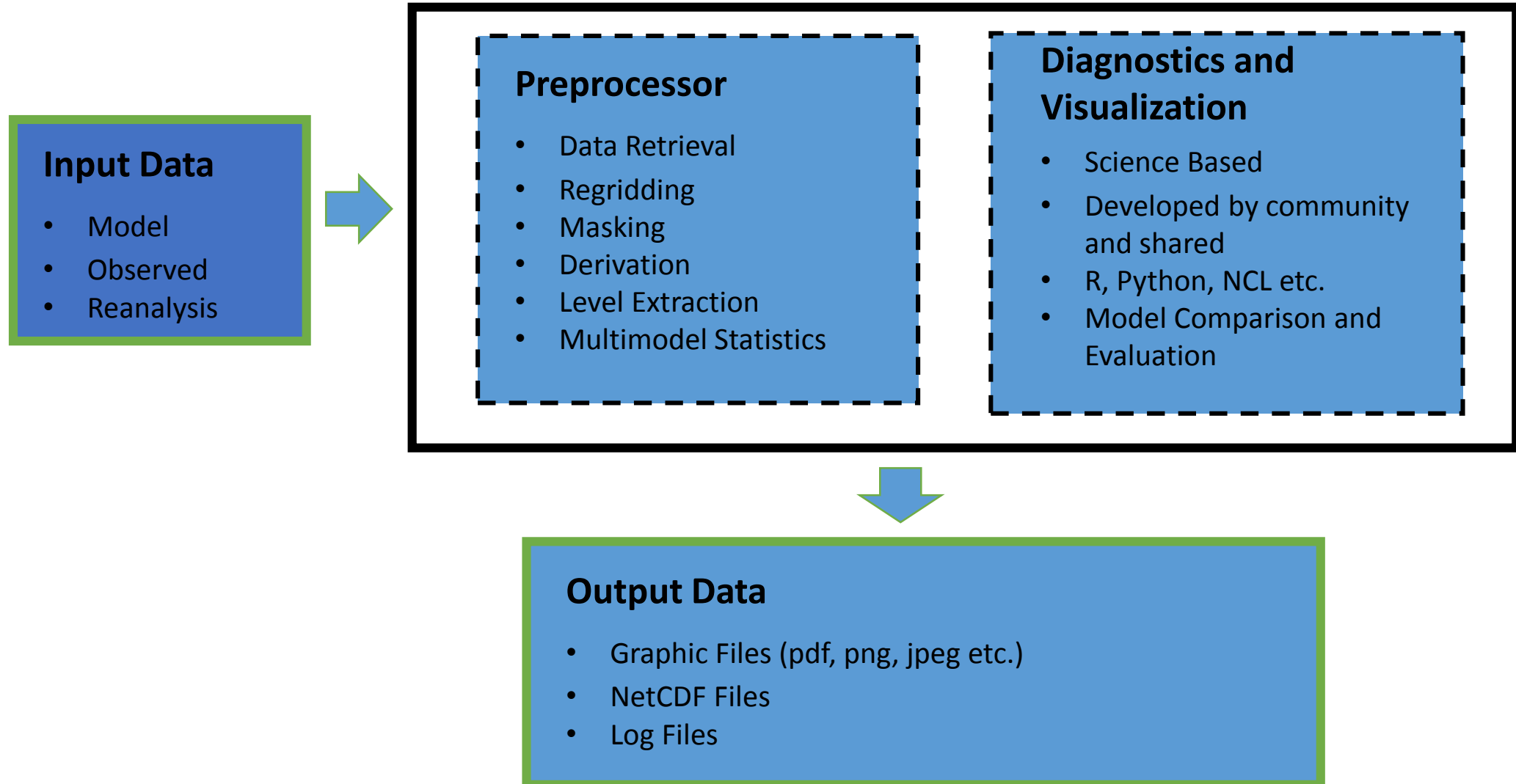
# ESMValTool

Tristan Quaife, Ranjini Swaminathan,  
Richard Allan and Valeriu Predoia.

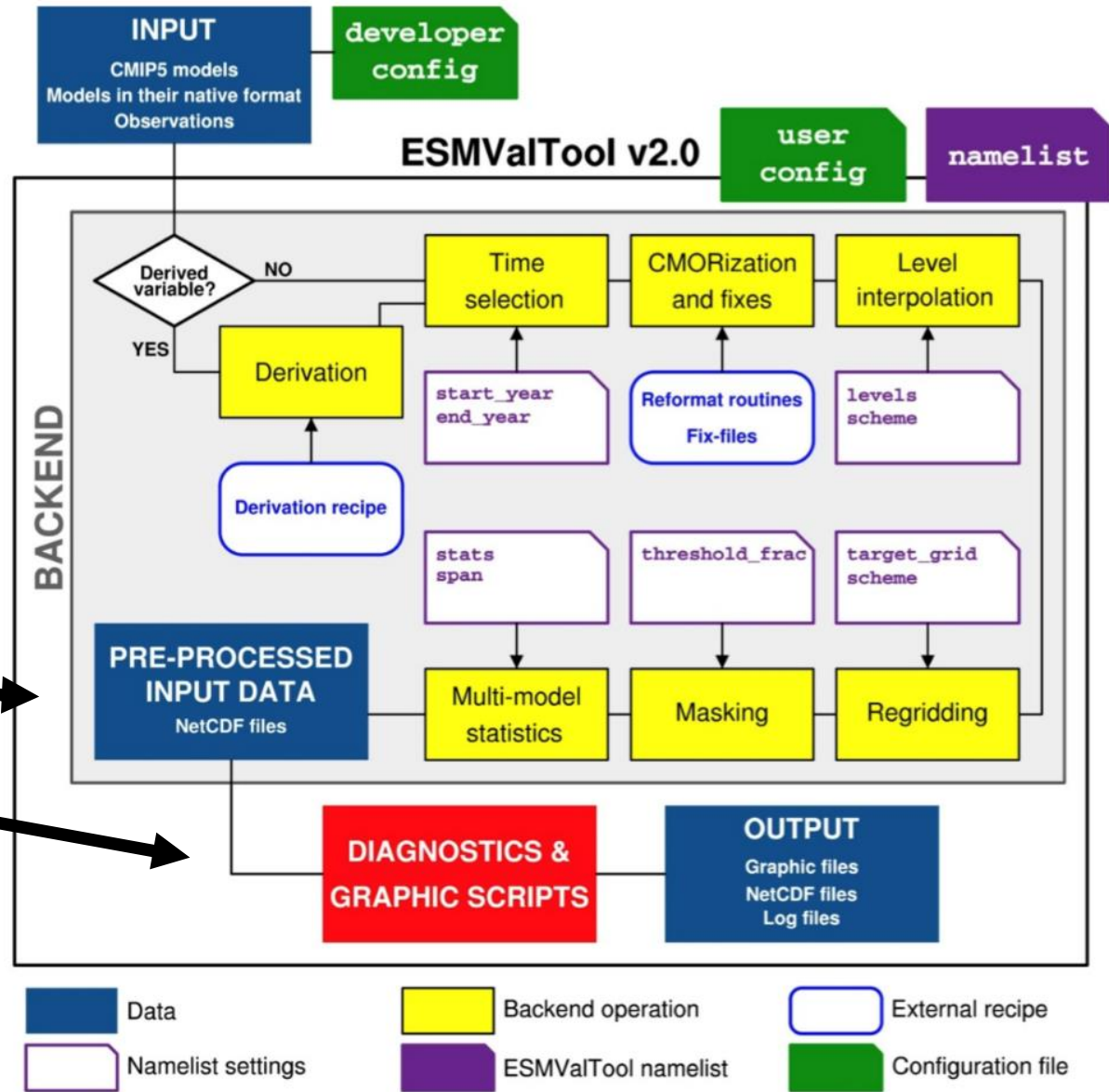
# NCEO, ESMValTool and JULES

- ESMValtool is a community developed diagnostic tool that enables comparisons between model output and observed data
  - Aimed at GCMs/climate models (not just land-surface)
- NCEO has a dedicated post looking at developing diagnostics in ESMValTool for UKESM using EO data
- Specific focus is land-atmosphere interactions
  - Hence direct relevance to JULES
- We have scope to include offline JULES runs (e.g. TRENDY)
- However focus is primarily on coupled runs for CMIP6

# ESMValTool overview and features



# Structure



NCEO is focused on delivering:

- 1) New EO data
- 2) New diagnostics

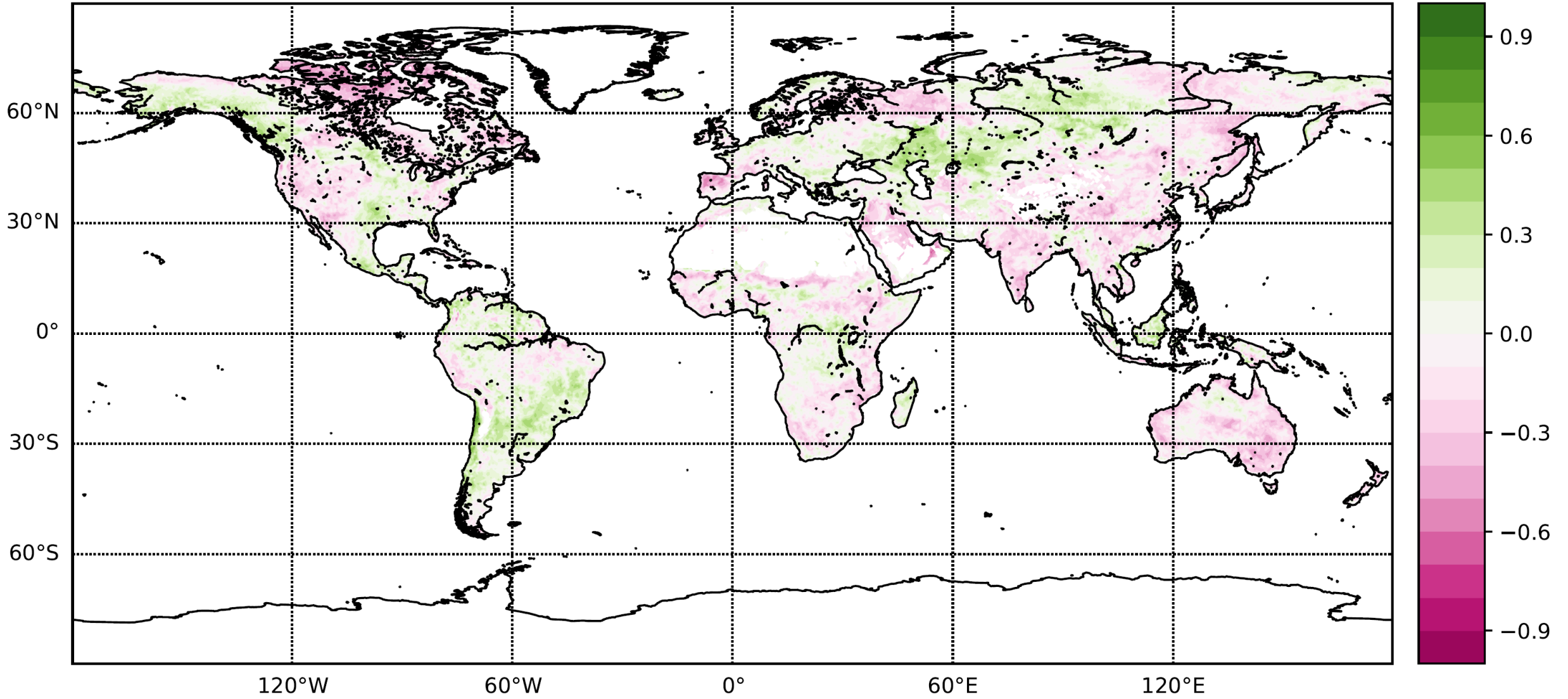
Other parts of the tool are being developed in the wider UKESM project.



# Current focus

- Currently implementing metrics to evaluate the interaction of the water and carbon cycles, e.g:
  - Relationship between GPP and latent heat
  - Response of GPP and precipitation to ENSO
- Working on implementing a large range of EO data into the tool
- We are very happy to discuss new EO data sets and diagnostics to be added especially where they clearly benefit UKESM

Seasonal Anomaly GPP data and ENSO (Nino3.4) index correlation -JJA



# ILAMB

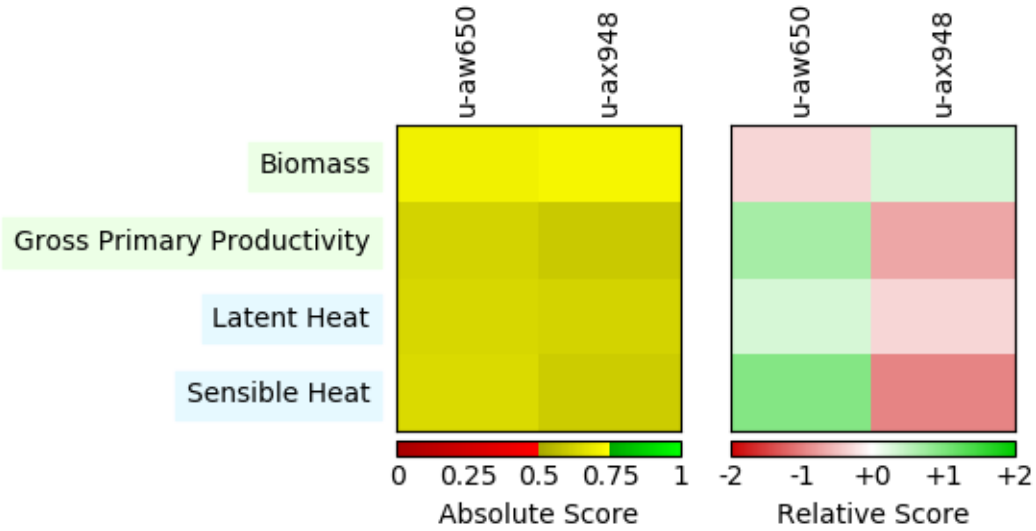
Eddy Robertson, Andy Wiltshire, Carolina Duran Rojas

# ILAMB and JULES

- ILAMB is an open source evaluation tool that enables comparisons between model output and observational estimates
  - Led from US, but gaining international traction
- Focused on land surface evaluation, including meteorological drivers and relationships between climate and land surface variables
- Mostly uses global monthly observational data
- Various groups are adding datasets and metrics (including better site-level evaluation)
- Available on JASMIN, rose suite to evaluate JULES runs on JASMIN available soon
- Will be used to evaluate TRENDY and CMIP6 results

# ILAMB on JASMIN

- The rose suite includes a python script to pre-process JULES data
  - Put data on lat-long grid
  - Some unit changes
  - Some variables derived, e.g. gridbox mean LAI
- Rose suite then calls ILAMB
- Results proved as webpage
- Processed netCDFs and .pngs also proved



ILAMB Benchmark Results - Mozilla Firefox

atch/eroberts/ILAMB/RESULTS/u-au081\_JES\_JC/index.html

ILAMB Benchmark Results

Mean State Results Table

	u-aw650	u-ax948
Biomass	0.71	0.72
Gross Primary Productivity	0.68	0.67
Latent Heat	0.76	0.79
Sensible Heat	0.70	0.70
USForest (11.1%)	0.71	0.75
Gross Primary Productivity	0.68	0.73
Fluxnet (37.5%)	0.63	0.60
GBAF (62.5%)	0.65	0.65
Ecosystem and Carbon Cycle Summary	0.61	0.57
Latent Heat	0.67	0.66
Fluxnet (25.0%)	0.64	0.63
GBAF (75.0%)	0.64	0.62
Sensible Heat	0.65	0.61
Fluxnet (37.5%)	0.66	0.63
GBAF (62.5%)	0.64	0.60
Hydrology Cycle Summary	0.64	0.62
Overall Summary	0.65	0.64

### Overview Figure

- Scores for each variable
- Scores relative to other models

### Summary table

- Scores for each observational dataset

u-aw650 = JULES-ES

u-ax948 = JULES-C

Mean State - Mozilla Firefox

ILAMB Benchmark Results x Mean State x +

file:///net/spice/scratch/roberts/ILAMB/RESULTS/u-au081\_JES\_JC/EcosystemandCarbonCycle/GrossPrimaryProductivity/GBAF/GBAF

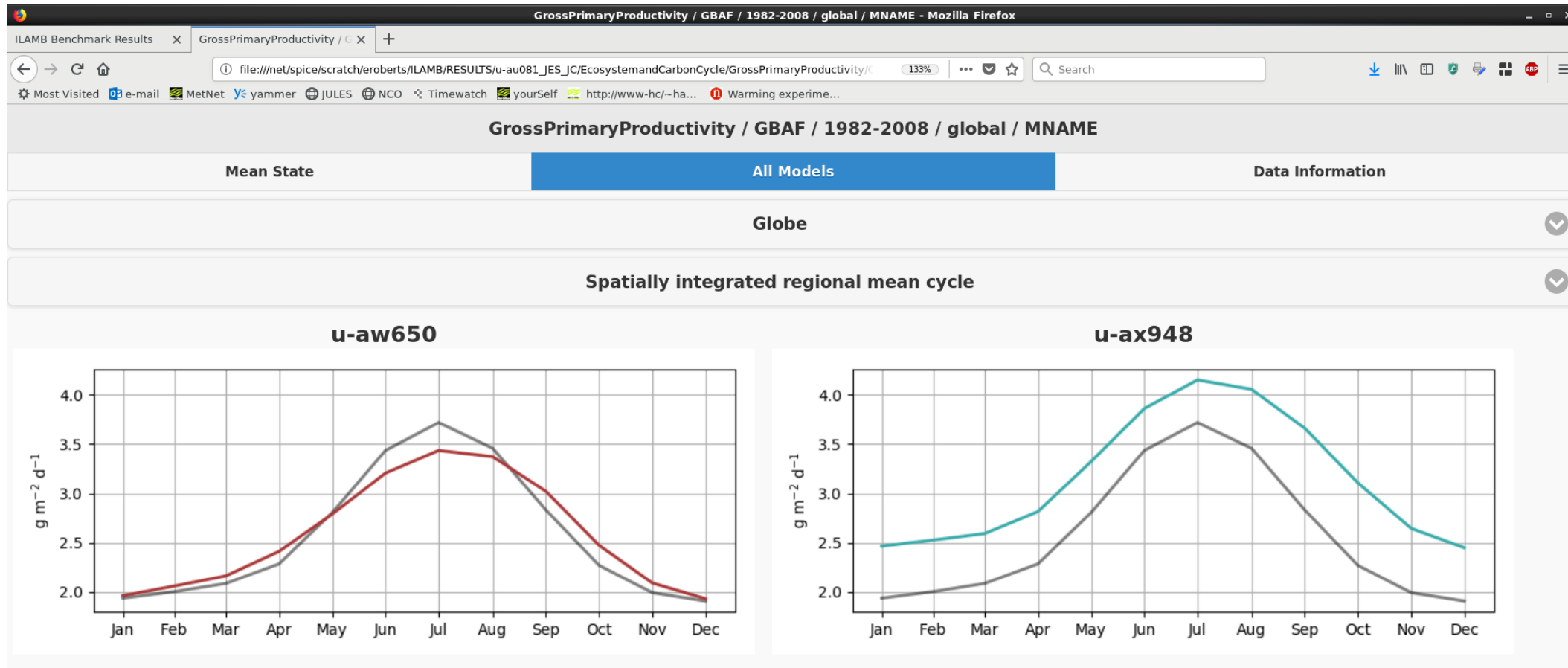
GrossPrimaryProductivity / GBAF / 1982-2008 / global / u-aw650

Mean State All Models Data Information

Globe

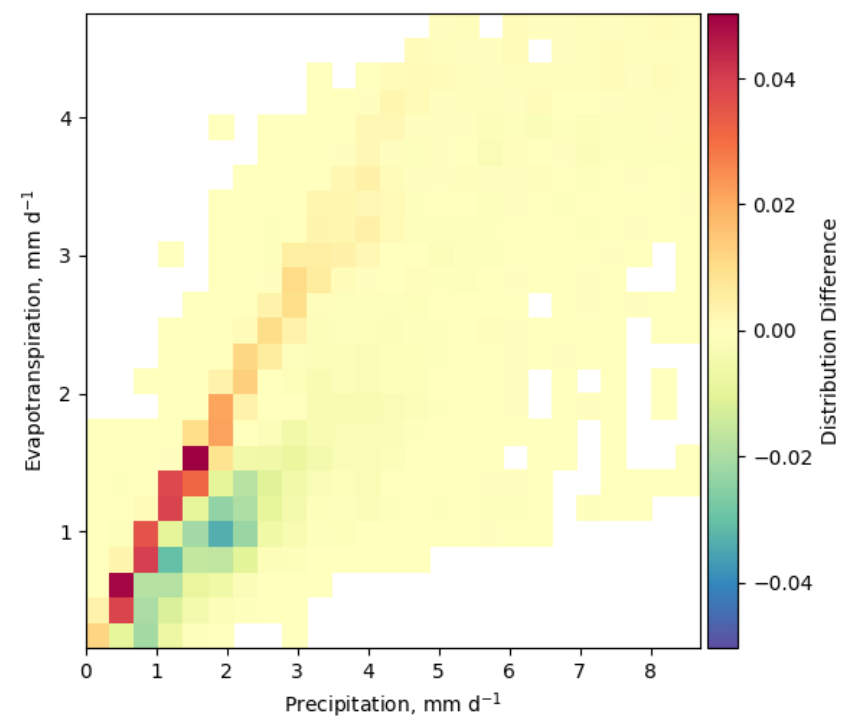
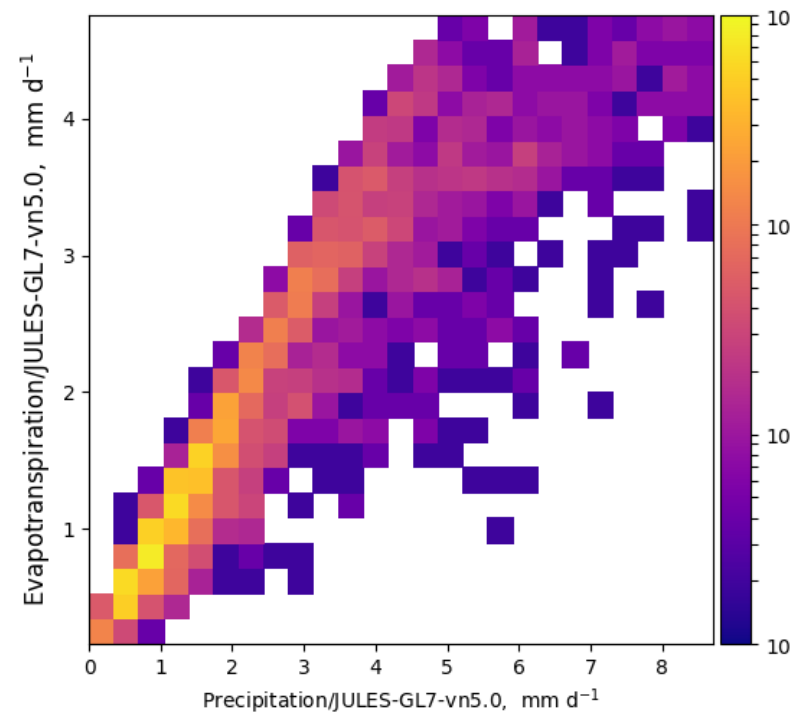
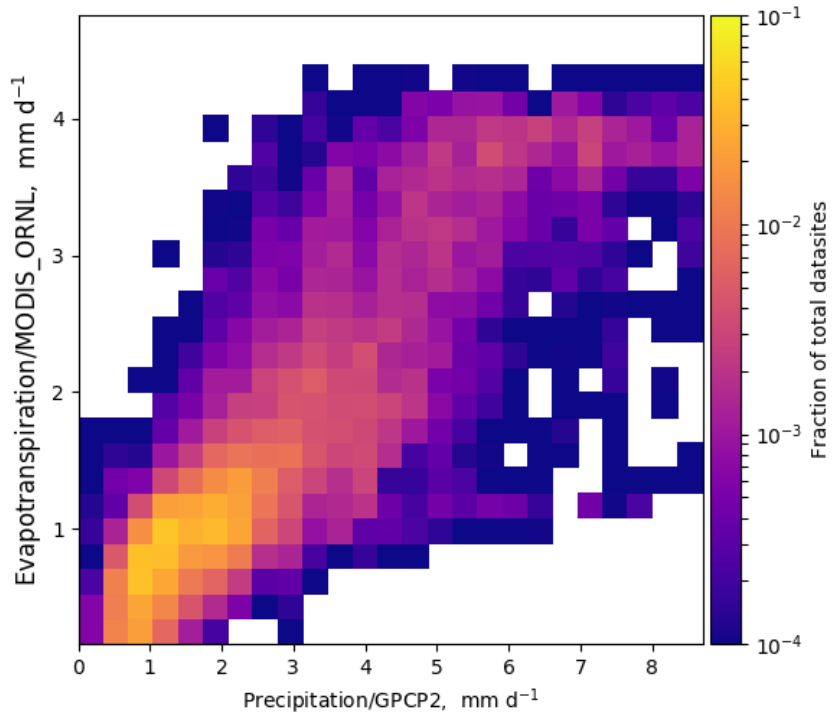
Benchmark	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (intersection) [Pg yr-1]	Benchmark Period Mean (complement) [Pg yr-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Overall Score [1]
u-aw650	119.	118.	111.	6.90	118.	0.623	0.108	1.44	1.18	0.50	0.39	0.81	0.95	0.61
u-ax948	142.	134.	7.67	118.	0.944	0.657	1.79	1.17	0.44	0.36	0.81	0.89	0.57	

Metrics and scores for each observational dataset



Figures for each model-data comparison





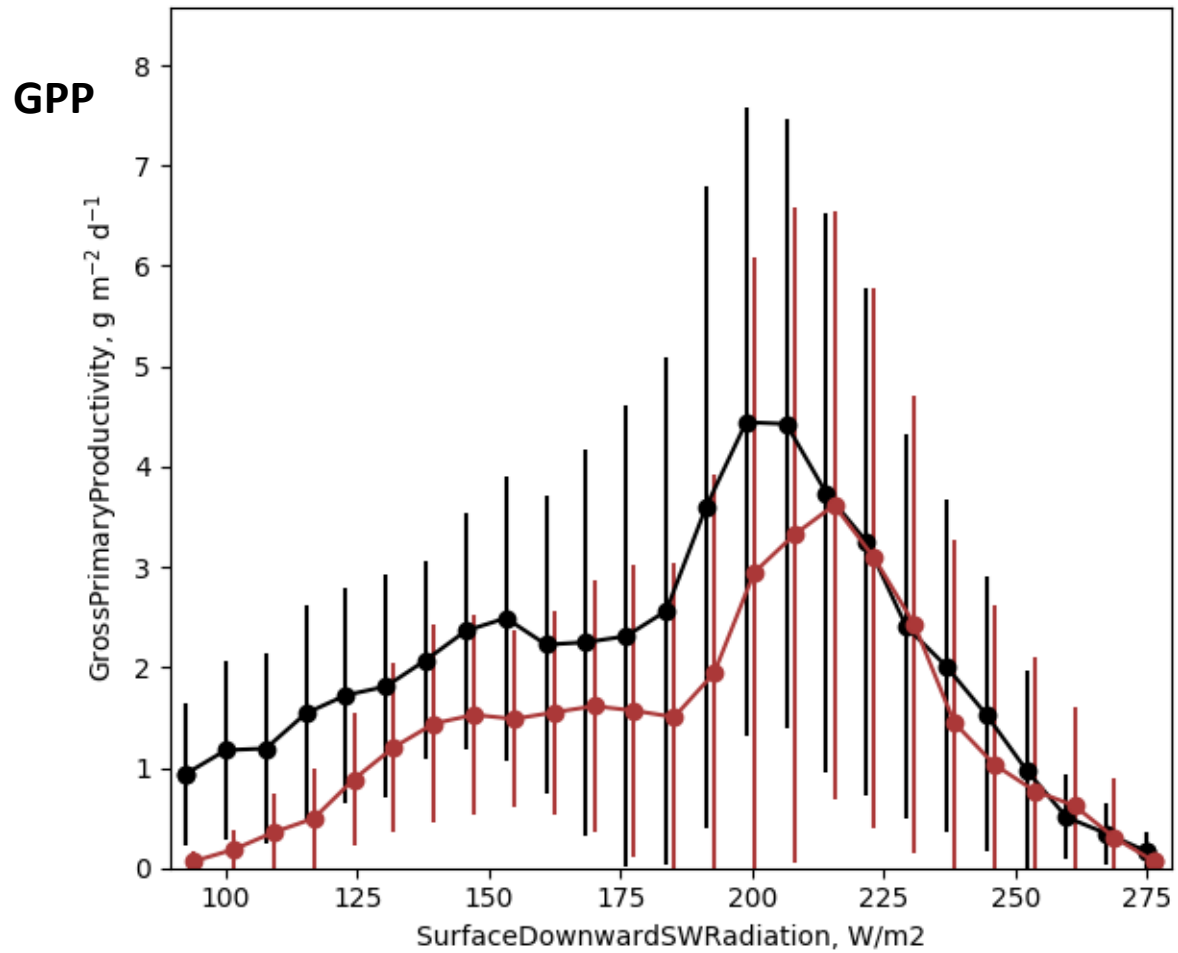
Observed

Model

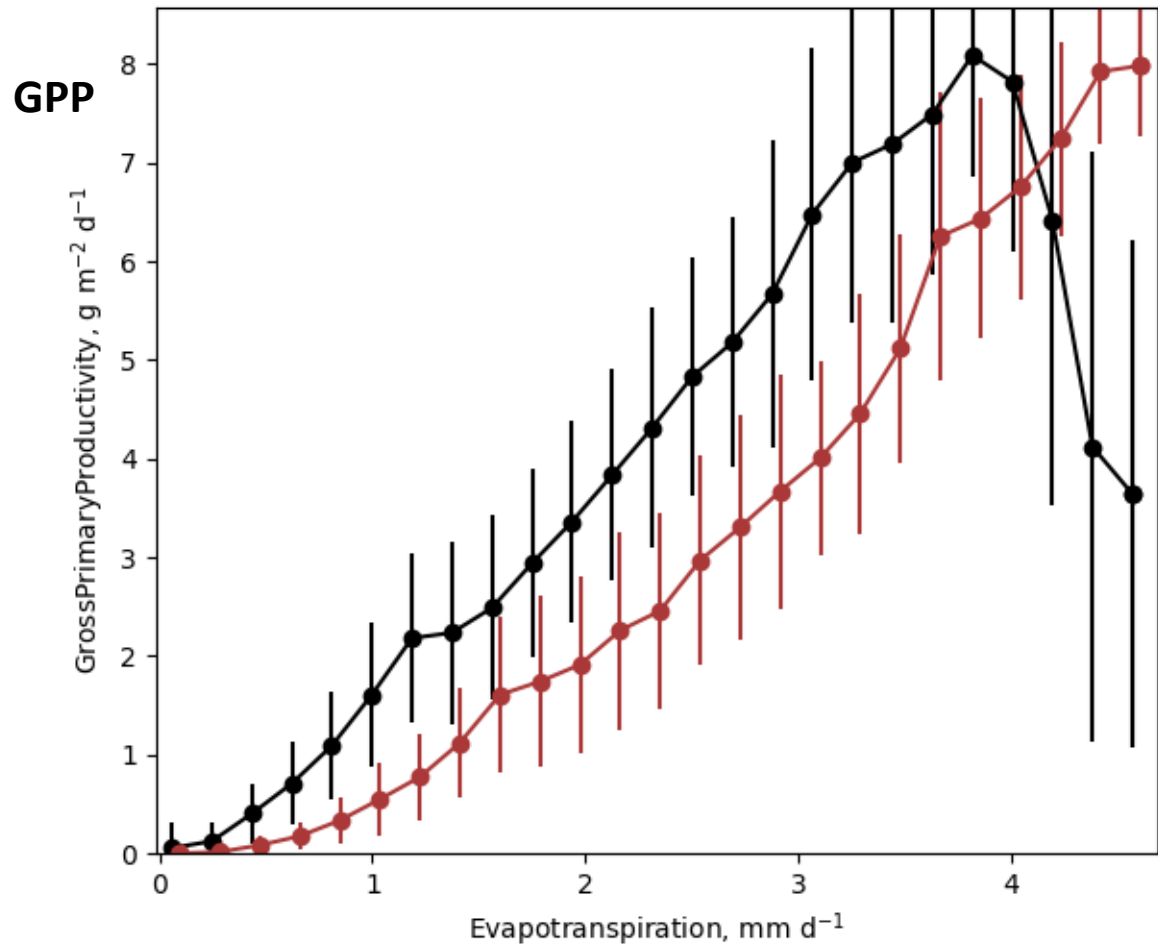
Difference

Evaluation of relationships between variables: e.g. ET and precip

(GL7 example)



**Downward Shortwave**



**ET**

# What will JLMP do?

- Early days so plans still in the making (comments welcome)
- Initially likely to be 'TRENDY' style historical simulations that can be benchmarked using ILAMB.
- Initially, likely to be JULES-GL7 and ILAMB running on JASMIN. Users will be able to their processes to JULES-GL7 and re-benchmark.
- Future will add JULES-ES configuration
- Aim is for users to be able to add obs/metric to ILAMB as well.
- Role of other evaluation tools to be discussed

*\*TRENDY is a set of experiments run annually as part of the Global Carbon Project.*

# The Community

- JLMP (CEH, NCAS, NCEO, Met Office) would like to encourage all users to take advantage of the configurations being made available.
- We would encourage:
  - Adopt use of evaluation and benchmarking tools
  - Routinely run using standard suites and contribute to their development
- Routine evaluation hoped to become standard part of JULES meeting/JPEG process

