# The PALS Land sUrface Model Benchmarking Evaluation pRoject (PLUMBER)

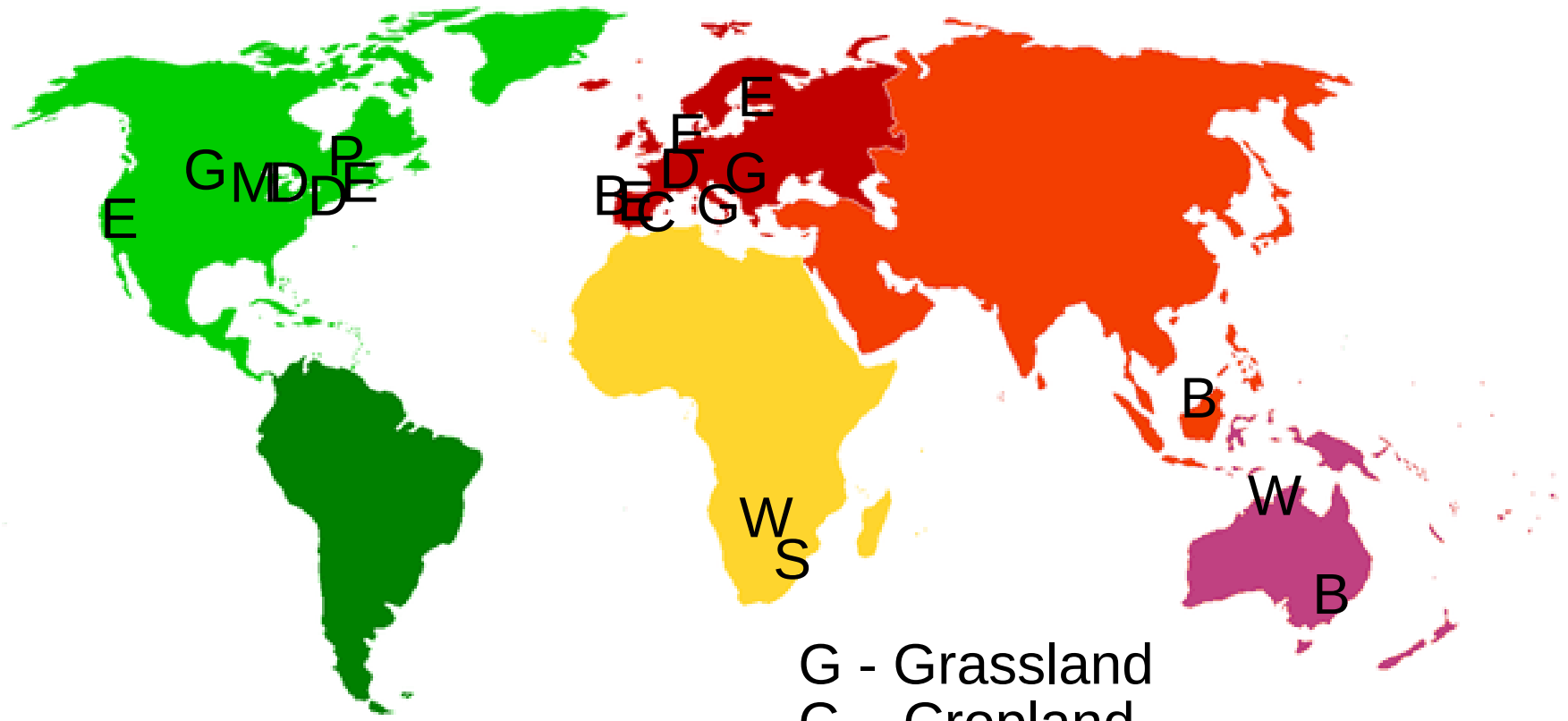
&

# CABLE's benchmarking environment

Gab Abramowitz

UNSW THE UNIVERSITY OF NEW SOUTH WALES
ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE
TERN Terrestrial Ecosystem Research Network

# PLUMBER – the 20 flux tower sites




E – Evergreen Needleleaf
B – Evergreen Broadleaf
D – Deciduous Broadleaf
M – Mixed Forest

G - Grassland
C – Cropland
W – Woody Savanna
S – Savanna
P – Permanent Wetlands

GLASS

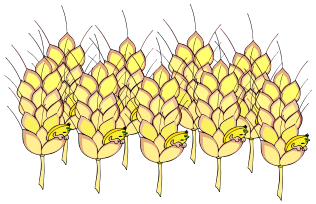
Met Office

ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE

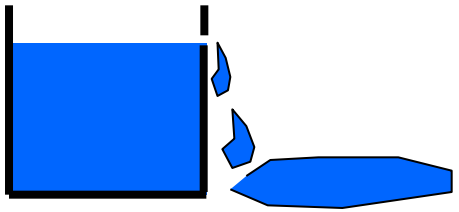
TERN

# The Benchmarks

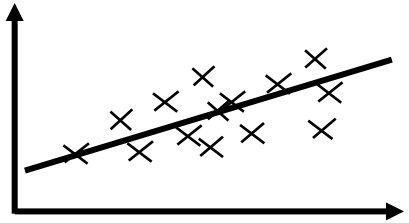
- Penman-Monteith



- Manabe bucket model



- Empirical regressions



# The three empirical models

- All 3 are automatically calculated and plotted alongside model and observed data on PALS
- All 3 empirical models are trained with data from sites other than the testing site (i.e. out of sample)
- They are each created for LE, H, NEE:
  - “1lin”: linear regression of flux against downward shortwave (SW)
  - “2lin”: as above but against SW and surface air temperature (T)
  - “3km27”: non-linear regression – 27-node k-means clustering + linear regression against SW, T and relative humidity at each node

All are instantaneous responses to met variables with no knowledge of vegetation type, soil type, soil moisture or temperature, C pools

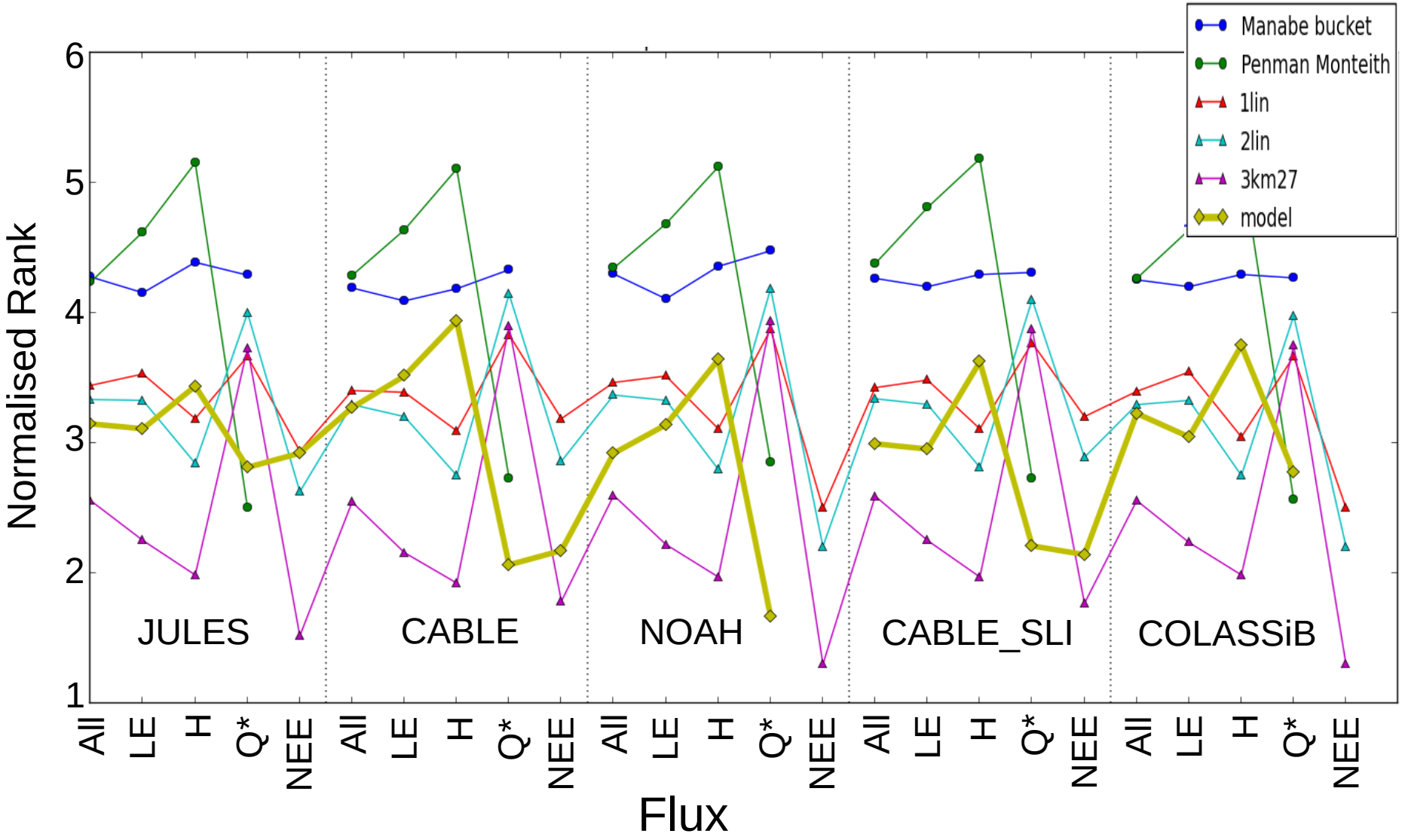
# Fluxes and metrics for normalised rank calculation (next slide)

| Sensible heat flux (H) | Latent Heat flux (LE) | Net Radiation (Q*) | Net Ecosystem Exchange (NEE) |
|---|---|---|---|

| | | |
|---|---|---|
| Mean Bias Error | MBE | $\left(\sum_{i=1}^{n}\left(M_i - O_i\right)\right)/n$ |
| Normalised Mean Error | NME | $\frac{\sum_{i=1}^{n}\lvert M_i - O_i \rvert}{\sum_{i=1}^{n}\lvert \overline{O} - O_i \rvert}$ |
| Standard Deviation | sd | $\sqrt{\frac{\sum_{i=1}^{n}\left(M_i - \overline{M}\right)^2}{n-1}}$ |
| Correlation coefficient | r | $\frac{n\sum_{i=1}^{n}(O_i M_i) - \left(\sum_{i=1}^{n} O_i \sum_{i=1}^{n} M_i\right)}{\sqrt{\left(n\sum_{i=1}^{n} O_i^2 - \left(\sum_{i=1}^{n} O_i\right)^2\right)\left(n\sum_{i=1}^{n} M_i^2 - \left(\sum_{i=1}^{n} M_i\right)^2\right)}}$ |

GLASS Met Office ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE TERN Terrestrial Ecosystem Research Network

# Ranks against benchmarks




GLASS

Met Office

ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE

TERN

# Benchmarking CABLE

Ideally, proposed changes to CABLE will need:

- Documentation to support the changes:
  - Motivation – added functionality? Improvement in performance? Valuable alternative representation of processes?
  - Scientific description
  - User guide changes – new switches, parameters?
- Benchmarking. Does the change degrade/improve CABLE's performance:
  - Offline at single sites, continental and global scales?
  - * Coupled: ACCESS, WRF/LIS…

Submit both to CABLE committee.

# Benchmarking CABLE

- We aim for PALS to eventually host all offline testing required for this process:
  - Benchmark tests to be part of standard CABLE configuration
  - Run updated code, upload to appropriate experiment on PALS
  - Generate PALS report comparing against current CABLE version
- Single site offline tests in PALS already
- What are they key offline tests we want for CABLE?
  - Variables – energy, water, carbon
  - Resolution – continental, global
  - Time step size, period
  - Which specific data sets + versions?
  - Metrics that speak to strengths of evaluation data
    - Means, variability, distributions, extremes, temporal and spatial correlation

# Benchmarking CABLE

The task for (at least some) working groups:

**Nominate at least one key benchmarking experiment**

At least: land use/disturbance; soil and hydrology; non-vegetated tiles

- Variables – energy, water, carbon
- Resolution – continental, global
- Time step size, period
- Which specific data sets + versions?
- Metrics that speak to strengths of evaluation data
  - Means, variability, distributions, extremes, temporal and spatial correlation