CABLE and the PLUMBER experiment: results and possible causes

Gab Abramowitz & Ned Haughton
Climate Change Research Centre, UNSW and
ARC Centre of Excellence for Climate Systems Science

+ PLUMBER co-authors: M. Best, H. Johnson, A. Pitman, G. Balsamo, A. Boone, M. Cuntz, B. Decharme, P.A. Dirmeyer, J. Dong, M. Ek, Z. Guo, V. Haverd, B. van den Hurk, G. Nearing, B. Pak, C. Peters-Lidard, J. Santanello, L. Stevens, N. Vuichard.

UNSW AUSTRALIA

ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE

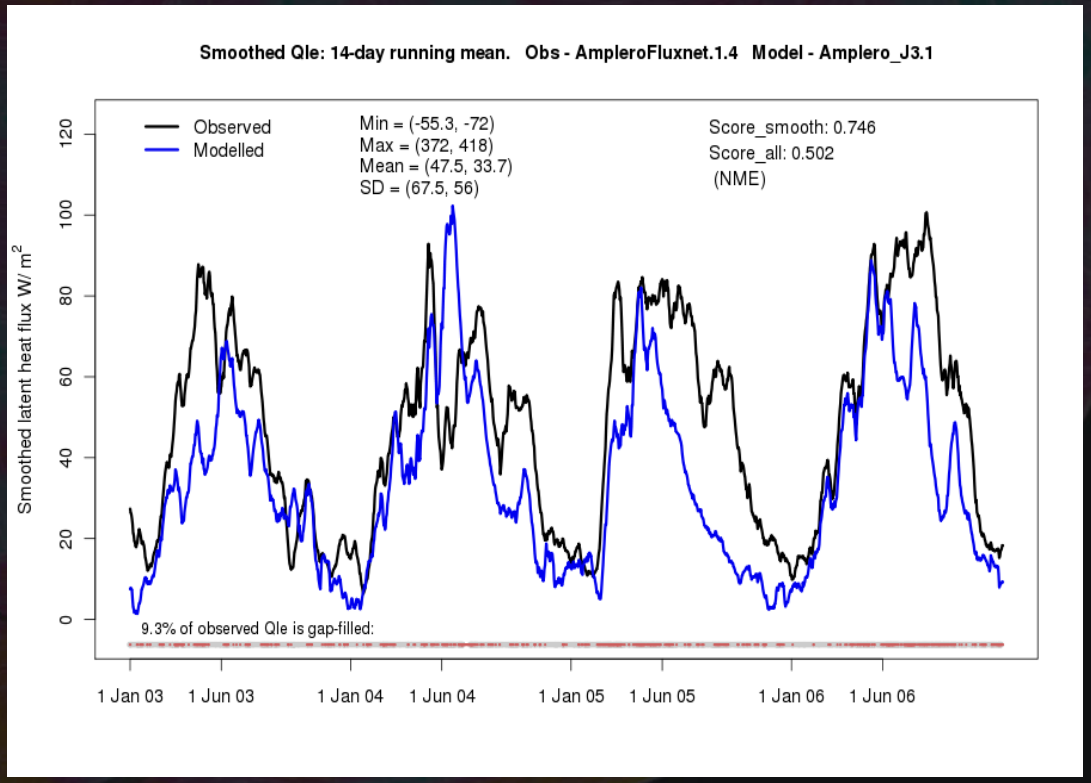
Benchmarking – a simple example

How well should we expect a LSM to predict latent heat (LH) flux at the Amplero site?

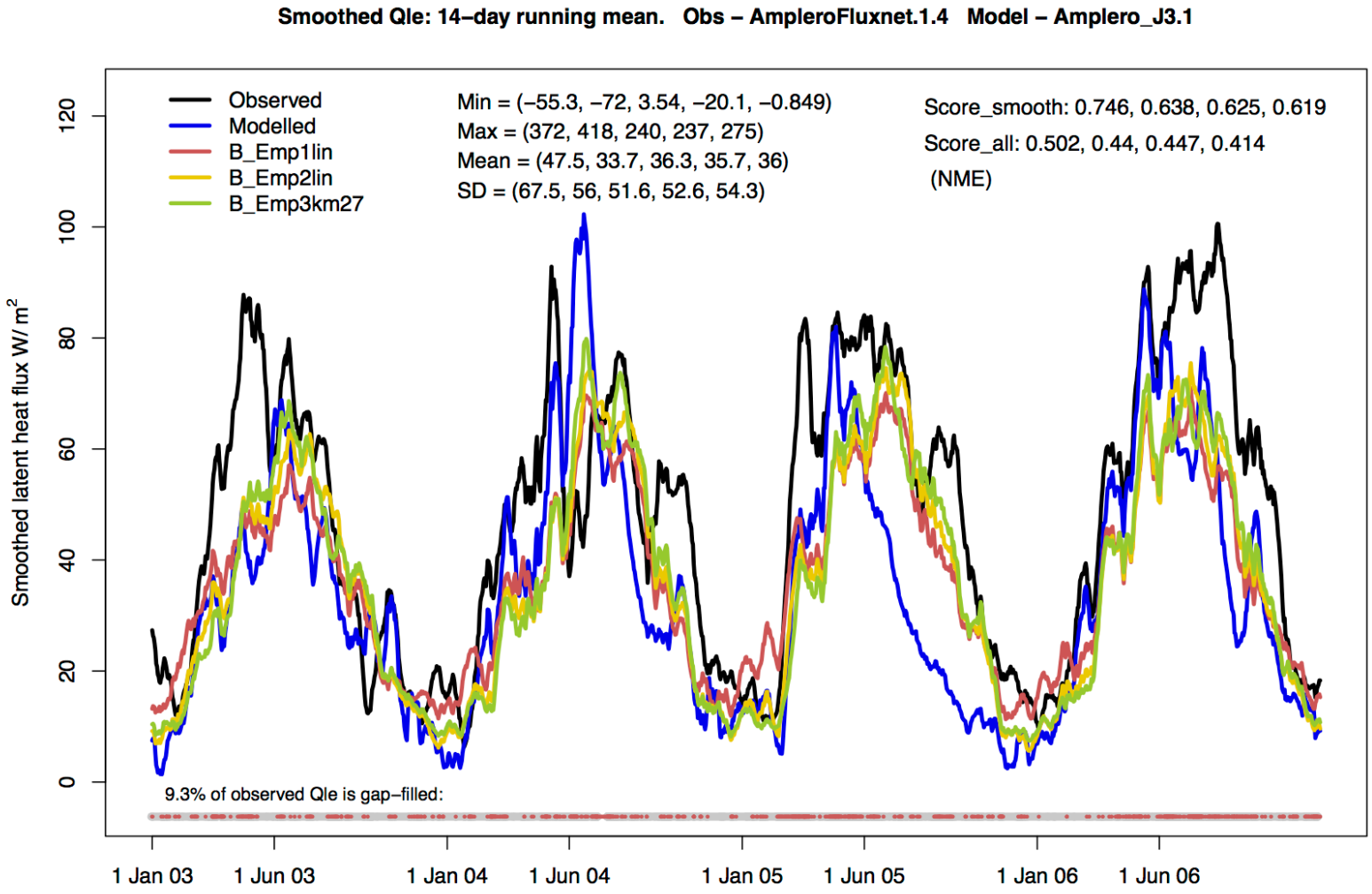
- Several (19) flux tower sites other than Amplero
- Train a linear regression between shortwave radiation and LH
- Use these regression parameters to predict LH at Amplero using site SW radiation

This will tell us:

- The extent to which LH is predictable from SWdown - just 1 model input variable
- How a very simple functional relationship would represent LH in our usual diagnostics
- How predictable LH at Amplero is, out-of-sample



Benchmarking – an example

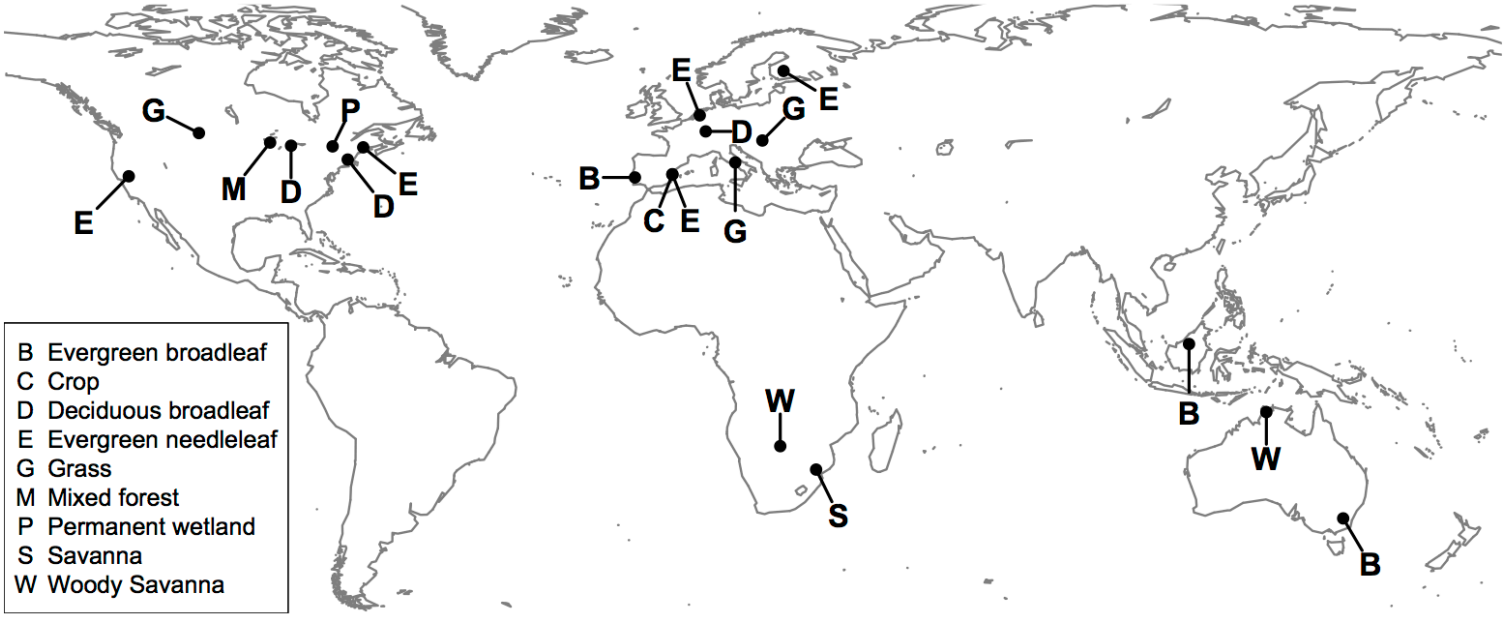


Benchmarking – an example

- Using empirical models (out of sample) as benchmarks can quantify the amount of information available to a model in its inputs about its prediction variables
- It gives one way to quantify how well we should *expect* a model to perform
- It provides a model-like time series, and so provides benchmark performance levels in any chosen metric

The PALS Land sUrface Model Benchmarking Evaluation pRoject (PLUMBER)

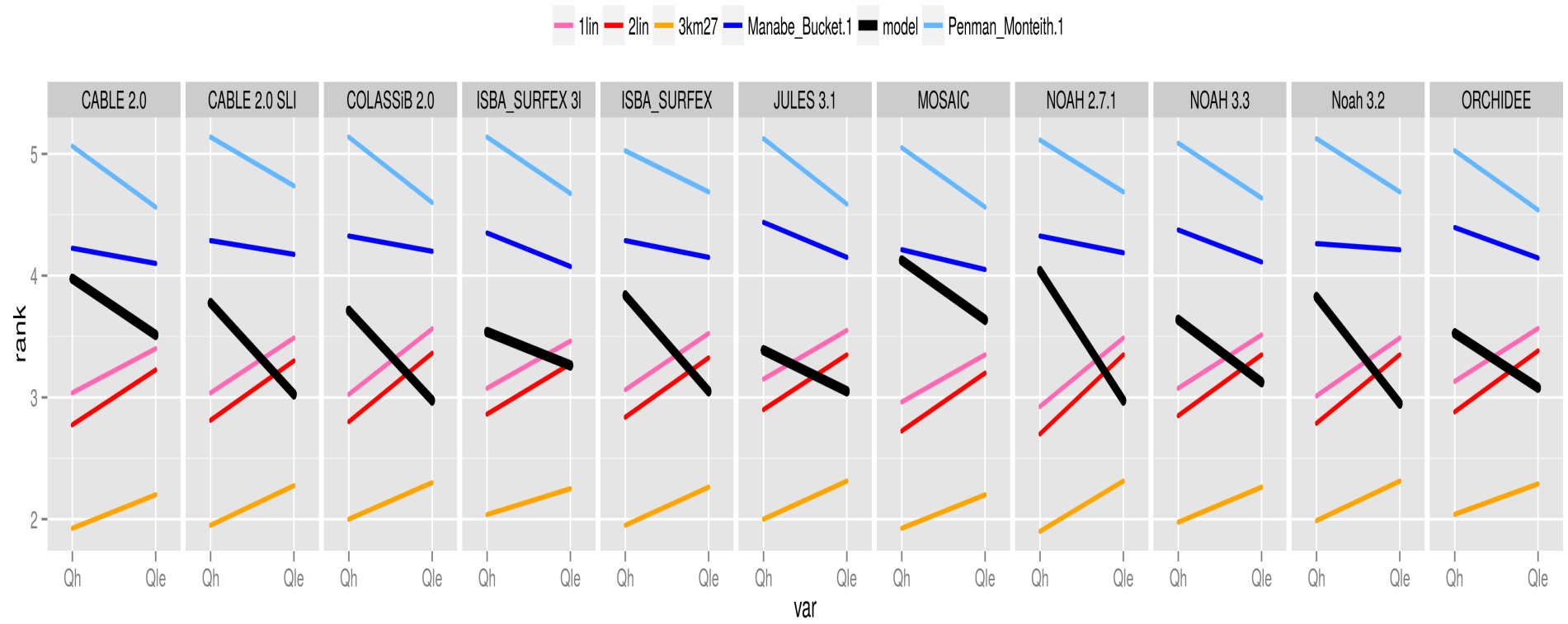
- 20 Flux tower sites; latent and sensible heat flux
- 4 metrics: bias, correlation, SD, normalised mean error
- 10+ LSMs
- Benchmarks: two 'physical' – PM and Manabe bucket; 3 empirical



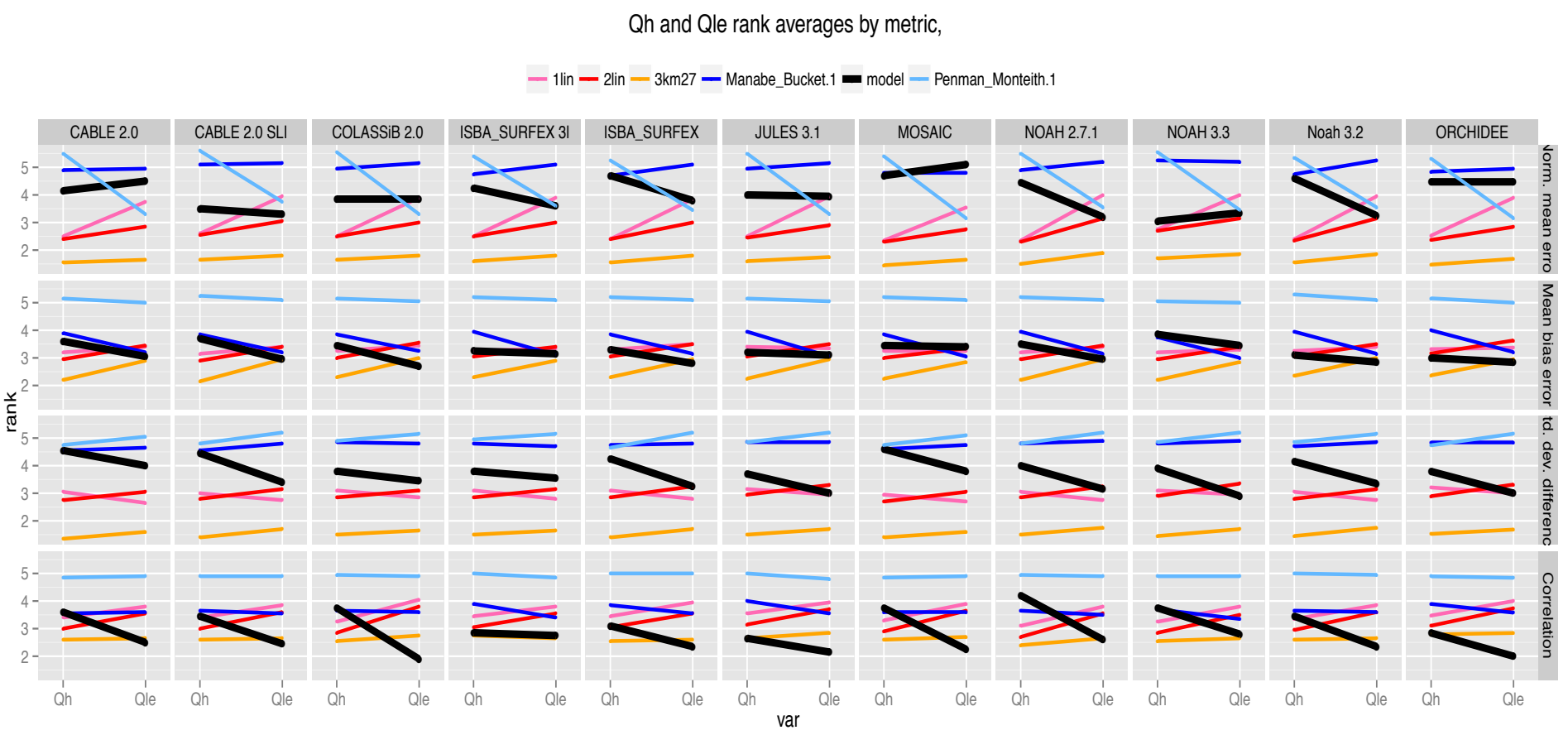
The three empirical benchmarks in PLUMBER

- All 3 empirical models relate met forcing and a flux and are trained with data from sites other than the testing site (i.e. out of sample)
- They are each created for LE, H:
  - "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.

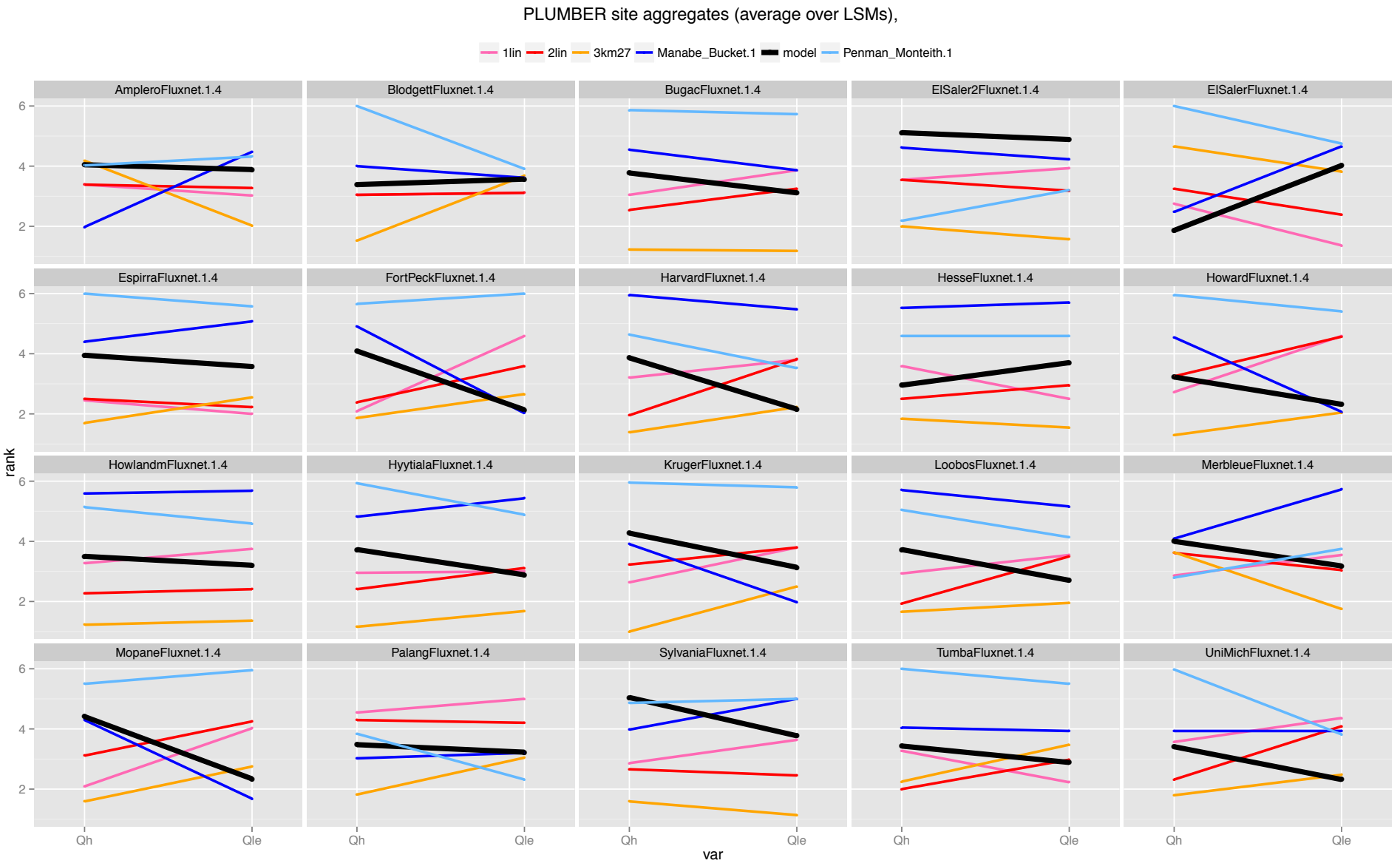
PLUMBER results



PLUMBER results – by metric



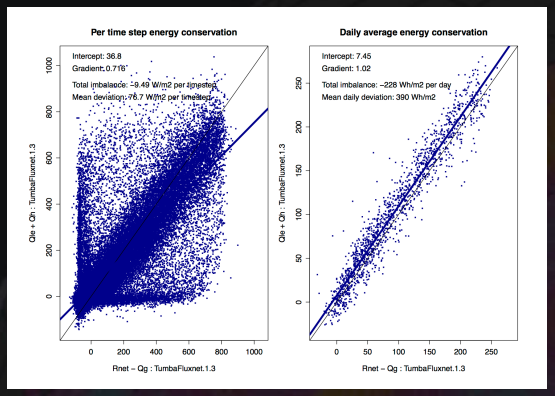
PLUMBER results – by site

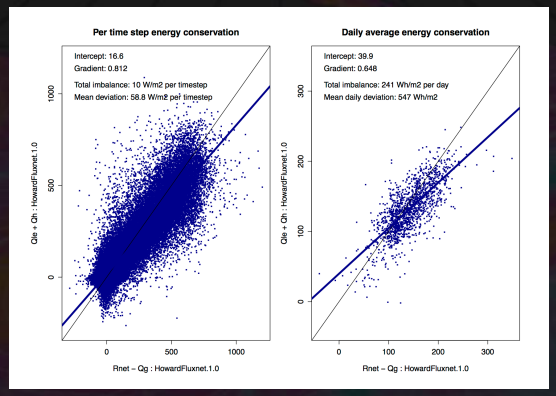


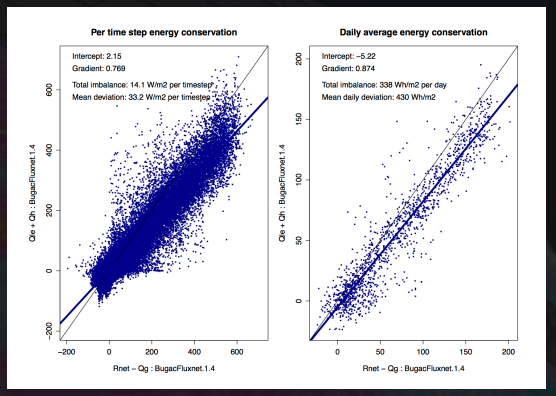
PLUMBER results – why?

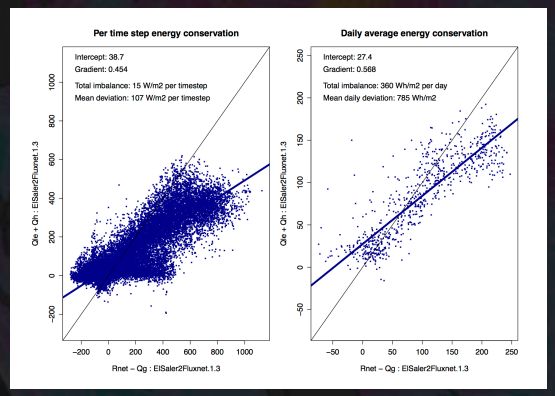
1. Flux tower measurements – conservation issues?
2. Is it because the PLUMBER analysis focuses on short timescales?
3. Are flux towers are at the wrong spatial scale?
4. Is the state initialisation inappropriate?
5. Time scale of state variables?
6. Over-parameterisation is hurting – calibration of unconstrained parameters inhibits predictive capacity?
7. LSMs are essentially conceptual models – too many processes not supported by data in the scope of their application

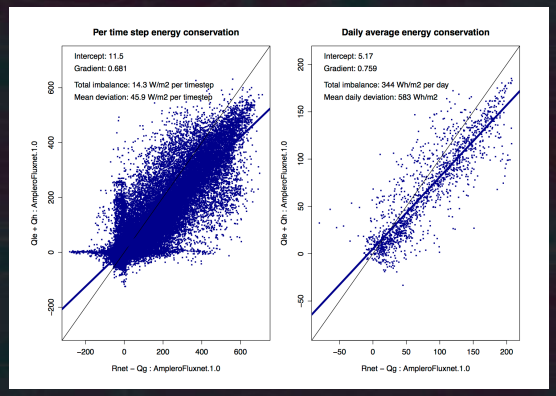
Flux tower systematic bias / conservation?

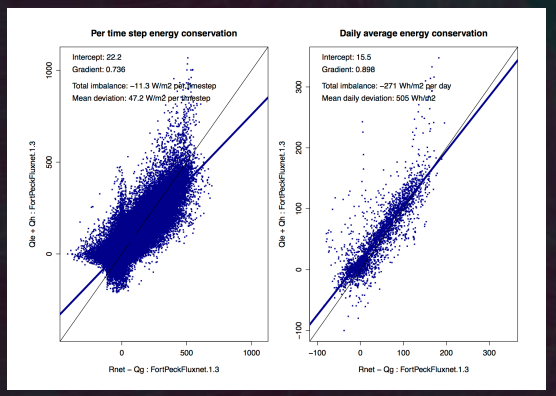


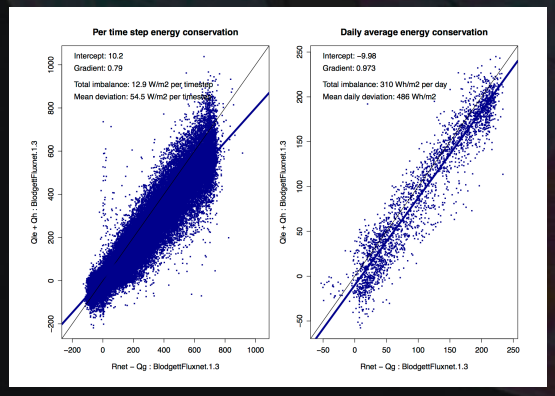


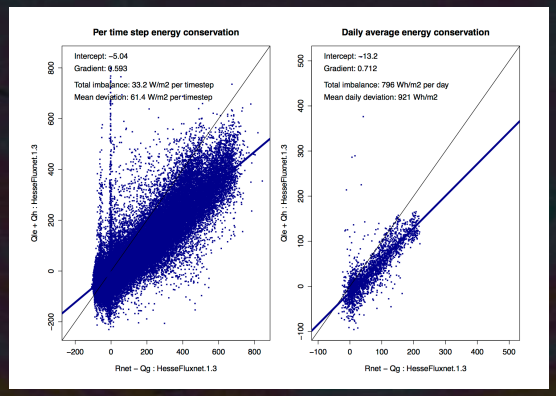


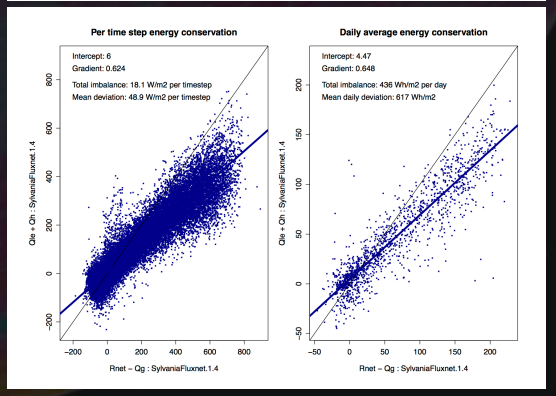






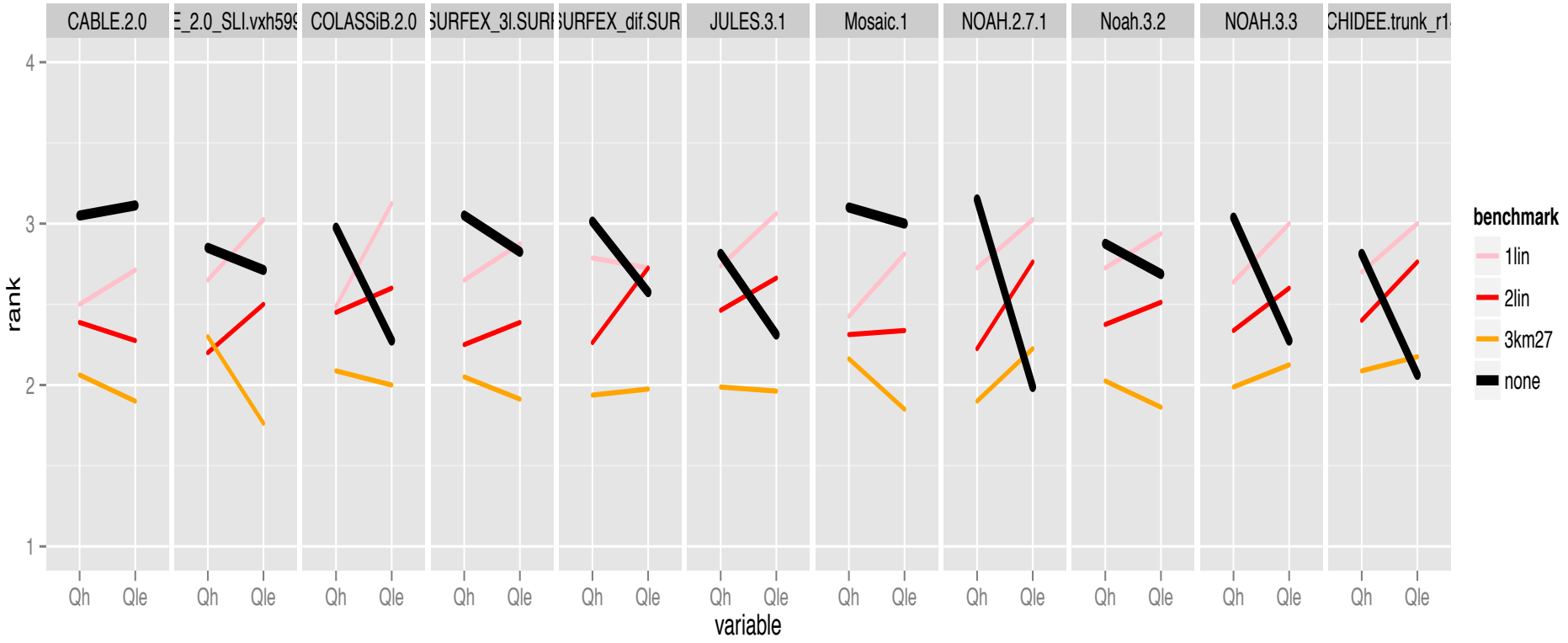




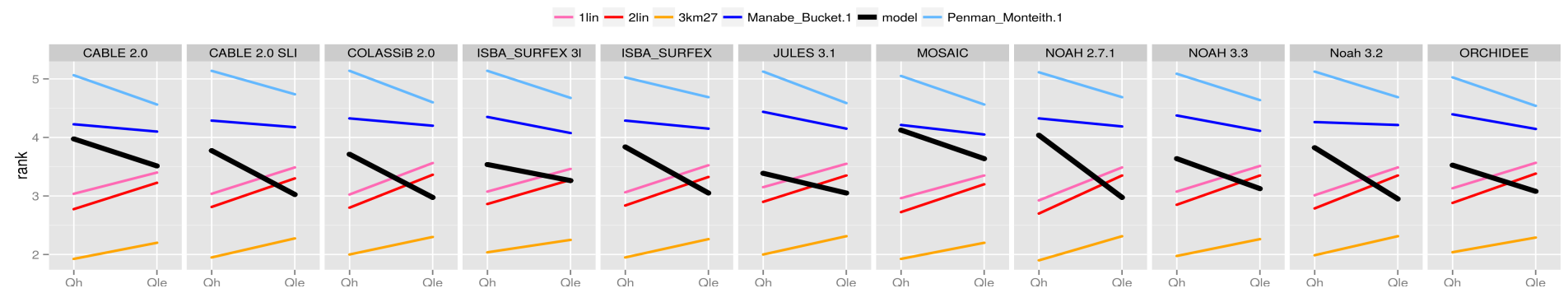


PLUMBER results – energy consevation?

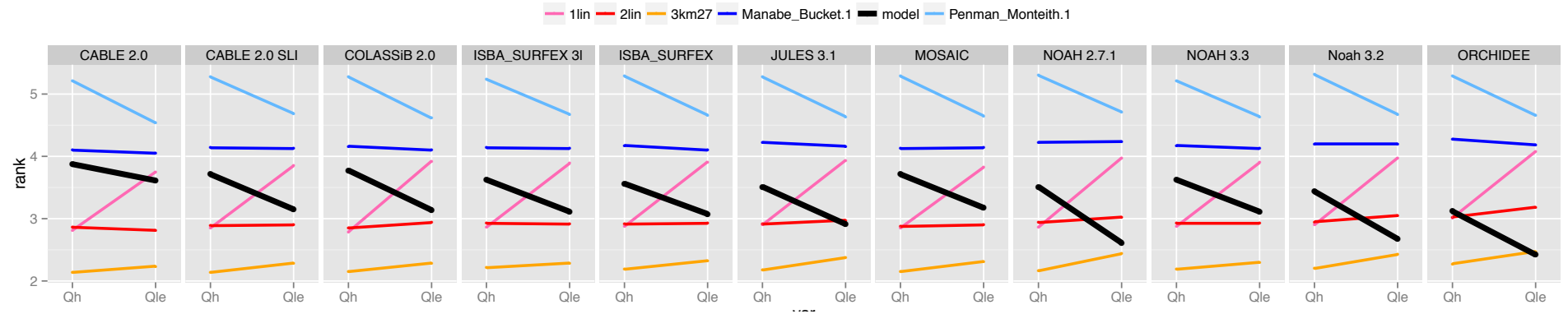
Force each empirical model's [Qle+Qh] to equal each LSM's [Qle+Qh] at every time step – i.e. empirical models now have same Rnet, Gflux as LSMs and conserve energy.



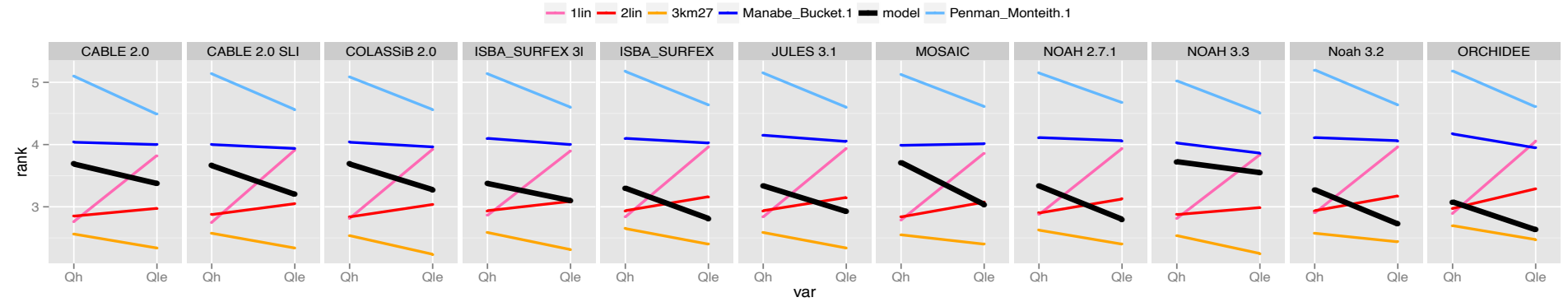
PLUMBER results – timescale?



Qh and Qle rank averages, daily



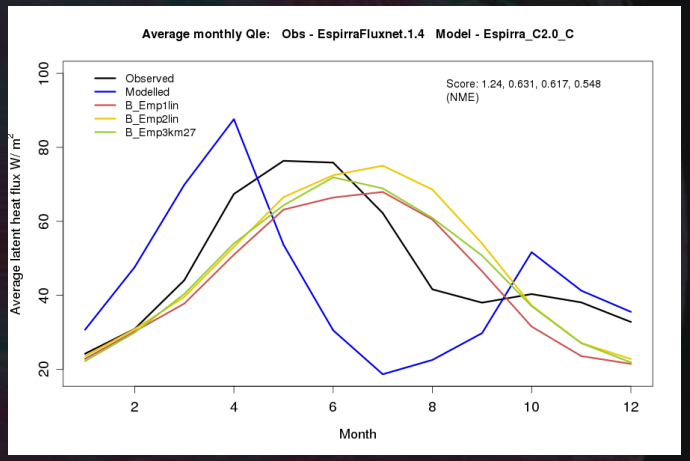
Qh and Qle rank averages, seasonal



PLUMBER results – initialisation?

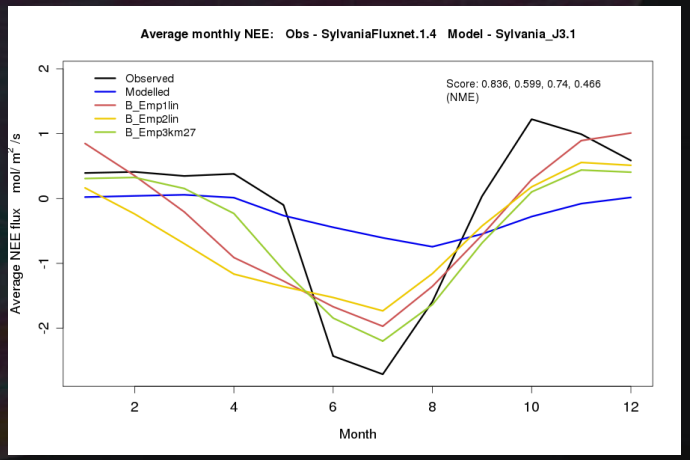
- Repeated spin-up on flux tower met data – no guarantee its representative
- Likely not perfect, but unlikely the major issue – very few cases of flux being consistently too high or too low for all months of average annual cycle:

Latent heat	Too high	Too low	neither
JULES	1	3	16
NOAH	3	1	16
COLA	1	1	18
CABLE	4	0	16



- More likely an issue for NEE?

NEE	Too high	Too low	neither
JULES	1	2	17
CABLE	3	3	16



PLUMBER results – why?

1. Flux tower measurements – conservation issues?
2. Is it because the PLUMBER analysis focuses on short timescales?
3. Are flux towers are at the wrong spatial scale?
4. Is the state initialisation inappropriate?
5. Time scale of state variables?
6. Over-parameterisation is hurting – calibration of unconstrained parameters inhibits predictive capacity?
7. LSMs are essentially conceptual models – too many processes not supported by data in the scope of their application