Protocol for the Analysis of Land Surface models (PALS)

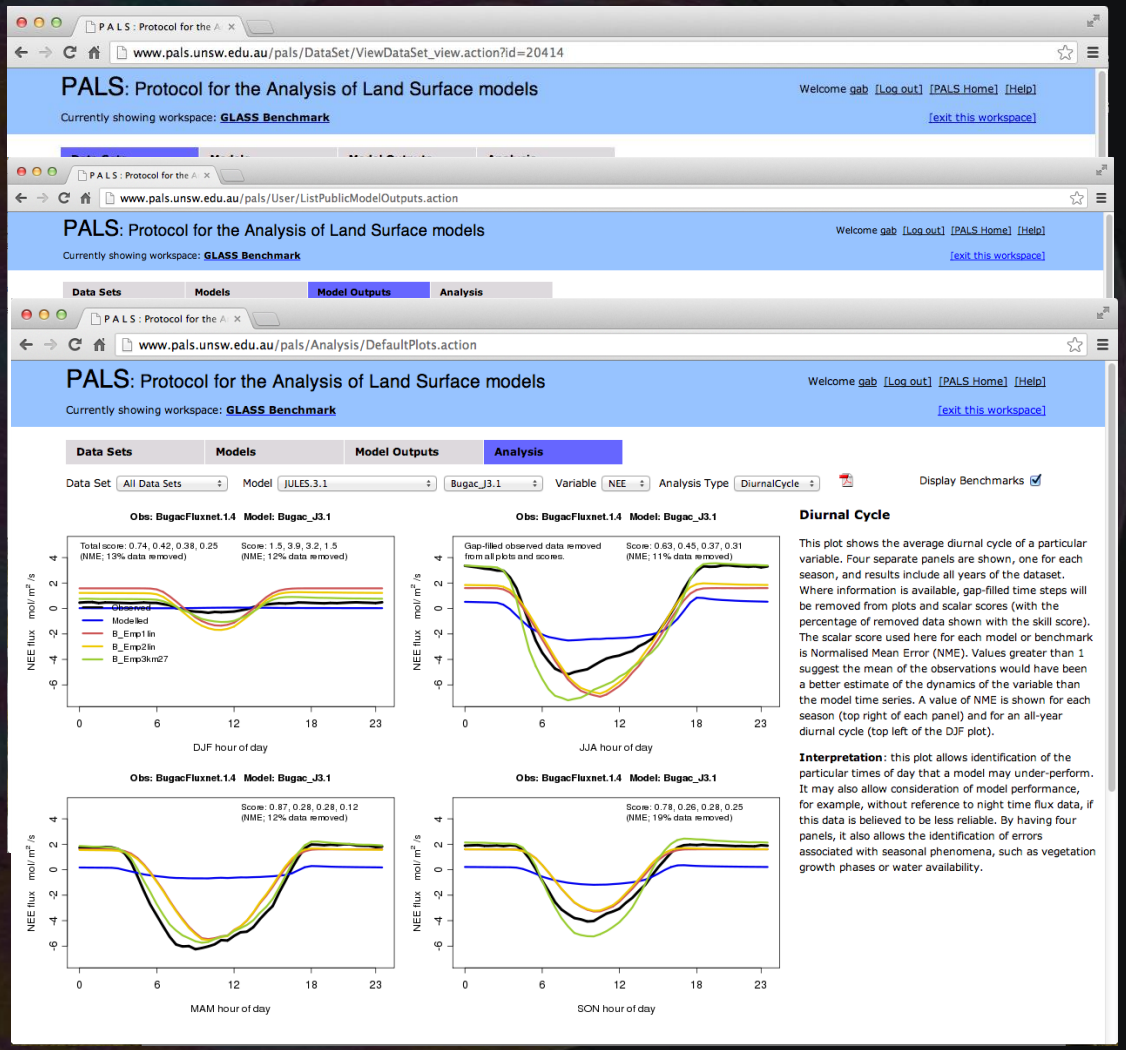
Gab Abramowitz

UNSW Australia /

ARC Centre of Excellence for Climate Systems Science

What is (was) PALS?

- A web application for evaluating land surface/ecosystem models
- PALS hosts Experiments:
  - Data sets required to drive/force a model for an experiment
  - Users run their models locally upload their model simulations for an experiment (including ancillary files)
  - PALS automatically runs analysis of the model output, comparing with evaluation data products, other models and empirical benchmarks



PALS motivations

- A place to run MIPs
- Do models share particular weaknesses?
- Can we understand why some models perform better than others in different environments?
- A platform to illustrate the value of benchmarking
- Around 230 users from 60+ institutions in 20 countries (~20% active)
- Facilitated two published MIPS – PLUMBER (Best et al, 2015; Haughton et al, in press) and SavMIP (Whitley et al, in review)

- Only worked with site-based flux tower data
- Down since 2014 – hacked due to Struts vulnerability.

What did we learn from PALS that was useful?

1. The importance of distinguishing between benchmarking and evaluation

Defining model performance expectations *a priori* – can lead to very different conclusions about model performance – Martin's talk next

2. Having a model evaluation package as a *web application*, rather than a collection of local scripts, is beneficial.

What if had an online environment that could utilise ILAMB, LVT and other evaluation packages in the same place?

A community web-based environment for model evaluation and benchmarking

- Not specific to any particular package / language (e.g. R, Python, NCL, Matlab etc all possible) – ILAMB, LVT, PALS
- Could be used privately as a model development tool or publicly as a MIP
  - Immediate sharing of results online
- Breadth of analysis, since automated
  - All plots could be viewed side-by-side in custom web pages – efficient sorting
  - potential for data mining (including ancillary data)
- Strict enforcement of provenance and ancillary data collection
  - Capture performance history throughout model development
  - Aid reproducibility
- Simplicity of MIP creation
  - MIPs are continuous and ongoing – rather than once every N years
  - Ability to include new analysis types retrospectively
- Imagine having GSWP phases, PILPS experiments, PLUMBER, GLACE (et al) data still available and analysable – quickly.

A community web-based environment for model evaluation and benchmarking

- Distributed architecture allows analysis to be co-located with big data:
  - 'Worker' nodes (e.g. R / Python analysis servers) can be installed locally across multiple locations, co-located with large data sets
  - 'Upload' of files to the system simply stores path: (a) if local worker node is present, files are not copied (b) local worker not present, files are uploaded
- Continuous integration testing of *science* in model, not just coding
  - API access could allow continuous integration testing elsewhere, e.g. Jenkins
- Requires standardised i/o protocols ALMA / CF / CMIP
  - May force adherence to standards
- All users would have equal access (no setup / local resources required)
  - Potentially many more users for participating packages

A community web-based environment for model evaluation and benchmarking

- We have started to design and build this
  - Thoughts on how it should be done?
  - Interested in collaborating? On the system itself? On getting a particular package in?
  - Poster A.1 details architecture and implementation – feedback please!
- Not specific to LSMs
- Will hopefully launch as modelevaluation.org
- It should help the community move forward much more quickly…