Protocol for the Analysis of Land Surface models (PALS)

Gab Abramowitz
UNSW Australia /
ARC Centre of Excellence for Climate Systems Science

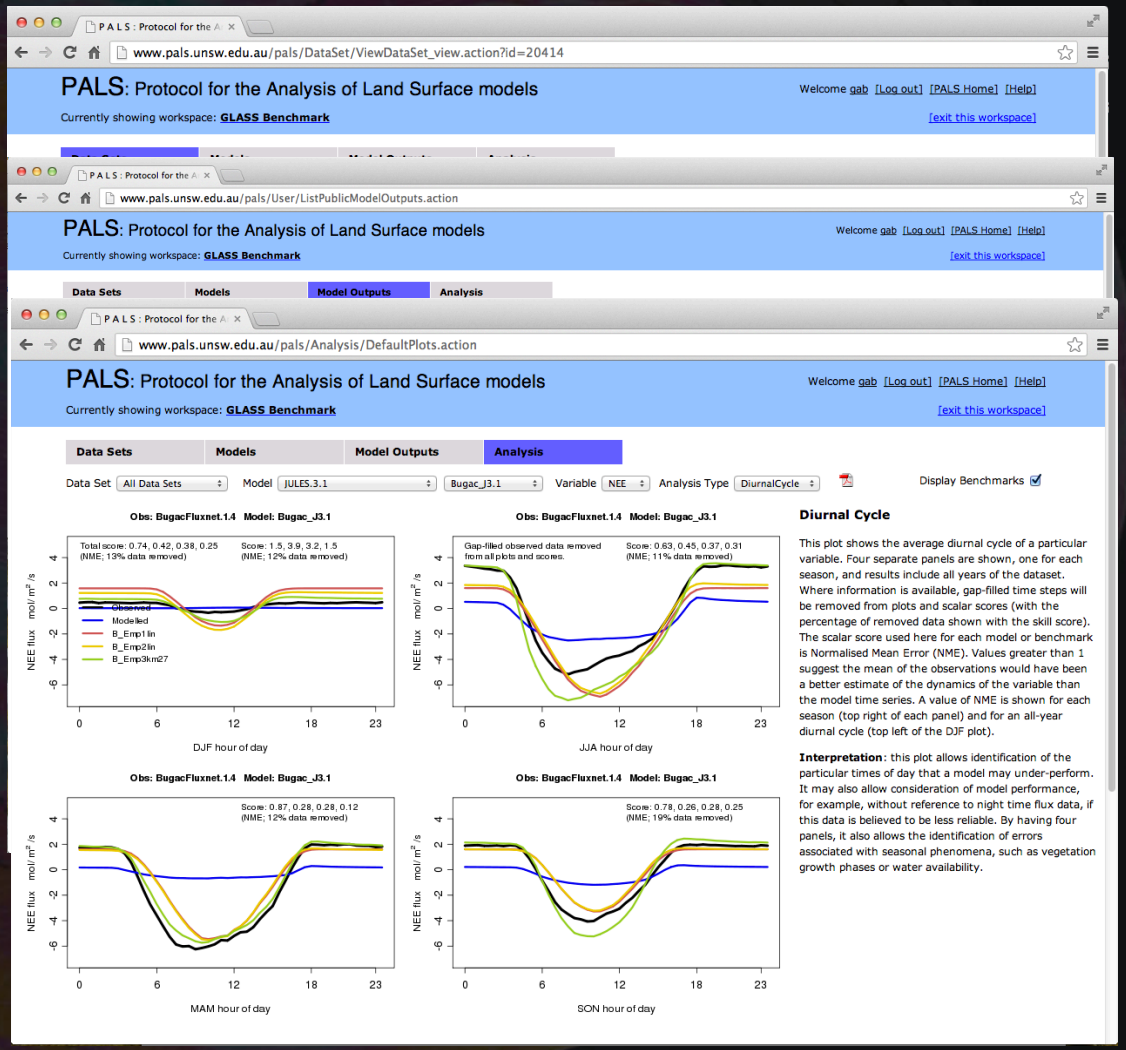
UNSW AUSTRALIA
ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE
TERN Terrestrial Ecosystem Research Network
NCI
GLOBAL LAND ATMOSPHERE SYSTEM STUDY GLASS

Outline

- PALS as it was
- What we learnt from it that was useful
- What's new in the 'PALS2' web application
- What's new in the 'PALS2' analysis suite
- Next steps
- Discussion and feedback

What is (was) PALS?

- A web application for evaluating land surface 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, 2016) and SavMIP (Whitley et al, 2016) – and model development within some LSM teams

- 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

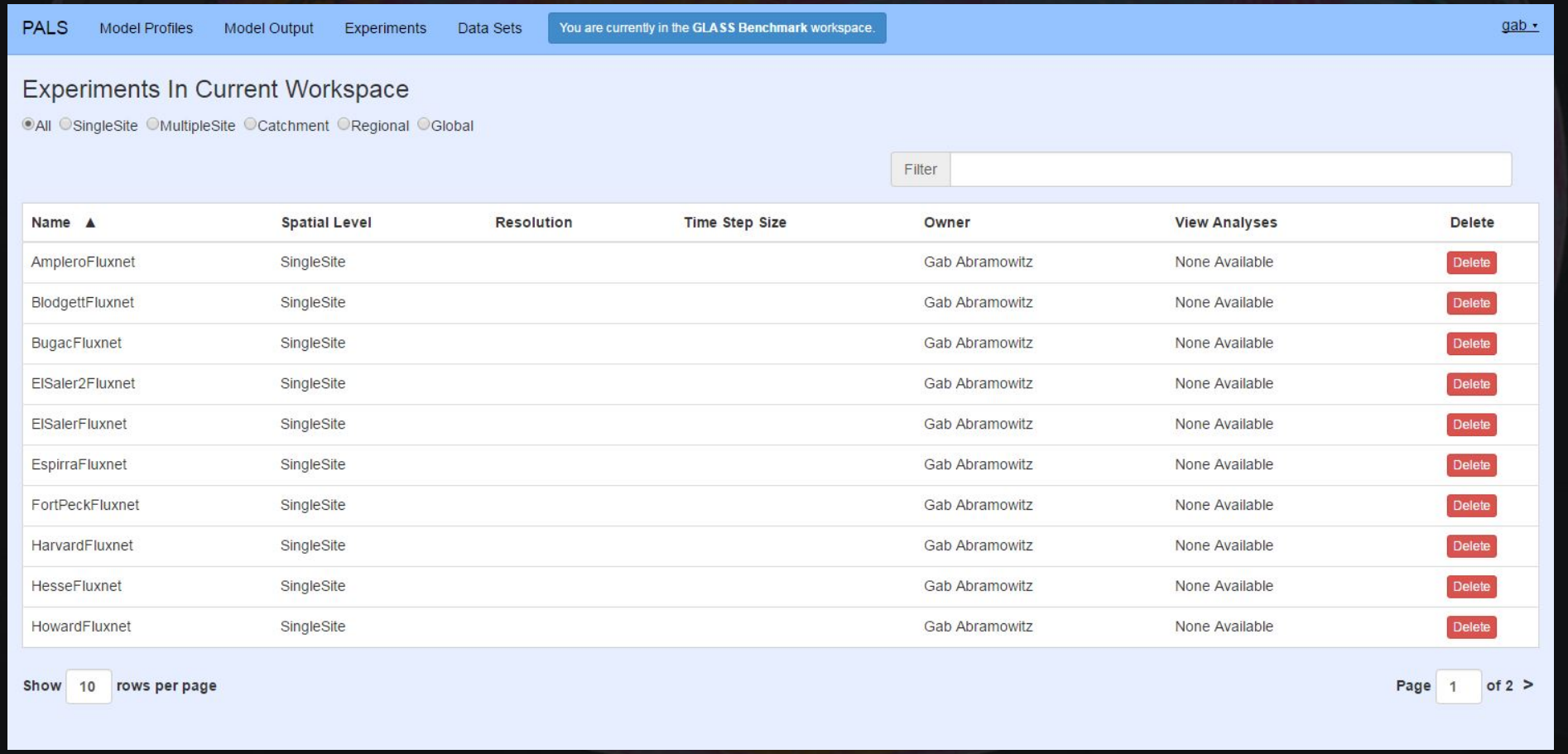
Defining model performance expectations *a priori* – can lead to very different conclusions about model performance

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

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

PALS 2 – “modelevaluation.org” – what’s new

- Currently in testing – release with basic functionality this year
- Not specific to LSMs – any modelling endeavour that uses reference datasets

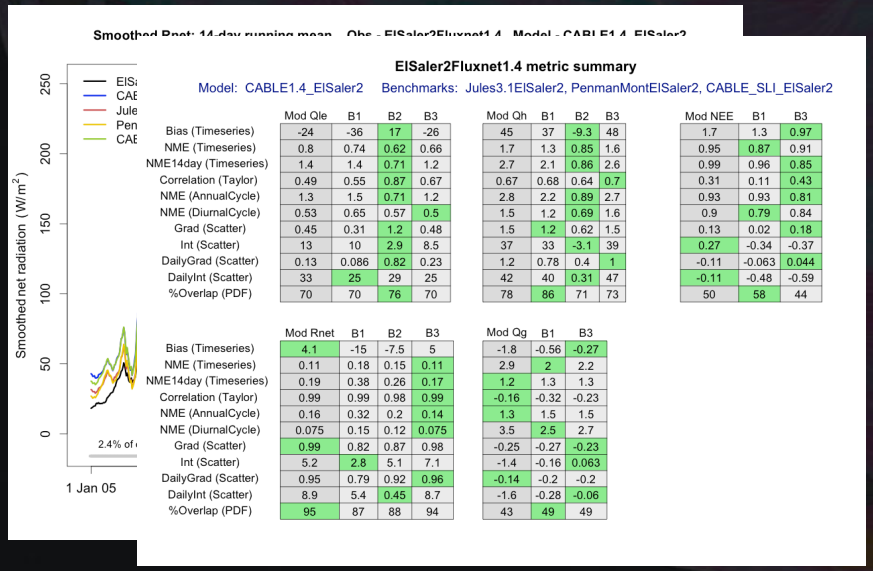


PALS 2 – “modelevaluation.org” – what’s new

Benchmarks

Now defined by the user – choice of up to 3 model outputs that has already been submitted to an Experiment:

- Previous model versions
- Other LSMs you’re jealous of
- Empirical benchmarks as before (but are now created and submitted manually)

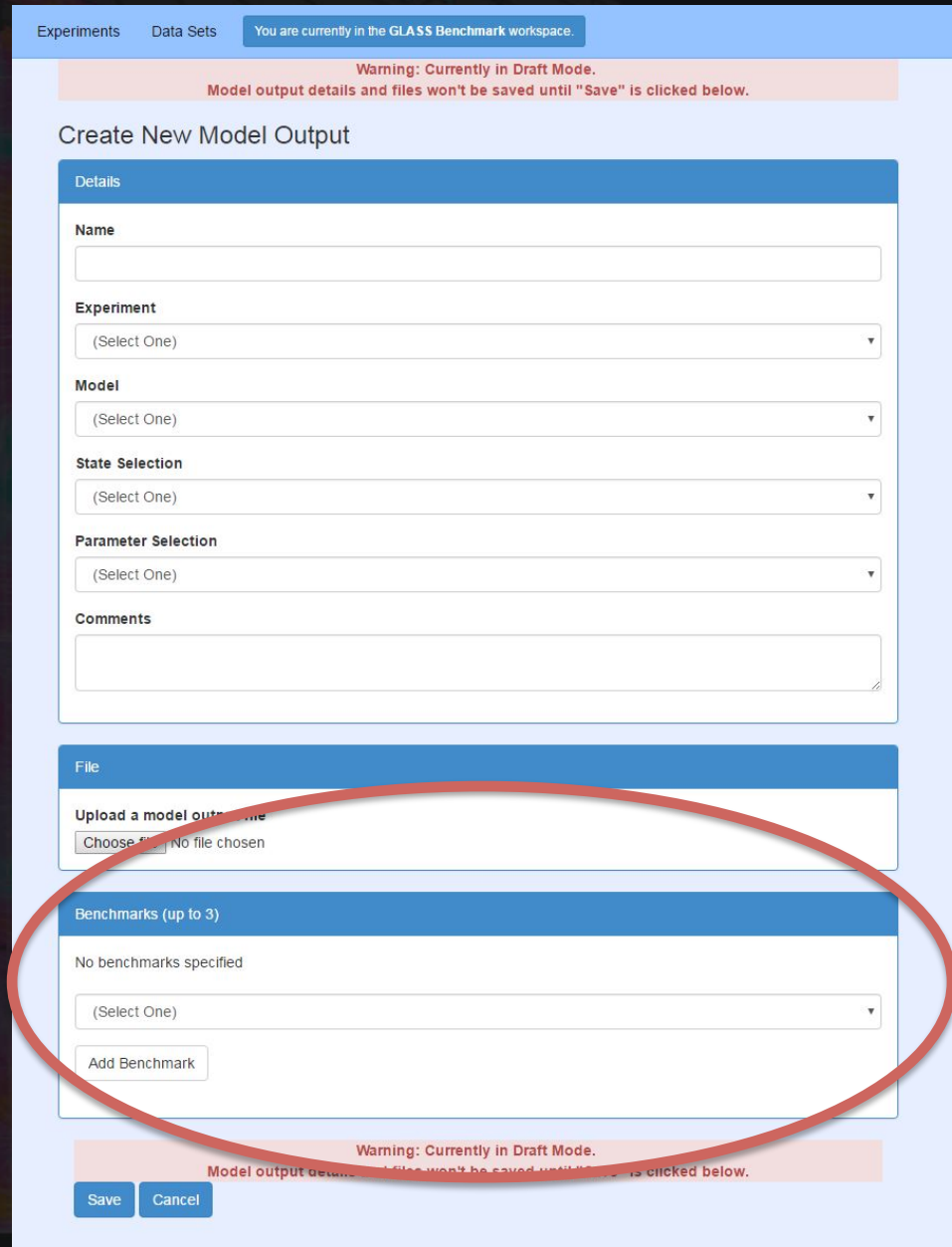


ElSaler2Fluxnet1.4 metric summary

Model: CABLE1.4_ElSaler2 Benchmarks: Jules3.1ElSaler2, PenmanMontElSaler2, CABLE_SLI_ElSaler2

	Mod Qle	B1	B2	B3	Mod Qh	B1	B2	B3	Mod NEE	B1	B3
Bias (Timeseries)	-24	-36	17	-26	45	37	-9.3	48	1.7	1.3	0.97
NME (Timeseries)	0.8	0.74	0.62	0.66	1.7	1.3	0.85	1.6	0.95	0.87	0.91
NME14day (Timeseries)	1.4	1.4	0.71	1.2	2.7	2.1	0.86	2.6	0.99	0.96	0.85
Correlation (Taylor)	0.49	0.55	0.87	0.67	0.67	0.68	0.64	0.7	0.31	0.11	0.43
NME (AnnualCycle)	1.3	1.5	0.71	1.2	2.8	2.2	0.89	2.7	0.93	0.93	0.81
NME (DiurnalCycle)	0.53	0.65	0.57	0.5	1.5	1.2	0.69	1.6	0.9	0.79	0.84
Grad (Scatter)	0.45	0.31	1.2	0.48	1.5	1.2	0.62	1.5	0.13	0.02	0.18
Int (Scatter)	13	10	2.9	8.5	37	33	-3.1	39	0.27	-0.34	-0.37
DailyGrad (Scatter)	0.13	0.086	0.82	0.23	1.2	0.78	0.4	1	-0.11	-0.063	0.044
DailyInt (Scatter)	33	25	29	25	42	40	0.31	47	-0.11	-0.48	-0.59
%Overlap (PDF)	70	70	76	70	78	86	71	73	50	58	44

	Mod Rnet	B1	B2	B3	Mod Qg	B1	B3
Bias (Timeseries)	4.1	-15	-7.5	5	-1.8	-0.56	-0.27
NME (Timeseries)	0.11	0.18	0.15	0.11	2.9	2	2.2
NME14day (Timeseries)	0.19	0.38	0.26	0.17	1.2	1.3	1.3
Correlation (Taylor)	0.99	0.99	0.98	0.99	-0.16	-0.32	-0.23
NME (AnnualCycle)	0.16	0.32	0.2	0.14	1.3	1.5	1.5
NME (DiurnalCycle)	0.075	0.15	0.12	0.075	3.5	2.5	2.7
Grad (Scatter)	0.99	0.82	0.87	0.98	-0.25	-0.27	-0.23
Int (Scatter)	5.2	2.8	5.1	7.1	-1.4	-0.16	0.063
DailyGrad (Scatter)	0.95	0.79	0.92	0.96	-0.14	-0.2	-0.2
DailyInt (Scatter)	8.9	5.4	0.45	8.7	-1.6	-0.28	-0.06
%Overlap (PDF)	95	87	88	94	43	49	49



PALS 2 – "modelevaluation.org" – what's new

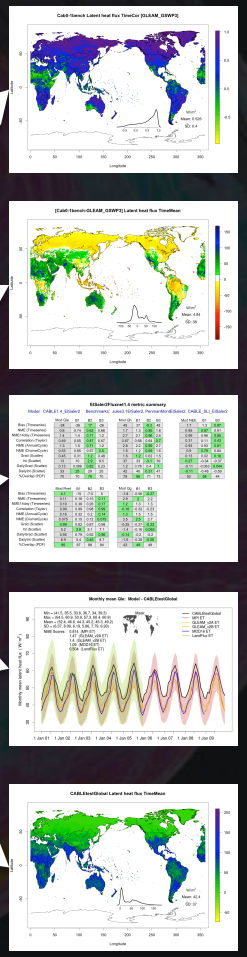
Not specific to any particular analysis package / language (e.g. R, Python, NCL, Matlab, Fortran etc all possible) – ILAMB, LVT, PALS

Information sent to the Analysis Script when a Model Output is uploaded to a given Experiment (including paths, meta-data):

- All **Data Sets** associated with the Experiment
- All **Model Outputs** that have been uploaded to the Experiment (within current Workspace)
- Model Output that is being uploaded and is triggering the analysis
- User's nominated **Benchmarks** associated with this Model Output

Master Analysis Script

(Editable by workspace owner)



PALS 2 – "modelevaluation.org" – what's new

Not specific to any particular analysis package / language (e.g. R, Python, NCL, Matlab, Fortran etc all possible) – ILAMB, LVT, PALS

Information sent to the Analysis Script when a Model Output is uploaded to a given Experiment (including paths, meta-data):

- All **Data Sets** associated with the Experiment
- All **Model Outputs** that have been uploaded to the Experiment (within current Workspace)
- Model Output that is being uploaded and is triggering the analysis
- User's nominated **Benchmarks** associated with this Model Output

Both JSON files

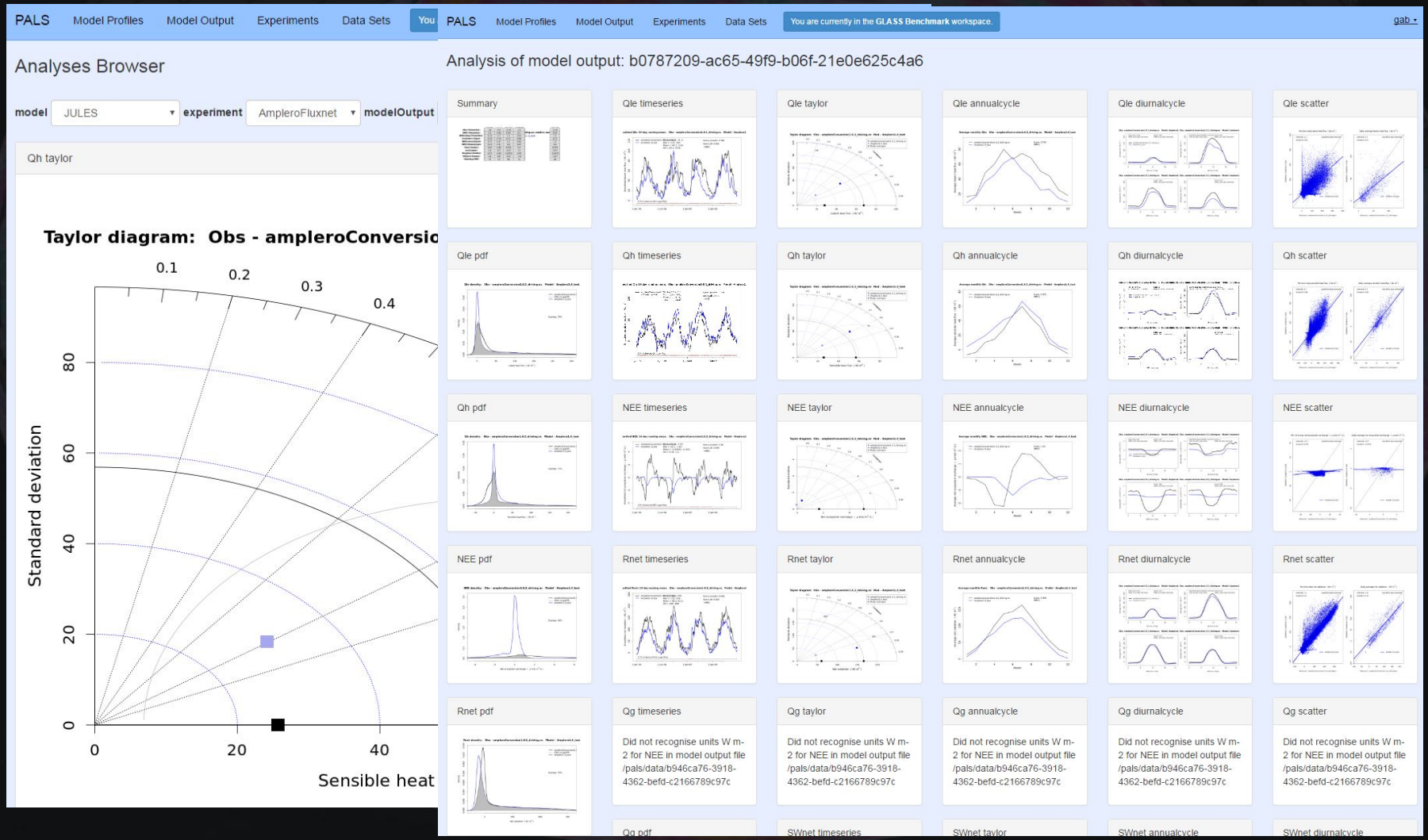
Each analysis returns:

- Image file path / location
- Variable name
- Analysis name
- Metric names and values
- Evaluation products used
- Model Outputs used
- Benchmarks used
- Analysis error messages
- Benchmark error messages

Analysis browser page uses these properties, as well as Experiment, Workspace to classify display panels

PALS 2 – “modelevaluation.org” – what’s new

Not specific to any particular analysis package / language (e.g. R, Python, NCL, Matlab, Fortran etc all possible) – ILAMB, LVT, PALS



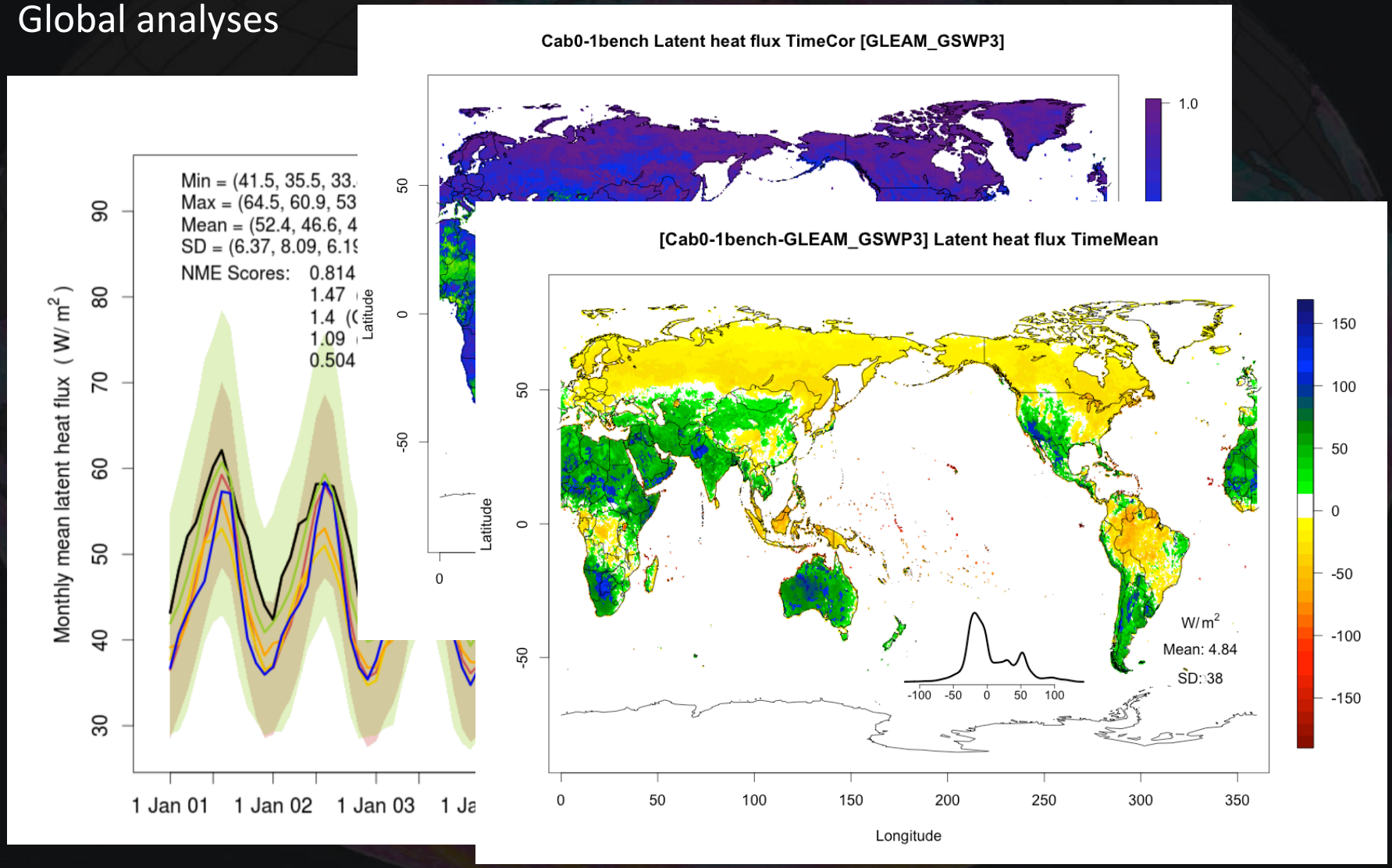
PALS 2 – "modelevaluation.org" – what's new

- Distributed architecture aims to allow analysis to be co-located with big model outputs:
  - 'Worker' nodes (e.g. R / Python analysis servers) can be installed on VMs 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

- Attempt to be increasingly strict about enforcement of provenance and ancillary data collection
  - Aid reproducibility
  - Capture what caused performance history throughout model development
  - Utilise ancillary / meta-data as part of automated analyses in MIPs

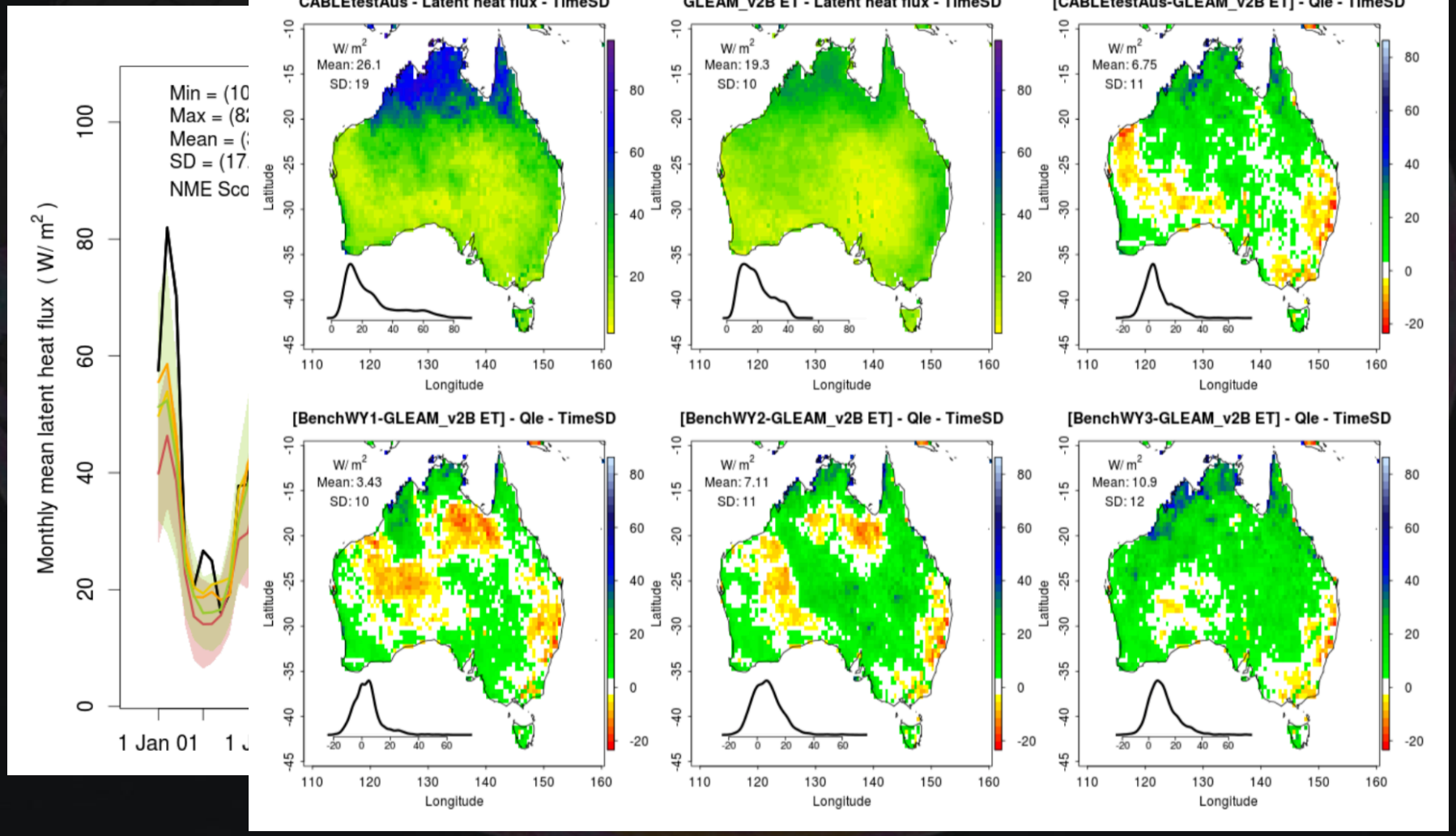
PALS 2 – “modelevaluation.org”: what’s new in the PALS R package

Global analyses



PALS 2 – “modelevaluation.org”: what’s new in the PALS R package

Regional analyses



PALS 2 – "modelevaluation.org": what's new in the PALS R package

Summary analysis

ElSaler2Fluxnet1.4 metric summary

Model: CABLE1.4_ElSaler2 Benchmarks: Jules3.1ElSaler2, PenmanMontElSaler2, CABLE_SLI_ElSaler2

	Mod Qle	B1	B2	B3
Bias (Timeseries)	-24	-36	17	-26
NME (Timeseries)	0.8	0.74	0.62	0.66
NME14day (Timeseries)	1.4	1.4	0.71	1.2
Correlation (Taylor)	0.49	0.55	0.87	0.67
NME (AnnualCycle)	1.3	1.5	0.71	1.2
NME (DiurnalCycle)	0.53	0.65	0.57	0.5
Grad (Scatter)	0.45	0.31	1.2	0.48
Int (Scatter)	13	10	2.9	8.5
DailyGrad (Scatter)	0.13	0.086	0.82	0.23
DailyInt (Scatter)	33	25	29	25
%Overlap (PDF)	70	70	76	70

	Mod Qh	B1	B2	B3
Bias (Timeseries)	45	37	-9.3	48
NME (Timeseries)	1.7	1.3	0.85	1.6
NME14day (Timeseries)	2.7	2.1	0.86	2.6
Correlation (Taylor)	0.67	0.68	0.64	0.7
NME (AnnualCycle)	2.8	2.2	0.89	2.7
NME (DiurnalCycle)	1.5	1.2	0.69	1.6
Grad (Scatter)	1.5	1.2	0.62	1.5
Int (Scatter)	37	33	-3.1	39
DailyGrad (Scatter)	1.2	0.78	0.4	1
DailyInt (Scatter)	42	40	0.31	47
%Overlap (PDF)	78	86	71	73

	Mod NEE	B1	B3
Bias (Timeseries)	1.7	1.3	0.97
NME (Timeseries)	0.95	0.87	0.91
NME14day (Timeseries)	0.99	0.96	0.85
Correlation (Taylor)	0.31	0.11	0.43
NME (AnnualCycle)	0.93	0.93	0.81
NME (DiurnalCycle)	0.9	0.79	0.84
Grad (Scatter)	0.13	0.02	0.18
Int (Scatter)	0.27	-0.34	-0.37
DailyGrad (Scatter)	-0.11	-0.063	0.044
DailyInt (Scatter)	-0.11	-0.48	-0.59
%Overlap (PDF)	50	58	44

	Mod Rnet	B1	B2	B3
Bias (Timeseries)	4.1	-15	-7.5	5
NME (Timeseries)	0.11	0.18	0.15	0.11
NME14day (Timeseries)	0.19	0.38	0.26	0.17
Correlation (Taylor)	0.99	0.99	0.98	0.99
NME (AnnualCycle)	0.16	0.32	0.2	0.14
NME (DiurnalCycle)	0.075	0.15	0.12	0.075
Grad (Scatter)	0.99	0.82	0.87	0.98
Int (Scatter)	5.2	2.8	5.1	7.1
DailyGrad (Scatter)	0.95	0.79	0.92	0.96
DailyInt (Scatter)	8.9	5.4	0.45	8.7
%Overlap (PDF)	95	87	88	94

	Mod Qg	B1	B3
Bias (Timeseries)	-1.8	-0.56	-0.27
NME (Timeseries)	2.9	2	2.2
NME14day (Timeseries)	1.2	1.3	1.3
Correlation (Taylor)	-0.16	-0.32	-0.23
NME (AnnualCycle)	1.3	1.5	1.5
NME (DiurnalCycle)	3.5	2.5	2.7
Grad (Scatter)	-0.25	-0.27	-0.23
Int (Scatter)	-1.4	-0.16	0.063
DailyGrad (Scatter)	-0.14	-0.2	-0.2
DailyInt (Scatter)	-1.6	-0.28	-0.06
%Overlap (PDF)	43	49	49

PALS 2 – “modelevaluation.org”: what’s new in the PALS R package

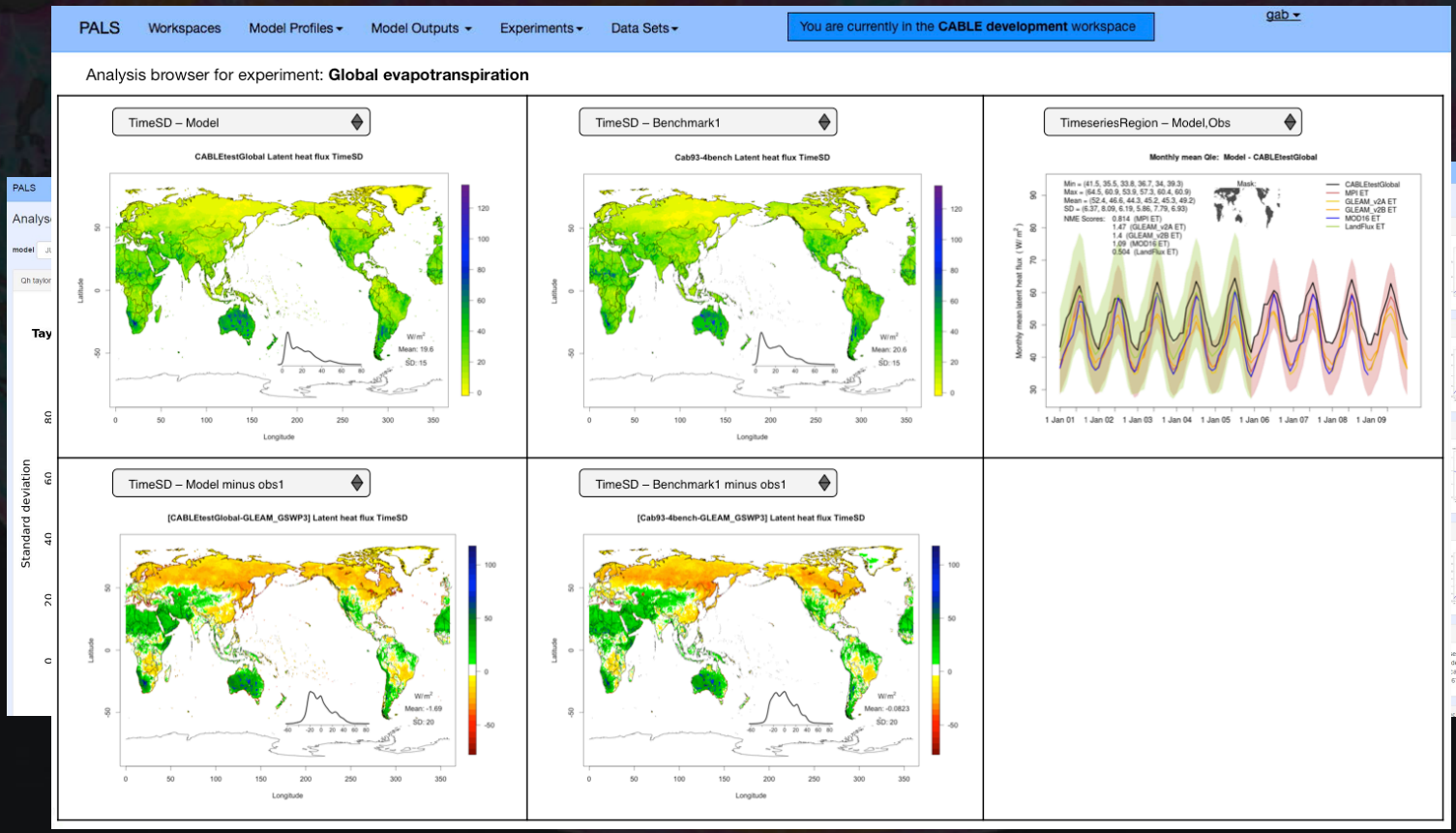
Much faster execution for single site analyses (~ 50 plots in ~ 5 seconds)

PALS 2 – "modelevaluation.org"

- As before – private/group use for model development or public use for MIPs
  - Immediate sharing of results online
- Flexible display of analyses
- Simplicity of MIP creation
  - MIPs are continuous and ongoing – rather than once every N years
  - Ability to include new analysis types retrospectively
- Potential e.g. for GSWP phases, PILPS experiments, PLUMBER, GLACE (et al) data still available and analysable – quickly.
- All users would have equal access (no setup / local resources required)
  - Greater opportunity for standardisation of evaluation

PALS 2 – “modelevaluation.org” – next steps

- Plans to visit ORNL (Nate, Forrest), GSFC (Sujay) and NCAR in December to work on ILAMB / LVT wrappers and display protocol.
- Resolve appropriate analysis sorting vs summary display

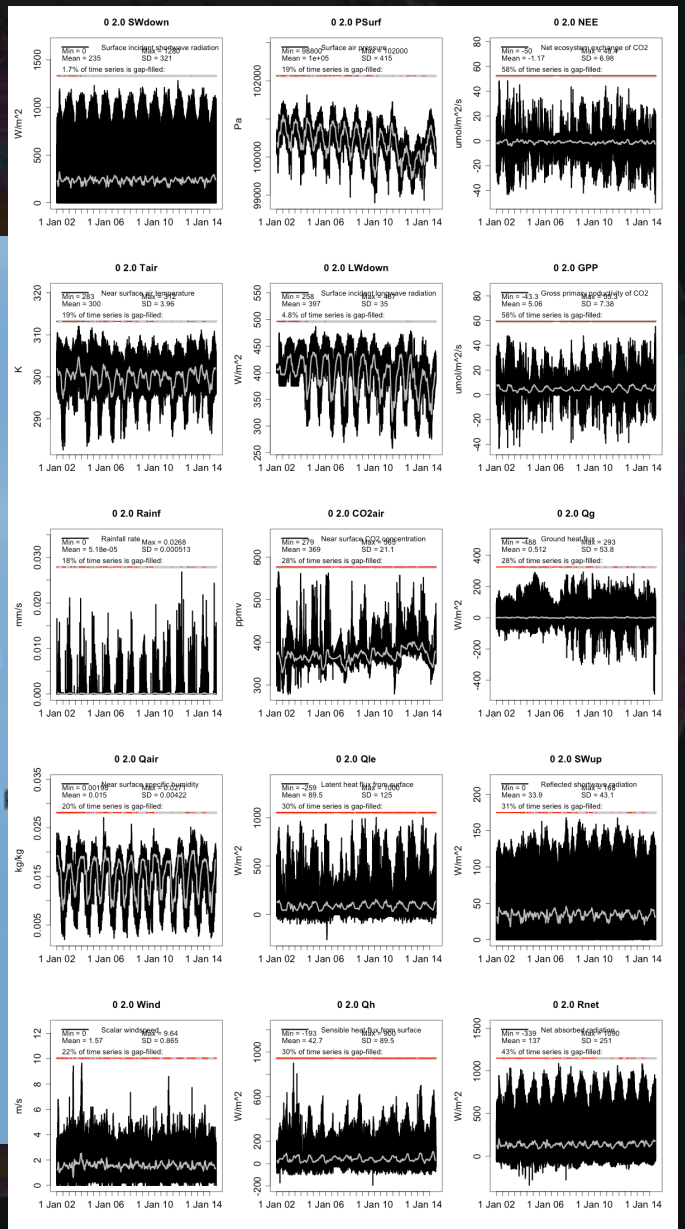


PALS 2 – "modelevaluation.org" – next steps

- Plans to visit ORNL (Nate, Forrest) and NCAR (+ spent time with Sujay) in December to work on ILAMB / LVT wrappers and display protocol.
- Complete and test distributed architecture
- Resolve appropriate analysis sorting vs summary display
- API development:
  - continuous integration testing of *science* in model, not just compilation and coding standards – possible integration with e.g. Jenkins
- Relies heavily on adherence to ALMA / CF / CMIP
  - Implement next ALMA phase into PALS2 database (perhaps separate API for this)
  - Expand scope anthropogenic water, urban, agriculture, BGC, non-grid catchment data?
  - How best to promote adherence?

PALS 2 – “modelevaluation.org” – next steps

- Already processed and QC’d 20+ OzFlux sites (especially the Northern Australian Tropical Transect):



Plan to keep site QC plots in PALS2 -->

PALS 2 – “modelevaluation.org” – next steps

- Already processed and QC’d 20+ OzFlux sites (especially the Northern Australian Tropical Transect).
- Process and QC Fluxnet2015 release.
- Much larger scale PLUMBER *actually* analysed through PALS2?
- Continue global, regional scale analyses in PALS analysis package

Questions

1. What kind of model / simulation meta-data would we want (especially for data mining in analyses)? How can we best formalise/codify this?

2. How can we bring the community along with ALMA netcdf and its expansion to hydrology, urban, agriculture and BGC?

3. How might we encourage groups to take community ownership and work together to co-develop this type of system?

4. Any thoughts on targeting funding to accelerate development?
   - ARC Linkage grants
   - Infrastructure, rather than research funding streams