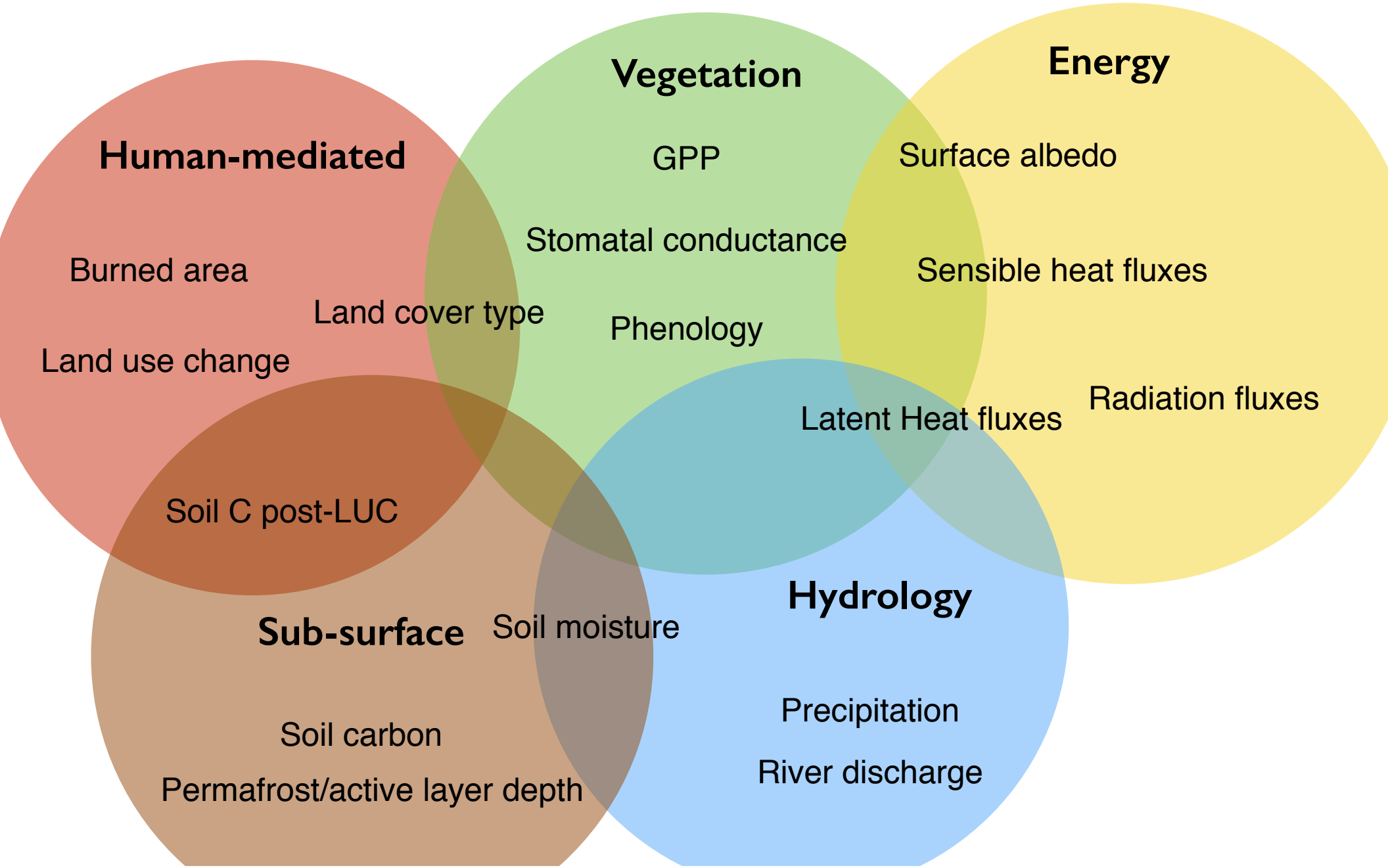
Summary of Evaluation Methods at Modeling Centers

Gretchen Keppel-Aleks
University of Michigan
gkeppela@umich.edu

Acknowledgements:
Martin Best, Victor Brovkin, Tomohiro Hajima, Nancy Kiang, Randy Koster, Dave Lawrence, Bill Riley, Yinping Wang, Qian Zhang

Evaluating land model processes requires diverse metrics



Does your modeling center have its own software package for evaluating land model output?

Most models have their own package for evaluating output

- one "no" response

Few responses indicated that metrics were quite general, rather than specific focus on land

Does the package include quantitative metrics/scores?

Roughly equal yes/no responses

To what extent do you rely on qualitative (expert judgment) versus quantitative comparisons of models and observations?

Most modeling centers (6) rely roughly equally on qualitative and quantitative comparisons

Two modeling centers rely more on quantitative metrics

Important caveat is data quality

Selecting metrics is based on both observational characteristics and importance of constraint

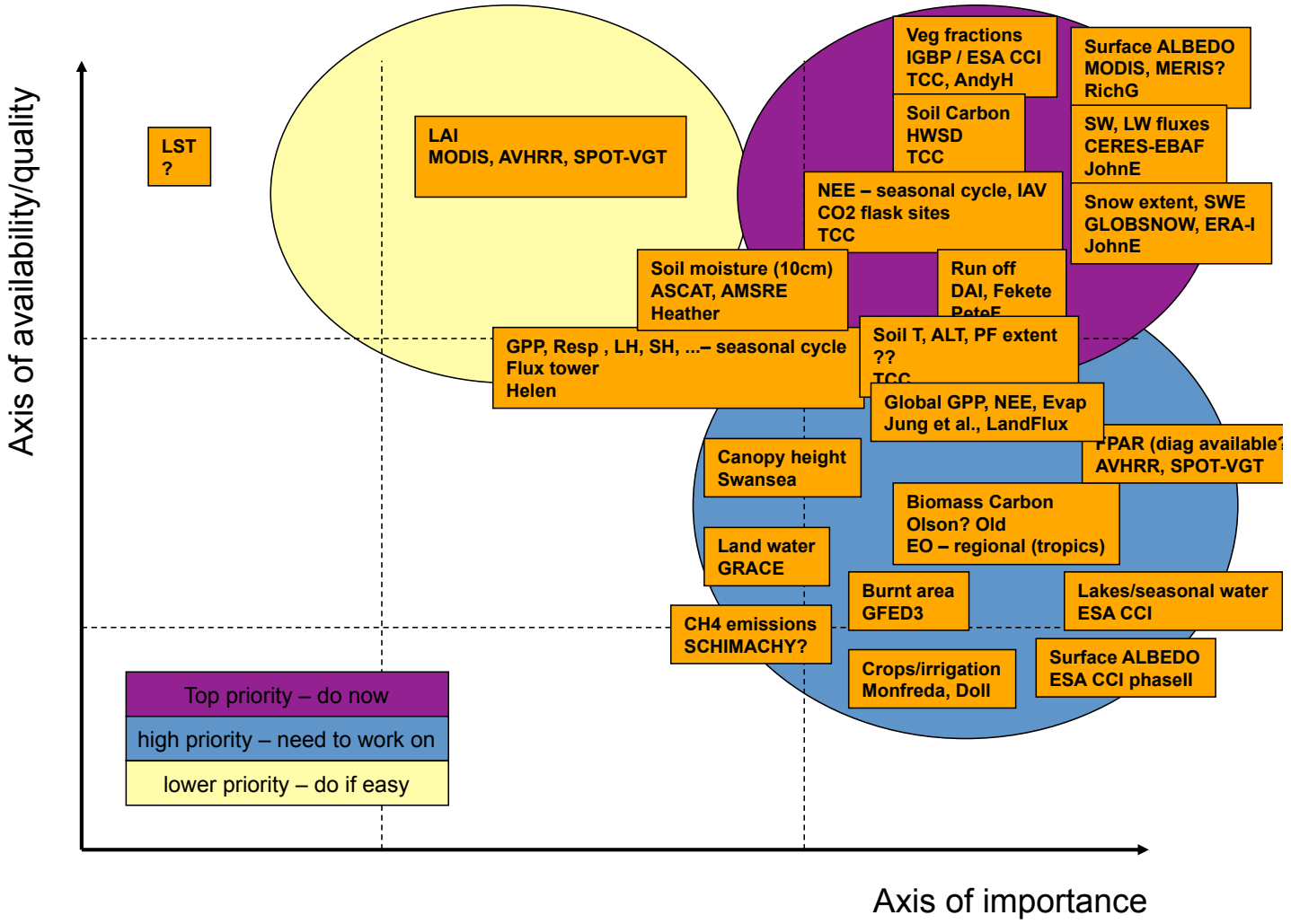


Figure Courtesy: Martin Best

Spatial comparisons leverage gridded observations to quantify regionally coherent biases

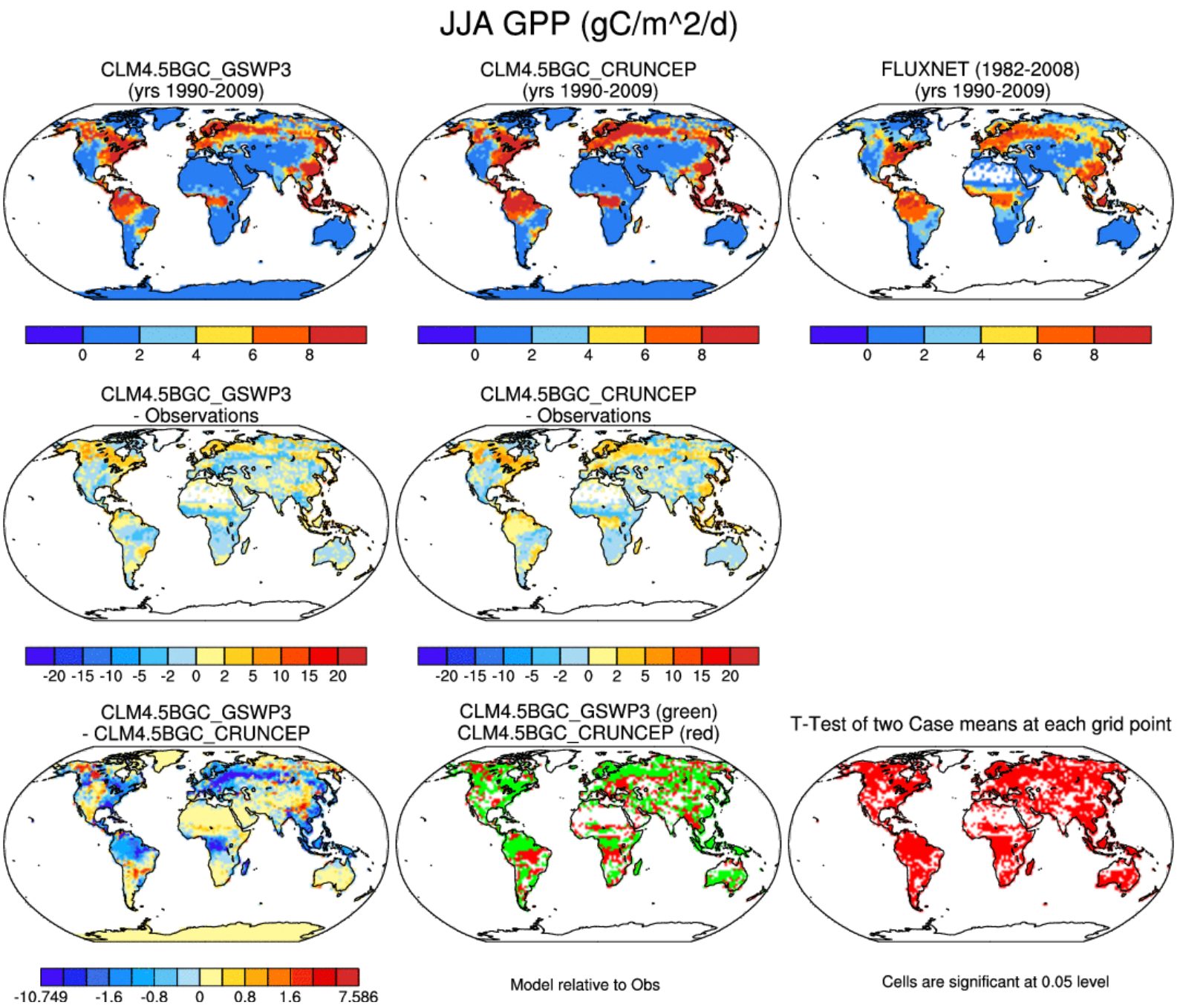


Figure Courtesy: Dave Lawrence

Do you primarily develop/evaluate your land model in uncoupled or coupled mode? (i.e, do you develop/evaluate in uncoupled mode and then couple or do you develop/evaluate primarily in coupled mode?)

Most modeling centers develop sequentially

- First focus on uncoupled simulations
- Subsequent adjustments and tuning for coupled simulations

One modeling center said rarely was model development/ evaluation done for uncoupled mode

One center reported that development/evaluation was primarily conducted only for uncoupled mode

Uncertainty in forcing data represents leaves significant imprint on and produces large uncertainty for land output

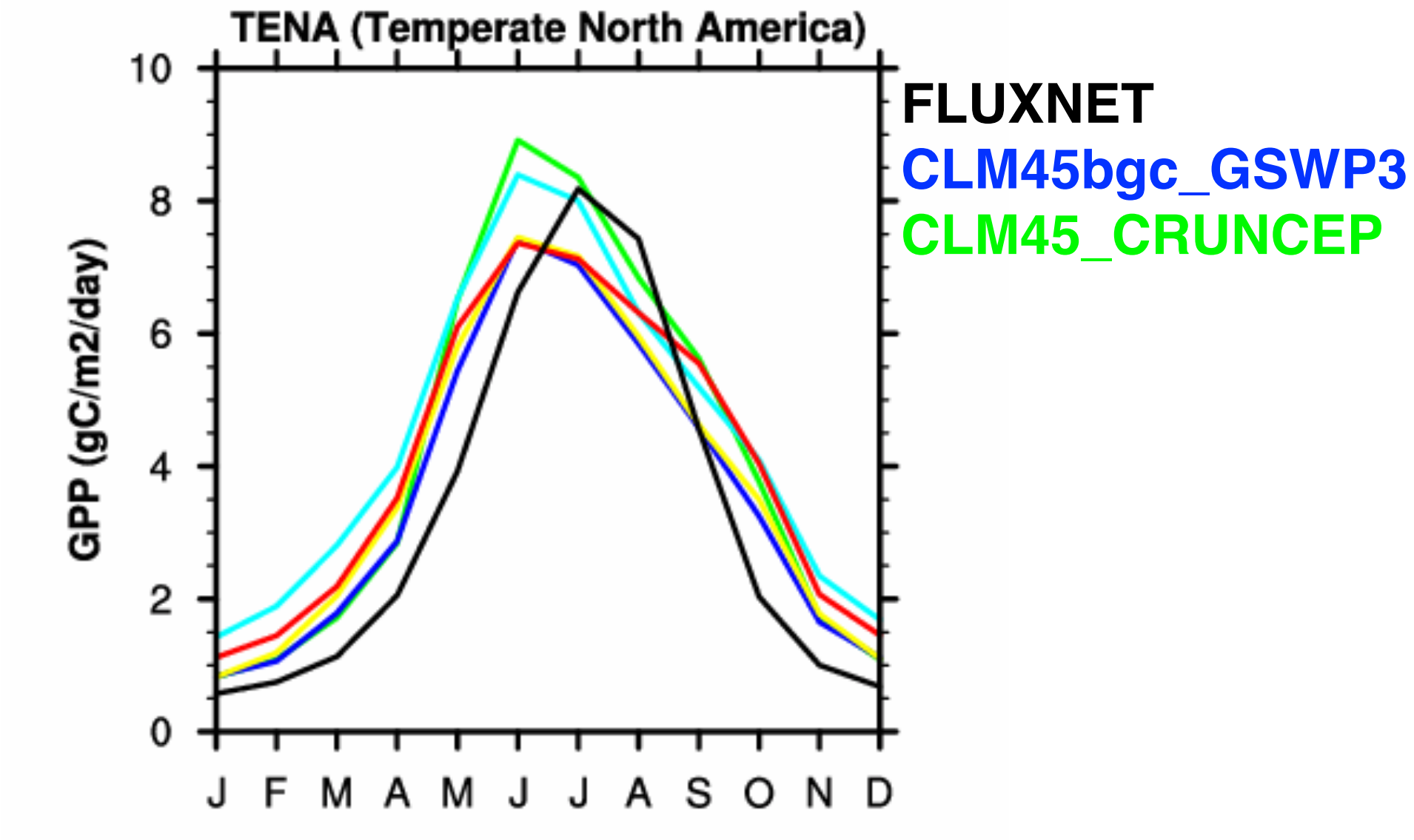


Figure Courtesy: Mingquan Mu

ESMs run in fully coupled mode won't capture the timing of interannual variability in the observations

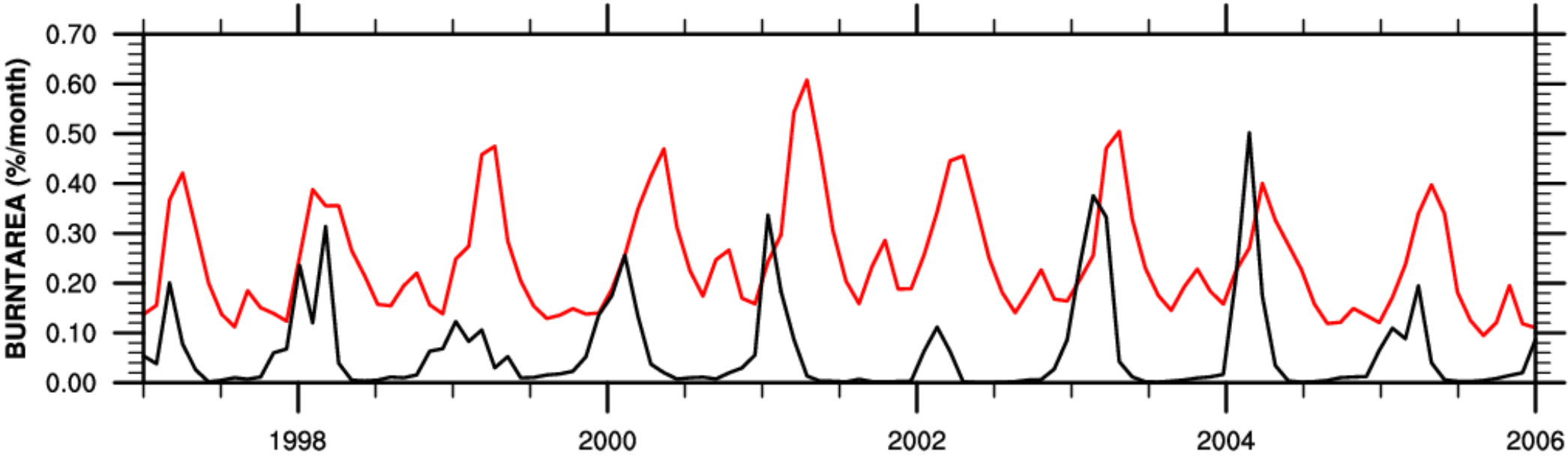
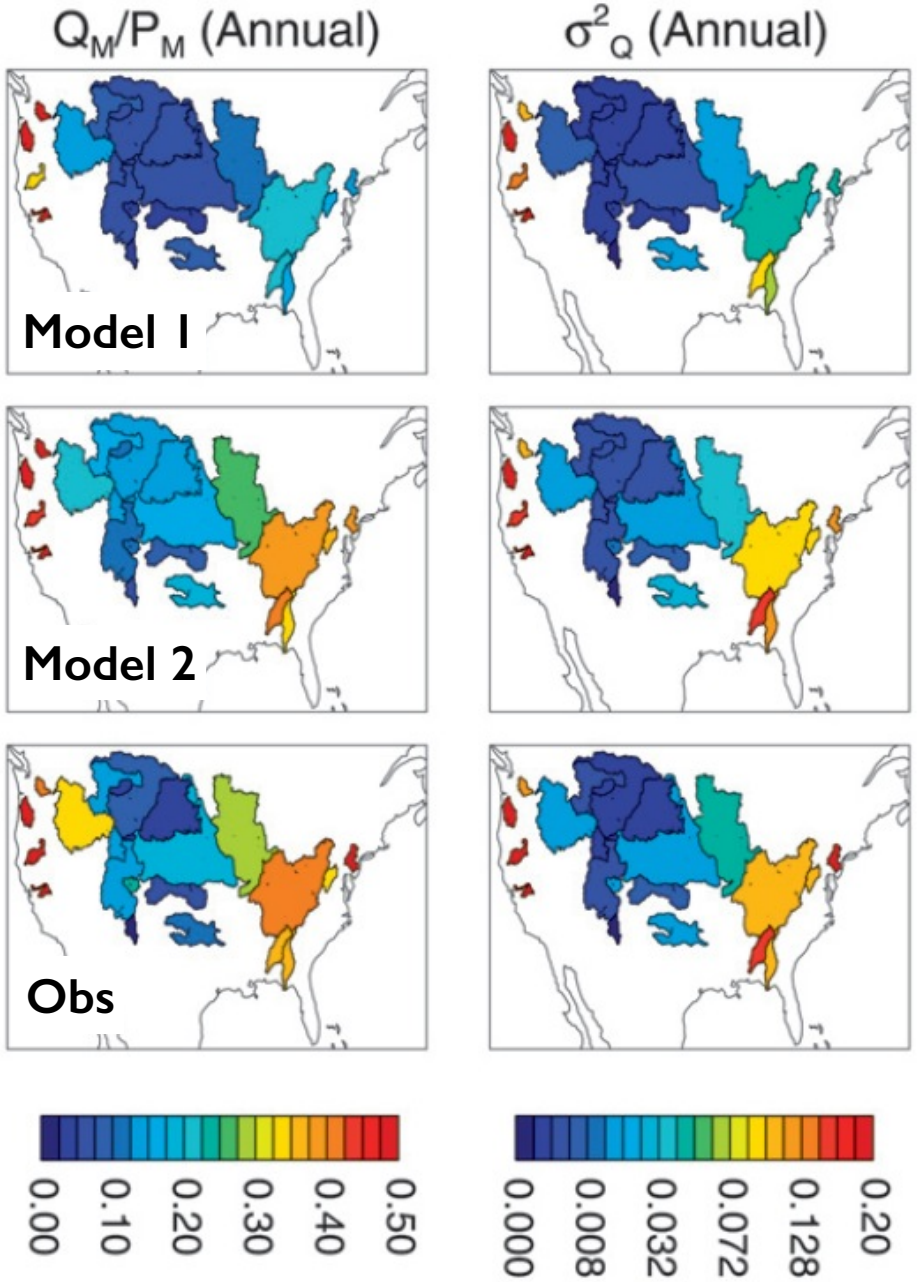


Figure Courtesy: Mingquan Mu

Analysis of variability in models does not require timing of anomalies to coincide with observations



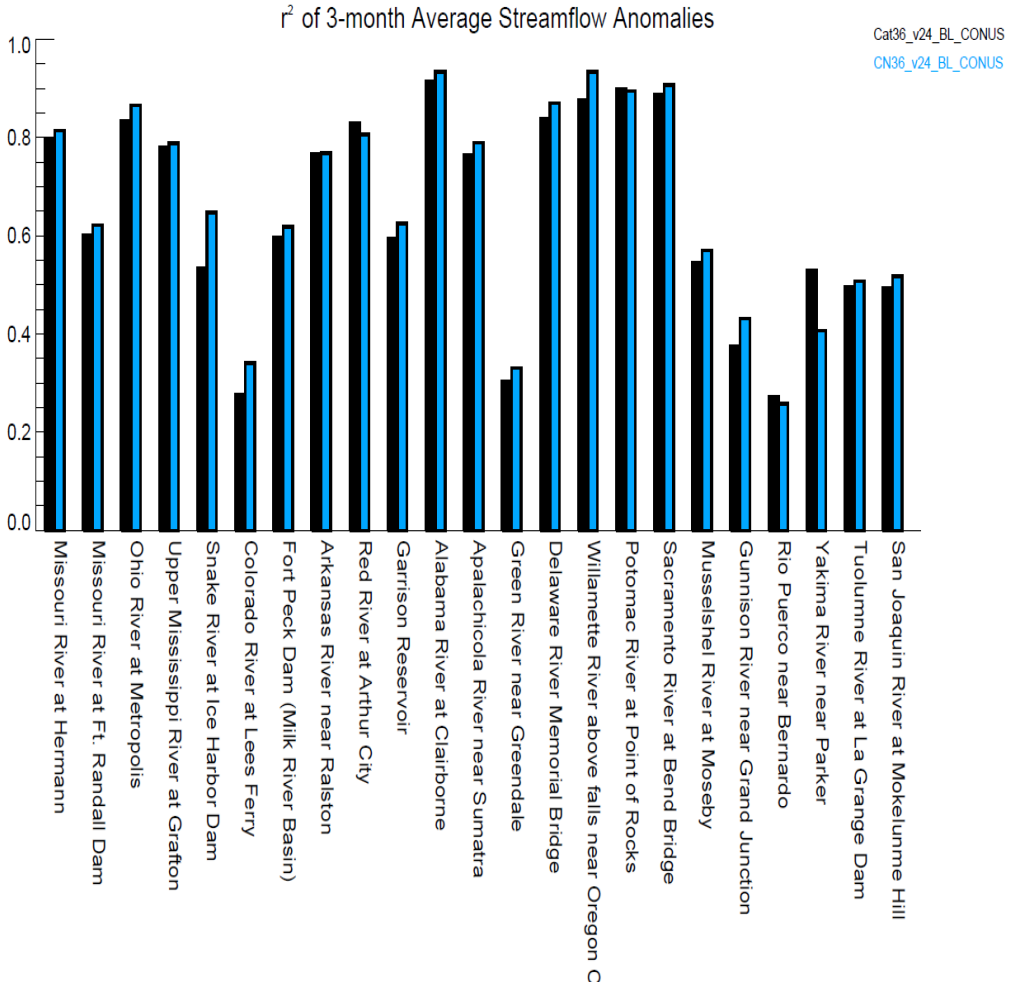


Figure Courtesy: Randy Koster

Do you use your evaluation package for
- formal model calibration
- to help with tuning
- as a diagnostic of model errors
- to aid in model analyses?

Diagnosing errors was leading response (4)

Tuning model parameters was second response (3)

Aid with model analyses (2)

Do you use an externally developed software package to evaluate your model?

Generally no sharing of packages

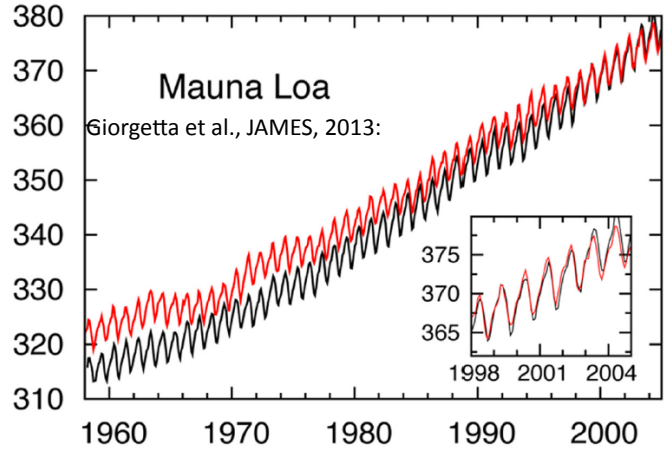
A few modeling centers are using ILAMB

Several modeling centers desire better integration of their system with others

A diversity of software is used for analysis:

NCL, Ferret, Fortran, R, Python

Open-source benchmarking packages shared across centers may facilitate better data-model integration



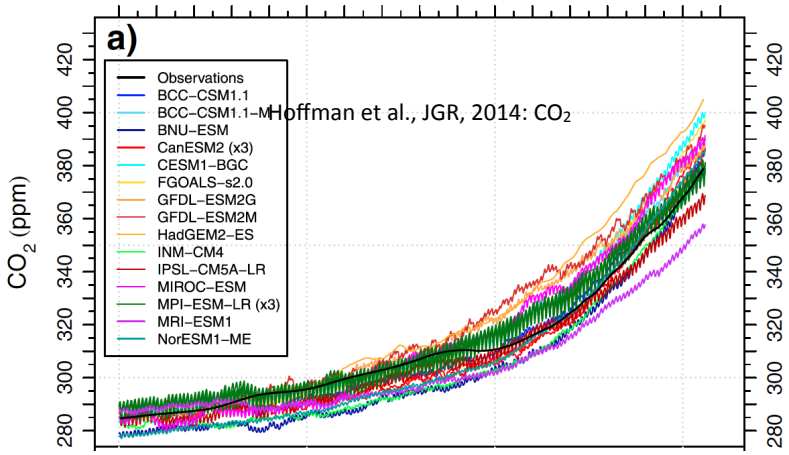


Figure Courtesy: Victor Brovkin

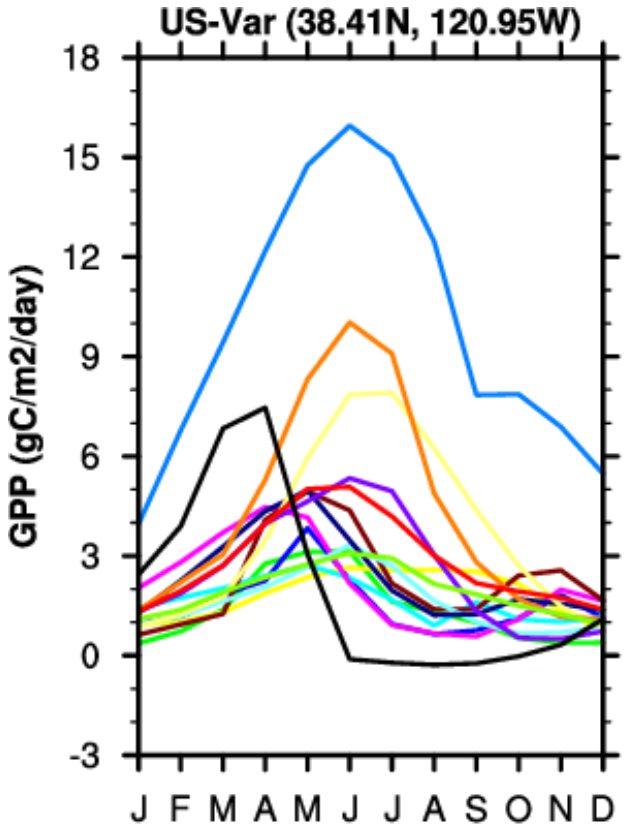


Figure Courtesy: Mingquan Mu

Discussion

An integrated framework will facilitate quantitative benchmarks

weighting of various spatial, temporal, and variability

Optimizing integration of benchmarking into work flow for model development remains a challenge

New metrics (e.g., functional response) may facilitate benchmarking across biogeochemistry, biogeophysics, and their drivers