Description of PILPS 2g Experiment: Model Comparison over Semi-Arid Areas

Luis A. Bastidas[1,2]
Bart Nijssen[2,3]
William Emmerich[4]
Hoshin V. Gupta[2,3]
Eric Small[3,5]

Proposed to the GLASS Science Panel
October 2002

[1] Department of Civil & Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan, Utah
[2] SAHRA – NSF Science and Technology Center for Sustainability of Semi-arid Hydrology and Riparian Areas, University of Arizona, Tucson, Arizona
[3] Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona
[4] U.S. Department of Agriculture, Agricultural Research Service, Tucson, Arizona
[5] Department of Geosciences, University of Colorado, Boulder, Colorado

1. INTRODUCTION

1.1 *Objective of PILPS 2g*

The PILPS experiments conducted so far have been important for the development and evaluation of land surface models developed by different research groups *[Pitman et al, 1993, Henderson-Sellers et al., 1993, Henderson-Sellers et al., 1995]*. However, these experiments have not included any on semi-arid lands, despite the fact that they constitute 1/3 of the global land surface of Earth. It is imperative, thus, to carry out a PILPS experiment over semi-arid lands.

The PILPS 2g experiment proposed here is an initiative within the GEWEX/GLASS (Global Land Atmosphere System Studies) panel. The objective of this study is the comparison of models that simulate water, energy, and CO_2 cycles with continuous observations at five different sites.

The availability of 4+ years of data at two sites and data from locations with similar vegetation coverage but hundreds of kilometers apart provide an exciting opportunity for cross-validation of the model results and for comparison of different models. The three different vegetation types existing at the data sites also provide a quick look of the diversity of environments in arid lands and will allow to establish whether or not further distinction is required to better represent the water, energy, and CO_2 exchanges taking place over such areas.

In previous PILPS studies *[Lettenmaier et al., 1996]*, it was shown that the calibration of model parameters yielded improvement in the models performance. For this reason, we propose to use the multi-criteria framework and a set of optimization codes for calibration of hydro-meteorological models that has been developed and successfully applied to a variety of land surface models at the University of Arizona [*Gupta et al, 1998, 1999; Bastidas et al., 1999, 2001, 2002; Vrugt et al., 2002*]. This framework is very appropriate for constraining the parameter estimation of land surface models to be consistent with observations and will allow for a comparison of "optimal" performances of the models. However, the use of this multi-criteria framework is not compulsory and the participants may carry out parameter estimation in the way they see fit.

Some of the science questions to be addressed by the PILPS 2g experiment are:

- What is the ability of the models to reproduce the water, energy, and carbon exchanges in semi-arid environments?
- Are the current (usually single) representations of semi-arid lands in the models enough to reproduce the different environments that exist in those areas?
- Does model calibration reduce the among-model range in the model simulations?
- How much influence does the model parameterization have on the parameter estimations of "physically meaningful" parameters?
- Do current carbon representations, developed for forests, properly reproduce carbon exchanges over vegetated arid lands?

It is clear the, that the proposed experiment has unique characteristics, not only due to the difference in environment from the previous PILPS experiments but because of the proposed use of appropriate system methods for parameter estimation that will help the

different groups in identifying parameter sets that make the models consistent with the data.

PILPS 2g is open to models with and without a representation of carbon fluxes. To guarantee comparisons under similar conditions, all participants will be required to carry out calibrations/optimizations that do not use carbon flux information. Modeling groups that represent carbon processes will be required to perform an additional set of calibration experiments to evaluate the changes and potential *improvements* due to inclusion of the carbon information.

1.2 *Description of the sites and instrumentation*

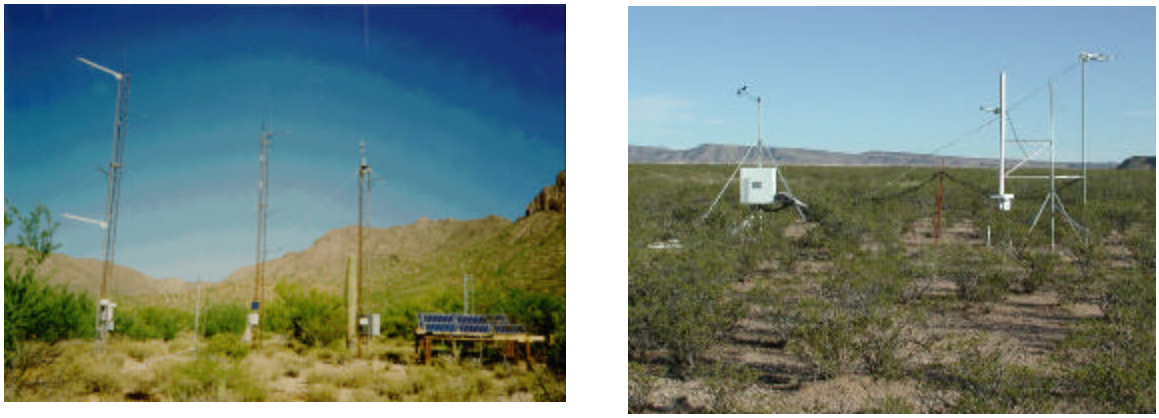
The proposed experiment will be carried out at five different sites located within the semi-arid Southwest USA, in the states of Arizona (3 sites) and New Mexico (2 sites) (See Figure 1 for locations). Two of the sites, Lucky Hills and Sevilleta Shrub, have a shrubby vegetation coverage with predominant species Acacia (*Acacia constricta*), tarbush (*Flourensia Cernua*), creosotebrush (*Larrea divaricata*), and desert zinnia (*zinnia pumila*). The Kendall and the other Sevilleta site are grasslands with predominant species sideoats gramma (*Bouteloua curtipendula*), black gramma (*Bouteloua eriopoda*), harry gramma (*Bouteloua hirsuta*) and lehmann lovergrass (*Eragrostis lehmanniana*). The Tucson site has a shrubs, grass, and saguaro cacti.

The data for the Lucky Hills and Kendall sites has been collected by the USDA-ARS Tucson from January 1996 till December 2000 using a Bowen ratio system with a tower height of 3 m *[Emmerich et al, 2002]*. It includes measurements of sensible and latent heat fluxes, CO_2 flux and soil temperature. The data from the Tucson site was collected by Jim Shuttleworth's group of the University of Arizona from May 1993 to June 1995 with and eddy covariance system on a tower 10 m height *[Unland et al, 1996]*. The measurements are of sensible and latent heat, and soil temperature. The data at the Sevilleta sites was collected by Eric Small of the University of Colorado with tower heights of 10m. Measurements include sensible, latent, and CO_2 fluxes, soil temperature and soil moisture at 5 cm depths.

Site	Longitude West	Latitude North	Elevation [m.a.s.l.]	Precipitation [mm/year]	Temperature [°C]
Lucky Hills Shrubland	110°03'05''	31°44'37"	1372	340	18.6
Kendall Grassland	109°56'28"	31°44'10"	1526	340	19.3
Tucson Shrub/cacti	111°49'48"	32°13'01"	730	305	20.2
Sevilleta Grassland			1730	270	17.2
Sevilleta Shrubland			1776	270	16.9

Measurements of the vegetation coverage and descriptions of the soil types at all the sites are also available. The soils tend to be coarse loams with high clay content. The detailed information will be provided to the participants.

Lucky Hills and Kendall sites



Tucson and Sevilleta sites

Grass at Sevilleta site.

2 EXPERIMENTAL SETUP

2.1 *General*

The experiment is focused on data obtained from observational towers that are located comparatively near to each other (1-5 km) to hundreds of kilometers apart, but within similar environments. We propose a set of offline experiments that will allow for a series of "cross-validations" or evaluations of the model performance. As an innovation, we propose the use of optimization routines for the identification of "calibrated/optimal" parameter sets for all the models. The model intercomparison will be among the "optimal" performances of the models. The optimizations will be carried out within the multi-criteria framework developed at the University of Arizona, which will provide the computer codes and training for such exercise. The participants are not obliged to use this framework for their parameter estimation procedures, if they so choose. The initialization of the models for all the cases will be left to the participants. To assess the impact of calibration in the model performance a *"default"* parameter set and the associated model run will be requested. The evaluation of the models will be carried out at annual, monthly, daily, and 30 minute time scales.

2.2 *Proposed intercomparison runs*

Because of the availability of 4 years of data, at the Lucky Hills and Kendall sites, that include both a "wet" and a "dry" year they can be used for performing temporal "split sample" tests of the model performances. At the same time, the availability of the New Mexico Sevilleta sites, with similar soil and vegetation characteristics, but hundreds of kilometers away, allow for spatial split sample testing and for checking the transferability of the parameters.

2.2.1 *Single location temporal "split-sample" test.*

The participants will be provided with a subset of the data that will contain wet and dry periods for the calibration of their models. The subsets will be both from the Lucky Hills and the Kendall data sets. The models will be compared using the provided data subset, the non-provided subset, and the whole set at each of the locations. The Tucson site data set will be also provided for the participants to check their parameter estimates with different vegetation coverage. The participants will be provided with the forcings for all the periods. However, the outputs to be used for model evaluation, i.e. latent and sensible heat fluxes, CO_2 fluxes, ground temperatures, and soil moisture, will be provided only for the non-evaluation periods. This part of the experiment will help in establishing the consistency of the parameter estimation procedures and the consistency of the models under different forcing conditions.

2.2.2 *Spatial "split-sample" test, transferability of parameters*

The Sevilleta shrub and grassland sites will be used to evaluate and compare the model performances when they use the parameters obtained from the Lucky Hills and

Kendall sites respectively. This exercise will check for the assumed general behavior of arid lands in the models and for the transferability of parameter estimates in similar but distant conditions. The availability of soil moisture measurements at the Sevilleta sites will allow for the testing of the consistency in the model estimation of state variables that were not used for the parameter estimation procedures.

2.2.3 *Carbon flux simulations*

A separate evaluation will be carried out for the carbon simulations using the same schemes of "split-sample" tests, i.e. temporal and spatial. Only models that simulate carbon fluxes will be requested to perform this runs.

3. MODEL FORCINGS

All the data will be provided using the NetCDF format and the ALMA conventions (www.lmd.jusssieu.fr/ALMA).

3.1 *Surface forcings*

The surface forcings will be provided with a 30 minute time step for all the sites, except Tucson, where the time step is 20 minutes. They include:

- Rainfall and snowfall. Only rainfall is provided.
- Wind speed
- Air temperature
- Specific humitidy, derived from relative humidity at the Lucky Hills and Kendall sites.
- Incident shortwave radiation
- Incident longwave radiation, derived.
- Surface pressure. Assumed constant according to altitude when measurements are not available.

3.2 *Soil characteristics*

The texture characteristics of the soil at each of the sites will be provided.

4. MODEL OUTPUT

The output information to be requested is limited to that matching the observations, and includes the following:

- Net shortwave radiation
- Net longwave radiation
- Latent heat flux
- Sensible heat flux
- Ground heat flux
- Carbon flux

- Soil moisture (@ 5cm depth)
- Ground temperature (@ 5cm depth)

The output information should include the results for all the five sites using a "default" parameter set and the estimated parameter sets. The corresponding parameter sets will also be requested.

5. PROPOSED ANALYSIS

As stated above the proposed experiment will attempt to test the models under the so called "split sample" framework and to establish the possible advantages of using parameter estimation procedures. The evaluation will include comparisons of the model outputs to the observations at the same site, but for a different time period; and comparisons at different locations, with similar physical characteristics.

The analysis will be carried out for the fluxes and state variables directly measured at the sites, and namely: latent heat flux, sensible heat flux, net shortwave radiation, net longwave radiation, soil temperature and soil moisture @ 5 cm depth, and the carbon flux.

Each of these variables will be compared to the observations for the following conditions using measures as the correlation coefficient, the Nash-Sutcliffe Efficiency, the root mean square error, the bias, the maximum distance, etc. In particular we will focus on the:

- Monthly mean
- Daily mean
- Daily amplitude
- Daily phase
- Min and max of the diurnal cycle
- Values at each time step

Based on the different error measures, we will attempt to estimate the usefulness of the parameter estimation procedures for the models. Codes for the optimization of the models and the required training in their use will be provided to the participants if they wish to do so. For this evaluation, the performances of the models using the *default* parameter sets will be used as benchmarks.

We will work within the framework proposed by Klemes (1986) for model evaluation, i.e. the split sample test will be used for both temporal and spatial evaluations. In addition to that, variables not used in the calibration procedures will be used for performance evaluation.

7. DATA PROTOCOLS

All data handling and format requirements will follow the ALMA-3 guidelines, as described in the ALMA website (www.lmd.jussieu.fr/polcher/ALMA/dataex_main.html).

6.1 *Data distribution and return*

The meteorological forcing data and the outputs will be provided via FTP or CD as the participants choose. The output variables that a given model cannot provide or does not produce should simply be omitted in the netCDF file. The specific instructions about the sites and the naming conventions for the files will be provided via email to each participant.

6.2 *Quality control*

ALMA has made available a screening program to check the correctness of the output netCDF files prior to return. This program based on those defined for the PILPS 2e Experiment will apply the annual water, energy, and carbon balance criteria; as well as ensuring that all variables are within reasonable ranges. The range requirements are not meant to comment on the appropriateness of model output, merely to verify unit correctness and sign. A number of utilities are freely available for plotting netCDF files, as listed on the ALMA web site, and we encourage their use as well. The screening program will be run after submission prior to any analysis. Any data that fail the screening will not be considered. So we encourage participants to run the program prior to submitting results.

6.3 *Results documentation*

Besides the model runs results some additional information will be requested from the participants:

- Short description of the model and the model structure. Include references.
- Description of the calibration procedure and which were the outputs used in the calibration.
- General impressions and comments on the results obtained based on the experience with the participants own models.
- Default parameter set. Which parameters were calibrated and the calibrated parameter sets.
- Specific problems or concerns experienced.
- Details of any modification to the provided information.

7. PROPOSED TIMELINE

- October 2002, Submission of experimental protocol
- January 2003, Distribution of forcing data to the participants
- February 2003, Workshop for training of participants in the use of the multi-criteria procedures
- April 2003, Deadline for submission of results
- July 2003, Workshop for analysis of preliminary results.

8. REFERENCES

Bastidas, L. A., H.V. Gupta, S. Sorooshian, W.J. Shuttleworth and Z.L. Yang "Sensitivity Analysis of a Land Surface Scheme using Multi-Criteria Methods", *Journal of Geophysical Research*, Vol. 104, No D16, p. 19,481-19,490, 1999.

Bastidas, L. A., H. V. Gupta, and S. Sorooshian, Bounding parameters of land surface schemes with observational data, in Observations and Modeling of the Land Surface Hydrological Processes, V. Lakshmi, J. Albertson & J. Schaake, Editors, Water Science and Application Volume 3, p. 65-76, American Geophysical Union, 2001.

Bastidas, L. A., H. V. Gupta, and S. Sorooshian, Parameter, Structure and Performance Evaluation for Land Surface Models, in Advances in the Calibration of Watershed Models, Q. Duan, H. V. Gupta, S. Sorooshian, A. Rousseau, and R. Turcotte, Editors, AGU, Expected Publication 2002

Gupta, H.V., L. A. Bastidas, L., S. Sorooshian, W.J. Shuttleworth and Z.L. Yang, Parameter Estimation of a Land Surface Scheme using Multi-Criteria Methods", *Journal of Geophysical Research*, Vol. 104, No D16, p. 19,491-19,504, 1999.

Gupta, H.V., S. Sorooshian, and P.O. Yapo, Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Water Resources Research*, *34* (4), 751-763, 1998.

Henderson-Sellers A., Pitman A., Love P., Irannejad P., Chen T. (1995) The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. Bull. of the Amer. Met. Soc. 76:489-503

Henderson-Sellers A., Yang Z., Dickinson R. (1993) The Project for Intercomparison of Land Surface Parameterization Schemes. Bull. of the Amer. Met. Soc. 74:1335-1349

Pitman A., Henderson-Sellers A., Abramopoulos F., Avissar R., Garratt J., Frech M., Hahmann, A., Koster R., Kowalczyk E., Laval K., Lean J., Lee T., Lettenmaier D., Liang X., Mahfouf J.-F., Mahrt L., Milly P., Mitchell K., de Noblet N., Noilhan J., Pan H., Pielke R., Robock A., Rosenzweig G., Schlosser C., Scott R., Suarez M., Thompson S., Verseghy D., Wetzel P., Wood E., Xue Y., Yang Z.-L., , Zhang L. (1993) Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS). Results from off-line control simulations (Phase 1A). Gewex report IPGO Publication

Lettenmaier, D., D. Lohmann, E. F. Wood, and X. Liang, *PILPS-2c Workshop Report*, Princeton Univ., Princeton, N. J., October 8–31, 1996.

Vrugt, J. A., H. V. Gupta, L.A. Bastidas, W. Bouten, and S.Sorooshian, A new effective and efficient algorithm for multi-objective optimization of hydrologic models, *submitted to Water Resources Research*, 2002

Yapo, P.O., H.V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *Journal of Hydrology*, *204*, 83-97, 1998.

Xia, Y., A.J. Pitman, H.V. Gupta, M. Lepastrier, A. Henderson-Sellers, and L.A. Bastidas, Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set, *Journal of Hydrometeorology*, V3(2) pp. 181-194, 2002.