Water budget in the Amazon basin and impacts on flood modelling

AUGUSTO GETIRANA, SUJAY KUMAR, CHRISTA PETERS-LIDARD & KRISTI ARSENAULT

Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, USA
augusto.getirana@nasa.gov

Abstract Although recent modelling and observational efforts have been performed to better understand the hydrological processes at the global scale, estimates of the water budget over the continents are still inaccurate. Several modelling attempts have been conducted trying to improve the simulation of water and energy cycles at different temporal and spatial scales worldwide. These attempts are based on numerous modelling approaches and meteorological forcings, resulting in contrasting water balance estimates. Considering the restricted availability of observed data to fully evaluate simulated water balances at large scales, remote sensing is revealed as an important source of information for model evaluation. The objective of this study is to assess the water budget in the Amazon basin simulated by land surface models (LSMs) and impacts on flood modelling. For that purpose, outputs of three LSMs currently implemented in the Land Information System (LIS) were considered. They are: Noah3.2, Mosaic and CLM2. LSMs were run for the 1980–2008 period using Princeton's meteorological forcings on a 3-hourly time step and at a 1° resolution. The precipitation was rescaled to match the monthly global GPCP dataset. Flood modelling is evaluated in this study by means of daily streamflows and monthly floodplain extent simulated by the Hydrological Modelling and Analysis Platform (HyMAP) river routing scheme using simulated surface and sub-surface runoffs as forcings. Results demonstrate that mean evapotranspiration rates vary from 2.5 to 3.3 mm/day, depending on the model. Noah3.2 had the best overall performance coefficients for streamflows, followed by Mosaic. CLM2 showed a considerable overestimation of mean streamflows and floodplain extent all over the basin.

Key words water budget; Amazon basin; floodplains; HyMAP; LIS

INTRODUCTION

Several studies have attempted to more accurately represent the water and energy cycles at different temporal and spatial scales worldwide, and consider various modelling approaches and meteorological forcings, resulting in contrasting evapotranspiration and runoff rate estimates. Considering the restricted availability of observed data to fully evaluate LSM water balance at large scales, the use of discharge observations at gauging stations is found as a straightforward way to assess the water budget at the catchment scale. In this sense, the Global Land Water Budget (GLWB) benchmarking initiative (Getirana *et al.*, 2014) has the main objective to evaluate the global water budget provided by state-of-the-art LSMs, trying (1) to understand how these models simulate the large-scale hydrology and the complex hydrological processes of large river basins, and (2) to identify limited representations of highly nonlinear physical processes at the global scale. In addition, GLWB investigates the impacts of precipitation uncertainties on LSM hydrological responses by forcing models with different global *in situ*-based rainfall datasets.

As a groundwork for GLWB analyses, this study presents the first results of the water budget in the Amazon basin as computed by three LSMs currently implemented in the Land Information System (LIS; Kumar *et al.*, 2006). They are: Noah3.3 (Ek *et al.*, 2003), Mosaic (Koster & Suarez, 1996) and the Community Land Model version 2 (CLM2: Zeng *et al.*, 2002). The impacts of simulated water budgets on flood modelling are evaluated using the Hydrological Modelling and Analysis Platform (HyMAP; Getirana *et al.*, 2012) river routing scheme. HyMAP is forced with LSM surface and sub-surface (or baseflow) runoffs, providing a complete diagnosis of surface water dynamics, including streamflows and floodplain extent.

DATASETS

Meteorological forcings

The meteorological forcings used as inputs in the LSMs are those provided by the Princeton University at a 3-hourly time step and 1° spatial resolution (Sheffield *et al.*, 2006). This dataset is based on the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) re-analysis. Sheffield *et al.* (2006) carried out corrections of the systematic biases in the 6-hourly NCEP–NCAR re-analyses via hybridization with global monthly gridded observations. In addition, the precipitation was disaggregated in both space and time at 1° spatial resolution via statistical downscaling and at a 3-hourly time step using information from the 3-hourly Tropical Rainfall Measuring Mission (TRMM) dataset. The 3-hourly precipitation from Sheffield *et al.* (2006) was then rescaled to match the monthly or daily precipitation values given by the Global Precipitation Climatology Project (GPCP) version 2.2 (Adler *et al.*, 2003).

Streamflow observations

Daily observed water discharge data at 176 gauging stations operated by the Brazilian Water Agency (ANA) were used to evaluate HyMAP streamflows. These gauging stations have time series with at least one year of observations within the 1980–2008 period.

Floodplain extent from multi-satellites

Floodplain extent simulated by HyMAP was evaluated against a product derived from the multi-satellite estimates of surface water extent dataset from Papa *et al.* (2010), called P10 hereafter. P10 is available at a monthly time step for 1993–2004, with a spatial resolution of 773 km^2. It was generated from complementary multiple satellite observations, including passive (Special Sensor Microwave Imager (SSM/I)) and active (European Remote Sensing (ERS) series satellites) microwaves, along with visible and near infrared imagery (Advanced Very High Resolution Radiometer (AVHRR)). In order to eliminate the impacts of other kinds of surface water bodies (including lakes, anthropogenic and natural reservoirs, or irrigated agriculture) on floodplain extent signals, we used a corrected version of P10 following a methodology suggested by Decharme *et al.* (2012).

METHODOLOGY

The Land Information System (LIS)

LIS is a land modelling system that operates several community LSMs with the required initial and boundary conditions. LIS is designed with advanced software engineering principles and it provides many user extensible interfaces to incorporate diverse data sets from different sources as inputs to the LSMs. LIS also includes generic support for high performance computing, enabling the use of LSMs at global scales with spatial resolutions as high as 1 km. More details about LIS can be found in the literature (e.g. Kumar *et al.*, 2006; Peters-Lidard *et al.*, 2007).

The HyMAP global flow routing scheme

HyMAP is a global scale flow routing scheme capable of simulating water discharge, flow velocity, depth and storage in both rivers and floodplains. Surface runoff (R) and baseflow (B) generated by LSMs are routed using a kinematic wave formulation through a prescribed river network to oceans or lakes. The model is fully described and evaluated in Getirana *et al.* (2012, 2013). Spatial and temporal resolutions can be flexible, according to the application. The model is composed of four modules accounting for: (1) the surface runoff and baseflow time delays; (2) flow routing in river channels; (3) flow routing in floodplains; and (4) evaporation from open water surfaces. In order to preserve the water budget derived from the LSMs, module (4) was not used in this study.

Model setup

In order to keep coherency in all experiments, a default model setup was defined and used in LSM runs. In this sense, except for the soil type and land cover parameters, which are those inherent to each model, LSMs were run from 1979 to 2008 at the 30-minute time step and 1° spatial resolution, globally, using Princeton's forcings and GPCP rescaled precipitation dataset. The first year of simulation was set for model spin-up and was not considered in the evaluation. LSM outputs were provided for evaluation at the daily time step. Daily surface runoff and baseflow derived from the LSMs were used as inputs in HyMAP that was run over the Amazon basin at the 0.25° spatial resolution during the same period used for LSMs. HyMAP time step was set as 15 min and outputs provided as daily averages.

Streamflows and floodplain extent evaluation procedure

The total runoff (which is the summation of R and B) simulated each LSM is evaluated by means of simulated streamflows derived from the HyMAP flow router. R and B are used to force HyMAP resulting in spatially distributed streamflows over the studied domain. The accuracy of simulated streamflows and floodplain extent was determined by using two performance coefficients: the Nash-Sutcliffe (NS) coefficient and the relative volume error of streamflows (RE). NS and RE are represented by the equations below:

$$NS = 1 - \frac{\sum_{t=1}^{nt} (y_t - x_t)^2}{\sum_{t=1}^{nt} (y_t - \bar{y})^2} \tag{1}$$

$$RE = \frac{\sum_{t=1}^{nt} x_t - \sum_{t=1}^{nt} y_t}{\sum_{t=1}^{nt} y_t} \tag{2}$$

where t is the time step, nt the total number of days disposing of observed data, x and y are, respectively, the simulated and target (observed) signals at time step t. $\bar{y}$ stands for the mean value of observations. NS ranges from $-\infty$ to 1, where 1 is the optimal case and zero is when simulations represent observed signals as well as the mean value. RE varies from -1 to $+\infty$, where zero is the optimal case. One can obtain RE values as a percentage by multiplying them by 100.

RESULTS

Water budget

Figure 1 shows the monthly mean time series of water budget variables, including surface runoff (R), baseflow (B), precipitation (P) and evapotranspiration (ET), resulting from the three LSMs. The averaged rainfall rate for the entire basin is 6 mm/day, which is in agreement with previous estimates found in the literature using different datasets (e.g. Costa & Foley, 1998; Marengo, 2005). A pronounced seasonality is noticed in both R and B for Noah3.2, Mosaic and CLM2, with peaks generally occurring between February and April. Mosaic delays the baseflow peak in one month in comparison to its runoff and a slight advance of both R and B peaks is noticed in CLM2 outputs. Seasonality of mean monthly ET is quite low in all LSM outputs, with mean rates varying from 2.5 mm/d (CLM2) to 3.2 mm/day (Noah3.2). R and B present non-negligible differences from one LSM to another. Noah3.2, for example, has the lowest mean surface runoff (R = 0.5 mm/day) and highest baseflow (B = 2.5mm/day). On the other hand, CLM2 generates low baseflow (1.7 mm/day), significantly increasing R values to 2 mm/day. The highest total runoff (TR = R + B) was provided by CLM2, with TR = 3.7 mm/day, against 3 mm/day (Noah3.2) and 3.2 mm/day (Mosaic).

Streamflow

Figure 2 shows *NS* coefficients and *RE* values for the three experiments. Both coefficients are presented as a function of the drainage area and spatially distributed over the basin. Based on the scatter plots of *NS* coefficients, one can notice that although Mosaic provides the best average of

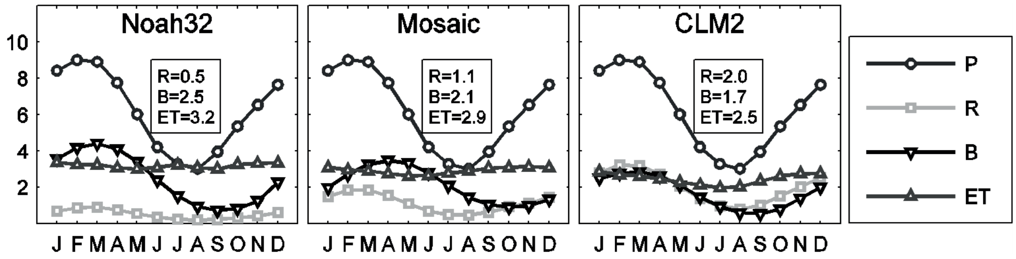

Fig 1 Monthly mean values of simulated water budget variables. Units are in mm/day.

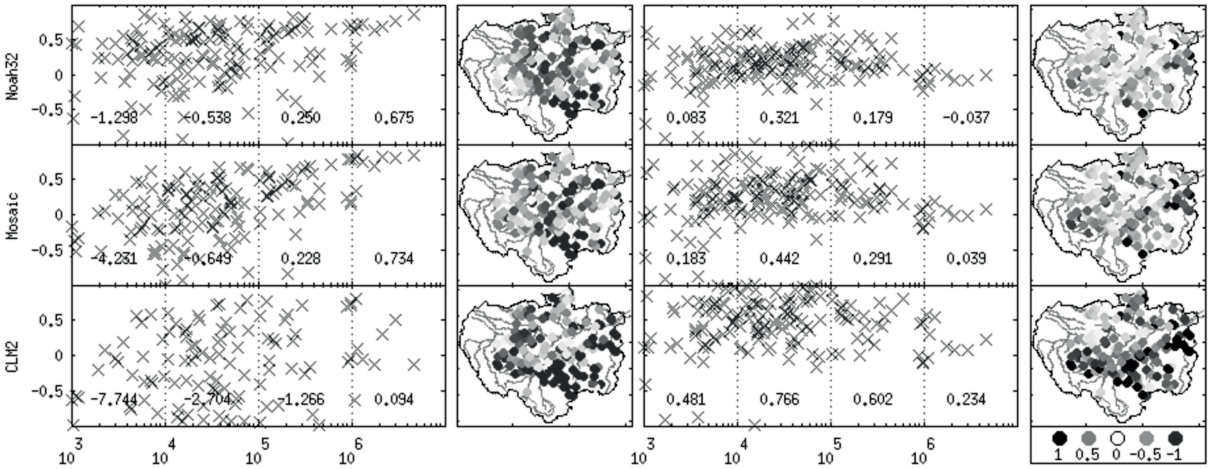

Fig. 2 Nash-Sutcliffe coefficient (NS; columns 1–2) and relative error (RE; columns 3–4) of simulated streamflows at gauging stations within the Amazon basin.

NS values at gauging stations draining catchments bigger than 10^6 km^2, Noah3.2 performed better in all other area thresholds. *RE* values obtained from Noah3.2 are also better than those achieved by all other models, with averaged values closer to zero and lower standard deviation. CLM2 resulted in negative *NS* values and highly overestimated *RE* in most parts of the basin. The recurrent low *NS* values noticed in the southern Amazon basin for the three experiments can be explained by inconsistencies in both meteorological forcings and HyMAP parameterization.

Results at the Obidos station, located about 800 km upstream from the river mouth and draining about 4 160 000 km^2, can determine how close simulated long-term basin-wide water budgets are to *in situ* observations. Noah3.2 provided the best *NS* and *RE* values (0.86 and –0.7%, respectively), followed by Mosaic (*NS* = 0.84 and *RE* = 8.1%). CLM2 had the worst performance at Obidos, with *NS* = –0.14 and overestimated total runoff with *RE* = 22%.

Floodplain extent

Basin-wide floodplain extents simulated by both Noah3.2 and Mosaic agreed generally well with P10 (Fig. 3). Both experiments provided floodplain extent time series in phase with satellite-based estimates, resulting in high NS (0.73 and 0.75, respectively) and low *RE* (–5% and zero). On the other hand, CLM2 provided unsatisfactory results for most selected areas, overestimating the floodplain extent in comparison to the other two models and with peaks rising one or two months in advance. For a more quantitative comparison, five other sub-regions are considered in order to evaluate the monthly averaged flooded areas over the 1993–2004 period: the central Amazon floodplains (defined as the rectangle from 08S–548W to 88S–728W), and the Negro, Madeira, Xingu, and upper Solimoes river basins. According to the results obtained in these sub-regions, the Negro River basin, located in the northern Amazon basin, is the only one where CLM2 performed better than other LSMs, with *NS* = 0.70 and *RE* = –11%, in comparison to 0.21 and –23% for Noah3.2, and 0.38 and –17% for Mosaic. Based on streamflows at two gauging stations (Serrinha

(294 000 km^2) and Caracarai (126 000 km^2) stations) covering about 60% of the Negro River basin, it can be said that simulated streamflows are generally overestimated in that basin. In particular, CLM2 is the model with the highest *RE* values (15% at Serrinha and 59% at Caracarai). Noah3.2 and Mosaic had 2% and 10% at Serrinha, and 18% and 20% at Caracarai, respectively. Although the streamflow simulations in the Negro River basin are overestimated, floodplain extents are considerably underestimated in comparison to P10. Also, a systematic overestimation of floodplain extent is evidenced in other areas, such as the Xingu and upper Solimoes river basins, with *RE* values varying from 52% to 86% in these basins. These disagreements suggest: (1) possible limitations in the elevation profile in HyMAP (see Getirana *et al.*, 2012 for more details) derived from the SRTM30 digital elevation model (DEM) and/or (2) poor P10 floodplain extent estimates.

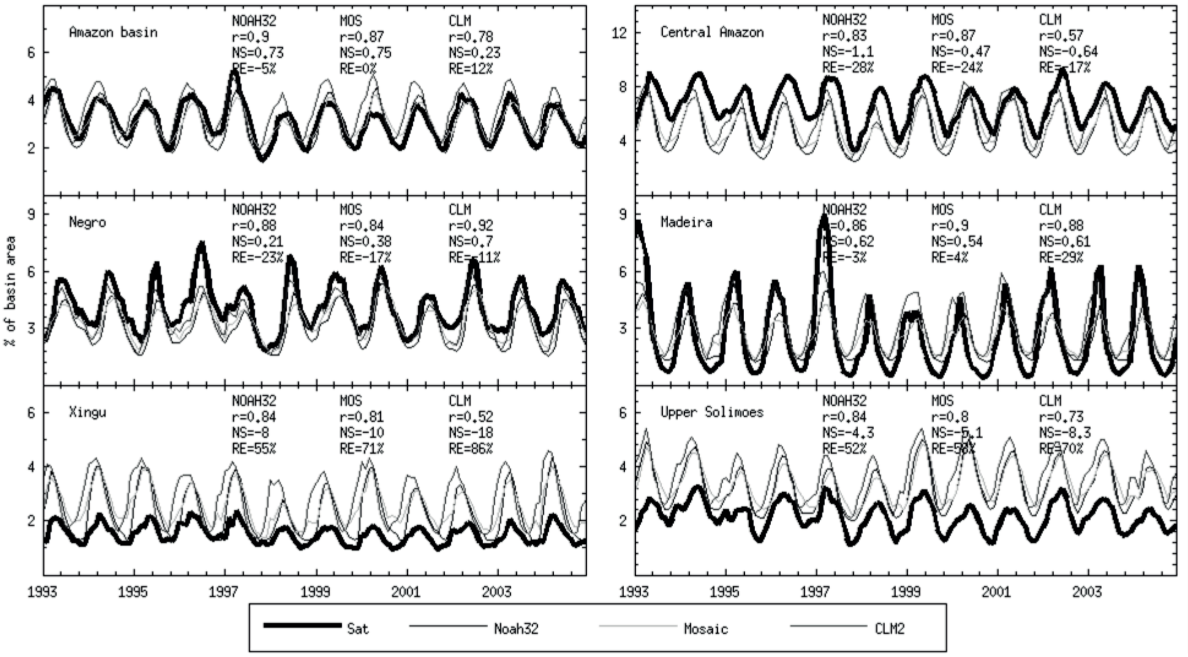

Fig. 3 Monthly averaged flooded extent over the 1993–2004 period for the six areas. Model outputs are in dashed grey lines and satellite observations in black. The correlation (r), Nash-Sutcliffe coefficient (NS), and relative error (RE) are given for each experiment.

CONCLUDING REMARKS

In the framework of the Global Land Water Budget benchmarking (GLWB) initiative (Getirana *et al.*, 2014), this paper presents a preliminary evaluation of how water budget simulated by LSMs impacts flood modelling. The experiments presented in this study considered only one rainfall dataset and three LSMs currently implemented in LIS. Daily streamflows and monthly floodplain extents were compared against *in situ* and satellite-based data. Overall, Noah3.2 and Mosaic had similar performances over the Amazon basin. A slight advantage in simulated streamflows is evident for Noah3.2 results in medium and relatively small catchments. High total runoff rates simulated by CLM2 resulted in overestimated streamflows all over the basin and early surface runoff and baseflow peaks impacted flood timing.

Although simulated and satellite-based floodplain extent interannual variability matches well at the basin scale, significant differences can be noticed in smaller areas. Getirana *et al.* (2013) showed that surface water dynamics simulated by HyMAP are closely dependent on river geometry and topography. It is worth noting that both simulated and satellite-based floodplain extent estimates have errors and such a comparison should be performed carefully. The development of more accurate DEMs that take into account canopy heights as well as better estimates of river geometry can reduce model errors. Also, future comparisons can benefit from higher resolution imagery (Jung *et al.*, 2010), providing more accurate datasets for model evaluation.

Finally, LSMs are in constant development and new versions are frequently released. More recent versions of Noah (Noah-MP: Niu *et al.*, 2011) and CLM (CLM4: Lawrence *et al.*, 2011) are available and should be included in the full comparison of LSMs. Also, numerous other models developed at different institutions worldwide are expected to be used in further evaluations.

Acknowledgements The first author is funded by the NASA Postdoctoral Program (NPP) managed by Oak Ridge Associated Universities (ORAU). The study benefited from data made available by *Agência Nacional de Águas* (ANA).

REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., & Arkin, P. (2003) The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present) *Journal of Hydrometeorology* 4, 1147–1167.

Costa, M. H. & Foley, J. A. (1998) A comparison of precipitation datasets for the Amazon basin. *Geophysical Research Letters* 25(2), 155–158.

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H. & Prigent, C. (2011) Global off-line evaluation of the ISBA-TRIP flood model *Climate Dynamics* 38, 1389–1412, doi:10.1007/s00382-011-1054-9.

Ek, M., Mitchell, K., Yin, L., Rogers, P., Grunmann, P., Koren, V., Gayno, G., & Tarpley J. (2003) Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model *Journal of Geophysical Research* 108(D22), 8851, doi:10.1029/2002JD003296.

Getirana, A. C. V. *et al.* (2014) The global water budget from a LSM intercomparison perspective (in preparation).

Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F. & Mognard, N. (2012) The Hydrological Modeling and Analysis Platform (HyMAP): evaluation in the Amazon basin *Journal of Hydrometeorology* 13, 1641–1665. doi: 10.1175/JHM-D-12-021.1.

Getirana, A. C. V., Boone, A., Yamazaki, D. & Mognard, N. (2013) Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon basin *Water Resources Research* doi: 10.1002/wrcr.20077.

Jung, H. C., Hamski, J., Durand, M., Alsdorf, D., Hossain, F., Lee, H., Hossain, A. K. M. A., Hasan, K., Khan, A. S., Hoque & A. K. M. Z. (2010) Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra Rivers *Earth Processes and Landforms* 35(3), 294–304, doi: 10.1002/esp.1914.

Koster, R. D. & Suarez, M. J. (1996) Energy and Water Balance Calculations in the Mosaic LSM NASA Technical Memorandum 104606, 9, 76 pp.

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., Lighty, L., Eastman, J. L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E. F. & Sheffield, J. (2006) Land Information System – An Interoperable Framework for High Resolution Land Surface Modeling *Environmental Modelling & Software* 21, 1402–1415, doi:10.1016/j.envsoft.2005.07.004.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B. & Slater, A. G. (2011) Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems* 3, doi: 10.1029/2011MS000045.

Marengo, J. A. (2005) Characteristics and spatio-temporal variability of the Amazon River Basin Water Budget *Climate Dynamics* 24, 11–22. doi 10.1007/s00382-004-0461-6.

Niu, G.-Y., *et al.* (2011) The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements *Journal of Geophysical Research* 116, D12109, doi:10.1029/2010JD015139.

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B. & Matthews, E. (2010) Interannual variability of surface water extent at the global scale, 1993–2004 *Journal of Geophysical Research* 115, D12111, doi:10.1029/2009JD012674.

Peters-Lidard, C. D., Houser, P. R.. Tian, Y., Kumar, S. V., Geiger, J., Olden, S., Lighty, L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E. F. & Sheffield, J. (2007) High Performance Earth System Modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering* 3(3), 157–165, doi:10.1007/s11334-007-0028-x

Sheffield, J., Goteti, G. & Wood E. F. (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modelling. *Journal of Climate* 19, 3088–3111.

Zeng, X. B., Shaikh, M., Dai, Y. J., Dickinson, R. E. & Myneni, R. (2002) Coupling of the Common Land Model to the NCAR Community Climate Model *Journal of Climate* 15(14), 1832–1854 doi: 10.1175/1520-0442(2002)015<1832: COTCLM>2.0.CO.