Assessing Evapotranspiration Estimates from the Second Global Soil Wetness Project (GSWP-2) Simulations

C. ADAM SCHLOSSER AND XIANG GAO

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 13 August 2009, in final form 11 January 2010)

ABSTRACT

This study assesses the simulations of global-scale evapotranspiration from the second Global Soil Wetness Project (GSWP-2) within a global water budget framework. The scatter in the GSWP-2 global evapotranspiration estimates from various land surface models can constrain the global annual water budget fluxes to within ±2.5% and, by using estimates of global precipitation, the residual ocean evaporation estimate falls within the range of other independently derived bulk estimates. The GSWP-2 scatter, however, cannot entirely explain the imbalance of the annual fluxes from a modern-era, observationally based global water budget assessment. Inconsistencies in the magnitude and timing of seasonal variations between the global water budget terms are also found. Intermodel inconsistencies in evapotranspiration are largest for high-latitude interannual variability as well as for interseasonal variations in the tropics, and analyses with field-scale data also highlight model disparity at estimating evapotranspiration in high-latitude regions. Analyses of the sensitivity simulations that replace uncertain forcings (i.e., radiation, precipitation, and meteorological variables) indicate that global (land) evapotranspiration is slightly more sensitive to precipitation than net radiation perturbations, and the majority of the GSWP-2 models, at a global scale, fall in a marginally moisture-limited evaporative condition. Lastly, the range of global evapotranspiration estimates among the models is larger than any bias caused by uncertainties in the GSWP-2 atmospheric forcing, indicating that model structure plays a more important role toward improving global land evaporation estimates (as opposed to improved atmospheric forcing).

1. Introduction

In the quest to accurately portray global hydroclimatological conditions as well as predict variations, potential changes, and effects of the climate system, evapotranspiration E is regarded as one of the critical fluxes that links the energy, water, and biogeochemical cycles of the terrestrial ecohydrological systems. With respect to our ability of direct measurement, however, evapotranspiration is a key, missing variable in global water balance assessments (e.g., Swenson and Wahr 2006) as well as for regional assessments of hydroclimatological variability and change (e.g., Werth and Avissar 2004). At large spatial scales for climate studies, it is an inherently difficult flux to measure directly, and a variety of other methods have been aimed to estimate and assess its mean state and variability. More recent observationally based residual estimates of evapotranspiration have been provided at basin (e.g., Rodell et al. 2004a) to continental scales (e.g., Karam and Bras 2008; Walter et al. 2004), and they show promise in the ability of these methods to estimate mean fluxes as well as their variability and possible trends. Other techniques for evapotranspiration estimation using remotely sensed data (e.g., Wang et al. 2007; Song et al. 2000) have been undergoing refinement and have been provisionally analyzed at a global scale (e.g., Wang and Liang 2008); however, data availability and sensitivity to retrieval and interpolation errors (in temperature and vegetation properties) continue to be significant issues with these sorts of techniques. As such, reliable and comprehensive direct and/or derived measurements of global- or large-scale evapotranspiration remain elusive.

In light of this, the climate research community has placed a heavy reliance upon modeling and assimilation techniques to estimate land evapotranspiration (as well as other land flux and state variables). Many such models are actively in use within the climate research community

Corresponding author address: C. Adam Schlosser, Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, E19-411k, 50 Ames St., Cambridge, MA 02139.
E-mail: casch@mit.edu

DOI: 10.1175/2010JHM1203.1

© 2010 American Meteorological Society

(e.g., Rodell et al. 2004b) and represent a variety of parameterization recipes to represent key biogeophysical and biogeochemical processes. Evaluation of these model simulations, wherever possible, is of considerable interest to document their reliability and consistency. Furthermore, with the multiple model-based estimates comes a degree of uncertainty that must also be quantified and preferably within the context of complementary, and wherever possible, directly comparable measurements of other water cycle storages and fluxes.

In previous studies, direct comparisons of models used to estimate evapotranspiration have proven quite useful in this regard (e.g., Chen et al. 1997; Werth and Avissar 2004; Su et al. 2005), yet most of these analyses were of limited spatial and/or temporal coverage. Recently, the second Global Soil Wetness Project (GSWP-2; Dirmeyer et al. 2002) has provided an unprecedented collection of global simulations spanning the 1986–95 period of land states and fluxes calculated from 13 participating biogeophysical models used in climate research and weather prediction. The simulations provide a baseline set of runs as well as additional subsets of sensitivity runs that consider sources of uncertainty in the required atmospheric inputs and land cover fields. The GSWP-2 simulation period also falls within the time domain of a recent modern-era assessment of the global water cycle (Schlosser and Houser 2007, hereafter SH07), in which an absence of uncertainty estimates for global land evapotranspiration was highlighted. In view of these issues, we have analyzed the outputs of evapotranspiration from the GSWP-2 model simulations to serve a few key purposes: 1) to provide global estimates of land evapotranspiration rates to complement a modern-era, observationally based global water cycle assessment; 2) to quantify the uncertainty in these evapotranspiration estimates; and 3) to determine the primary sources of these uncertainties (i.e., from models or inputs) as well as areas where evapotranspiration estimates are in most need for improvement. In the section that follows, we describe the GSWP-2 model experiments that include outputs of a baseline and sensitivity runs used for this study. In addition, we also describe the data taken from a global water budget assessment employed for our analysis as well as field data used for a complementary evaluation of the GSWP-2 simulations. Section 3 describes the results of our analysis, and lastly, in section 4, we present our conclusions and closing remarks for continued research.

2. Data

a. GSWP-2

The GSWP is an element of the Global Land–Atmosphere System Study and a study of the Global Energy and Water Cycle Experiment (GEWEX) Modelling and Prediction Panel (GMPP), both contributing projects of GEWEX. GSWP is charged with producing large-scale datasets of soil moisture, temperature, runoff, and surface fluxes by integrating one-way offline land surface schemes (LSSs) using externally specified surface forcing and standardized soil and vegetation distributions. The GSWP-2 (see Dirmeyer et al. 2006 for details) produced a 10-yr daily global gridded dataset of land surface state variables and fluxes—excluding Antarctica. To gauge the effect of this omission in this global-scale modeling effort, we have also obtained an estimate of annual evaporation over Antarctica using the technique described by Loewe (1957). GSWP-2 is closely linked to the International Satellite Land Surface Climatology Project Initiative II data effort (Hall et al. 2006), and the LSSs simulations in GSWP-2 encompass the same 10-yr core period (1986–95). The model simulations are conducted on a $1° \times 1°$ grid, and each model is driven by identical meteorological forcings. The 3-hourly near-surface meteorological forcing datasets are derived from the regridding of the National Centers for Environmental Prediction (NCEP)–Department of Energy (DOE) reanalyses (Kanamitsu et al. 2002), with corrections to the systematic biases in the reanalysis fields made by hybridization with global observationally based gridded datasets (Zhao and Dirmeyer 2003). This provides the land models with some of the most accurate forcing data available.

Thirteen LSSs in use today within the climate modeling community have participated in the baseline (B0) simulation for GSWP-2 (Table 1), and they constitute a broad cross section of numerical recipes to parameterize biogeophysical land processes. All the participating models adhere to the same land mask and as closely as possible to the supplied datasets of vegetation distribution and properties, soil properties, and surface albedos, among others. They also follow the same procedure for the spin-up process (see Dirmeyer et al. 2002 for details) with the same initial condition (soil temperature, soil moisture, and snow cover) and report a standard set of output data for the 10-yr core period 1986–95. The results from the land surface models were checked for quality, consistency, and conservation of mass and energy; corrected when problems were detected; and then combined to produce a multimodel land surface analysis (Dirmeyer et al. 2006). This analysis has been validated and shown to be superior to any individual model in terms of its representation of soil moisture variations (Guo et al. 2007; Gao and Dirmeyer 2006); however, an explicit evaluation of the evapotranspiration against direct or complementary observations has not been performed. The bulk of the GSWP-2 output

TABLE 1. List of model acronyms used in figures and tables to indicate the participating models in the B0 and sensitivity simulations (Table 2) of GSWP-2. The left-hand column provides the model acronym (in some cases modified for brevity), and the right-hand column gives the corresponding institute and nation from Table 1 of Dirmeyer et al. (2006, refer to reference for further model details).

Acronym*	GSWP-2 model information
CLM2-TOP	University of Texas
HySSiB	NASA Goddard Space Flight Center (GSFC)
ISBA	Météo-France/Centre National de Recherches Météorologiques (CNRM), France
LaD	NOAA/Geophysical Fluid Dynamics Laboratory (GFDL)
Mosaic	NASA GSFC Hydrological Service Branch (HSB)
MOSES 2	Met Office, United Kingdom
Noah	NOAA/NCEP/Environmental Modeling Center (EMC)
NSIPP	NSIPP Catchment: NASA GSFC Global Modeling and Assimilation Office (GMAO)
ORCHIDEE	Institut Pierre Simon Laplace (ISPL), France
SiBUC	Kyoto University, Japan
SSiBCOLA	Institute of Global Environment and Society (IGES)/COLA
SWAP	Russian Academy of Sciences/Institute of Water Problems (IWP), Russia
VISA	University of Texas

* CLM2-TOP is the Community Land Model, version 2-TOPMODEL; HySSiB is the Hydrological improvements to the Simplified version of the Simple Biosphere Model (HySSiB) ISBA is the Interactions between the Soil, Biosphere, and Atmosphere model; LaD is the Land Dynamics model; MOSES2 is the Met Office Surface Exchange Scheme 2 model; NSIPP is the NASA Seasonal to Interannual Prediction Project; ORCHIDEE is the Organizing Carbon and Hydrology in Dynamic Ecosystems model; SiBUC is the SiB including Urban Canopy model; and VISA is Versatile Integrator of Surface–Atmospheric Processes model.

data—including baseline simulations, multimodel analyses, and sensitivity studies—are reported at a daily interval. There exist also subdiurnal outputs at 3-h intervals from the models, which were logged (as instructed by the GSWP-2 exercise) during the last year (1995) for all the baseline simulations.

Another essential component of GSWP-2 involves a suite of sensitivity studies (Table 2) by the participating LSSs, where forcing data or boundary conditions are altered to examine the response of the models to uncertainties in those parameters. GSWP-2 provides various alternates of meteorological forcing variables and land surface parameters for designated sensitivities studies (Dirmeyer et al. 2002). Participation in the sensitivity studies by each modeling group was optional. Table 3 lists all the sensitivity simulations that the models performed and the outputs collected. These simulations include substitutions to precipitation P, radiation R, all meteorological forcing, and vegetation properties. The sensitivities of different LSSs to uncertainties in the precipitation data (i.e., runs P1, P2, P3, P4, and PE) specifically address the effects of bias correction by hybridization, choice of different reanalysis products, the range in observational estimates, and rain gauge undercatch. The radiation series (i.e., runs R1, R2, and R3) provide a similar evaluation for the effect of the systematic differences between the reanalyses and First International Satellite Cloud Climatology Project (ISCCP) radiation. The all-meteorological study (i.e., runs M1 and M2) gives the broadest assessment as to the effect of differences between the two reanalyses. The sensitivity with vegetation properties (run I1) examines the effect of the interannual variability versus mean seasonal cycle of vegetation phenology. Since reanalysis products are widely used as a proxy for true atmospheric conditions, these sensitivity studies have important implications, such that we can gauge the certitude of scientific results achieved using these datasets (i.e., for global hydrological cycle studies).

b. Observations

1) GLOBAL-SCALE DATA

For our global-scale assessment of the GSWP-2 evapotranspiration estimates, we draw upon data and results from a recent global water budget analysis (SH07). The SH07 study combined global fields of precipitation, evaporation (separate land and ocean estimates), and water vapor to perform an atmospheric-based water budget assessment via six core datasets:

1) The Global Precipitation Climatology Project (GPCP), version 2 (Adler et al. 2003)
2) The Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997)
3) Goddard Satellite-based Surface Turbulent Fluxes, version 2 (GSSTF; Chou et al. 2003)
4) Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) data (Bentamy et al. 2003)
5) Center for Ocean–Land–Atmosphere Studies (COLA) Global Offline Land Surface Dataset (GOLD; Dirmeyer and Tan 2001)
6) National Aeronautics and Space Administration (NASA) Water Vapor Project (NVAP; Vonder Harr et al. 2003)

SH07 provides further details regarding these datasets, and the period of overlap between these datasets and the GSWP-2 data covers the years 1988–95. Missing from the SH07 study was an explicit estimate of the uncertainty in the global land evapotranspiration, and therefore we will use the GSWP-2 results to provide a scatter

TABLE 2. Description of various GSWP-2 sensitivity experiments (meteorological forcing and vegetation datasets used in the sensitivity experiments are the same as the B0 baseline integration unless otherwise specified).

Sensitivity experiment	Description of meteorological forcing and vegetation datasets
B0	Tair and Qair: NCEP–DOE hybridized with Climatic Research Unit (CRU) Wind: NCEP–DOE SWdown and LWdown: SRB PSurf: NCEP/DOE with altitude [Earth Resources Observation and Science (EROS) Data Center (EDC) topography] correction Rainf, Rainf_C (Convective), and Snowf: NCEP–DOE hybridized with GPCC gridded gauge analysis, corrected for wind-caused gauge undercatch, and blended with GPCP where the gauge density is low Vegetation: Observed interannually varying monthly vegetation parameters
M1	All original NCEP–DOE meteorological data (no hybridization with observational data)
M2	All original ECMWF (ERA-40) meteorological data (no hybridization with observational data)
P1	Original ERA-40 precipitation (no hybridization with observational data)
PE	ERA-40 precipitation hybridized with GPCC gridded gauge analysis, corrected for wind-caused gauge undercatch, and blended with GPCP where the gauge density is low
P2	NCEP–DOE precipitation hybridized with GPCC gauge analysis and corrected for wind-caused gauge undercatch
P3	NCEP–DOE precipitation hybridized with GPCC gauge analysis only
P4	Original NCEP–DOE precipitation (no hybridization with observational data)
R1	Radiation from NCEP–DOE reanalysis
R2	Radiation from ERA-40 reanalysis
R3	Radiation from ISCCP (Rossow and Dueñas 2004)
I1	Climatological annual cycle of vegetation parameters

of land evapotranspiration within the global water balance. In addition, we have augmented the data collection of SH07 in our analysis to include the latest version of the HOAPS ocean evaporation estimate (HOAPS3, available online at http://www.hoaps.zmaw.de/) as well as a gap-filled version of CMAP using the National Center for Atmospheric Research reanalysis precipitation values [CMAPr, provided by the National Oceanic and Atmospheric Administration Office of Atmospheric Research Earth System Research Laboratory Physical Sciences Division (NOAA/OAR/ESRL/PSD), available online at http://www.cdc.noaa.gov/].

2) FIELD DATA

To evaluate the performances of evapotranspiration simulations from various land surface models as well as the quality of the precipitation forcing in GSWP-2, observations of precipitation and evapotranspiration (or latent heat flux) have been collected. Four sites have been identified for this study, whose data temporally overlap the GSWP-2 period. Table 4 summarizes the characteristics of each dataset used in this study. Some of these observational sites have a relatively short record of overlap with the GSWP-2, but they all have at least one year of data for comparison. The GSWP-2 grid values corresponding to the individual validation site have been extracted from the various model baseline simulations, multimodel analyses, and sensitivity experiments for evaluation with the observations.

Our most complete source of field data (in terms of temporal domain) is from the North Appalachian Experimental Watershed (NAEW; Harmel et al. 2007), which is located near Coshocton in east-central Ohio, an unglaciated portion of the state with rolling uplands. Its 1050-acre outdoor laboratory facility is operated by the U.S. Department of Agriculture's Agricultural Research Services. The NAEW consists of a network of 22 instrumented watersheds, 11 large lysimeters, meteorological stations, and rain gauges for surface water and groundwater hydrology and water quality studies. The experimental watersheds with natural setting range in size from 1 to 300 acres and five of them are larger than 40 acres. The NAEW is one of only two hydrologic stations worldwide with more than 60 years of continuous data collected from small watersheds and groundwater lysimeters. The Coshocton site was selected because it represented land conditions prevalent in many states in the Appalachian region. There are 11 active rain gauges

TABLE 3. Summary of conducted sensitivity experiments for each of the participating GSWP-2 models, with an "X" indicating that the simulation was performed.

	B0	I1	M1	M2	P1	P2	P3	P4	PE	R1	R2	R3
SSiBCOLA	X	X	X	X	X	X	X	X	X	X	X	X
NSIPP	X	X	X	X	X	X	X	X	X	X		X
SWAP	X	X	X		X	X	X			X		
Noah	X	X			X	X	X		X	X		
MOSES2	X						X			X		

TABLE 4. Site characteristics of field data selected in the local validation of this study. Shown are the vegetation types specified for the site-corresponding grids in the SiB, Biosphere–Atmosphere Transfer Scheme (BATS), and International Geosphere–Biosphere Programme (IGBP)-based vegetation classifications upon which the GSWP-2 model simulations are based. Map labels used in Fig. 5 are also indicated, but note that because of proximity, ''A'' identifies all ARM stations.

Site (map label)	Lat, lon, elev	Site vegetation	Vegetation type	Measurements (sampling interval)	Time domain
NAEW (C)	40.37°N, 81.79°W, 243 m	Rangeland	SiB, BATS, IGBP: Broadleaf deciduous	Lysimeter (daily), precipitation (irregular, subdaily)	1986–95
Harvard Forest EMS (H)	42.54°N, 72.17°W, 340 m	Temperate deciduous	SiB, BATS, IGBP: Broadleaf deciduous	Water vapor eddy covariance flux (hourly), precipitation (daily)	1992–95
BOREAS NOBS (B)	55.88°N, 98.48°W, 259 m	Needleleaf evergreen	SiB, BATS, IGBP: Needleleaf evergreen	Latent heat flux (30 min), precipitation (daily)	1994–95
Walker Branch (W)	35.96°N, 84.29°W, 372 m	Deciduous broadleaf temperate	SiB: Deciduous and evergreen, BATS and IGBP: Mixed forest	Latent heat flux (30 min), precipitation (daily)	1995
ARM E8 Coldwater (A)	37.33°N, 99.31°W, 664 m	Rangeland (grazed)	SiB: C3 grass, BATS: Cropland, IGBP: Cropland	Latent heat flux (30 min), precipitation (30 min)	1994–95
ARM E9 Ashton (A)	37.13°N, 97.27°W, 386 m	Pasture	SiB: C3 grass, BATS: Cropland, IGBP: Cropland	Latent heat flux (30 min), precipitation (30 min)	1994–95
ARM E13 Lamont (A)	36.61°N, 97.49°W, 318 m	Pasture and wheat	SiB: C3 grass, BATS: Cropland, IGBP: Cropland	Latent heat flux (30 min), precipitation (30 min)	1994–95
ARM E15 Ringwood (A)	36.43°N, 98.28°W, 418 m	Pasture	SiB: Groundcover, BATS: Short grass, IGBP: Grassland	Latent heat flux (30 min), precipitation (30 min)	1994–95
ARM E20 Meeker (A)	35.56°N, 97.00°W, 309 m	Pasture	SiB: Groundcover with trees and shrubs, BATS: Forest/field, IGBP: Woody savanna	Latent heat flux (30 min), precipitation (30 min)	1994–95

distributed across the watershed area. Analyses (not shown) indicate little spatial variability in the watershed precipitation with the temporal cross correlations among the 60-yr daily precipitation time series of 11 rain gauges all larger than 0.95. Therefore, all these rain gauges are averaged to approximately represent the scales of the GSWP-2 LSS grid box at 1° resolution. There is only one weighing lysimeter to record the evapotranspiration. All the observations are aggregated to monthly interval for comparisons with GSWP-2 model simulations.

The second source of data comes from the FLUXNET network of micrometeorological tower sites (Baldocchi et al. 2001), designed primarily to measure the exchanges of carbon dioxide, water vapor, and energy between the terrestrial ecosystem and atmosphere. Specifically, the level 3 data from the AmeriFlux regional networks are available for a number of years overlapping with the GSWP-2 period. This level of data has gone through consistency checks for units, naming conventions, reporting intervals, and formatting with quality flag assigned but without filling in the missing values. We have chosen to use the unfilled data instead of gap-filled data because of the questionable quality of the model-based gap-filling procedure (B. Munger 2007, personal communication). Three sites have multiyear records of fluxes and precipitation within the GSWP-2 period. Data from the Harvard Forest Environmental Measurement Site (EMS) was established in October 1989 but the quality-assured dataset started in 1992. Data collection at the Northern Study Area Old Black Spruce site (NOBS), located near Thompson, Manitoba, Canada, started in 1994 during the Boreal Ecosystem–Atmosphere Study (BOREAS) experiment in the northern boreal forests of Canada. The meteorological tower in the Walker Branch Watershed near Oak Ridge, Tennessee, was established in 1979, and flux data collection started in 1994.

There are gaps in the precipitation data of 1994 at the BOREAS NOBS site. One reason is that the rain gauge did not seem to work well for snow, which is a major part of the precipitation at this site. As a result, the data gaps are not random and measurements are somewhat biased toward convective precipitation (A. Dunn 2007, personal communication). Therefore, in this study, we use precipitation data from nearby Thompson Airport, Manitoba, Canada (55.8°N, 97.86°W, 223.1-m elevation, available online at http://www.climate.weatheroffice.ec.gc.ca/climateData/canada_e.html), to complement the available flux measurements for the evaluation exercise. The Thompson site reports both rainfall Rainf (amount of all liquid precipitation, such as rain, drizzle, freezing rain, and hail) and snowfall Snowf (amount of frozen/solid precipitation, such as snow and ice pellets). The sum of rainfall and the water equivalent of the snowfall is used here.

During the years overlapping with the GSWP-2 period, data collection in all three AmeriFlux sites experienced technical difficulties and instrumentation failure. As a result, temporal coverage for the relevant flux measurements is, at times, irregular (although it has improved in recent years). For our analyses, these gaps in half-hourly or hourly data are addressed in the following manner. We first derive the climatology of diurnal cycle for each calendar month based on the available observations of that month. Then we fill in missing measurements with the derived month-specific diurnal cycle climatology. The half-hourly or hourly data are aggregated to 3-hourly (1995 only), daily, and monthly whenever necessary for comparisons with the model simulations.

The U.S. Department of Energy operates the Atmospheric Radiation Measurement Program (ARM; Ackerman and Stokes 2003). In particular, the southern Great Plains site consists of a central facility and a number of extended facilities across a large area of Oklahoma and southern Kansas, each having instrument clusters to measure radiation, near-surface meteorology, and surface fluxes. For our study, data from the Energy Balance Bowen Ratio (EBBR; Cook 2005) system and the Surface Meteorological Observation System (SMOS) at the extended facility is appropriate. The EBBR uses observations of net radiation, soil surface heat flux, and the vertical gradients of temperature and relative humidity to estimate the vertical heat flux at the local surface. The SMOS mostly uses conventional in situ sensors to obtain averages of surface wind speed, wind direction, air temperature, relative humidity, barometric pressure, and precipitation at the 1-min, 30-min, and daily intervals. Data archives of 10 stations exist for the EBBR and of 5 stations for the SMOS during the 1994–95 period. Herein, we use the A1-level data (Table 4), in which calibration factors are applied. The data are provided as 30-min averages, and we apply the same procedure as for the FLUXNET (month-specific diurnal cycle climatology) to fill in any missing measurements. The resulting half-hourly data are further averaged to 3-hourly (1995 only), daily, and monthly for consistency with the model output from the GSWP-2.

3. Analyses

a. Global-scale evaluation

1) ANNUAL MEAN AND VARIABILITY

For the global, mean annual estimates of evapotranspiration, the GSWP-2 models exhibit a range of values in the B0 simulation of 49–75 trillion metric tons per year (TMT yr^{-1}; 10^{15} kg yr^{-1}). The model-mean value is 65 TMT yr^{-1} (Fig. 1a) with a notable clustering of model results (i.e., 7 of the 13 models are within ±2.5%). In terms of a unit-area flux, 1 TMT is equivalent to 6.67-mm depth of water distributed equally across all land areas, and thus the model-mean, global land annual evapotranspiration flux is 434 mm yr^{-1} or 1.19 mm day^{-1}. The intermodel scatter seen in the baseline simulations is largely preserved in the sensitivity experiments. Even though fewer of the participating models conducted these sensitivity runs (Table 3), the range between COLA's Simplified Simple Biosphere Model (SSiBCOLA) and the Soil–Water–Atmosphere–Plant (SWAP) model remains fairly constant across all sensitivity runs. The total range (i.e., highest to lowest) of the baseline simulations of global evapotranspiration is 26 TMT yr^{-1}. With respect to the modern-era observationally based global water budget assessments by SH07, this range is comparable to the global imbalance of precipitation and evaporation (approximately 24 TMT yr^{-1} or 5% of the global precipitation rate). This result could presumably be regarded as evidence that the GSWP-2 scatter could potentially "explain" the (mean annual) global imbalance of water budget observations; however, the analyses that follow will show this explanation to be unlikely. The range is considerably larger than the interannual variability of any particular GSWP-2 model's annual evapotranspiration, which is approximately 0.65 TMT yr^{-1} [taken as the value of σ_{total} from Table 4 of Dirmeyer et al. (2006)]. Furthermore, the choice of atmospheric forcing (discussed in more detail later) is seen to shift the model-mean estimate by as much as ±5 TMT yr^{-1} (or approximately ±8% of the baseline simulation model-mean value), and that the largest shifts result from changes in the precipitation forcing. We interpret the scatter among the baseline simulations as an indication of "structural uncertainty" in the GSWP-2 modeled evapotranspiration and thus a result of the models' differences in parametric complexity, parameter values, as well as hydrothermal discretization of the soil [summarized in Table 1 of Dirmeyer et al. (2006)]. As such, the results indicate that model structure plays a more important role than uncertainty in atmospheric forcing for these global evapotranspiration estimates.

We can use the GSWP-2 model-mean estimate of global land evaporation (and the intermodel standard deviation) together with the global precipitation estimates and sampling error (from SH07) to obtain as a residual an estimate for global, mean ocean evaporation (Table 5). To perform this calculation, an estimate for the evaporation rate over Antarctica (not considered in the GSWP-2 simulations) is also required. For this, we used the approach as given by Loewe (1957), which provides evaporation flux rates as a function of latitude, and

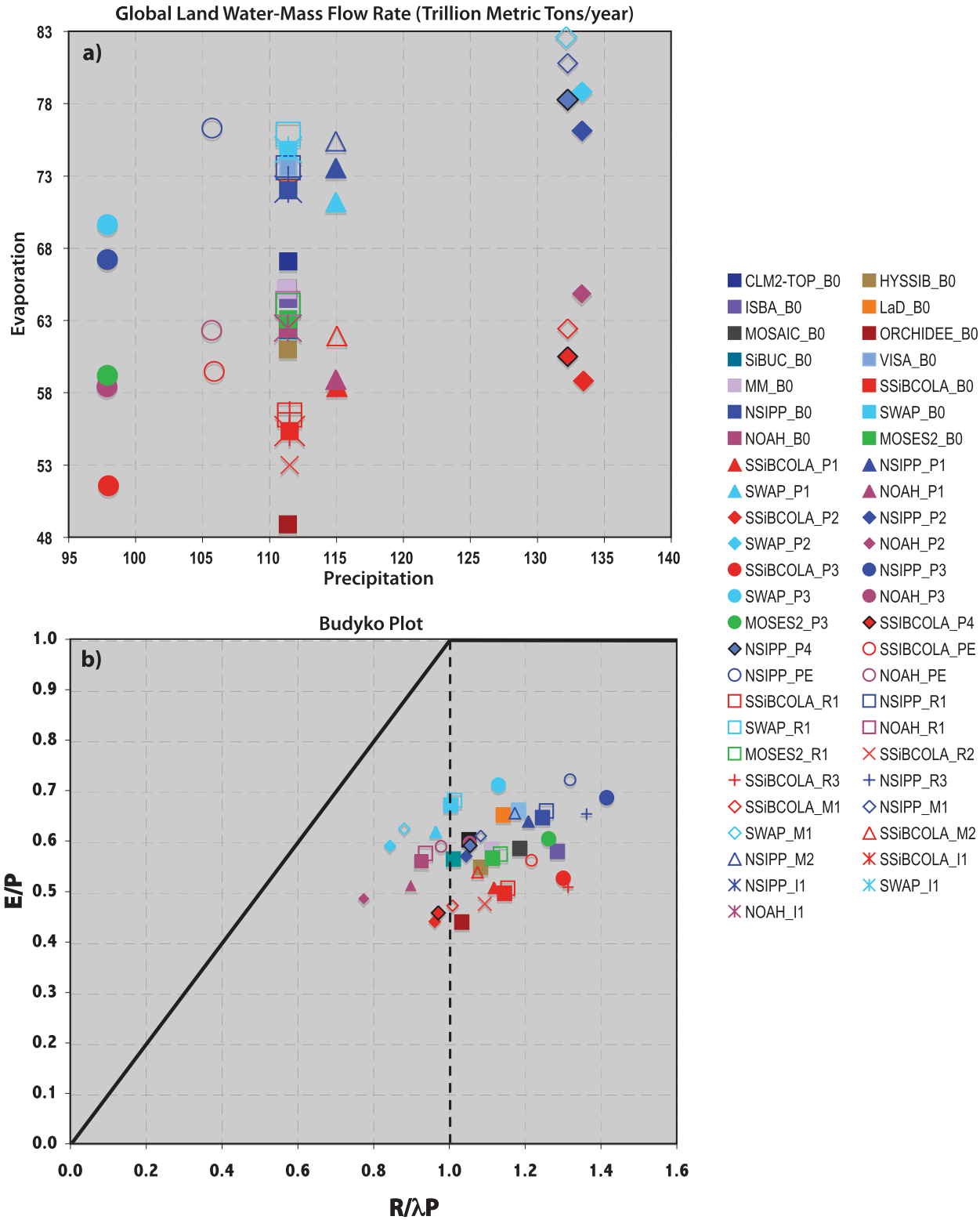

FIG. 1. (a) Scatterplot of global annual mean land E, as simulated by the GSWP-2 models, against global P prescribed by the various simulation experiments (experiments denoted in the figure key by a postfix label to each model acronym; see Table 2 for each simulation description). The vertically stacked model points for each simulation experiment illustrate that the models were driven by the same P data (as expected; TMT yr^{-1}). In terms of flux units, note that 1 TMT yr^{-1} is equivalent to 6.67 mm yr^{-1} (distributed equally across all land areas). (b) Budyko diagram of GSWP-2 simulations showing the ratio E:P as a function of potential evaporation (estimated as net R) over P (energy equivalent). Results are based on global mean annual values.

TABLE 5. Global (excluding Antarctica) annual land evaporation rates and standard deviations based on the GSWP-2 models (for the period 1986–95). Also shown is an estimate of total annual evaporation for Antarctica based on Loewe (1957). The residual calculations of annual ocean evaporation are shown using the global GPCP and CMAP P rates (from Schlosser and Houser 2007) together with the GSWP-2 evaporation plus the Antarctica evaporation (denoted by asterisk). Residual error of ocean evaporation calculated using sampling errors from GPCP and CMAP (Schlosser and Houser 2007) and the standard deviation of the annual GSWP-2 evaporation (Dirmeyer et al. 2006). Units are in kilograms per year.

Land evaporation		Global precipitation	
Antarctica	GSWP-2	GPCP	CMAP
7.41×10^{14}	$6.51 \times 10^{16} \pm 0.08 \times 10^{16}$	$4.9 \times 10^{17} \pm 0.15 \times 10^{17}$	$4.94 \times 10^{17} \pm 0.09 \times 10^{17}$

Ocean evaporation			
GPCP–GSWP-2*	CMAP–GSWP-2*	HOAPS3	GSSTF2
$4.24 \times 10^{17} \pm 0.15 \times 10^{17}$	$4.28 \times 10^{17} \pm 0.09 \times 10^{17}$	3.95×10^{17}	4.30×10^{17}

we integrated these rates over the Antarctic land area. Inclusion of this Antarctic flux estimate increases the global GSWP-2 evapotranspiration by approximately 1% (Table 5). On the basis of these estimates, we find that the implied mean evaporation from the global oceans to be 426 ± 12 TMT yr^{-1}. The GSWP-2 residual estimate is more consistent to the GSSTF2 estimate (430 TMT yr^{-1}) as opposed to the HOAPS estimate (395 TMT yr^{-1}); however, uncertainty bounds for both the GSSTF2 and HOAPS estimates are not available (and beyond the scope of this study), and thus an unequivocal assessment in this regard is not possible. It is encouraging that the GSWP-2 residual falls in between the more explicit and widely used estimates of global ocean evaporation rates.

Looking further at the disparity among these global-scale evaporation estimates (Fig. 2), the spread in the annual land estimates from all of the participating GSWP-2 models (13 B0 simulations) is approximately half of—and never greater than—the difference between the GSSTF and HOAPS ocean estimates. Considering that the ocean covers approximately twice as much of the earth's surface as the land, this twofold increase in the difference between the global ocean evaporation rates (compared to the GSWP-2 range) is not surprising. Yet, it is worth noting that, generally speaking, the two ocean estimates considered in this study use very similar bulk aerodynamic algorithms but with different sources of atmospheric data to satisfy their formulas' requirements, whereas the GSWP-2 spread is a result of structural differences among the models, but each one is forced by identical atmospheric conditions. There is also a notable increase in the spread of the global evaporation estimates (constructed by the GSWP-2 B0 estimates and the ocean evaporation algorithms) starting in 1991. As noted in SH07 (see their Fig. 8), this increase is primarily a result of a sharp decrease in the HOAPS humidity gradient fields (derived from Advanced Very High Resolution Radiometer data) throughout the tropics following the Mt. Pinatubo eruption. Then, the persistently smaller values of HOAPS (compared to the GSSTF estimate) in subsequent years are primarily attributed to weaker tropical wind fields (Fig. 8 of SH07). Nevertheless, in choosing any of the two ocean evaporation datasets considered (and widely used in the climate research community), the GSWP-2 scatter cannot account for the global imbalance between evaporation and precipitation for all years considered in this study, which according to SH07, should only be on the order of 10^{14} kg, as indicated by annual global water vapor tendencies (Fig. 6 of SH07).

2) MEAN ANNUAL CYCLE

One of the more considerable discrepancies among the global water budget terms in the SH07 study is seen in the depiction of the mean annual cycles. For this study, none of the combinations of water flux terms (i.e., precipitation and evaporation), which include the addition of the GSWP-2 estimates, were able to produce global $E - P$ values that matched consistently with observed variations in global atmospheric water vapor storage (Fig. 3). When considering the GSWP-2 model-mean estimate for global land evapotranspiration, as well as the model spread about the mean (Fig. 3, gray shaded region), only marginal consistency can be inferred between monthly tendencies of global $E - P$ and water vapor storage during the Northern Hemisphere warm-season months; however, for the remaining months of the annual cycle, none of the GSWP-2 model results can account for the substantial bias that exists between global $E - P$ and the monthly changes in atmospheric water storage. Additionally, the relative maximum of net atmospheric water gain (occurs in June) is one month earlier than that inferred from the $E - P$ estimates (occurs in July) and similar—but mixed—results are seen for the relative minimum. Moreover, all $E - P$ estimates show notably higher magnitudes of their annual cycles as

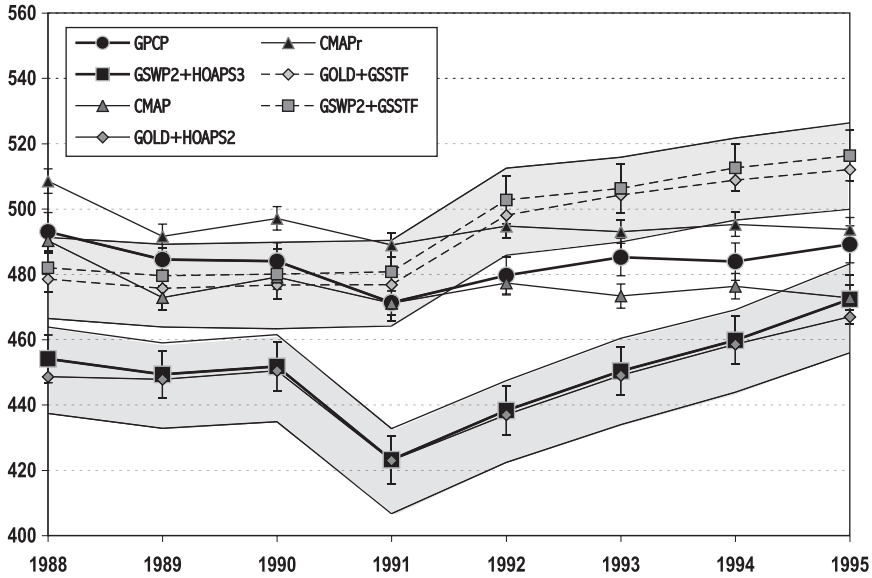

FIG. 2. Annual time series of global water mass flows (TMT yr^{-1}; refer Fig. 1 caption for conversion to flux units). Shown are the combinations of various datasets (see text for details) to obtain global evaporation, as well as the global P rates from GPCP and CMAP (that also include CMAPr). The gray shaded region indicates the extent of the GSWP-2 model scatter about the global evaporation estimate (the model-mean GSWP-2 is used for the marked curves).

compared to the atmospheric water storage changes. The inconsistent timing of the relative maxima/minima and magnitude of the $E - P$ annual cycle are closely aligned with the corresponding features of the GSWP-2 global evapotranspiration (Fig. 4). This does not necessarily prove that all the GSWP-2 estimates are wrong, but it does implicate that its interplay with observationally based estimates of global precipitation and ocean

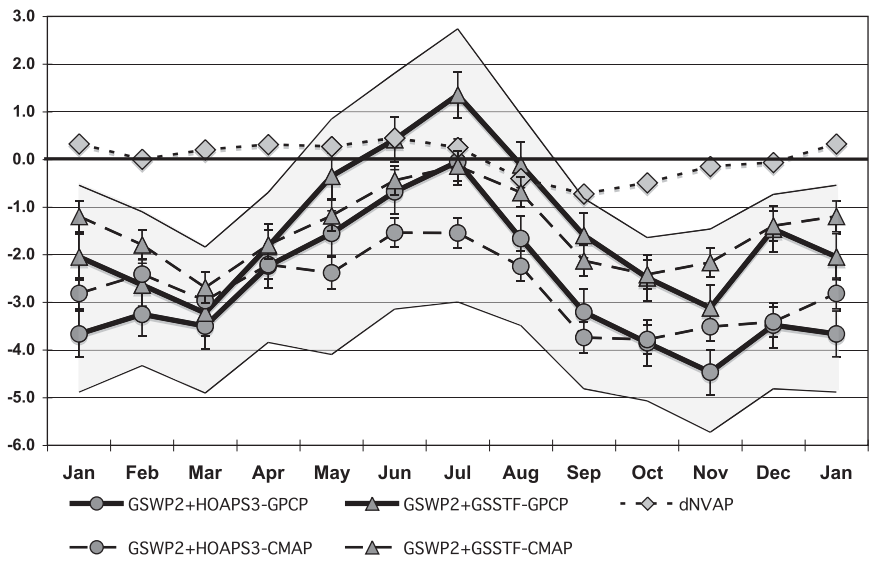

FIG. 3. Global mean annual cycles, for the GSWP-2 simulation period, of $E - P$ (TMT $month^{-1}$) for selected combinations of evaporation estimates (from Fig. 2) as well as GPCP and CMAP P estimates. As in Fig. 2, the gray shaded region indicates the total range of the $E - P$ estimates from the GSWP-2 model collection. Also shown is the corresponding monthly change in total atmospheric water vapor storage.

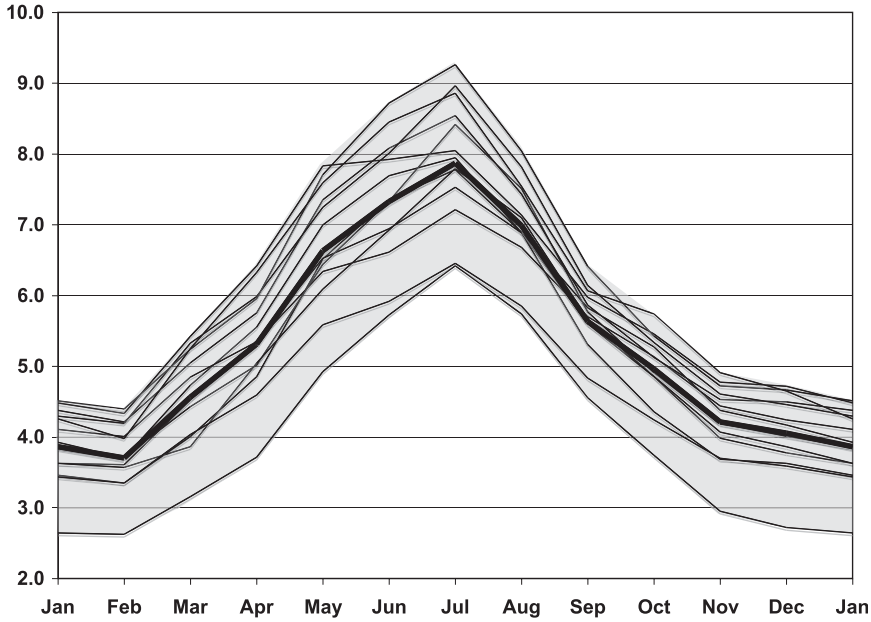

FIG. 4. Global mean annual cycles of E from all the GSWP-2 simulations (TMT month^{-1}). The thick curve indicates the GSWP-2 model mean, and thin curves indicate individual model results. The total scatter of the GSWP-2 models is highlighted in gray.

evaporation is not consistent with observations of global water vapor.

The systematically lower values of $E - P$ (and in some months, opposite sign) to atmospheric water vapor changes, particularly from October through May, imply substantial biases between E and P and/or a measurement error in water vapor. Unfortunately, the uncertainty estimates of the monthly atmospheric water vapor were not readily obtainable for evaluation in this study. Nevertheless, given these large systematic differences (between 2 and 5 TMT, depending on the choice of E and P estimates), the measurement error in global water vapor would need to be on the order of 20% (i.e., noting that from Fig. 6 of SH07, global water vapor storage is approximately 10 TMT or 10^{16} kg) to partially explain these discrepancies; however, in doing so, this would also consume most—if not all—of its annual cycle signal (seen in Fig. 3). Furthermore, in the absence of water vapor trends, the annual mean of the $E - P$ tendencies should be zero. The NVAP observations indicate a decrease in global water vapor storage of approximately 0.03 TMT through the 1988–95 period (Fig. 6 of SH07). While this trend implies a mean negative rate (or bias) of global $E - P$ through the period, it is orders of magnitude smaller than the systematic bias of approximately 2 TMT month^{-1} seen here. In addition, the range of GSWP-2 evapotranspiration (Fig. 4) cannot account for this inconsistency throughout the entire annual cycle. Thus, refinements in the global precipitation and ocean evaporation estimates and error estimates of water vapor measurement are needed to clarify these inconsistencies.

b. *Sensitivity to precipitation and radiation forcing*

Taking advantage of the suite of sensitivity experiments (Table 2) run by a subset of the GSWP-2 models for which baseline runs were also submitted (Table 3), we assess the global-scale sensitivity of evapotranspiration to two primary atmospheric forcing terms: P and net R. Comparison between these two sensitivities can indicate whether the GSWP-2 models are more sensitive to global changes in water or energy availability. For every model, we calculate the change in global evapotranspiration with respect to all combinations of changes in the two forcing terms considered (Table 6); however, to create directly comparable sensitivities to precipitation or net radiation changes, changes in R (W m^{-2}) are converted to millimeters per day for these calculations as given by Dirmeyer et al. 2004 (1 W m^{-2} = 0.034 55 mm day^{-1}). Using units of millimeters per day for E and P, these sensitivities $d\overline{E}/d\overline{P}$ and $d\overline{E}/d\overline{R}$ (Table 7 and overbar denotes global area-weighted mean) are unitless and, in principle, calculable given that each of the sensitivity experiments changes these forcings one at a time in a consistent fashion. As will be shown, however, care must be taken in the interpretation of these results.

The precipitation sensitivity results provide five model samples, with four of the models reporting runs for at least three of the five possible experiments (i.e., runs

TABLE 6. On the basis of Table 1, the differences that exist between the P and R sensitivity simulations as well as the B0 experiment are described. Table rows are grouped to distinguish certain run comparisons (refer to text)—such as SRB, NCEP, or ERA-40 R forcing terms being replaced by ISCCP—or the changes made to P are modifications to the NCEP–DOE product used in the B0 run. The bottom-most delineated rows indicate that at least one of the two sensitivity runs considered uses ERA-40 P but with NCEP–DOE meteorology. The rows with boldface font indicate the runs used to assess overall sensitivity (see text for details).

Expt	Relative differences
R1–B0	**NCEP–DOE vs SRB radiation**
R2–B0	**ERA-40 vs SRB radiation**
R2–R1	**ERA-40 vs NCEP–DOE radiation**
R3–B0	ISCCP vs SRB radiation
R3–R1	ISSCP vs NCEP–DOE radiation
R3–R2	ISCCP vs ERA-40 radiation
P4–P3	**Add GPCC gauge analysis to original NCEP–DOE**
P3–P2	Add wind undercatch correction to NCEP–DOE with GPCC gauge analysis
P2–B0	Add GPCP relaxation to complete the NCEP–DOE hybridization
P3–B0	**Add wind undercatch correction and GPCP relaxation to complete the NCEP–DOE hybridization**
P4–B0	**Original vs hybridized NCEP–DOE**
P4–P2	Add wind undercatch correction and GPCC gauge analysis to original NCEP–DOE
P1–PE	Original vs hybridized ERA-40
P1–B0	Original ERA-40 vs hybridized NCEP–DOE precipitation
PE–B0	Hybridized ERA-40 vs hybridized NCEP–DOE precipitation
P4–P1	Original NCEP–DOE vs original ERA-40 precipitation
P4–PE	Original NCEP–DOE vs hybridized ERA-40 precipitation
P2–P1	Hybridized NCEP–DOE without GPCP relaxation vs original ERA-40
P3–P1	NCEP–DOE with GPCC analysis only vs original ERA-40
P2–PE	Hybridized NCEP–DOE without GPCP relaxation vs hybridized ERA-40
P3–PE	NCEP–DOE with GPCC analysis only vs hybridized ERA-40

TABLE 7. Global evapotranspiration sensitivities $d\overline{E}/d\overline{P}$ and $d\overline{E}/d\overline{R}$ as calculated from the R and P sensitivity experiments for the participating models. Here, R (W m^{-2}) is converted to millimeters per day as given by Dirmeyer et al. 2004 (see section 3b). The tendency terms are calculated based on 10-yr averaged global mean values. Row delineations are consistent with those shown in Table 6. The rows with boldface font indicate the runs used to assess overall sensitivity (see text for details).

		$d\overline{E}/d\overline{P}$ or $d\overline{E}/d\overline{R}$				
Sensitivity	Expt	SSiB	NSIPP	SWAP	MOSES2	Noah
$d\overline{E}/d\overline{R}$	**R1–B0**	**0.24**	**0.31**	**0.21**	**0.19**	**0.38**
	R2–B0	**0.27**				
	R2–R1	**0.26**				
	R3–B0	0.02	0.03			
	R3–R1	−0.03	−0.03			
	R3–R2	0.09				
$d\overline{E}/d\overline{P}$	**P4–P3**	**0.26**	**0.32**			
	P3–P2	0.21	0.25	0.26		0.18
	P2–B0	0.16	0.18	0.18		0.11
	P3–B0	**0.28**	**0.36**	**0.39**	**0.29**	**0.30**
	P4–B0	**0.25**	**0.30**			
	P4–P2	−1.55	−2.00			
	P1–PE	−0.11	−0.29			−0.37
	P1–B0	0.86	0.41	−1.04		−0.97
	PE–B0	−0.73	−0.74			0.01
	P4–P1	0.12	0.27			
	P4–PE	0.04	0.07			
	P2–P1	0.02	0.14	0.41		0.32
	P3–P1	0.40	0.37	0.09		0.03
	P2–PE	−0.02	−0.01			0.09
	P3–PE	1.02	1.17			0.50

P1, P2, P3, P4, PE, see Table 3). First, we focus on the runs that change—but do not substitute—the NCEP–DOE precipitation (used in the B0 run), which are runs P2, P3, and P4. For the most part, the evapotranspiration sensitivities in this group (third group in Tables 6 and 7) show a reasonable consistency in the sign and magnitude. The notable exception is found for the P4–P2 result, which shows an exaggerated negative sensitivity to a small change in global precipitation from the NCEP–DOE product as a result of the Global Precipitation Climatology Centre (GPCC) analysis plus the wind undercatchment adjustment. Recent evidence suggests that the wind undercatchment adjustment is likely to have been excessive and erroneous, resulting in questionable quality of the P2 precipitation field [Decharme and Douville (2006) and see next section]. We also note that, for all GSWP-2 models performing these sensitivity runs, the evapotranspiration sensitivities obtained from the P2–B0 change (i.e., effect of GPCP blending at low gauge density) consistently show the lowest, nonnegative value compared to all other NCEP–DOE precipitation modifications [i.e., excluding *substitution* with the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) precipitation]. In view of these results, we must call into question the sensitivity quantifications that result from the P2 simulations.

What is perhaps more striking is that the sensitivities obtained from either the P1 or PE runs, which *substitute* the NCEP–DOE with the ERA-40 precipitation (bottom

group of Tables 6 and 7), show a wide ride of values with no apparent consistency or clustering. For these P1 and PE runs, the consistency of the substituted ERA-40 precipitation (hybridized or not) with the remaining meteorological fields[1] (i.e., radiation, surface-air temperature Tair, winds, humidity, and air pressure Psurf) of the NCEP–DOE product is not assured. In other words, we are referring to the condition in which the timing, duration, and/or amount of (ERA-40) precipitation at any grid cell may not necessarily correspond to the (NCEP–DOE) radiation or atmospheric state variables (noted earlier). Therefore, it is reasonable to expect that any degree of inconsistency between the precipitation and remaining meteorological fields will cause spurious sensitivities and inconsistent behavior from the models (and seen in these results).

For the radiation sensitivity runs, we have a much smaller sample size of model results (Table 3). Nevertheless, we are able to make some characterizations among the modeled evapotranspiration sensitivities obtained. There is a notable difference between those sensitivities obtained with the ISCCP radiation substitution (denoted by the 2nd group in Tables 6 and 7 mean value of approximately 0.02 and with values of opposite sign) as opposed to those that result from a substitution of the B0 radiation fields with the ERA-40 or NCEP reanalyses radiation (mean value of approximately 0.27). This disparity is not necessarily a reflection of differences in quality between any of the radiation products, but more likely consistency issues with the remaining meteorological data (as seen in the precipitation sensitivities). The B0 radiation field (shortwave down SWdown and longwave down LWdown) is a hybridization of the Surface Radiation Budget (SRB) data with the NCEP reanalyses (Dirmeyer et al. 2006), while the R3 radiation field is a result of replacing the 3-hourly ISCCP product with no hybridization. Further, the R1 radiation is the NCEP reanalysis (used in the B0 hybridization), and the spatiotemporal patterns of the R1 and R2 radiation fields (not shown) are quite similar. While this does not quantify the extent of inconsistency in the R3 radiation (to the remaining meteorological variables), it does call into question its suitability for this sort of sensitivity assessment, and that further analysis (beyond the scope of this study) is warranted.

Therefore, in considering these results to characterize overall evapotranspiration sensitivity (to uncertainties in forcing), we consider only the simulations with NCEP–DOE precipitation, and we further exclude any runs that involve the wind undercatchment adjustment (i.e., the P2 run). For sensitivities with respect to radiation, we have chosen not to consider any of the R3 simulations given the aforementioned considerations. This leaves us with three combinations of runs to pool for sensitivity to precipitation (i.e., P4–P3, P3–B0, and P4–B0), and three combinations for sensitivity to net radiation (i.e., R1–B0, R2–B0, and R2–R1). These runs are shown in boldface in Tables 6 and 7. As such, we find that global evapotranspiration's sensitivity to precipitation is 0.31, and the averaged sensitivity of evapotranspiration to radiation is approximately 0.27. The difference between these two mean sensitivities, while small, is consistent with the characterization that most of the GSWP-2 model simulations are marginally located on the ''water limited'' region of the Budyko curve (Fig. 1, bottom panel); however, looking further at the results for Noah, we find that the sensitivity for evapotranspiration with respect to radiation is higher than that with respect to precipitation. This result is, nevertheless, consistent with the positioning of its global evaporability and index of dryness values that place it predominantly within an ''energy limited'' categorization.

c. *Intermodel consistency*

Our findings indicate that model structure plays a more substantial role than the meteorological inputs in the uncertainty of the GSWP-2 evapotranspiration estimates. Given this information, we use a simple metric to quantify the degree to which the models perform consistently (or not) among themselves, as a guide for further model analyses and development. We perform pointwise temporal correlations R^2 between all possible combinations of models for the B0 simulations (a total of 78) and then take the average of these correlations. The strongest and most ubiquitous agreement among the models lies in the simulation of the annual cycle. The most notable exception to this characterization (i.e., low correlation) is seen in tropical regions (top panel, Fig. 5) and can be associated to the locations of the ''broadleaf evergreen'' vegetation type prescribed by the GSWP-2 models. We also find that the models show their largest and most widespread inconsistency among evapotranspiration variations at interannual time scales in many boreal regions (bottom panel, Fig. 5); however, consistency among the model simulations is not necessarily indicative of their fidelity. For example, while the GSWP-2 models may agree in the timing of the seasonal maximum of global evapotranspiration (Fig. 4), it may very well be contributing to an inconsistent seasonal variation between the global balance of $E - P$ and atmospheric water vapor (Fig. 3). As shown (Fig. 5), regions where the GSWP-2

[1] Hereafter, the term ''remaining meteorological fields'' refers to all of the atmospheric variables of the GSWP-2 forcing but excludes the variables to which they are made reference.

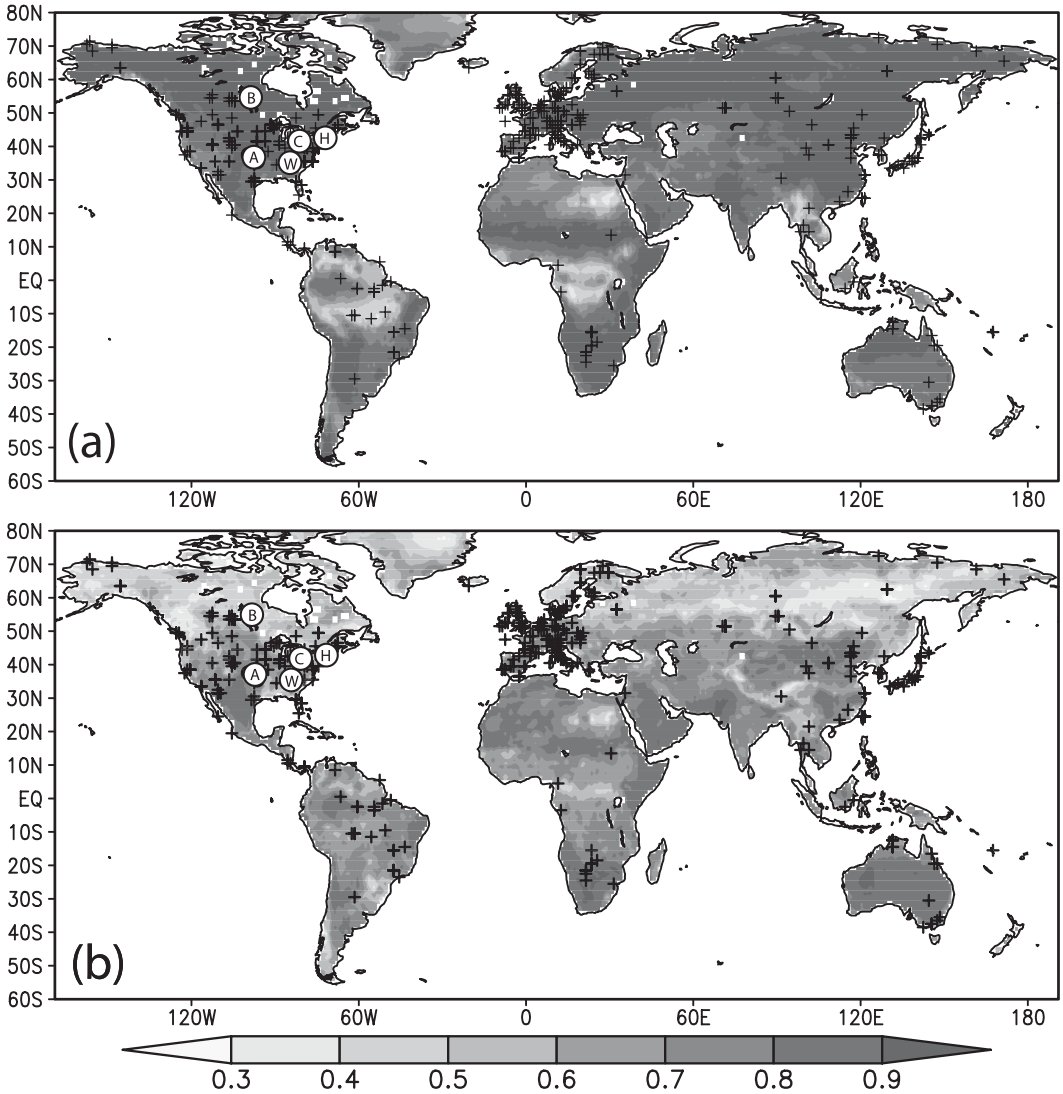

FIG. 5. Maps summarizing the global, pointwise temporal consistency among the monthly E estimates of the GSWP-2 models. Consistency is defined as the averaged R^2 among all the possible combinations of models. (a) The result of this metric with the annual cycle included. (b) The result with the annual cycle removed (i.e., anomaly R^2). Encircled letters indicate locations of field data used in subsequent local-scale evaluations (Figs. 6–7; see text for key/details of field data). Crosshairs indicate locations of field data for which E estimates are available, but they do not coincide with the GSWP-2 period (and are not considered explicitly in this study for evaluation).

models indicate some of the largest model disparities (northern high latitudes) cannot be comprehensively evaluated because of the absence of field data. Nevertheless, we are able to partially address these issues with a small collection of complementary field data (Table 4).

With the available field data, we calculate monthly correlation and RMSE of the models against observed evapotranspiration, and we display these two metrics as scatterplots (Fig. 6). First and foremost, the results tend to corroborate the global assessment provided by Fig. 5, that the ability of the GSWP-2 models to reproduce the observed interannual variability of evapotranspiration at higher latitude locations is not as robust. For the three highest latitude sites, all correlations are reduced and a considerable portion of the correlations becomes negative when the annual cycle is removed from the time series (bottom panel, Fig. 6). While the RMSE is reduced in these cases, this is caused mostly by the fact that the magnitude of the interannual variations is smaller than the annual cycle [cf. Fig. 4b of Dirmeyer et al. (2006)]. For the lower latitude points (Fig. 6), the results are qualitatively consistent—but the diminished correlations, when removing the annual cycle, are not as dramatic.

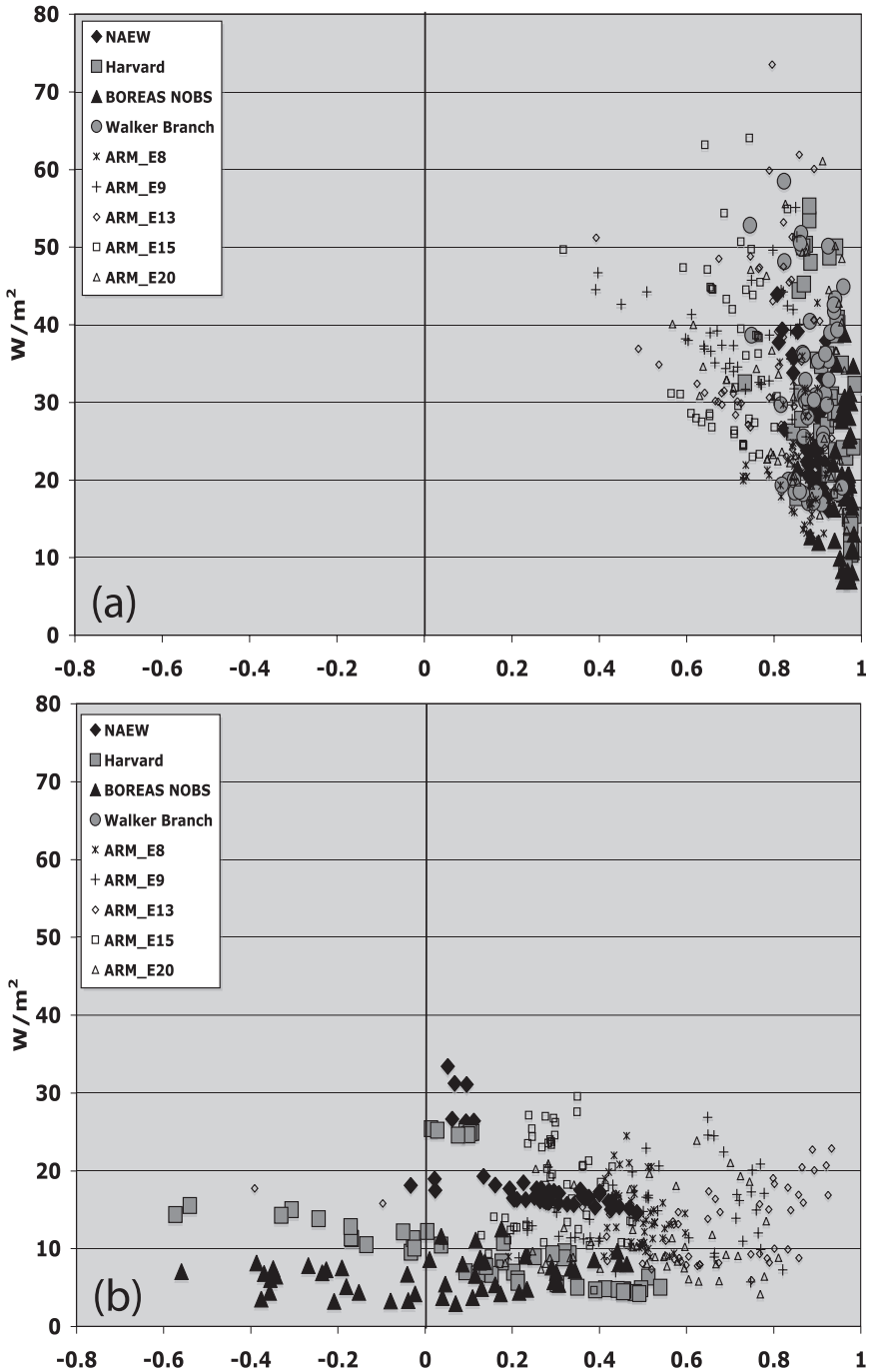

FIG. 6. Scatterplots of RMSE (W m^{-2}) vs temporal R^2 of GSWP-2 modeled monthly latent heat flux vs observations from a selection of field data collections (see text for details). (a) The results when considering the annual cycle. (b) The results with the mean annual cycle removed. The larger and filled plot marks indicate those field sites that are at (relatively) higher latitudes. The Walker Branch results are not included in (b) because of insufficient data length (only 1 yr of data).

Evaluation of the models' monthly averaged diurnal cycle of latent heat flux (Fig. 7, excluding NAEW, data not available) indicates that the models' collective inability to reproduce the observed values is greatest during the middle of the day during the warm-season months (April–October) of 1995. Additionally, we find that at the Walker Branch site, the models show the greatest RMSE during April and May, while the Harvard Forest and BOREAS

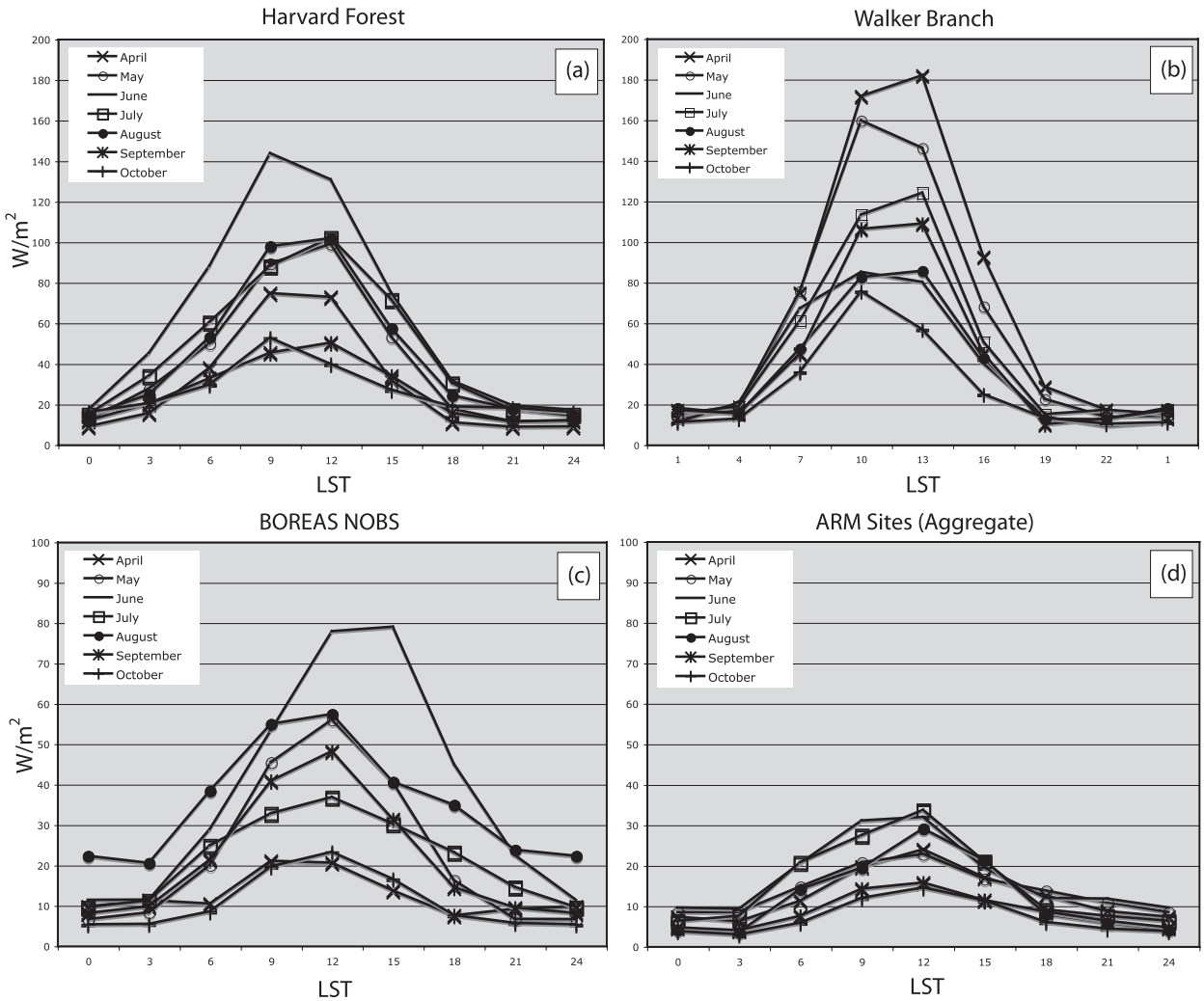

FIG. 7. Monthly aggregates of the RMSE of simulated diurnal cycle of latent heat flux (W m^{-2}) for (a) Harvard Forest, (b) Walker Branch, (c) BOREAS NOBS, and (d) average of all of the ARM Cloud and Radiation Testbed sites. Results based on the 1995 simulation year from the GSWP-2 models. Each curve represents the RMSE obtained by pooling each monthly averaged diurnal cycle (at 3-h time step) from the GSWP-2 B0 simulation data for a given month and calculating the RMSE against the corresponding observations for that site.

sites indicate June as the most problematic month for the modeled estimates. For the aggregated ARM sites, June and July show the highest peaks in RMSE but only marginally so compared to other months. Similar results (not shown) are found for the individual ARM sites (Table 4) as well.

Aside from model deficiencies, the errors shown between the models and the field observations may have also (partially) resulted from inconsistencies between the GSWP-2 grid aggregate and field site conditions. The largest errors (in the diurnal cycles) are found at the Harvard Forest and Walker Branch sites, and it is also these sites where the locally observed vegetation conditions show a weaker correspondence (compared to the BOREAS and ARM sites) to the vegetation type described at the GSWP-2 model grids (Table 4). An additional concern is whether the local meteorological conditions at these field sites have any consistency to the corresponding GSWP-2 grid. Available precipitation data at these sites indicate that the baseline simulation (as well as the P1, P3, and PE sensitivity runs) shows a strong degree of consistency in the seasonal-to-interannual variations (Fig. 8), and therefore the evaporation errors at these sites are likely not a result of inconsistent precipitation provided by the GSWP-2 gridded data. Conversely, the correlation and/or RMSEs of the P2 and P4 precipitation to the field observations are considerably degraded, which is consistent with previous evaluations (e.g., Decharme and Douville 2006) and the interpretations of our own findings in the evapotranspiration sensitivities (Table 7).

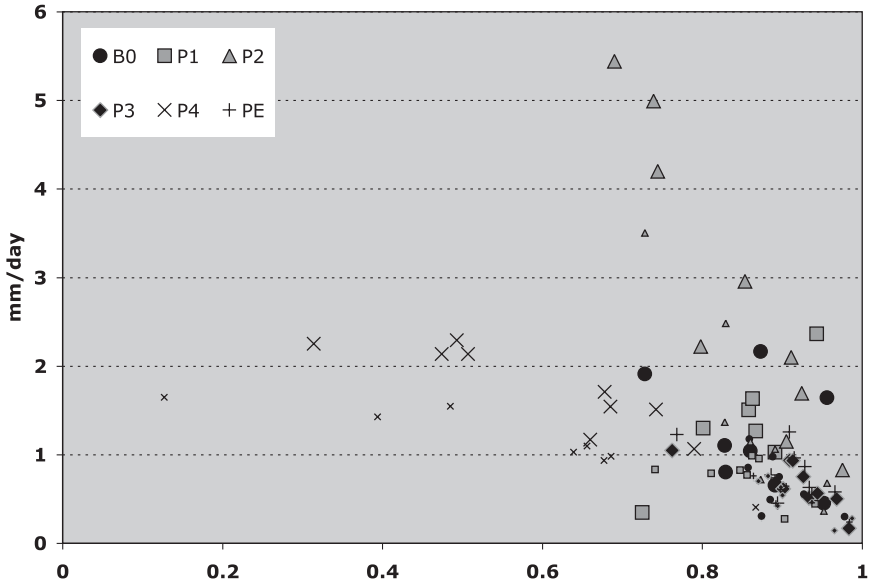

FIG. 8. Temporal consistency of the (monthly) gridded values of GSWP-2 *P* that correspond to the location of the field sites selected for this study (Table 4). Shown is a scatterplot of the RMSE (mm day^{-1}) vs R^2. Each mark denotes the choice of the GSWP-2 *P* forcing data (i.e., B0, P1, P2, P3, P4, and PE). The larger marks indicate results from the raw time series; the smaller marks show results with the averaged annual cycle removed.

4. Closing remarks

We have assessed the simulations of global-scale evapotranspiration from the second Global Soil Wetness Project (GSWP-2). We find that at a global scale, the scatter of GSWP-2 evapotranspiration estimates can constrain a modern-era water budget assessment to within ±2.5%, but it cannot unequivocally explain the imbalance between the global (i.e., ocean plus land) precipitation and evaporation annual variations. In addition, inconsistencies in the magnitude and timing of seasonal variations of the global water budget terms are also found to be associated with the GSWP-2 estimates. The scatter among the GSWP-2 global evapotranspiration estimates shows a weak sensitivity to the choice of atmospheric forcing prescribed to the models, and the intermodel temporal inconsistencies are largest for high-latitude interannual variations as well as for the interseasonal variations in the tropics. Evaluation of corresponding field-scale data also confirms the models' discrepancy for estimating evapotranspiration in high-latitude regions. Analyses of sensitivity simulations that replace uncertain forcings (i.e., radiation and precipitation) indicate that most models' evapotranspiration is slightly more sensitive to precipitation than to net radiation perturbations, and that the majority of the GSWP-2 models, at a global scale, are in a slightly moisture-limited evaporative condition.

In the context to faithfully quantify the global water budget, global water vapor variations from the SH07 study, as well as from the results of this study, indicate that variations of atmospheric storage are roughly 0.01% of global precipitation or evaporation. Thus, the scatter of the GSWP-2 evapotranspiration (2.5%) seems quite unsatisfactory. Rigorous error estimates in water vapor retrievals appear to remain elusive, yet more recent data from the AMSR-E and AIRS satellite instruments show great promise in providing a more comprehensive assessment in this regard. Nevertheless, the GSWP-2 results have clarified that improvements in model-based estimates will not be delivered through improvements in the atmospheric data used for inputs. Rather, refinements in the numerical recipes of these land models hold the most promise toward constraining our global water budgets.

This evaluation of the GSWP-2 modeled evapotranspiration places an emphasis on improving our estimates for high-latitude (cold season) processes and tropical areas. For areas of the tropics, the regions showing the largest degree of model disparity are collocated with the widespread regions of the broadleaf evergreen vegetation type (Dirmeyer et al. 2006). The seasonal discrepancies seen could likely be a result of the treatment of the prescribed vegetation phenology among the GSWP-2 models, both in the way they employ algorithms to represent phenology and their effect as well as the datasets required as inputs (i.e., leaf area index/stem area index, fraction of photosynthetically active radiation, greenness, and others). For the high-latitude regions, we find

that a small set of data currently exists to rectify the model discrepancies, and therefore future field experiments need to augment the low density of data. Furthermore, in these regions, many other processes are important for the controls on evapotranspiration that involve complex interactions with carbon cycling and the biogeochemistry of peatlands (e.g., Frolking et al. 2009). At the time of the GSWP-2 exercise, none of the models employed had the capability to represent the dominant plant type of peatlands: bryophytes (i.e., nonvascular plants with no roots or vascular systems), which may potentially be an additional key issue in the subsequent analyses and model development, as well as supporting field observations, to rectify the disparity seen in the GSWP-2 simulations and for modeling evapotranspiration in general. Furthermore, for these regions, which are dominated by cold-season processes, the modeling challenges of snow cover (e.g., Slater et al. 2001) and seasonally frozen soil (e.g., Luo et al. 2004) as well as their interplay with nonfrozen soil hydrothermal processes also contribute substantially to the evapotranspiration simulations. Thus, any subsequent field experiments will need to satisfy a multitude of observational requirements that span across many subdisciplines of biogeophysical and biogeochemical processes.

Acknowledgments. The authors thank Paul A. Dirmeyer and Taikan Oki for their invaluable contributions toward the planning and execution of the second Global Soil Wetness Project, without which this study would not be possible. In addition, the authors also wish to thank all the participants of GSWP-2 who contributed their time and resources to provide their model simulations, without which the fruits of GSWP-2 would have never been realized. The authors also wish to thank Andy Pitman for his valuable review and comments of an earlier version of this manuscript. This work was supported by the NASA Energy and Water Cycle Study (NEWS; Grant NNX06AC30A), under the NEWS Science and Integration Team activities.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today,* **56,** 38–44.

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Baldocchi, D., and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteor. Soc.,* **82,** 2415–2434.

Bentamy, A., K. B. Katsaros, A. M. Mestas-Nuñez, W. M. Drennan, E. B. Forde, and H. Roquet, 2003: Satellite estimates of wind speed and latent heat flux over the global oceans. *J. Climate,* **16,** 637–656.

Chen, T. H., and Coauthors, 1997: Cabauw experimental results from the Project for Intercomparison of Land-Surface Parameterization Schemes. *J. Climate,* **10,** 1194–1215.

Chou, S.-H., C.-L. Shie, B. Atlas, and J. Ardizzone, 2003: Surface turbulent heat and momentum fluxes over global oceans based on the Goddard satellite retrievals, version 2 (GSSTF2). *J. Climate,* **16,** 3256–3273.

Cook, D. R., 2005: Energy Balance Bowen Ratio (EBBR) handbook. ARM Tech. Rep. TR-037, 23 pp.

Decharme, B., and H. Douville, 2006: Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations. *Climate Dyn.,* **27,** 695–713.

Dirmeyer, P. A., and L. Tan, 2001: A multi-decadal global land-surface data set of state variables and fluxes. COLA Tech. Rep. 102, 43 pp. [Available from the Center for Ocean–Land–Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705.]

——, X. Gao, and T. Oki, 2002: The Second Global Soil Wetness Project science and implementation plan. IGPO Publication Series 37, 69 pp.

——, Z. Guo, and X. Gao, 2004: Comparison, validation, and transferability of eight multiyear global soil wetness products. *J. Hydrometeor.,* **5,** 1011–1033.

——, X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.,* **87,** 1381–1397.

Frolking, S., N. Roulet, and D. Lawrence, 2009: Issues related to incorporating northern peatlands into global climate models. *Carbon Cycling in Northern Peatlands, Geophys. Monogr.,* Vol. 184, Amer. Geophys. Union, 19–35.

Gao, X., and P. A. Dirmeyer, 2006: A multimodel analysis, validation, and transferability study of global soil wetness products. *J. Hydrometeor.,* **7,** 1218–1236.

Guo, Z.-C., P. A. Dirmeyer, X. Gao, and M. Zhao, 2007: Improving the quality of simulated soil moisture with a multi-model ensemble approach. *Quart. J. Roy. Meteor. Soc.,* **133,** 731–747.

Hall, F. G., and Coauthors, 2006: ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies. *J. Geophys. Res.,* **111,** D22S01, doi:10.1029/2006JD007366.

Harmel, R. D., J. V. Bonta, and C. W. Richardson, 2007: The original USDA-ARS experimental watersheds in Texas and Ohio: Contributions from the past and visions for the future. *Trans. Amer. Soc. Agric. Biol. Eng.,* **50,** 1669–1675.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1648.

Karam, H. N., and R. L. Bras, 2008: Climatological basin-scale Amazonian evapotranspiration estimated through a water budget analysis. *J. Hydrometeor.,* **9,** 1048–1060.

Loewe, F., 1957: Precipitation and evaporation in the Antarctic. *Meteor. Antarct.,* **11,** 71–89.

Luo, L., and Coauthors, 2004: Effects of frozen soil on soil temperature, spring infiltration, and runoff: Results from the PILPS 2(d) experiment at Valdai, Russia. *J. Hydrometeor.,* **5,** 334–351.

Rodell, M., J. S. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004a: Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophys. Res. Lett.,* **31,** L20504, doi:10.1029/2004GL020873.

——, and Coauthors, 2004b: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.,* **85,** 381–394.

Rossow, W. B., and E. N. Dueñas, 2004: The International Satellite Cloud Climatology Project (ISCCP) Web site: An online resource for research. *Bull. Amer. Meteor. Soc.,* **85,** 167–172.

Schlosser, C. A., and P. Houser, 2007: Assessing a satellite-era perspective of the global water cycle. *J. Climate,* **20,** 1316–1338.

Slater, A. J., and Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeor.,* **2,** 7–25.

Song, J., M. L. Wesely, M. A. LeMone, and R. L. Grossman, 2000: Estimating watershed evapotranspiration with PASS. Part II: Moisture budgets during drydown periods. *J. Hydrometeor.,* **1,** 462–473.

Su, H., M. F. McCabe, E. F. Wood, Z. Su, and J. H. Prueger, 2005: Modeling evapotranspiration during SMACEX: Comparing two approaches for local- and regional-scale prediction. *J. Hydrometeor.,* **6,** 910–922.

Swenson, S., and J. Wahr, 2006: Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. *J. Hydrometeor.,* **7,** 252–270.

Vonder Harr, T. H., J. M. Forsythe, D. McKague, D. L. Randel, B. C. Ruston, and S. Woo, 2003: Continuation of the NVAP global water vapor data sets for Pathfinder science analysis. STC Tech. Rep. 3333, 53 pp. [Available online at http://eosweb.larc.nasa.gov/PRODOCS/nvap/sci_tech_report_3333.pdf.]

Walter, M. T., D. S. Wilks, J.-Y. Parlange, and R. L. Schneider, 2004: Increasing evapotranspiration from the conterminous United States. *J. Hydrometeor.,* **5,** 405–408.

Wang, K., and S. Liang, 2008: An improved method for estimating global evapotranspiration based on satellite determination of net radiation, vegetation index, temperature, and soil moisture. *J. Hydrometeor.,* **9,** 712–727.

——, P. Wang, Z. Li, M. Cribb, and M. Sparrow, 2007: A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *J. Geophys. Res.,* **112,** D15107, doi:10.1029/2006JD008351.

Werth, D., and R. Avissar, 2004: The regional evapotranspiration of the Amazon. *J. Hydrometeor.,* **5,** 100–109.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.

Zhao, M., and P. Dirmeyer, 2003: Production and analysis of GSWP-2 near-surface meteorology data sets. COLA Tech. Rep. 159, 22 pp. [Available from the Center for Ocean–Land–Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705.]

Copyright of Journal of Hydrometeorology is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.