Assimilation of satellite-derived skin temperature observations into land surface models

Rolf H. Reichle[1], Sujay V. Kumar[2,3], Sarith P. P. Mahanama[1,4], Randal D. Koster[1] and Q. Liu[1,2]

Manuscript submitted to Journal of Hydrometeorology

19 January 2010, 2:07 PM

1 Global Modeling and Assimilation Office
NASA Goddard Space Flight Center
Greenbelt, MD, USA

2 Science Applications International Corporation
Beltsville, MD, USA

3 Hydrological Sciences Branch
NASA Goddard Space Flight Center
Greenbelt, MD, USA

4 Goddard Earth Sciences and Technology Center
University of Maryland, Baltimore County
Baltimore, MD, USA

Corresponding author address

Rolf H. Reichle
NASA Goddard Space Flight Center
Mail Code 610.1
8800 Greenbelt Road
Greenbelt, MD 20771
USA

Tel.: 301-614-5693
FAX: 301-614-6297
Email: rolf.reichle@nasa.gov

Abstract

Land surface (or "skin") temperature (LST) lies at the heart of the surface energy balance and is a key variable in weather and climate models. Here we assimilate LST retrievals from the International Satellite Cloud Climatology Project (ISCCP) into the Noah and Catchment (CLSM) land surface models using an ensemble-based, off-line land data assimilation system. LST is described very differently in the two models. A priori scaling and dynamic bias estimation approaches are applied because satellite and model LST typically exhibit different mean values and variability. Performance is measured against 27 months of in situ measurements from the Coordinated Energy and Water Cycle Observations Project at 48 stations. LST estimates from Noah and CLSM without data assimilation ("open loop") are comparable to each other and superior to that of ISCCP retrievals. For LST, RMSE values are 4.9 K (CLSM), 5.6 K (Noah), and 7.6 K (ISCCP), and anomaly correlation coefficients (R) are 0.62 (CLSM), 0.61 (Noah), and 0.52 (ISCCP). Assimilation of ISCCP retrievals provides modest yet statistically significant improvements (over open loop) of up to 0.7 K in RMSE and 0.05 in anomaly R. The skill of surface turbulent flux estimates from the assimilation integrations is essentially identical to the corresponding open loop skill. Noah assimilation estimates of ground heat flux, however, can be significantly worse than open loop estimates. Provided the assimilation system is properly adapted to each land model, the benefits from the assimilation of LST retrievals are comparable for both models.

1. Introduction

Land surface conditions are intimately connected with the global climate system and have been associated, through different pathways, with atmospheric predictability. Land surface temperature (LST) lies at the heart of the surface energy balance and is therefore a key variable in weather and climate models. LST influences the latent and sensible heat fluxes to the atmosphere through which it affects the planetary boundary layer and atmospheric convection. LST also plays an important role in the assimilation of atmospheric remote sensing observations because atmospheric retrieval algorithms (or forward radiative transfer modeling) for surface-sensitive (window) channels require information about land surface conditions. Accurate LST specification is therefore critical to improving estimates of the surface water, energy, and radiation balance as well as atmospheric quantities, which in turn are all critical to improving weather and climate forecast accuracy.

Satellite retrievals of LST (also referred to as "skin temperature") are available from a variety of polar orbiting and geostationary platforms carrying infrared and microwave sensors (Aires et al. 2004, Jin 2004, Minnis and Khaiyer 2000, Pinheiro et al. 2004, Rossow and Schiffer 1991, 1999, Trigo and Viterbo 2003, Wan and Li 1997). Land surface models (driven by observed meteorological forcing data or coupled to an atmospheric model) offer estimates of global land surface conditions, including LST. Errors in the forcing fields, however, along with the imperfect parameterization of land-atmosphere interactions can lead to considerable drifts in modeled land surface states. Land data assimilation systems combine the complementary information from modeled and observed land surface fields and produce dynamically consistent, spatially complete and temporally continuous estimates of global land surface conditions.

45 Assimilating LST retrievals into a land surface model should, in concept, improve model
46 estimates of land surface conditions.
47
48 There has been considerable progress in the methodological development and application of land
49 data assimilation algorithms (Andreadis and Lettenmaier 2005, Balsamo et al. 2007, Crow and
50 Wood 2003, Drusch 2007, Dunne and Entekhabi 2006, Mahfouf et al. 2009, Margulis et al. 2002,
51 Pan and Wood 2006, Reichle et al. 2009, Seuffert et al. 2003, Slater and Clark 2006, Walker et
52 al. 2001, Zaitchik and Rodell 2009, Zhou et al. 2006), with ensemble-based Kalman filtering and
53 smoothing algorithms emerging as a common and promising method for land data assimilation.
54 Development and applications of land data assimilation, however, have largely focused on
55 assimilating observations of surface soil moisture, snow cover, and snow water equivalent, with
56 less effort devoted to LST assimilation.
57
58 The goal of this study is to investigate the potential for assimilating satellite retrievals of LST
59 within a state-of-the-art land surface data assimilation system. Specifically, LST retrievals from
60 the International Satellite Cloud Climatology Project (ISCCP) are assimilated into the NASA
61 Catchment land surface model (CLSM; Koster et al. 2000) and into the Noah land surface model
62 (Ek et al. 2003) with the ensemble Kalman filter (EnKF) developed at the NASA Global
63 Modeling and Assimilation Office (Reichle et al. 2009). For validation of the assimilation
64 products we use in situ observations from the Coordinated Energy and Water Cycle Observations
65 Project (CEOP). We pay particular attention to bias between observed and modeled LST and
66 have fitted the EnKF with several bias estimation algorithms designed specifically to address
67 LST biases. It will be shown that the assimilation algorithm must be customized for the model-
68 specific representation of LST.

2. Background

LST can be retrieved from infrared and microwave sensors on geostationary and polar-orbiting platforms, including the NOAA Geostationary and Polar-orbiting Operational Environmental Satellite series dating back to the early 1980's. Given the extensive global and multi-decadal record of satellite-based LST retrievals, and given the importance of accurate LST estimation in particular for global atmospheric data assimilation systems, it is telling that the challenge of operational LST assimilation has been largely unmet.

The difficulties of LST data assimilation are rooted in the nature of LST retrievals and modeling. LST data from retrievals and land surface models typically exhibit strong biases that depend on observation time and location and that have been well documented (see, for instance, Jin et al. 1997, Trigo and Viterbo 2003, Jin 2004; see also section 6 for examples). Biases arise for a variety of reasons. For instance, LST modeling is fraught with numerical stability problems because in nature the effective heat capacity associated with LST is very small. Land modelers are thus forced to approximate the corresponding heat capacity as zero or to use a surface temperature prognostic variable that represents more than just a very thin layer. The first approach, used for example in Noah, derives LST as a diagnostic variable from the surface energy balance. The second approach, used for example in CLSM, lumps the vegetation and top few centimeters of soil matter into a single model prognostic variable with a small but non-zero heat capacity. The latter approach is obviously at odds with satellite retrievals of LST, which describe the temperature in a much shallower layer at the land-atmosphere interface (vegetation or soil, as viewed from the satellite sensor). On the other hand, the zero-heat capacity approach requires an additional connection between the diagnostic model LST variable and a model

prognostic variable to which data assimilation increments can be applied (so that they can alter the model forecast). Additional discrepancies between the LST observed by the satellite and that computed by the land model stem from the inability of global land models to resolve the land surface at the same fine horizontal resolution as infrared satellite sensors.

Satellite-based LST retrievals suffer from their own set of disadvantages and problems. Infrared LST retrievals are only available under clear-sky conditions and are notoriously prone to cloud contamination (Jin 2004). Microwave LST retrievals are available under cloudy conditions (Aires et al. 2004) but depend on uncertain estimates of microwave land surface emissivity. Both infrared and microwave LST retrievals depend on knowledge of atmospheric conditions above the LST footprint. LST retrievals also depend on the look-angle, which is particularly important for retrievals at high viewing angles (Minnis and Khaiyer 2000, Pinheiro et al. 2004). The benefit of having a great variety and long record of different platforms from which LST can be retrieved is partly negated by the corresponding variety of sensor characteristics and sensor-specific LST retrieval algorithms that make it difficult to achieve a homogeneous satellite LST record.

Additional complications arise when model and satellite LST are combined in a data assimilation system. The strong seasonal and diurnal cycles of LST must be considered because error characteristics may depend on time-of-day and season. Moreover, obvious problems result when (clear-sky) LST retrievals are assimilated into a model at a time and location for which the model state or forcing indicate cloudy conditions. Because the infrared and microwave emissivities of the land surface are not well known, it is difficult to compare the radiometric temperature

observed by the satellite with the physical temperature of the land model. Last but not least, the dearth of validating in situ observations of LST and land surface fluxes is a severe impediment to the validation of data products from satellite observations, modeling, and data assimilation.

One major development path is to assimilate LST retrievals into a simple representation of the land surface energy balance using an adjoint-based variational assimilation approach (Castelli et al. 1999; Boni et al. 2001; Caparrini et al. 2004; Sini et al. 2008). This elegant method requires only a minimal amount of ancillary data and provides robust estimates of evaporative fraction. It is not, however, easily applicable to existing global atmospheric or land data assimilation systems because it is very difficult to develop and maintain adjoint models for the complex land surface model components in such systems. Recently, Meng et al. (2009) developed the adjoint model of just the surface energy balance component of the Common Land Model. Using the variational method, the authors assimilated in situ LST observations from four AmeriFlux sites for up to 20 days and report improvements in evapotranspiration estimates when verified against coincident in situ observations.

Other off-line surface temperature assimilation studies used filtering techniques. Kumar and Kaleita (2003) used the Extended Kalman filter to assimilate in situ observations of surface soil temperature from a single site for one month into a soil heat transfer model based on the discretized diffusion equation. Lakshmi (2000) merged one year of satellite retrievals of LST over the Red-Arkansas basin into a simple two-layer model of the surface water and energy balance.

A few attempts have been made to adjust terms in the surface energy balance of atmospheric models in response to satellite LST retrievals. McNider et al. (1994) describe a technique to assimilate satellite LST into the surface energy budget of a regional-scale atmospheric model. Their method assumes that surface soil moisture is the least known parameter in the early-morning surface energy budget. Van den Hurk et al. (2002) assimilated satellite LST and near-surface relative humidity measurements into a regional weather forecast model. By adjusting the root zone soil moisture and the roughness length for heat the authors find small improvements in the surface energy balance. Garand (2003) outlines a variational method for a unified land and ocean surface skin temperature analysis, including a linear a priori bias correction of the assimilated radiances.

Bosilovich et al. (2007) developed an algorithm for LST assimilation into a global model that introduces an incremental bias correction term into the model's surface energy budget. In contrast to the McNider et al. (1994) approach, all temperature-dependent terms in the surface energy budget respond directly to the LST retrievals. In its simplest form, the Bosilovich et al. (2007) algorithm estimates and corrects a constant time mean bias for each grid point; additional benefits are attained with a refined version of the algorithm that allows for a correction of the mean diurnal cycle. The results of Bosilovich et al. (2007) indicate that LST assimilation improves estimates of 2m air temperature, both in mean and variability, in a coupled land-atmosphere model. Neglecting the diurnal cycle of the LST bias causes degradation of the diurnal amplitude of background model air temperature in many regions. In situ measurements of energy fluxes at several locations were used to inspect the surface energy budget more closely. LST assimilation generally improves the sensible heat flux and, in some cases, it improves the

Bowen ratio. At many stations, however, LST assimilation increases slightly the bias in the monthly latent heat flux. A critical limitation of the method of Bosilovich et al. (2007) is the assumption of unbiased LST retrievals.

In this paper, we restrict ourselves to uncoupled land data assimilation and test several variants of the bias estimation strategy of Bosilovich et al. (2007). We also explore an alternative strategy of scaling the LST retrievals into the climatology of the modeled LST. As we will show in section 6, not scaling the LST retrievals prior to (uncoupled land) data assimilation can create serious imbalances in the model-generated mass and energy fluxes and can lead to entirely unrealistic land surface fluxes. We test these approaches with two land surface models that represent LST very differently: CLSM and Noah (section 4). Our results are directly linked to weather and climate prediction applications because these two land models are used in the atmospheric data assimilation systems of the NASA Global Modeling and Assimilation Office (GMAO) and the NOAA National Centers for Environmental Prediction (NCEP), respectively.

3. Data

In this study, we assimilate LST retrievals from ISCCP (http://isccp.giss.nasa.gov; Rossow and Schiffer 1991, 1999). The ISCCP archive contains satellite-based estimates of global cloud cover and radiative properties from 1983 through the present (recent data are added with a latency of about 1 year), and is based on observations from an international network of meteorological satellites. Specifically, the ISCCP 30 km Pixel Level Cloud Product (DX) includes global, 3-hourly, clear-sky LST retrievals from infrared radiances. For this study, we extracted LST retrievals from the DX archive for the geostationary platforms and aggregated the data to a global latitude-longitude grid with 1 degree resolution for assimilation into our system.

The availability of validating land surface temperature and flux data is very limited. In this study, we use the comparably large collection of such data provided by CEOP (http://www.ceop.net) to validate the data assimilation products. Specifically, we obtained hourly data from the network of CEOP Reference Sites from 1 October 2002 through 31 December 2004 (Figure 1, Table 1). Sufficient data for validation are available at 48 distinct sites, of which 19 sites have LST data, 30 have latent heat (LH) and sensible heat (SH) flux data, and 20 sites have ground heat (GH) flux data. Only 4 stations have LST as well as LH, SH, and GH observations sufficient for validation (Cabauw, Bondville, Lindenberg Falkenberg, and Lindenberg Forest). The hourly CEOP data were aggregated to 3-hourly averages for comparison with the 3-hourly retrieval, model, and assimilation products.

The surface meteorological forcing data for the two land models are from the Global Land Data Assimilation Systems (GLDAS) project (Rodell et al. 2003; http://ldas.gsfc.nasa.gov) and were

provided at 3-hourly time steps and at 2º and 2.5º resolution in latitude and longitude, respectively. The GLDAS data stream is based on output from the global atmospheric data assimilation system at the NASA Global Modeling and Assimilation Office (Bloom et al. 2005) adjusted with pentad precipitation observations from the Climate Prediction Center Merged Analysis of Precipitation (CMAP; http://www.cdc.noaa.gov/cdc/data.cmap.html) and daily estimates of surface radiation from the Air Force Weather Agency (AFWA) Agricultural Meteorology (AGRMET) system. The observation-based corrections ensure that the forcing data and hence the land model output are as close to reality as is possible (without the benefit of assimilating the LST retrievals).

4. Data assimilation system

4.a. The Catchment and Noah land surface models

Modeled LST and land surface fluxes are from integrations of CLSM (Koster et al. 2000) and Noah (Ek et al. 2003). Again, these two models are the land model components of the atmospheric data assimilation and forecasting systems at NASA GMAO and NOAA NCEP, respectively. Both models dynamically predict land surface water and energy fluxes in response to surface meteorological forcing but follow markedly different approaches to modeling soil moisture and LST.

CLSM's basic computational unit is the hydrological catchment (or watershed). The global land surface is divided into catchments (excluding inland water and ice-covered areas) with a mean linear scale of around 50 km (ranging from a few km to 250 km). Unlike traditional, layer-based models, CLSM includes an explicit treatment of the spatial variation of soil water and water table depth within each hydrological catchment based on the statistics of the catchment topography. The surface energy balance is computed separately for the (dynamically varying) saturated, transpiring, and wilting sub-areas of each catchment. In each of these three sub-areas, the bulk temperature of the vegetation canopy and the top 5 cm of the soil column is modeled with a "surface temperature" (TSURF) prognostic variable that is specific to the soil moisture regime. The three TSURF prognostic variables interact with an underlying heat diffusion model for soil temperature (consisting of six layers with depths equal to about 10, 20, 40, 75, 150, 1000 cm from top to bottom) that is common to the three sub-areas (see Figure 5 of Koster et al. 2000). In the absence of snow, the area-weighted average of the three prognostic TSURF variables (hereinafter also referred to as the "surface temperature" in CLSM) is the most appropriate

quantity to compare to satellite-based LST retrievals (Figure 2). CLSM integrations were conducted using the GMAO land data assimilation system (Reichle et al. 2009) with a model time step of 20 minutes.

Noah is a more traditional, layer-based model. Four soil layers of increasing thicknesses of 10, 30, 60 and 100 cm are used to model soil temperature and moisture dynamics with layer-based formulations of the heat diffusion equation (for energy) and of the standard diffusion and gravity drainage equations (for unsaturated water flow). LST in Noah is diagnosed from the surface energy balance equation and, unlike in CLSM, is *not* a prognostic variable and has no associated heat capacity (Figure 2). In this study, we use Noah version 2.7.1 on a 0.5 degree grid with a 30 minute time step. Noah integrations were carried out with the Land Information System (Kumar et al. 2008) fitted with the GMAO ensemble data assimilation and bias estimation modules (Reichle et al. 2009).

4.b. Data assimilation method and parameters

In a data assimilation system, the model-generated land surface estimates are corrected toward observational estimates, with the degree of correction determined by the levels of error associated with each. The assimilation system used here is based on the Ensemble Kalman filter (EnKF), which is well suited to the non-linear and intermittent character of land surface processes (Reichle et al. 2002a, 2002b). The EnKF works sequentially by performing in turn a model forecast step and a filter update step. Formally, the forecast step for ensemble member i can be written as

(1) $x_{t,i}^{-} = f(x_{t-1,i}^{+}, q_{t,i})$

where $x_{t,i}^{-}$ and $x_{t-1,i}^{+}$ are the forecast (denoted with $^{-}$) and analysis (denoted with $^{+}$) state vectors at times t and t-1, respectively. The model error (or perturbation vector) is denoted with $q_{t,i}$ and its covariance with Q_t. The filter update produces the analyzed state vector $x_{t,i}^{+}$ at time t and can be written as

(2) $x_{t,i}^{+} = x_{t,i}^{-} + K_{x,t} (y_{t,i} - H_t x_{t,i}^{-})$

where $y_{t,i}$ denotes the observation vector (suitably perturbed) and H_t is the observation operator (which is written as if it were linear for ease of notation, but in practice the update is solved without explicitly computing H_t; Keppenne et al. 2000). The Kalman gain matrix $K_{x,t}$ is given by

(3) $K_{x,t} = P_{x,t} H_t^{T} (H_t P_{x,t} H_t^{T} + R_t)^{-1}$

where $P_{x,t}$ is the state forecast error covariance (diagnosed from the ensemble $x_{t,i}^{-}$), R_t is the observation error covariance, and superscript T denotes the matrix transpose. Simply put, the Kalman gain $K_{x,t}$ represents the relative weights given to the model forecast and the observations, based on their respective uncertainties and based on the error correlations between the elements of the state vector and the model prediction of the observed variable. In this paper, we use 12 ensemble members with a "one-dimensional" (1d) EnKF that processes each location independently of all other locations (see, for example, Reichle and Koster (2003) for 1d versus

277 3d assimilation). The key feature of the EnKF is that error estimates of the model-generated
278 results are dynamically derived from an ensemble of non-linear model integrations.
279
280 Perturbation fields were generated and applied to surface air temperature and radiation which
281 represent the dominant forcing inputs for LST. Perturbations were also applied to model
282 prognostic variables (the Catchment surface temperature TSURF and the Noah top layer soil
283 temperature TSOIL1), reflecting errors in the modeling of the surface energy balance.
284 Collectively, the perturbations allow us to maintain an ensemble of land surface conditions that
285 represents the uncertainty in modeled LST. An overview of the perturbation parameters is given
286 in Table 2. Depending on the variable, normally distributed additive perturbations or log-
287 normally distributed multiplicative perturbations were applied. The ensemble mean for all
288 perturbations was constrained to zero for additive perturbations and to one for multiplicative
289 perturbations. Moreover, time series correlations were imposed via a first-order auto-regressive
290 model (AR(1)) for all fields. Since we used a one-dimensional EnKF in this study, the
291 perturbation fields were not spatially correlated. At hourly and daily time scales, the
292 meteorological forcing fields are ultimately based on output from atmospheric modeling and
293 analysis systems and not on direct observations of surface forcings. We imposed error cross-
294 correlations that are motivated by the assumption that the atmospheric forcing fields represent a
295 realistic balance between radiation, clouds, and air temperature. Under that assumption, for
296 example, a positive perturbation to the downward shortwave radiation tends to be associated with
297 negative perturbations to the longwave radiation and with a positive perturbation to air
298 temperature.
299

Model errors are difficult to quantify at the global scale. The parameter values listed in Table 2 are largely based on experience and are partially supported by earlier studies (Reichle et al. 2002b; Reichle and Koster 2003; Reichle et al. 2007). The success of the assimilation presented here (section 6) suggests that these values are acceptable. In any case, further calibration of the filter parameters would, in theory, only improve the assimilation results. Clearly, more research is needed on the exact nature of the model and forcing errors. Recently developed adaptive filtering methods for land assimilation may assist with error parameter estimation (Reichle et al. 2008, Crow and Reichle 2008).

The mapping of the satellite information from the observation space into the space of the model states is accomplished through the Kalman gain during the EnKF update step. Equation (2) linearly relates "innovations" (observations minus corresponding model estimates before EnKF update, that is, $y_{t,i} - H_t\, x_{t,i}^-$) to "increments" (model states after EnKF update minus same before EnKF update, that is, $x_{t,i}^+ - x_{t,i}^-$). In this study, we use CLSM's area-average TSURF variable and Noah's diagnostic TSKIN variable to compute the (observation-space) innovations (Figure 2 and Section 4a). The EnKF state vector for CLSM consists of the three TSURF prognostic variables (specific to each soil moisture regime, section 4a), while the EnKF state vector for Noah consists of the top layer soil temperature TSOIL1. The key ingredients to the Kalman gain are the error correlations between the LST variables in observation space and the EnKF state variables (Reichle et al. 2002b). For Noah, the relevant error correlation is between the diagnostic TSKIN variable (that has no associated heat capacity) and the temperature in the top 10 cm soil layer. Therefore, in the case of Noah the error correlation is affected by a small phase shift between the diurnal cycle of the diagnostic TSKIN and the top layer soil temperature

TSOIL1. Observations are not assimilated when precipitation is falling or when the ground is covered with snow (as indicated by the land model integration).

4.c. Bias estimation

For the reasons outlined in section 2 there are considerable differences between the temporal moments of the satellite and model LST (see section 6 for examples). Such biases need to be addressed in the data assimilation system. For this study, we implemented two different strategies. The first strategy is to scale the satellite observations to the model's climatology so that the first and second moments of the satellite LST and the model LST match. This strategy is a simplified version of the cumulative distribution function matching (Reichle and Koster 2004) that has been used successfully for soil moisture assimilation (Reichle et al. 2009). Because of the strong diurnal and seasonal cycles of LST, we estimated the multi-year LST mean and variance separately for each calendar month and for eight different times-of-day (0z, 3z, …, 21z). This strategy is very easy to implement through preprocessing of the LST retrievals and makes no assumptions regarding whether the model's or the observations' climatology is more correct. Although the assimilation estimates are by design produced in the model's climatology, they could be scaled back to the observational climatology if desired. The scaling strategy can be applied to the assimilation of retrievals from a variety of satellite datasets with different climatologies. An obvious disadvantage is the fact that the a priori scaling is static and cannot automatically adjust to dynamic changes in bias.

The second strategy is to dynamically estimate bias parameters along the lines of the algorithm developed by Dee (2005), which was used for LST by Bosilovich et al. (2007) and for soil

346 moisture by De Lannoy et al. (2007). This dynamic bias estimation approach is based on a
347 second Kalman filter for bias estimation (in addition to the Kalman filter for state estimation).
348 Assume that we have a bias estimate b_{t-1}^{+} at time t-1. Furthermore, assume that this bias
349 estimate can be propagated to time t with a simple bias evolution model
350
351 (4) $b_t^{-} = \alpha\, b_{t-1}^{+}$
352
353 that relaxes the bias estimates to zero ($0 < \alpha < 1$). In our experiments, we chose α to correspond
354 to an e-folding scale of 1 day. The use of a relaxation factor is different from the implementation
355 of Bosilovich et al. (2007) and is a prudent strategy for experiments that cover many seasons.
356 Because observations may not be available for extended periods, relaxing the bias estimate to
357 zero is safer than keeping the latest bias estimate through seasons for which it may not be
358 appropriate.
359
360 Next, we compute a bias-corrected model forecast
361
362 (5) $\xi_{t,i}^{-} = x_{t,i}^{-} - b_t^{-}$
363
364 that is used in the state update equation (2) (instead of the biased model forecast $x_{t,i}^{-}$). From
365 ensemble average innovations (computed as $y_t - H_t\, \xi_t^{-} \equiv E\{y_{t,i} - H_t\, \xi_{t,i}^{-}\}$, where $E\{\cdot\}$ is the
366 ensemble mean operator), we can then update the bias via
367
368 (6) $b_t^{+} = b_t^{-} - \gamma\, K_{x,t}\, (y_t - H_t\, \xi_t^{-})$

A key assumption of this algorithm is that the bias error covariance $P_{b,t}$ is a small fraction of the state error covariance, that is $P_{b,t} = \gamma P_{x,t}$, which implies that the gain for the bias ($K_{b,t}$) can be computed as a fraction of the gain for the state, which has already been computed. Here, we use $\gamma = 0.2$ for Catchment and $\gamma = 0.05$ for Noah. The difference in γ is motivated by the different surface layer thicknesses in the two models, but like the perturbations parameters that govern the model error, these bias parameters have not been optimized and are justified primarily by the success of the assimilation.

As formulated above, the bias algorithm estimates a single bias parameter (per state and per location). Here, we implemented a variant that estimates a separate bias parameter for eight different times-of-day (0z, 3z, …, 21z). Because this requires the estimation of eight bias parameters per state and per location, we refer to this algorithm as "b8". Additional variants of the dynamic bias algorithm are discussed in the appendix.

Equation (6) implies that in practice, the bias estimates can be thought of as an exponential moving (time) average of the LST increments. Unlike the scaling approach, the dynamic bias estimation strategy adapts to slow changes in bias over time. A major disadvantage of this strategy is the implicit assumption that only the model is biased, which contradicts previous findings that retrievals from different sensors may be biased against each other (for example, Trigo and Viterbo 2003). It is therefore critical that any bias between retrievals from different sensors is small compared to the bias between retrievals and the model estimates.

For this study we implemented the a priori scaling method and the dynamic bias estimation schemes independently. The latter can be invoked with and without a priori scaling. If invoked without a priori scaling, the dynamic bias estimation corrects for static (long-term) biases as well as shorter-term "bias" that results from transient differences between model and observational estimates. If invoked after a priori scaling, the dynamic bias estimation mostly corrects for transient bias. It can also be considered a tool for remembering assimilation increments that would otherwise be forgotten within a single model time step (because of the small heat capacity associated with LST). This is particularly important for CLSM as section 6 will show.

5. Experiment design and skill metrics

The experiment domain consists of the catchments (or grid cells) that contain the CEOP station locations (Figure 1). The land models were spun up by cycling 10 times through the 4-year period from 1 January 2001 to 1 January 2005. The models were then integrated in ensemble mode (12 members; using the perturbations settings of Table 2) for the same 4-year period. These open loop integrations also recorded the LST innovations (without applying any increments) for the computation of the model and retrieval statistics that are required for the a priori scaling approach. The (ensemble) assimilation integrations covered the same 4-year period and were validated against the 27 months of CEOP observations from 1 October 2002 through 1 January 2005.

For each land model, we conducted one open loop (no assimilation) ensemble integration and four different experiments in which ISCCP retrievals were assimilated assuming an observation error standard deviation of 2 K. Two of the four assimilation integrations (per model) were performed with the (unscaled) ISCCP retrievals ("s0" for "no scaling"), the other two utilized ISCCP retrievals that were scaled to each model's LST climatology prior to assimilation ("s1"; section 4c). In each set of two assimilation integrations, one was done without bias correction ("b0"), and the other used the dynamic bias algorithms ("b8"; section 4c). For each model, we thus compare four assimilation integrations: "s0b0", "s0b8", "s1b0", and "s1b8".

All integrations were analyzed by computing RMSE values (from raw time series) and anomaly correlation coefficients (R; from anomaly time series) for LST, LH, SH, and GH with respect to the available in situ CEOP observations. These performance metrics were first computed

separately for each station from 3-hourly time series and then averaged over the available stations. Anomaly time series were computed by subtracting the monthly mean seasonal and diurnal cycle (climatology) from the raw data. The climatologies and metrics were computed only from data at times and locations where ISCCP retrievals were available so that only clear-sky conditions are compared to the extent possible. Mean values for a given calendar month and time-of-day are computed only if a minimum of 20 data points are available; mean monthly values are computed only if mean values are available for all eight times-of-day for the month in question; and performance metrics are based on at least 100 data points.

We analyze two performance metrics, RMSE and anomaly R, to highlight the advantages and disadvantages of the various assimilation approaches. Given the typically strong seasonal and diurnal cycles of LST and land surface fluxes, metrics based on raw data are dominated by errors in the climatology. Metrics based on anomalies, by contrast, primarily capture day-to-day variations. RMSE values measure how closely the data agree in their original units and are affected by a mean bias or a mean difference in the amplitude of variations. R values, on the other hand, are not affected by such biases and only capture the phasing between the estimates and the validating observations. The choice of metric depends on the application at hand (Entekhabi et al. 2010). RMSE values are most relevant if absolute errors matter most. In other cases, anomaly R values may be of most relevance, for example in model-based applications (such as Numerical Weather Prediction) that could correct for known biases in the mean and variance.

6. Results

6.a. Aggregate performance

The station-average RMSE and R metrics, evaluated against the CEOP in situ observations as discussed in the previous section, measure the aggregate performance of the satellite, model, and assimilation estimates. In this section, we highlight the performance of select assimilation integrations in terms of select metrics (see appendix for a complete table of metrics and algorithms). Before turning to the assimilation integrations, however, we first assess the skill of the satellite retrievals and of the model integrations without assimilation (open loop). The top panel of Figure 3 illustrates the RMSE values computed from the raw LST estimates, which are 4.9 K for CLSM (yellow bar), 5.6 K for Noah (light blue bar), and 7.6 K for ISCCP (black bar). The bottom panel shows corresponding RMSE values for model estimates of LH, SH, and GH, which range from 50 to 67 W m^{-2} and are within a factor of two of typical measurement errors (around 30 W m^{-2}) for surface turbulent fluxes (Finkelstein and Sims 2001, Hollinger and Richardson 2005). The first important results of Figure 3 are therefore that (i) the CLSM and Noah open loop integrations show similar skill when compared to CEOP in situ observations and that (ii) the model estimates of LST are significantly better than ISCCP retrievals.

The situation is similar for the anomaly R metric, shown in Figure 4 for LST only. The models show reasonable skill in terms of reproducing standardized anomalies, with anomaly R values of about 0.6. Again, ISCCP retrievals are significantly less skillful than model estimates (anomaly R value of 0.52). An analysis of sampling error reveals that 95 % confidence intervals for all R values discussed here are less than ±0.01. For RMSE values, 95 % confidence intervals are less than ±0.1 K for LST and less than ±1 W m^{-2} for surface fluxes. In the following, RMSE and

anomaly R values are rounded accordingly (skill difference values rounded after computing the difference).

Obviously, the superior skill of the model LST estimates relative to the skill of the ISCCP retrievals limits the improvements that can be expected from assimilating the ISCCP data. Nevertheless, as will be shown, modest yet statistically significant improvements can be achieved through the assimilation of ISCCP LST retrievals. We expect that the use of a priori scaling produces better anomaly estimates, whereas omitting a priori scaling should yield better improvements in terms of absolute numbers (that is, raw data) due to likely biases in model climatology. Thus, for RMSE computed from raw data (Figure 3), we focus on assimilation without a priori scaling (s0b0, s0b8). The CLSM assimilation integration without dynamic bias correction (s0b0; orange bars) is characterized by only minor improvements in LST (0.2 K for RMSE) and virtually no changes in the land surface flux estimates (when compared to the open loop skill). Recall that CLSM's prognostic "surface temperature" represents the canopy and top 5 cm of soil and is associated with a very small heat capacity. Consequently, without dynamic bias estimation and correction, increments from the assimilation of ISCCP retrievals dissipate quickly and have little impact on the model state in CLSM. Increments applied to the Noah model's top layer (10 cm) soil temperature, on the other hand, have a somewhat more noticeable effect on the model state. For assimilation without a priori scaling and without dynamic bias correction (s0b0), the RMSE for Noah LST estimates is improved by 0.5 K (medium blue bar in Figure 3). As for CLSM, the Noah assimilation estimates for the latent and sensible heat fluxes are comparable to the open loop estimates. However, the RMSE value for Noah s0b0 estimates of GH increases by 11 W m^{-2} (more on this later).

493 Next, we analyze the skill of the assimilation integrations with dynamic bias estimation and
494 correction (s0b8), also illustrated in Figure 3. For CLSM, adding dynamic bias estimation (red
495 bar) enhances the LST improvements (over the open loop) to 0.7 K in terms of RMSE. Using a
496 bias-corrected model forecast at every model time step (equation (5)) enhances the impact of
497 LST increments in the CLSM assimilation integrations and thereby yields enhanced
498 improvements from the assimilation of the ISCCP retrievals (relative to not using the dynamic
499 bias algorithm). For Noah, using the dynamic bias algorithm (s0b8; dark blue bar in Figure 3)
500 yields only slightly better LST than the s0b0 assimilation integration without dynamic bias
501 correction (RMSE now reduced by 0.6 K over the open loop). At the same time, however, the
502 deterioration of the GH estimates is exacerbated in s0b8. The RMSE value for assimilation
503 estimates of GH increases by 28 W m^{-2} when compared with the open loop RMSE. LH and SH
504 estimates from s0b8 are again comparable to open loop estimates.
505
506 For the analysis of assimilation integrations *with* a priori scaling (s1b0, s1b8) we focus on the
507 anomaly R metric, shown in Figure 4. Qualitatively, the results for these integrations are similar
508 to those obtained without a priori scaling. In CLSM, a priori scaling alone (s1b0) yields only
509 small improvements in LST (anomaly R increases by 0.02). Most of the impact is realized
510 through dynamic bias estimation (anomaly R for LST increases by 0.05 over the open loop). For
511 Noah, on the other hand, the anomaly R for LST already increases by 0.04 (over the open loop)
512 when only a priori scaling is applied. The estimates get only slightly better when dynamic bias
513 correction is added (increase of 0.05 in anomaly R for LST). There is also a deterioration in the
514 Noah GH estimates (relative to the open loop) when a priori scaling is used (with or without bias
515 correction), but generally the loss of skill in GH is mitigated through a priori scaling (not shown,
516 see appendix). Noah assimilation estimates of LH and SH have marginally better anomaly R

values than open loop estimates (not shown), but these improvements are so small that we do not consider them to be relevant.

In summary, using dynamic bias estimation for CLSM (s0b8, s1b8) provides the best assimilation estimates and enables modest LST improvements (over the open loop) of up to 0.7 K in terms of RMSE and up to 0.05 in terms of anomaly R. Flux estimates from these CLSM assimilation integrations are essentially identical to open loop estimates – assimilation of LST does not lead to improved flux estimation. For Noah, assimilation without a priori scaling and without dynamic bias estimation (s0b0) already yields most of the benefit of assimilating the satellite retrievals. Using a priori scaling and/or dynamic bias estimation yields only small additional improvements. For Noah, LST improvements (over the open loop) are similar to those for CLSM: up to 0.6 K for RMSE and up to 0.05 in terms of anomaly R. Noah assimilation estimates of SH and LH are similar to the open loop estimates, but the assimilation estimates of GH are considerably worse than the open loop estimates (up to 28 W m^{-2} increase in RMSE), with lesser degradation seen when a priori scaling is used. Without a priori scaling, the worsening of the GH estimates in Noah may well outweigh the benefits of the LST improvements.

As shown in Figure 1 and Table 1, most CEOP stations either have LST and GH measurements or have LH and SH measurements. Only four stations measure LST and all three fluxes. The results in this section must be interpreted with this caveat in mind. Conceivably, if a large number of stations were able to provide LST, LH, and SH measurements together, and if our analyses were limited to that set of stations, we might indeed be able to show that improved LST estimates from assimilation correspond to improved estimates of the turbulent fluxes. Given the

limitations of the available in situ measurements, we cannot know for sure. Note, however, that 15 out of the 19 stations that have LST measurements also have GH measurements, suggesting that for Noah the LST improvements from assimilation are connected to the worsening of GH estimates.

As discussed in section 5, the RMSE and anomaly R values shown here are based on 3-h average data, including nighttime and wintertime when fluxes are small and noise may overwhelm the signal. In a separate analysis (not shown) we also computed the performance metrics from daily average data. While RMSE values are generally lower when based on daily averages (as can be expected from increased averaging) and R values are also somewhat different in overall magnitude, the relative performance is similar regardless of whether metrics are computed from daily or 3-h average data, and the main conclusions of this section thus remain unchanged. In a second separate analysis (not shown), we assessed the performance based strictly on summertime data and again find that our conclusions remain the same.

6.b. Seasonal and diurnal cycles

Estimates of the mean seasonal and diurnal cycles provide additional insights into the modeling and assimilation of LST. Figure 5 (left panels) shows the mean seasonal cycle of LST at two locations, Bondville in the US Midwest and BJ-SAWS3 in Tibet. At both locations, the seasonal cycle estimates of the CLSM and Noah open loop integrations agree fairly closely with each other (to within 2 K), primarily because both models are driven with the same surface meteorological forcing data. At Bondville, the open loop estimates of the seasonal cycle also agree closely with the in situ CEOP observations. At the Tibetan station, however, the open loop estimates are biased low by about 5 K (relative to the in situ observations). In contrast, ISCCP

565 estimates of the seasonal cycle are biased high by about 3 K at BJ-SAWS3 and differ by up to 5
566 K in the first half of the year at Bondville.
567
568 By construction, the seasonal and diurnal cycle estimates of assimilation integrations with a
569 priori scaling closely match those of the open loop integration (not shown). Figure 5 also shows
570 the seasonal climatology of assimilation integrations without a priori scaling and with dynamic
571 bias estimation and correction (s0b8). As expected, the s0b8 assimilation integrations draw more
572 closely to the ISCCP retrievals (when compared to the open loop). At BJ-SAWS3, this
573 fortuitously brings the assimilation estimates of the seasonal cycle into better agreement with the
574 in situ observations than either the ISCCP or the open loop estimates.
575
576 Figure 5 (right panels) also illustrates the mean August diurnal cycle estimates at Bondville and
577 BJ-SAWS3. At Bondville, the Noah open loop estimates have a slightly higher diurnal
578 amplitude than the CLSM estimates (because LST for Noah exceeds that of CLSM by up to 2 K
579 during the day and by less than 1K during the night). At BJ-SAWS3, the open loop integrations
580 agree closely with each other. Similar to the seasonal cycle estimates, the open loop estimates at
581 Bondville are in reasonable agreement with the in situ observations but are biased low (by about
582 5 K) at BJ-SAWS3. ISCCP retrievals, on the other hand, exhibit a weaker diurnal amplitude at
583 Bondville than model or in situ observations. At BJ-SAWS3, ISCCP retrievals are biased high
584 (compared to the in situ observations) in the morning and mid-day but biased low in the evening
585 and at nighttime.
586
587 Again, the LST diurnal cycle estimates of the CLSM s0b8 assimilation integrations draw towards
588 the ISCCP retrievals by construction (Figure 5, right panels). This brings them closer to the in

589 situ observations at BJ-SAWS3 but makes CLSM nighttime estimates at Bondville worse when
590 compared to the open loop. By contrast, LST from the Noah s0b8 assimilation integration is
591 similar to the open loop integration during daytime. Only during the evening hours (at Bondville
592 and BJ-SAWS3) are the Noah s0b8 LST estimates noticeably closer to the ISCCP retrievals.
593 The delayed impact of the LST assimilation is probably a result of the phase lag between the
594 diagnostic LST (observations space) and the top layer soil temperature (state space) in Noah.
595
596 **6.c. Filter diagnostics**
597 Internal filter diagnostics offer further clues about the performance of the assimilation
598 algorithms. For a filter that operates according to its underlying assumptions (that various
599 linearizations hold, that model and observation errors are unbiased, uncorrelated and normally
600 distributed), the time average of the (ensemble mean) innovations sequence ($y_t - H_t \bar{x_t}$) equals
601 zero. Moreover, the standard deviation of the "normalized" innovations ($y_t - H_t \bar{x_t}$)·($H_t P_{x,t} H_t^T$
602 + R_t)$^{-0.5}$ equals one (Reichle et al. 2002a). The latter diagnostic compares the actual spread in the
603 innovations to what the filter expects. A simple interpretation is that the assumed error bars of a
604 model forecast and its corresponding observation must have an appropriate overlap.
605
606 Figure 6 displays the distribution of these two internal filter diagnostics across the CEOP stations
607 listed in Table 1. The top panel indicates that without a priori scaling of the ISCCP observations
608 and without dynamic bias estimation (s0b0), biases of several Kelvin typically persist in the
609 model forecast and are reflected in the mean of the innovations. (The innovations statistics of
610 the open loop integrations are essentially the same as those of the s0b0 integrations and are not
611 shown in the figure.) A very modest reduction of the bias can be achieved with the "b8"
612 dynamic bias estimation algorithm. If in addition to dynamic bias estimation the observations

613 are also scaled prior to data assimilation (s1b8), the innovations mean at all stations is, by
614 construction, much closer to zero. These results hold for assimilation into both land models,
615 CLSM and Noah.
616
617 The standard deviation of the normalized innovations, shown in the bottom panel of Figure 6,
618 exceeds the target value of one at almost all stations and for almost all algorithms. This indicates
619 that the model and/or the observation error standard deviations were underestimated. Like for
620 the innovations mean, a priori scaling brings the standard deviation of the normalized
621 innovations much closer to its expected value of one. This implies that a large part of the
622 mismatch between the actual spread in the innovations and the expected spread is simply due to
623 bias. Finally, the fact that the innovations diagnostics are comparable for the CLSM and Noah
624 assimilation integrations indicates that the assimilation performance (relative to its unknown
625 optimum) is comparable for the two models, which lends further support to the broad
626 conclusions reached in this paper. In summary, a priori scaling in combination with dynamic
627 bias estimation exhibits the best performance in terms of the innovations diagnostics,
628 independent of the land model used.
629
630 **6.d. Impact of bias in data assimilation**
631 Estimates from a properly designed assimilation system should be no worse than open loop
632 estimates. The example in Figure 7 further illustrates the potentially serious detrimental impact
633 of not addressing bias properly in data assimilation. The figure shows LST and land surface flux
634 time series from select Noah integrations for a few days in August of 2003 at the MGS station in
635 Mongolia. At this location, LST and GH estimates from the open loop integrations agree fairly
636 well with CEOP in situ observations. The daytime peak LST estimates from ISCCP, however,

637 are warmer by as much as 30 K. This extreme bias may be due to one or more of the reasons
638 discussed in section 2.
639
640 At any rate, when the unscaled ISCCP retrievals are assimilated without a priori scaling (s0b0,
641 s0b8), LST assimilation estimates are drawn toward the extreme temperatures in the ISCCP
642 retrievals (Figure 7). However, since Noah is not designed to accommodate such extreme
643 temperatures, and because the surface meteorological forcing remains unchanged in the system,
644 the Noah assimilation integrations without a priori scaling respond with unrealistic and excessive
645 estimates of sensible and ground heat flux, most notably on 13 August 2003. Because the impact
646 on LST of assimilating unscaled ISCCP retrievals without bias correction (s0b0) is more limited,
647 the corresponding flux estimates are less pathological than in the s0b8 case with dynamic bias
648 correction. For reference, Figure 7 also shows an assimilation integration with a priori scaling
649 and dynamic bias correction (s1b8), which does not produce such unrealistic flux estimates.
650
651 The situation is similar for LST from CLSM assimilation integrations for the same location and
652 time period (not shown), but for CLSM we obtain unrealistic estimates of the latent heat flux for
653 select assimilation integrations without a priori scaling. To summarize, Figure 7 illustrates the
654 pitfalls of assimilating LST retrievals that are severely biased against model LST. While the
655 assimilation can be designed to produce LST estimates that are closer to the satellite retrievals,
656 there may be unintended and undesirable side effects in terms of the land surface fluxes and the
657 surface energy balance.
658

7. Summary and conclusions

An ensemble-based data assimilation method, the EnKF, was used with and without a priori scaling of observations and/or dynamic bias estimation methods to assimilate satellite retrievals of LST into two different land surface models at 48 CEOP sites scattered around the globe. The two land models, CLSM and Noah, follow distinct modeling approaches for land surface temperature. CLSM has model prognostic "surface temperature" variables, whereas Noah diagnoses the surface temperature from the surface energy balance. The LST, sensible, latent, and ground heat flux estimates from the data assimilation integrations were validated against 27 months of CEOP in situ observations.

The main conclusions from the experiments are as follows.

(1) There are strong biases between LST estimates from in situ observations, land modeling, and satellite retrievals that vary with season and time-of-day. Biases of a few Kelvin are typical, with larger values exceeding 10 K.

(2) The skill of LST estimates from the CLSM and Noah land model integrations is superior to that of the ISCCP satellite retrievals.

(3) Assimilation of ISCCP LST retrievals into the land surface models can improve LST estimates by up to 0.7 K for RMSE and by up to 0.05 for anomaly R, while not making surface turbulent fluxes worse.

(4) Gross errors in surface flux estimates can result if biases are not taken into account properly, with a combination of a priori scaling and dynamic bias estimation methods yielding the best overall results.

(5) Assimilation diagnostics for integrations without a priori scaling strongly reflect the underlying biases, indicating that without a priori scaling the assimilation system is far from operating in accordance with its underlying assumptions.

(6) Provided the assimilation system is properly configured for each land model, the benefits from the assimilation of LST retrievals are comparable for both land models.

There are many reasons why the improvements from the assimilation of satellite LST, while statistically significant, turn out to be modest. First and foremost, the skill of the satellite data is modest and much lower than that of the model to begin with. The information gained by assimilating the satellite data into the model is therefore naturally limited. In the present study, the parameters of the assimilation system, including the perturbations (or model error) parameters and the parameters of the bias algorithm (α, γ) were not optimized. Additional calibration may further improve the results and may also reveal differences in what can be achieved with a given land model structure. Finally, even if the assimilation estimates were perfect, the performance metrics would not show it because of errors in the in situ data and because of the mismatch of the spatial and temporal characteristics of the satellite, model, and in situ data sets.

By design, the present study was limited to LST (state) estimation in an uncoupled land modeling system. Ideally, the land model parameters would be calibrated to minimize LST biases prior to data assimilation, or perhaps even dynamically within the data assimilation system. Such a calibration could rely on sophisticated parameter estimation methods (Vrugt et al. 2003). Perhaps more importantly, though, is that the surface meteorological forcings in the

present experiments were fixed. In other words, LST increments did not feed back onto the atmospheric state. Future experiments will explore LST assimilation into a coupled land-atmosphere model with the methods proposed here. The relative performance of the algorithms may very well change in the coupled environment.

Another path for future research is to investigate further the role of specific aspects of LST modeling. In the present paper, Noah was integrated in its default configuration, including a 10 cm thick surface layer, which implied a small phase shift between the Noah diagnostic LST (used in the computations of the innovations) and the Noah top soil temperature (to which the increments were applied). Use of a thinner soil layer may alleviate the problems related to the phase shift in Noah between LST (observation space) and the top layer soil temperature (state space).

Even if the assimilation manages to improve LST only modestly and fluxes not at all, the impact may be significant because minor improvements in LST may already increase the number of atmospheric retrievals that can be assimilated in coupled systems, thereby possibly providing substantial indirect benefits. Obviously, the present study only scratches the surface of a very complex problem that has been a challenge for many years. Nevertheless, given the relative abundance of LST observations from satellites and the importance of accurate LST estimates, in particular in the context of atmospheric data assimilation, the results of the present study offer an encouraging step forward in land data assimilation.

Appendix

In addition to the dynamic bias algorithm "b8" discussed in section 4c we also tested additional variants of the algorithm used by Bosilovich et al. (2007). Specifically, we tested the simplest possible variant that estimates a single, "mean" bias estimation parameter per state and per location, here referred to as "b1". Bosilovich et al. (2007) also introduced a bias parameterization with sine and cosine functions that accounts for variations in the diurnal cycle of the bias. Here we implemented two variants: a "diurnal" bias parameterization (constant term plus sine and cosine waves with a period of one day) and a "semi-diurnal" bias parameterization ("diurnal" terms plus sine and cosine waves with a period of one half day). The "diurnal" and "semi-diurnal" algorithms estimate three and five bias parameters, respectively, per state and location, and are referred to as "b3" and "b5".

For the "b8" integrations discussed in the main text we always applied both the state increments (equation (2)) and the bias increments (equation (6)). More generally, though, any assimilation integration that uses dynamic bias estimation can also be done without applying the state increments (as, in fact, implemented by Bosilovich et al. (2007)). We also tested these variants. Per model, we therefore tested a total of 18 different assimilation integrations, listed in Table 3.

A close examination of Table 3 reveals that generally, the assimilation integrations show more skill when more bias parameters are used. Reductions in RMSE values for LST are greater by up to 0.4 K for CLSM and by up to 0.2 K for Noah when comparing "b8" and "b1"integrations. Corresponding differences in anomaly R values for LST are up to 0.03, respectively, for both models. As can be expected, these differences across bias estimation algorithms of varying

complexity tend to be greater without a priori scaling. Also in line with expectations, taking into account the diurnal cycle of the bias (as in "b3", "b5", and "b8", as opposed to "b1") has the biggest impact.

Finally, applying the state increments in addition to the bias increments contributes only a small amount of skill, typically less than 0.1 K in terms of RMSE reduction (or 0.01 in terms of R increase) for LST estimates. In the case of Noah, applying the state increments contributes commensurately to the worsening of the GH estimates.

To summarize, as long as the dynamic bias algorithm takes the diurnal cycle into account, the differences that result from the exact number of bias parameters used or that result from not applying the state increments are much smaller than the assimilation improvements over the open loop. In other words, the lessons learned in the main text about assimilation of LST retrievals in general and about using a priori scaling and/or dynamic bias correction are insensitive to the details of the dynamic bias estimation algorithm, provided the algorithm considers the diurnal cycle of the bias.

Acknowledgments

Thanks to Michael Bosilovich, John Eylander, and Christa Peters-Lidard for helpful comments. Partial support for this study was provided through NASA (NNX08AH36G) and the Air Force Weather Agency (F2BBBJ7080G001). We are grateful for access to the many data sets that supported this work, in particular the CEOP and ISCCP projects. Surface meteorological data were provided by H. Kato, J. Meng, and M. Rodell of the GLDAS project, with data contributions from the NOAA Earth System Research Laboratory. Computing was supported by the NASA High End Computing Program.

References

Aires, F., C. Prigent, and W. B. Rossow, 2004: Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions, *Journal of Geophysical Research, 109*, D06214, doi:10.1029/2003JD003527.

Andreadis, K. and D. Lettenmaier, 2005: Assimilating remotely sensed snow observations into a macroscale hydrology model, *Advances in Water Resources*, 29, 872–886.

Balsamo, G., J.-F. Mahfouf, and S. Belair, 2007: A land data assimilation system for soil moisture and temperature: An information content study, *Journal of Hydrometeorology*, 8, 1225-1242, doi:10.1175/2007JHM819.1.

Bloom, S., A. da Silva, D. Dee, M. Bosilovich, J.-D. Chern, S. Pawson, S. Schubert, M. Sienkiewicz, I. Stajner, W.-W. Tan, and M.-L. Wu, 2005: Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 4, Technical Report Series on Global Modeling and Data Assimilation, Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, 187 pp., Document number 104606, 26.

Boni, G., D. Entekhabi, and F. Castelli, 2001: Land data assimilation with satellite measurements for the estimation of surface energy balance components and surface control on evaporation, *Water Resources Research*, 37, 1713-1722.

Bosilovich, M., J. Radakovich, A. da Silva, R. Todling, and F. Verter, 2007: Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system, *J. Meteo. Soc. Japan*, 85A, 205–228.

797 Caparrini, F., F. Castelli, and D. Entekhabi, 2004: Variational estimation of soil and vegetation
798 turbulent transfer and heat flux parameters from sequences of multisensor imagery, *Water*
799 *Resources Research*, 40, W12515, doi:10.1029/2004WR003358.
800 Castelli, F., D. Entekhabi, and E. Caporali, 1999) Estimation of surface heat flux and an index of
801 soil moisture using adjoint-state surface energy balance, *Water Resources Research*, 35,
802 3115-3125.
803 Crow, W. T., and R. H. Reichle, 2008: Comparison of adaptive filtering techniques for land
804 surface data assimilation, *Water Resources Research, 44*, W08423,
805 doi:10.1029/2008WR006883.
806 Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness
807 temperature imagery into a land surface model using ensemble Kalman filtering: A case
808 study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, 26, 137-149.
809 De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest,
810 2007: Correcting for Forecast Bias in Soil Moisture Assimilation with the Ensemble Kalman
811 Filter, *Water Resources Research, 43*, W09410, doi:10.1029/2006WR005449.
812 Dee, D. P., 2005: Bias and data assimilation, *Q. J. R. Meterol. Soc.*, 131, 3323-3343,
813 doi:10.1256/qj.05.137.
814 Drusch, M., 2007: Initializing numerical weather prediction models with satellite derived surface
815 soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and
816 the TMI soil moisture dataset, *J. Geophys. Res.*, 112 (D3), Art. No. D03102,
817 doi:10.1029/2006JD007478.

818 Dunne, S., and D. Entekhabi, 2006: Land surface state and flux estimation using the ensemble
819 Kalman smoother during the Southern Great Plains 1997 field experiment, *Water Resour.*
820 *Res.*, 42, W01407, doi:10.1029/2005WR004334.
821 Ek, M., K. Mitchell, L. Yin, P. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley,
822 2003: Implementation of Noah land-surface model advances in the NCEP operational
823 mesoscale Eta model, *J Geophys Res.*, 108, doi:10.1029/2002JD003296.
824 Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance Metrics for Soil
825 Moisture Retrievals and Application Requirements, *J. Hydrometeor.*, submittted.
826 Finkelstein, P. L., and P. F. Sims, 2001: Sampling error in eddy correlation flux measurements,
827 *J. Geophys. Res.,* 106(D4), 3503–3509.
828 Garand, L., 2003: Toward an Integrated Land-Ocean Surface Skin Temperature Analysis from
829 the Variational Assimilation of Infrared Radiances, *J. Appl. Meteorol., 42,* 570-583.
830 Hollinger, D. Y., and A. D. Richardson, 2005: Uncertainty in eddy covariance measurements and
831 its application to physiological models, *Tree Physiology*, 25, 873-885.
832 Jin, M., 2004: Analysis of land skin temperature using AVHRR observations, *Bulletin of the*
833 *American Meteorological Society*, 85, 587-+, doi:10.1175/BAMS-85-4-587.
834 Jin, M. L., R. E. Dickinson, and A. M. Vogelmann, 1997: A comparison of CCM2-BATS skin
835 temperature and surface-air temperature with satellite and surface observations, *Journal of*
836 *Climate, 10*, 1505-1524.
837 Keppenne C. L., 2000: Data assimilation into a primitive-equation model with a parallel
838 ensemble Kalman filter, *Mon. Wea. Rev., 128*, 1971–1981

839 Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based
840 approach to modeling land surface processes in a general circulation model, 1: Model
841 structure. *Journal of Geophysical Research*, 105, 24809-24822.
842 Kumar, P., and A. L. Kaleita, 2003: Assimilation of near-surface temperature using extended
843 Kalman filter, *Advances in Water Resources*, 26, 79-93.
844 Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B.
845 Eylander, and P. R. Houser, 2008: A Land Surface Data Assimilation Framework using the
846 Land Information System: Description and Applications, *Advances in Water Resources*, *31*,
847 1419-1432, doi:10.1016/j.advwatres.2008.01.013.
848 Lakshmi, V., 2000: A simple surface temperature assimilation scheme for use in land surface
849 models, *Water Resources Research*, 36, 3687-3700.
850 Mahfouf, J.-F., K. Bergaoui, C. Draper, F. Bouyssel, F. Taillefer, and L. Taseva, 2009: A
851 comparison of two off-line soil analysis schemes for assimilation of screen level
852 observations, *J. Geophys. Res.*, 114, D08105, doi:10.1029/2008JD011077.
853 Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne, 2002: Land data assimilation and
854 estimation of soil moisture using measurements from the Southern Great Plains 1997 field
855 experiment, *Water Resour. Res.*, 38, 1299, doi:10.1029/2001WR001114.
856 McNider, R. T., A. J. Song, D. M. Casey, P. J. Wetzel, W. L. Crosson, and R. M. Rabin, 1994:
857 Toward a dynamic-thermodynamic assimilation of satellite surface-temperature in numerical
858 atmospheric models, *Monthly Weather Review*, 122, 2784-2787.
859 Meng, C. L., Z.-L. Li, X. Zhan, J. C. Shi, and C. Y. Liu, 2009: Land surface temperature data
860 assimilation and its impact on evapotranspiration estimates from the Common Land Model,
861 *Water Resour. Res.*, 45, W02421, doi:10.1029/2008WR006971.

862 Minnis, P., and M. M. Khaiyer, 2000: Anisotropy of Land Surface Skin Temperature Derived
863 from Satellite Data., *J. Appl. Meteor.*, 39, 1117–1129.
864 Pan, M., and E. F. Wood, 2006: Data Assimilation for Estimating the Terrestrial Water Budget
865 Using a Constrained Ensemble Kalman Filter, *J. Hydrometeorol.*, 7, 534-547.
866 Pinheiro, A. C. T., J. L. Privette, R. Mahoney, and C. J. Tucker, 2004: Directional effects in a
867 daily AVHRR land surface temperature dataset over Africa, *IEEE Transactions on*
868 *Geosciences Remote Sensing, 42*, 1941-1954, doi:10.1109/TGRS.2004.831886.
869 Reichle, R. H., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in
870 background fields on soil moisture estimation, *J. Hydrometeorol.*, 4 (6), 1229-1242.
871 Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture,
872 *Geophys. Res. Lett.*, 31, doi:10.1029/2004GL020938.
873 Reichle, R. H., D. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the
874 Ensemble Kalman filter, *Mon. Weather Rev.* 130(1), 103-114.
875 Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended versus Ensemble
876 Kalman filtering for land data assimilation, *J. Hydrometeorol.*, 3(6), 728-740.
877 Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007:
878 Comparison and assimilation of global soil moisture retrievals from the Advanced
879 Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the
880 Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, 112, D09108,
881 doi:10.1029/2006JD008033.
882 Reichle, R. H., W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble Kalman filter for
883 soil moisture data assimilation, *Water Resources Research, 44*, W03423,
884 doi:10.1029/2007WR006357.

885 Reichle, R. H., M. G. Bosilovich, W. T. Crow, R. D. Koster, S. V. Kumar, S. P. P. Mahanama,
886 and B. F. Zaitchik, 2009: Recent Advances in Land Data Assimilation at the NASA Global
887 Modeling and Assimilation Office, in *Data Assimilation for Atmospheric, Oceanic and*
888 *Hydrologic Applications*, edited by Seon K. Park and Liang Xu, 407-428, Springer Verlag,
889 New York, doi:10.1007/978-3-540-71056-1.
890 Rodell, M., and P. R. Houser, 2004: Updating a land surface model with MODIS-derived snow
891 cover, *Journal of Hydrometeorology, 5*, 1064-1075.
892 Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B.
893 Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, and D. L. Toll, 2003: The
894 Global Land Data Assimilation System, *Bull. Amer. Meteorol. Soc.*, 85, 381-394,
895 doi:10.1175/BAMS-85-3-381.
896 Rossow, W. B., and R. A. Schiffer, 1991: ISCCP cloud data products, *Bulletin of the American*
897 *Meteorological Society, 72*, 2-20.
898 Rossow, W. B., and R. A. Schiffer, 1999: Advances in understanding clouds from ISCCP,
899 *Bulletin of the American Meteorological Society, 80*, 2261-2287.
900 Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet, 2003: Soil
901 moisture analysis combining screen-level parameters and microwave brightness temperature:
902 A test with field data, *Geophys. Res. Lett., 30*, 1498, doi:10.1029/2063GL017128.
903 Sini, F., G. Boni, F. Caparrini, and D. Entekhabi, 2008: Estimation of large-scale evaporation
904 fields based on assimilation of remotely sensed land temperature, *Water Resources Research,*
905 *44*, W06410, doi:10.1029/2006WR005574.
906 Slater, A., and M. Clark, 2006: Snow data assimilation via an ensemble Kalman filter, *J.*
907 *Hydrometeorol., 7*, 478–493.

908 Trigo, I. F., and P. Viterbo, 2003: Clear-sky window channel radiances: A comparison between
909 observations and the ECMWF model, *Journal of Applied Meteorology, 42,* 1463-1479.
910 Walker, J. P., G. R. Willgoose, and J. D. Kalma, 2001: One-dimensional soil moisture profile
911 retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms,
912 *Adv. Water Resour. 24*, 631-650.
913 Van den Hurk, B. J. J. M., L. Jia, C. Jacobs, M. Menenti, and Z. L. Li, 2002: Assimilation of
914 land surface temperature data from ATSR in an NWP environment - a case study,
915 *International Journal of Remote Sensing, 23*, 5193-5209.
916 Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, S. Sorooshian, 2003: Effective and
917 efficient algorithm for multiobjective optimization of hydrologic models, *Water Resources*
918 *Research, 39*, W1214, doi: 10.1029/2002WR001746.
919 Wan, Z. and Z.-L. Li,, 1997: A physics-based algorithm for retrieving land-surface emissivity
920 and temperature from EOS/MODIS data, *IEEE Trans. Geosci. Remote Sens.*, 35, 980-996.
921 Zaitchik, B.F., and M. Rodell, 2009: Forward-Looking Assimilation of MODIS-Derived Snow-
922 Covered Area into a Land Surface Model, *J. Hydrometeor., 10,* 130–148,
923 doi:10.1175/2008JHM1042.1.
924 Zhou, Y., D. McLaughlin, and D. Entekhabi, 2006: Assessing the Performance of the Ensemble
925 Kalman Filter for Land Surface Data Assimilation, *Mon. Weather Rev. 134*, 2128-2142.
926

927 **Table captions**

928

929 **Table 1.** CEOP stations with sufficient in situ LST, LH, SH, or GH observations for validation
930 (CSE=Continental-Scale Experiment).

931

932 **Table 2.** Parameters for perturbations to meteorological forcing inputs and model prognostic
933 variables.

934

935 **Table 3.** Skill of model and assimilation integrations versus CEOP in situ observations.
936 Integrations shown in bold face are those discussed in the main text. Anomaly R for LST from
937 ISCCP is 0.52, RMSE is 7.6 K.

938

939

Figure captions

Fig. 1. Location of CEOP stations. Stations suitable for validation are indicated with circles (SH, LH, 30 stations), plus signs (GH, 20 stations), and crosses (LST, 19 stations).

Fig. 2. LST is described in (left) CLSM with a prognostic variable ("TSURF") and in (right) Noah with a diagnostic variable ("TSKIN"). LST increments are applied to "TSURF" in CLSM and to "TSOIL1" in Noah (section 4.b.).

Fig. 3. RMSE versus CEOP in situ observations for (top) LST and (bottom) flux estimates from ISCCP retrievals (LST only), model integrations, and select assimilation integrations *without* a priori scaling.

Fig. 4. R versus CEOP in situ observations for LST anomalies from ISCCP retrievals, model integrations, and select assimilation integrations *with* a priori scaling.

Fig. 5. LST (left) annual seasonal and (right) August diurnal cycle at (top) Bondville and (bottom) BJ-SAWS3 for CEOP, ISCCP, model, and assimilation data.

Fig. 6. (Top) Mean of innovations [K] and (bottom) standard deviation of normalized innovations [dimensionless] for (C) Catchment and (N) Noah assimilation integrations. The box plots indicate the average, standard deviation, minimum and maximum of the respective innovations diagnostic across the stations listed in Table 1.

963

964 **Fig. 7.** (Top) LST, (upper middle) LH, (lower middle) SH, and (bottom) GH for Noah

965 integrations, ISCCP retrievals, and CEOP observations at the MGS station.

966

CEOP Identifier			Coordinates		Data Availability		
CSE	**Reference Site**	**Station**	**Lat**	**Lon**	**LST**	**LH+SH**	**GH**
BALTEX	Cabauw	Cabauw	51.97	4.93	yes	yes	yes
BALTEX	Lindenberg	Falkenberg	52.17	14.12	yes	yes	yes
BALTEX	Lindenberg	Forest	52.18	13.95	yes	yes	yes
CAMP	ChaoPhrayaRiver	Lampang	18.40	99.47	no	no	yes
CAMP	Himalayas	Pyramid	27.96	86.81	no	no	yes
CAMP	Mongolia	BTS	46.78	107.14	yes	no	yes
CAMP	Mongolia	DGS	46.13	106.37	yes	no	yes
CAMP	Mongolia	DRS	46.21	106.71	yes	no	yes
CAMP	Mongolia	MGS	45.74	106.26	yes	no	yes
CAMP	NorthEastThai	Nakhonrachasima	14.47	102.38	yes	no	yes
CAMP	Tibet	Amdo-Tower	32.24	91.62	no	no	yes
CAMP	Tibet	ANNI-AWS	31.25	92.17	yes	no	yes
CAMP	Tibet	BJ-SAWS1	31.37	91.90	yes	no	no
CAMP	Tibet	BJ-SAWS2	31.37	91.90	yes	no	no
CAMP	Tibet	BJ-SAWS3	31.37	91.90	yes	no	no
CAMP	Tibet	BJ-Tower	31.37	91.90	yes	no	yes
CAMP	Tibet	D105-AWS	33.06	91.94	yes	no	yes
CAMP	Tibet	D66-AWS	35.52	93.78	yes	no	no
CAMP	Tibet	Gaize	32.30	84.05	yes	no	yes
CAMP	Tibet	MS3478-AWS	31.93	91.71	yes	no	yes
CAMP	Tibet	MS3608-AWS	31.23	91.78	yes	no	no
CAMP	Tongyu	Cropland	44.42	122.87	no	yes	yes
CAMP	Tongyu	Grassland	44.42	122.87	no	yes	yes
GAPP	Bondville	Bondville	40.01	-88.29	yes	yes	yes
GAPP	SGP	E 1 Larned	38.20	-99.32	no	yes	no
GAPP	SGP	E 2 Hillsboro	38.31	-97.30	no	yes	no
GAPP	SGP	E 3 Le_Roy	38.20	-95.60	no	yes	no
GAPP	SGP	E 4 Plevna	37.95	-98.33	no	yes	no
GAPP	SGP	E 5 Halstead	38.11	-97.51	no	yes	no
GAPP	SGP	E 6 Towanda	37.84	-97.02	no	yes	no
GAPP	SGP	E 7 Elk_Falls	37.38	-96.18	no	yes	no
GAPP	SGP	E 8 Coldwater	37.33	-99.31	no	yes	no
GAPP	SGP	E 9 Ashton	37.13	-97.27	no	yes	no
GAPP	SGP	E 10 Tyro	37.07	-95.79	no	yes	no
GAPP	SGP	E 12 Pawhuska	36.84	-96.43	no	yes	no
GAPP	SGP	E 13 Lamont	36.60	-97.48	no	yes	no
GAPP	SGP	E 14 Lamont	36.61	-97.49	no	yes	no
GAPP	SGP	E 15 Ringwood	36.43	-98.28	no	yes	no
GAPP	SGP	E 16 Vici	36.06	-99.13	no	yes	no
GAPP	SGP	E 18 Morris	35.69	-95.86	no	yes	no
GAPP	SGP	E 19 El_Reno	35.55	-98.02	no	yes	no
GAPP	SGP	E 20 Meeker	35.56	-96.99	no	yes	no
GAPP	SGP	E 21 Okmulgee	35.62	-96.06	no	yes	no
GAPP	SGP	E 22 Cordell	35.35	-98.98	no	yes	no
GAPP	SGP	E 24 Cyril	34.88	-98.20	no	yes	no
GAPP	SGP	E 26 Cement	34.96	-98.08	no	yes	no
GAPP	SGP	E 27 Earlsboro	35.27	-96.74	no	yes	no
MDB	Tumbarumba	Tumbarumba	-35.65	148.15	no	yes	yes

Table 1. CEOP stations with sufficient in situ LST, LH, SH, or GH observations for validation (CSE=Continental-Scale Experiment).

Perturbation	Additive (A) or multiplicative (M)?	Standard deviation	AR(1) time series correlation scale	Cross-correlation with perturbations in		
				T2M	SW	LW
Air temperature (T2M)	A	1 K	1 day	n/a	0.4	0.4
Downward shortwave (SW)	M	0.3	1 day	0.4	n/a	-0.6
Downward longwave (LW)	A	20 W m^{-2}	1 day	0.4	-0.6	n/a
Soil temperature prognostic variables (Catchment: TSURF; Noah: TSOIL1)	A	0.2 K	12 h	0	0	0

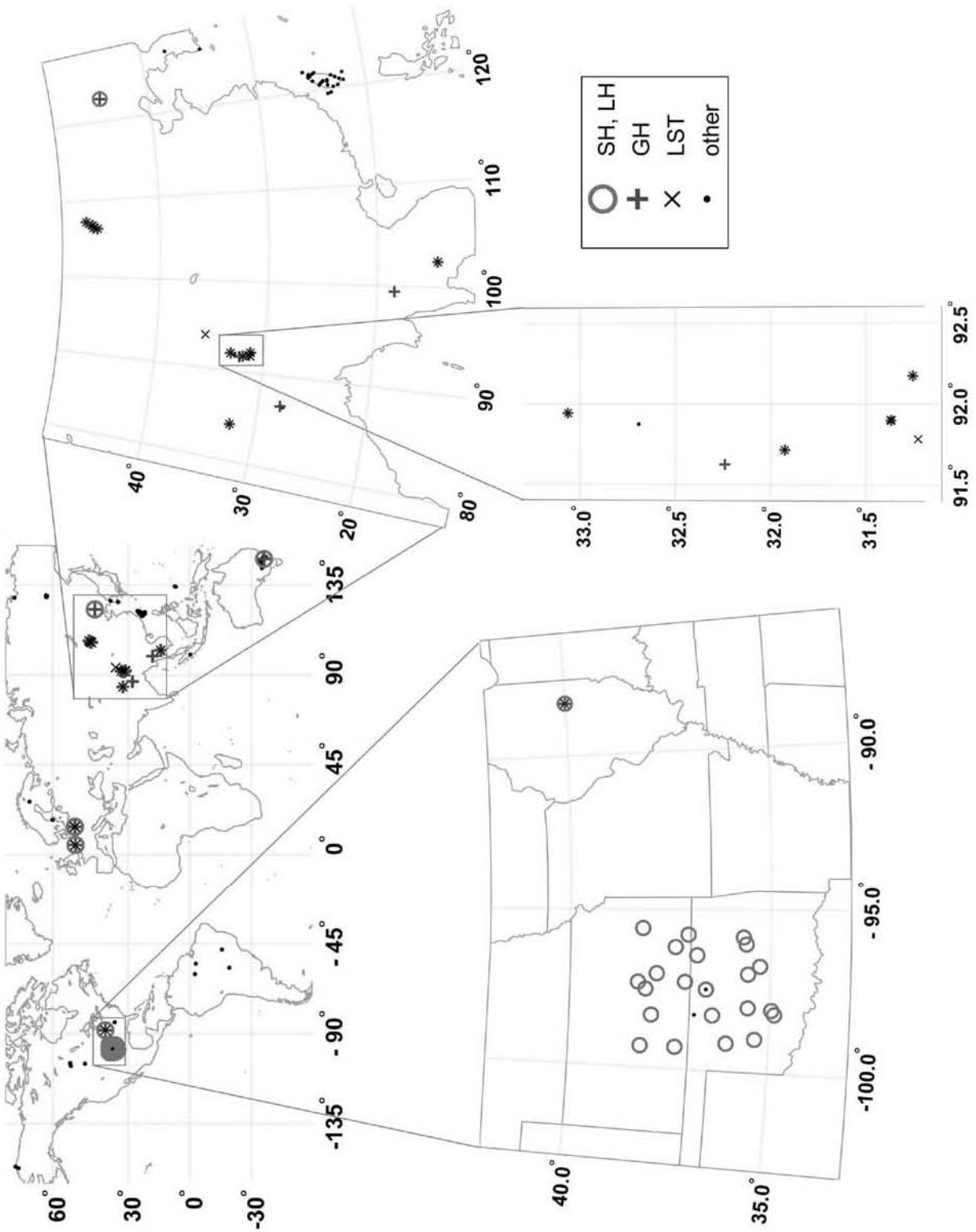
970

971 **Table 2.** Parameters for perturbations to meteorological forcing inputs and model prognostic

972 variables.

Assimilation	Apply State Increments	A Priori Scaling	Bias Algorithm	CLSM								Noah							
				Anomaly R				RMSE				Anomaly R				RMSE			
				LST	LH	SH	GH	LST	LH	SH	GH	LST	LH	SH	GH	LST	LH	SH	GH
				[-]	[-]			[K]	[W m^{-2}]			[-]	[-]			[K]	[W m^{-2}]		
no	**n/a**	**n/a**	**n/a**	**0.61**	**0.31**	**0.21**	**0.25**	**4.9**	**54**	**67**	**54**	**0.62**	**0.26**	**0.12**	**0.24**	**5.6**	**50**	**62**	**50**
yes	**yes**	**s0**	**b0**	**0.63**	**0.31**	**0.21**	**0.25**	**4.7**	**54**	**67**	**54**	**0.66**	**0.27**	**0.14**	**0.14**	**5.1**	**50**	**61**	**61**
yes	yes	s0	b1	0.64	0.31	0.21	0.26	4.6	54	67	53	0.64	0.28	0.14	0.08	5.2	50	62	81
yes	yes	s0	b3	0.66	0.31	0.21	0.26	4.4	54	68	54	0.65	0.28	0.14	0.08	5.1	50	62	84
yes	yes	s0	b5	0.66	0.31	0.21	0.25	4.4	54	68	54	0.65	0.27	0.14	0.08	5.1	50	62	84
yes	**yes**	**s0**	**b8**	**0.67**	**0.31**	**0.21**	**0.25**	**4.2**	**54**	**68**	**54**	**0.67**	**0.28**	**0.14**	**0.11**	**5.0**	**50**	**62**	**78**
yes	no	s0	b1	0.63	0.31	0.21	0.26	4.7	54	67	53	0.63	0.28	0.13	0.10	5.3	50	63	77
yes	no	s0	b3	0.65	0.31	0.21	0.26	4.4	54	68	54	0.64	0.28	0.13	0.09	5.2	50	63	78
yes	no	s0	b5	0.66	0.31	0.21	0.26	4.4	54	68	54	0.65	0.28	0.13	0.09	5.1	50	63	78
yes	no	s0	b8	0.66	0.31	0.21	0.25	4.3	54	68	54	0.66	0.28	0.13	0.14	5.0	50	62	73
yes	**yes**	**s1**	**b0**	**0.63**	**0.31**	**0.21**	**0.25**	**4.8**	**54**	**67**	**54**	**0.66**	**0.26**	**0.13**	**0.20**	**5.3**	**50**	**61**	**53**
yes	yes	s1	b1	0.65	0.31	0.21	0.25	4.6	54	67	54	0.66	0.26	0.13	0.15	5.2	50	62	58
yes	yes	s1	b3	0.67	0.31	0.21	0.25	4.5	54	67	54	0.66	0.26	0.13	0.15	5.2	50	62	58
yes	yes	s1	b5	0.67	0.31	0.21	0.25	4.5	54	67	54	0.66	0.26	0.13	0.15	5.2	50	62	58
yes	**yes**	**s1**	**b8**	**0.66**	**0.31**	**0.21**	**0.25**	**4.6**	**54**	**67**	**54**	**0.67**	**0.27**	**0.13**	**0.17**	**5.2**	**49**	**61**	**56**
yes	no	s1	b1	0.64	0.31	0.21	0.25	4.6	54	67	54	0.65	0.26	0.13	0.15	5.2	50	62	57
yes	no	s1	b3	0.66	0.31	0.21	0.25	4.6	54	67	54	0.65	0.27	0.13	0.15	5.2	50	62	57
yes	no	s1	b5	0.66	0.31	0.21	0.25	4.5	54	67	54	0.66	0.27	0.13	0.15	5.2	50	62	57
yes	no	s1	b8	0.65	0.31	0.21	0.25	4.6	54	67	54	0.66	0.27	0.12	0.17	5.2	50	62	55

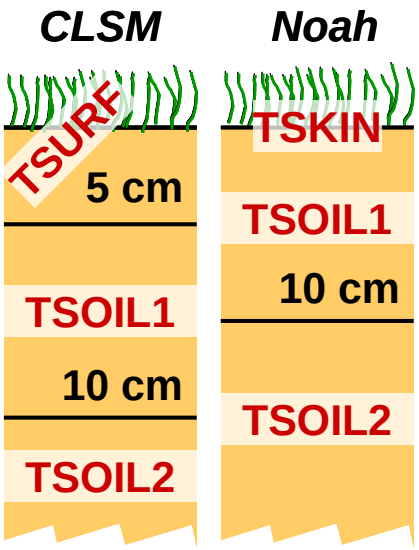
Table 3. Skill of model and assimilation integrations versus CEOP in situ observations. Integrations shown in bold face are those discussed in the main text. Anomaly R for LST from ISCCP is 0.52, RMSE is 7.6 K.



978

979 **Fig. 1.** Location of CEOP stations. Stations suitable for validation are indicated with circles

980 (SH, LH, 30 stations), plus signs (GH, 20 stations), and crosses (LST, 19 stations).



981

982 **Fig. 2.** LST is described in (left) CLSM with a prognostic variable ("TSURF") and in (right)

983 Noah with a diagnostic variable ("TSKIN"). LST increments are applied to "TSURF" in CLSM

984 and to "TSOIL1" in Noah (section 4.b.).

985

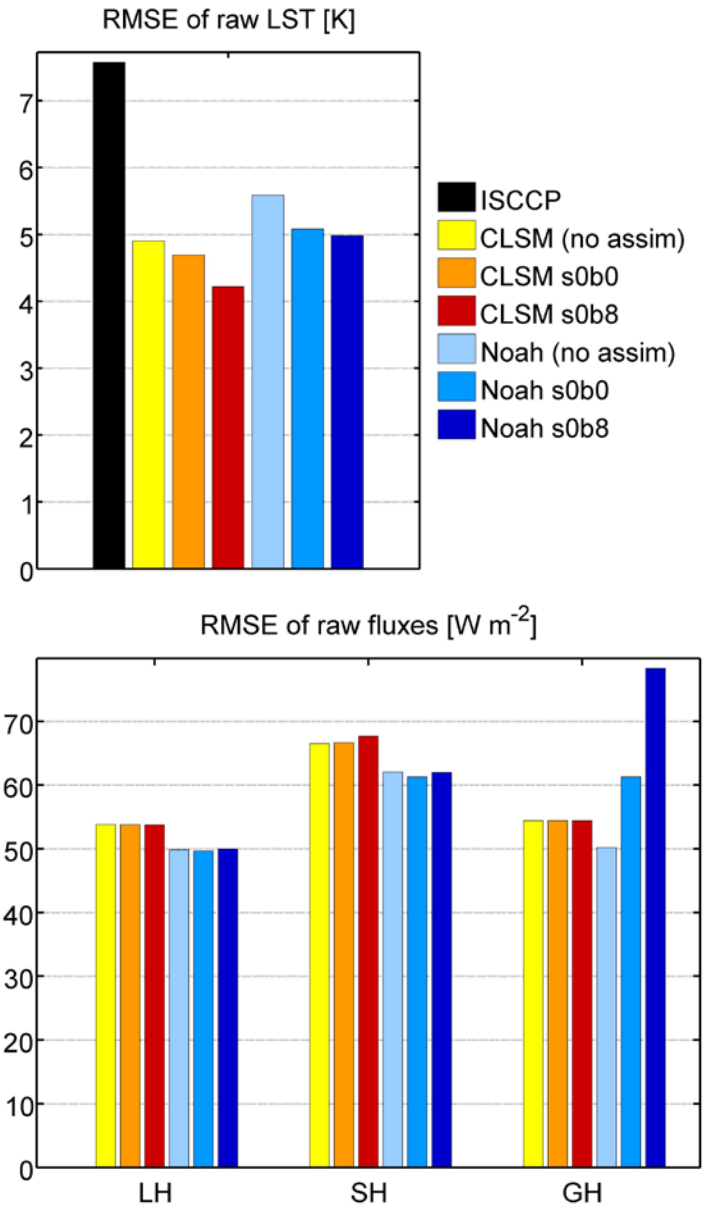
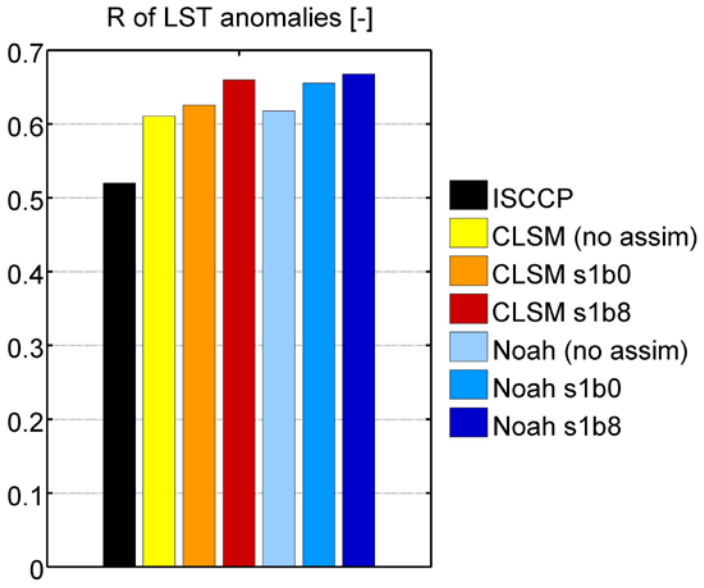


Fig. 3. RMSE versus CEOP in situ observations for (top) LST and (bottom) flux estimates from ISCCP retrievals (LST only), model integrations, and select assimilation integrations *without* a priori scaling.

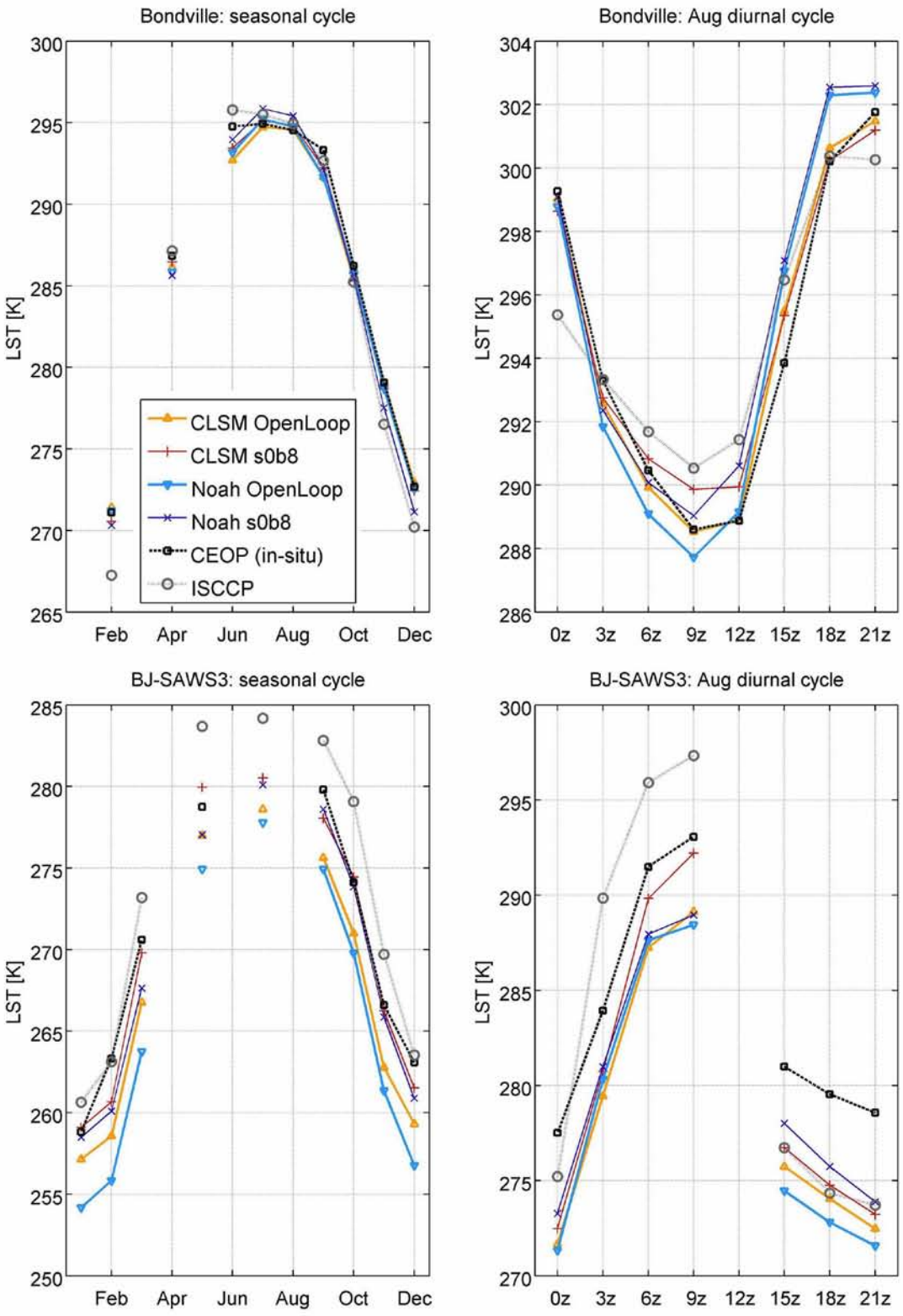


993

994 **Fig. 4.** R versus CEOP in situ observations for LST anomalies from ISCCP retrievals, model

995 integrations, and select assimilation integrations *with* a priori scaling.

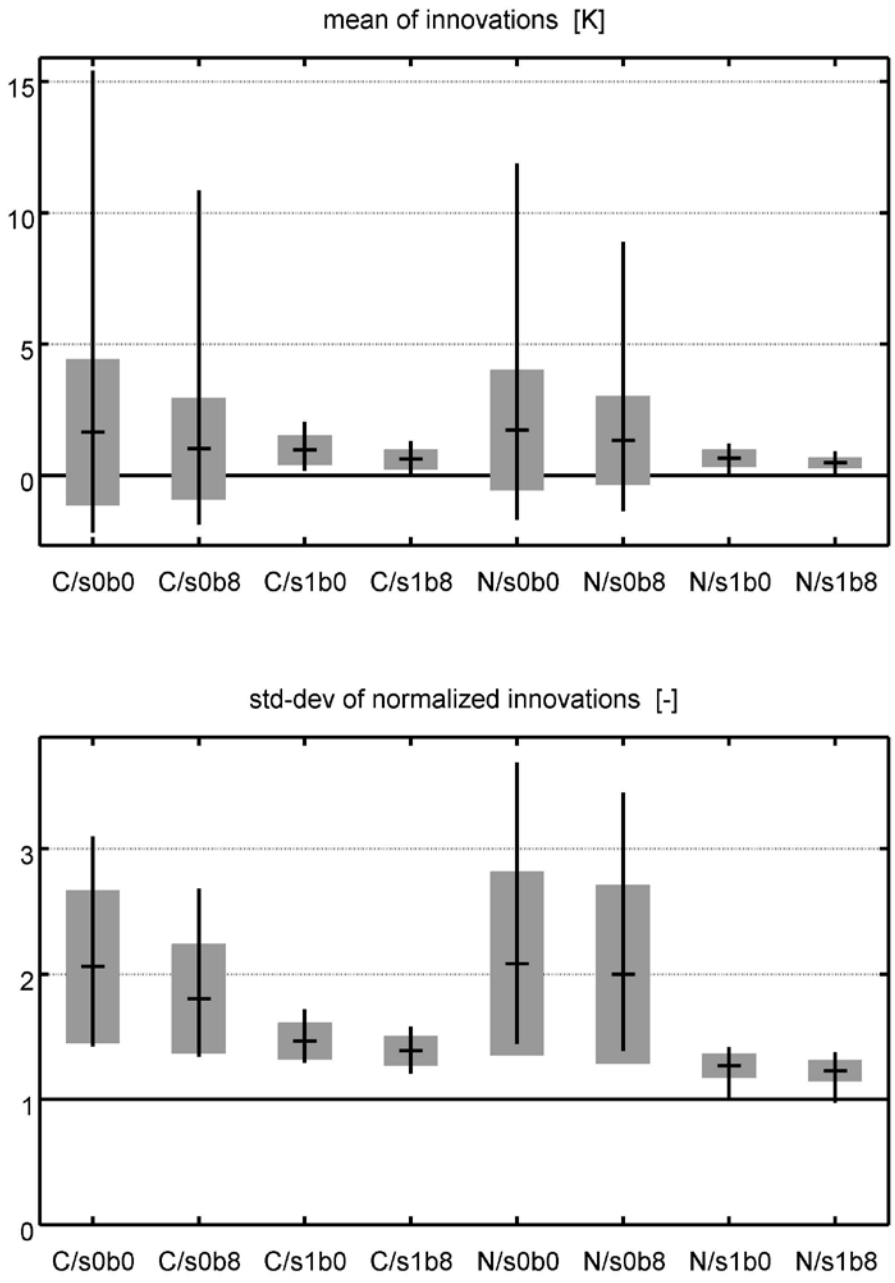
996



997

998 **Fig. 5.** LST (left) annual seasonal and (right) August diurnal cycle at (top) Bondville and

999 (bottom) BJ-SAWS3 for CEOP, ISCCP, model, and assimilation data.



1000

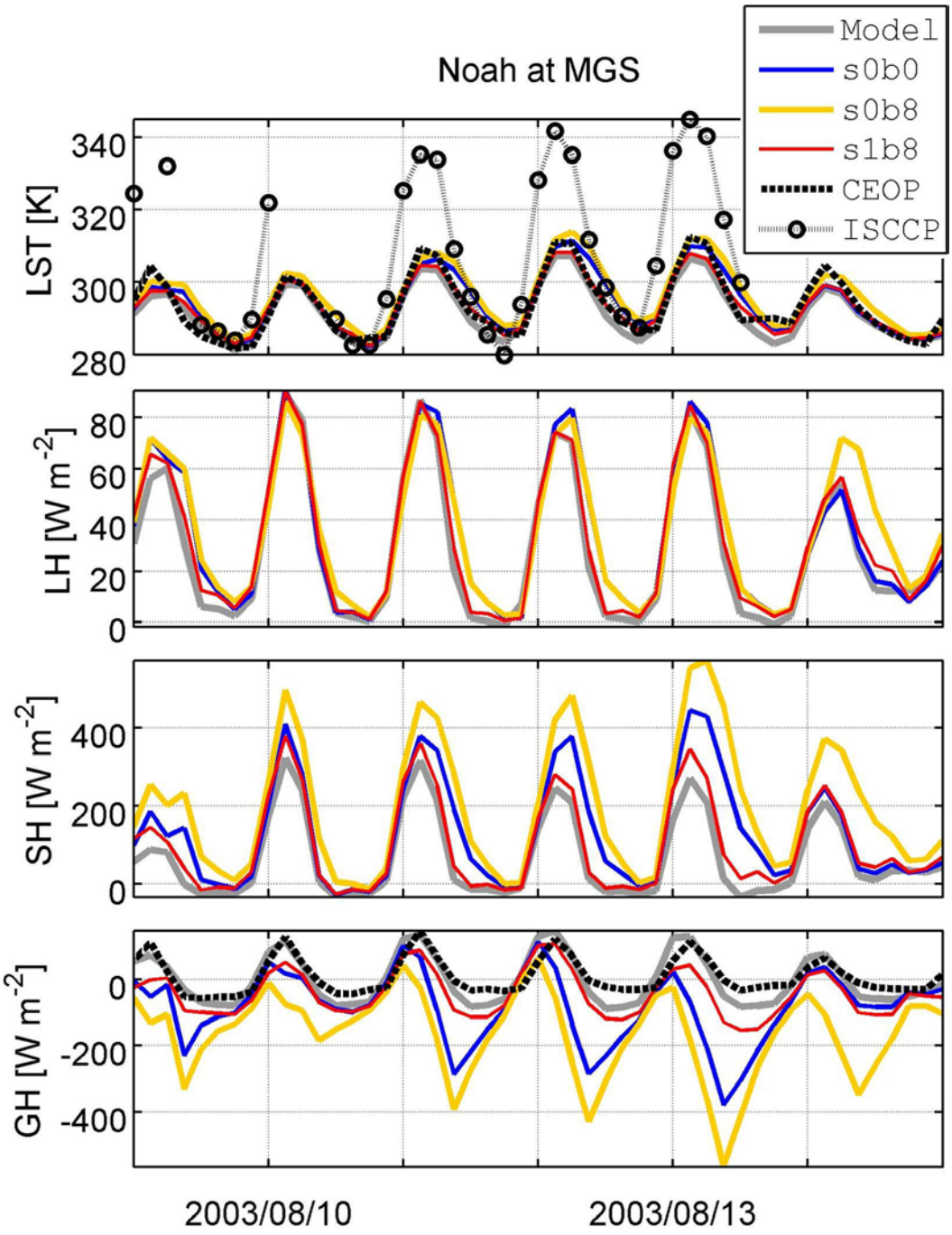
1001 **Fig. 6.** (Top) Mean of innovations [K] and (bottom) standard deviation of normalized

1002 innovations [dimensionless] for (C) Catchment and (N) Noah assimilation integrations. The box

1003 plots indicate the average, standard deviation, minimum and maximum of the respective

1004 innovations diagnostic across the stations listed in Table 1.

1005



1006

1007 **Fig. 7.** (Top) LST, (upper middle) LH, (lower middle) SH, and (bottom) GH for Noah

1008 integrations, ISCCP retrievals, and CEOP observations at the MGS station.