ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 22, NO. 1, 2005, 142–157

Optimal Parameter and Uncertainty Estimation of a Land Surface Model: Sensitivity to Parameter Ranges and Model Complexities

Youlong XIA*[1], Zong-Liang YANG[2], Paul L. STOFFA[1], and Mrinal K. SEN[1]

[1]*Institute for Geophysics, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin*, 4412 *Spicewood Spring Road, Austin, TX* 78759-8500, *USA*

[2]*Department of Geological Sciences, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, USA*

(Received 8 March 2004; revised 23 October 2004)

ABSTRACT

Most previous land-surface model calibration studies have defined global ranges for their parameters to search for optimal parameter sets. Little work has been conducted to study the impacts of realistic versus global ranges as well as model complexities on the calibration and uncertainty estimates. The primary purpose of this paper is to investigate these impacts by employing Bayesian Stochastic Inversion (BSI) to the Chameleon Surface Model (CHASM). The CHASM was designed to explore the general aspects of land-surface energy balance representation within a common modeling framework that can be run from a simple energy balance formulation to a complex mosaic type structure. The BSI is an uncertainty estimation technique based on Bayes theorem, importance sampling, and very fast simulated annealing. The model forcing data and surface flux data were collected at seven sites representing a wide range of climate and vegetation conditions. For each site, four experiments were performed with simple and complex CHASM formulations as well as realistic and global parameter ranges. Twenty eight experiments were conducted and 50 000 parameter sets were used for each run. The results show that the use of global and realistic ranges gives similar simulations for both modes for most sites, but the global ranges tend to produce some unreasonable optimal parameter values. Comparison of simple and complex modes shows that the simple mode has more parameters with unreasonable optimal values. Use of parameter ranges and model complexities have significant impacts on frequency distribution of parameters, marginal posterior probability density functions, and estimates of uncertainty of simulated sensible and latent heat fluxes. Comparison between model complexity and parameter ranges shows that the former has more significant impacts on parameter and uncertainty estimations.

Key words: optimal parameters, uncertainty estimation, CHASM model, bayesian stochastic inversion, parameter ranges, model complexities

1. Introduction

The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS, Henderson-Sellers et al., 1995, 1996) has led to the identification of large differences in land surface schemes in the partitioning of available water between runoff and evaporation and in the partitioning of available energy between sensible and latent heat fluxes (Henderson-Sellers, 1996; Shao and Henderson-Sellers, 1996; Chen et al., 1997; Wood et al., 1998; Dirmeyer et al., 1999; Schlosser et al., 2000; Bowling et al., 2003). The differences among model outputs mainly result from differences in model structures (e.g., degree of model complexity in representing biophysical and hydrological processes) and inappropriate values of model parameters if errors in forcing and calibrated data are assumed to be negligible. Although intercomparison efforts have attempted

*E-mail: youlong.xia@noaa.gov

Affiliation: NOAA Geophysical Fluid Dynamics Laboratory & Atmospheric and Oceanic Science Program, Princeton University, Forrestal Campus, US RTE 1, Princeton, NJ 08542, USA

to remove differences resulting from parameters by assigning a common set of parameters for all schemes, no effective mechanism existed to ensure that these parameter values were optimal. One way to obtain optimal parameter values among different schemes is to use methods of parameter calibration.

Automated methods for identifying optimal parameter sets have been developed over the past two decades. Sellers et al. (1989) used an iterative loop driven by a least square reduction program and micrometeorological measurements taken over the Amazonian tropical forest to estimate and optimize physiological parameters in the Simple Biosphere Model. Their results showed that the specification of optimal parameters improved the simulation of latent heat fluxes. Franks and Beven (1997) used the generalized likelihood uncertainty estimation technique to estimate the uncertainty in the fluxes simulated by a simple soil-vegetation-atmosphere transfer scheme. Gupta et al. (1999) used a multi-criteria (MC) calibration method to estimate acceptable optimal parameter sets for the Biosphere Atmosphere Transfer Scheme (BATS). The results showed that the BATS performed better when its parameters were optimized using the MC method. Xia et al. (2002) used the MC method to investigate the relationship between model complexity and simulation performance for one measurement site (viz., Cabauw). Their results showed that complex models performed better than the simple models when optimal model parameters were used. Recently, Bayesian Stochastic Inversion (BSI) has also been used to search for optimal parameters along with their uncertainties for the Chameleon Surface Model (CHASM) using data from Cabauw, Netherlands (Jackson et al., 2003).

More recently, Bastidas et al. (2004) used four land surface models with different complexities at five measurement sites to evaluate the ability of the MC method for optimal parameter estimation. Xia et al. (2004a) not only evaluated the ability of the BSI method for estimating optimal parameters and uncertainties of CHASM at seven sites but also made a comparison with the MC method. Their results showed that both BSI and MC are effective methods to estimate optimal parameters in terms of improving simulations of sensible and latent heat fluxes. Besides estimating optimal parameters, BSI was also used to estimate uncertainty of land surface parameters in the previous study (Xia et al., 2004a).

However, most of the previous studies used a global range set or an approximate global range to search for optimal parameter sets. The global range includes the largest possible scope of variations for a parameter. It should be noted that the global optimization methods (e.g., BSI, MC) are only mathematical tools for searching a set of optimal parameters. This set of optimal parameters is only mathematically optimal in that it has a minimum error function or cost function. If a land surface model were 'perfect' in representing the 'real world' land surface, and if the forcing data and energy fluxes had no observational errors, these optimization methods could find a set of optimal parameters which are physically reasonable (optimal), given a global range. However, due to model defects and observational errors, an arbitrary global range could result in physically unreasonable or unrealistic optimal parameter values although they are mathematically optimal. Parameter ranges are valuable only if the values of a parameter set found for a specific case are physically meaningful. At the Cabauw site, observed vegetation cover is close to unity all year round and vegetation roughness length is less than 0.2 m (see Beljaars and Bosveld, 1997). However, calibrated optimal values are 0.3–0.7 for vegetation cover and 0.2 m–1.5 m for vegetation roughness length for the six CHASM modes (Xia et al., 2002). These calibrated values are obviously unrealistic at the Cabauw site. Similar problems also existed in Gupta et al. (1999), Jackson et al. (2003), Bastidas et al. (2004) and Xia et al. (2004b) regardless of optimization methods and land surface models. Therefore, obtaining these unreasonable optimal parameter values may be a result of the improper selection of parameter ranges. In addition, model defect is another important source of uncertainty which can influence the selection of optimal parameters and their uncertainty estimates. In this study we used different CHASM modes (viz. simple and complex) and different parameter ranges (viz. realistic and global) to investigate their impacts on optimal parameter and uncertainty estimates at seven field observation sites.

2. Sites, model, method and experiment design

2.1 *Sites*

The model forcing data and surface flux data used in this study were collected at seven sites. These sites were chosen based upon data availability and different climate and vegetation characteristics. They represent mid-latitude grasslands, mid-latitude crops, tropical grasslands, tropical forests and mid-latitude forests. As suggested by Sen et al. (2001), these typical vegetations cover over 50% of the world's land area. At all sites, forcing data include downward longwave radiation (DLR), air temperature (T), relative humidity (q), wind speed (V), precipitation (P) and incoming solar radiation (ISR) or net radiation ($R_{\rm net}$). The energy flux data include sensible and latent heat fluxes.

Table 1. Description of seven sites and observed data.

Site Name	Site Location Lat Lon Elevation (m)	Observational Period	Observational Interval (min)	Vegetation Type	Annual Mean Precipitation (mm)	Annual Mean Temperature (K)	Input Data
Abracof	10°5′S 61°5′W 120 m	Jun 1992–Dec 1993	60	Tropical rain forest	1990	298	ISR, Rnet, T, q, V, P
Abracop	10°45′S 62°22′W 220 m	Jun 1992–Dec 1993	60	Tropical rain pasture	1985	297	ISR, Rnet, T, q, V, P
Amazon	2°57′S 59°57′W 80 m	Jan 1997–Dec 1998	30	Tropical rain forest	1990	298	ISR, DIR, T, q, V, P
Armcart	36°36′N 97°29′W 318 m	Apr 1995–Aug 1995	30	Mid-latitude crops	884	284	ISR, Rnet, T, q, V, P
Cabauw	51°58′N 4°56′E –0.7 m	Jan 1987–Dec 1987	30	Mid-latitude grassland	776	282	ISR, DLR, T, q, V, P
Loobos	52°10′N 5°44′E 52 m	Jan 1997–Dec 1998	30	Mid-latitude Scott pine	786	283	ISR, DLR, T, q, V, P
Tucson	32°13′N 111°5′W 730 m	May 1993–May 1994	20	Semi-arid grass and shrubs	275	293	ISR, DLR, T, q, V, P

Detailed description of site locations, site climates and site observations are given in Table 1. For more details, see Xia et al. (2004a, 2004b).

2.2 *Model*

The CHASM (Chamelon Surface Model, Desborough, 1999; Pitman et al., 2003) has been used for offline intercomparison of the PILPS phase 2d (Schlosser et al., 2000; Slater et al., 2001) and 2e (Bowling et al., 2003) and simulations of the coupled general circulation model (Desborough et al., 2001) and the limited area model (Zhang et al., 2001). It was designed to explore the general aspects of land-surface energy balance representation within a common modeling framework (Desborough, 1999) that can be run in a variety of surface energy balance modes ranging from the simplest energy balance formulation (Manabe, 1969) to a complex mosaic type structure (see Koster and Suarez, 1992). Two CHASM modes [Simple Land-Atmosphere Mosaic (SLAM) and Surface Resistance (RS)] used in this study are shown in Table 2 and are described below. Within the SLAM, the land-atmosphere interface is divided into two tiles. The first tile is a combination of bare ground and exposed snow with the second tile consisting of dense vegetation. The tiles may be of different sizes and the energy fluxes of each tile are area-weighted. Because a separate surface balance is calculated for each tile, temperature variations may exist across the land-atmosphere interface. A prognostic bulk temperature for the storage of energy and a diagnostic skin temperature for the computation of surface energy fluxes are calculated for each tile. Snow covering fractions for both ground and foliage surfaces are calculated as functions of the snowpack depth, density, and the vegetation roughness length. The vegetation fraction is further divided into wet and dry fractions if canopy interception is considered. This model has explicit parameterizations including canopy resistance, canopy interception, vegetation transpiration and bare ground evaporation, but has no explicit canopy-air space (see Pitman et al., 2003). The RS mode is also designed as a two-tile CHASM model with the simplest physical representation, such as the aerodynamic resistance to turbulent transport for heat and moisture and a temporally invariant surface resistance.

Table 2. Summary of the two CHASM modes.

Surface mode	Stability correction	Surface resistance	Canopy interception	Bare-ground evaporation	Canopy resistance	Temperature difference
RS	yes	yes	no	no	no	yes
SLAM	yes	yes	yes	yes	yes	yes

CHASM uses a common hydrological module originally described by Manabe (1969) as the hydrologic component of the land surface in which the root zone is treated as a bucket with finite water holding capacity. Any water accumulation beyond this capacity is assumed to be runoff. Except for moisture in the root zone, water can be stored as snow on the ground or on the canopy. Soil temperature is calculated within four soil layers using a finite difference method and zero-flux boundary condition. Each tile has four evaporation sources: canopy evaporation, transpiration, bare ground evaporation, and snow sublimation.

2.3 *Bayesian Stochastic Inversion* (*BSI*)

The BSI methodology is based on Bayes theorem and, usually, a stochastic method to select sets of parameter values from a distribution of realistic choices for model parameters. It has been used in solid geophysics (Sen and Stoffa, 1995, 1996) and land surface model simulations (Jackson et al., 2003; Xia et al., 2004a). Its detailed description can be found in Jackson et al. (2004). Here we give a brief description only. The basic idea of this method, based on very fast simulated annealing, is to strike a balance between identifying the optimal parameter set and mapping the entire multi-dimensional parameter probability distribution. Within the Bayesian nomenclature, the relative probability for each combination of parameter values is expressed as a 'posterior' probability density function (PPD) and is given mathematically as

$$\sigma(\boldsymbol{m} \mid \boldsymbol{d}_{\rm obs}) = \frac{\exp[-sE(\boldsymbol{m})]p(\boldsymbol{m})}{\int \exp[-sE(\boldsymbol{m})]p(\boldsymbol{m})d\boldsymbol{m}} , \qquad (1)$$

where the domain of integration spans the entire model parameter space $\boldsymbol{m}, \sigma(\boldsymbol{m} \mid \boldsymbol{d}_{\rm obs})$ is the PPD, vector $\boldsymbol{d}_{\rm obs}$ is the observational data, $E(\boldsymbol{m})$ is the error function, $\exp[-sE(\boldsymbol{m})]$ is the likelihood function, and $p(\boldsymbol{m})$ is the 'prior' probability density function for $\boldsymbol{m}$. The s is a shaping factor discussed in Xia et al. (2004a). Because only the range for each model parameter in $\boldsymbol{m}$ is known, a uniform distribution within the ranges is used as the 'prior' probability density function. This selection provides a non-informative constraint for any parameter values within the parameter search window.

Because the PPD calculated using Eq. (1) is multidimensional, it is difficult to visualize. Therefore, a one-dimensional projection of the PPD is usually displayed (Sen and Stoffa, 1995, 1996). This projection is called marginal PPD of a parameter. The marginal PPD shape of a parameter (e.g., Fig. 10) is proportional to the frequency distribution of the parameter and the distribution of $\exp(-sE)$ when shaping factor s is given. After the marginal PPD of a parameter is calculated, the parameter range at different confidence levels (e.g., 95%) can be used to estimate uncertainty of the parameter, or one can directly use the marginal PPD distribution to the discuss uncertainty of the parameter according to the shape of the PPD (a sharper-peaked PPD means smaller uncertainty, and a wider-peaked PPD means larger uncertainty).

2.4 *Experimental design*

In their previous sensitivity analysis, Xia et al. (2004a) identified ten sensitive CHASM parameters for the seven sites. The relative importance of these 10 parameters and their descriptions are listed in Table 3. The sensitivity of the CHASM parameters is closely related to the sites studied. Table 4 gives a list of 10 parameters in CHASM as well as their global ranges and realistic ranges for the seven sites based on the relevant literature (Henderson-Sellers et al., 1986; Oke, 1987; Dorman and Sellers, 1989; Pitman, 1994; Gulf et al., 1995; Bonan, 1996; Gulf et al., 1996; Sellers et al., 1996a, 1996b; Unland et al., 1996; Wright et al., 1996; Beljaars and Bosveld, 1997; Cash and Nobre, 1997; Chen et al., 1997; Yang et al., 1998; Desborough et al., 1999; Gupta et al., 1999; Sen et al., 2001; Xia et al., 2002; Dai et al., 2003; Xia et al., 2004a). Differences of maximum and minimum air temperatures at each site are used as realistic ranges of initial soil temperature. Sensible and latent heat fluxes are used as the 'target' observations for all seven sites. A ratio of the variance of the errors (RVE) to the variance of observations is used to define the mismatch between observations and model simulations for the BSI selection process (Xia et al., 2004a). The RVE is defined as

$$R_{\rm VE} = \frac{\sum_{n=1}^{N} (D_{{\rm obs},n} - D_{{\rm sim},n})^2}{\sum_{n=1}^{N} (D_{{\rm obs},n} - \overline{D}_{\rm obs})^2} , \qquad (2)$$

where N is the number of observational data, $D_{{\rm obs},n}$ is the observed data, $D_{{\rm sim},n}$ is the simulated data, and $\overline{D}_{\rm obs}$ is mean value of the observed data. An arithmetically-averaged error function for sensible and latent heat fluxes is used in this study. BSI is used for the seven sites to identify an optimal parameter set which has a minimum RVE and to estimate parameter uncertainty ranges for two CHASM modes (viz., SLAM and RS). Therefore, four basic experiments were designed for each site and a total of 28 runs were performed at the seven sites. 50 000 parameter sets for each run were used to calculate marginal PPDs. The four basic experiments were Expt 1 (SLAM with global range), Expt 2 (SLAM with realistic range), Expt 3 (RS with global range), and Expt 4 (RS with

Table 3. Parameter description and the relative importance of CHASM parameters for seven sites [ticks indicate the important parameters, taken from Xia et al. (2004a)].

Parameter	Abracof	Abracop	Amazon	Armcart	Cabauw	Loobos	Tucson	Description
ALBG						√	√	Bare ground albedo
ALBV	√	√	√		√	√	√	Vegetation albedo
LEFM			√					Maximum LAI
VEGM	√	√	√	√		√	√	Maximum fractional vegetation cover
RCMIN	√	√	√	√	√	√	√	Maximum canopy resistance (s m^{-1})
WRMAX	√	√	√				√	Available water hoding capacity (mm)
Z0G	√	√	√	√	√			Bare ground roughness length (m)
Z0V	√	√	√	√	√		√	Vegetation roughness length (m)
TS	√	√						Initial surface temperature (K)
WET	√	√	√	√	√	√	√	Initial soil wetness

Table 4. Realistic and global ranges of 10 CHASM parameters selected from different references.

Parameter	Realistic ranges by site							Global Ranges
	Abracof	Abracop	Amazon	Armcart	Cabauw	Loobos	Tucson	All Sites
ALBG	0.15–0.25	0.15–0.25	0.15–0.25	0.15–0.25	0.15–0.25	0.15–0.25	0.2–0.3	0.05–0.40
ALBV	0.10–0.25	0.10–0.25	0.10–0.25	0.20–0.30	0.20–0.30	0.15–0.25	0.15–0.35	0.05–0.40
LEFM	4.0–6.0	1.0–4.0	4.0–6.0	1.0–4.0	3.0–5.0	2.5–3.5	0.05–3.0	0.05–6.0
VEGM	0.9–1.0	0.8–0.9	0.9–1.0	0.80–1.0	0.9–1.0	0.9–1.0	0.2–0.6	0.0–1.0
RCMIN	40.00–200.0	40.0–200.0	40.0–200.0	40.0–200.0	40.0–200.0	40.0–200.0	40.0–200.0	40.0–300.0
WRMAX	200.0–300.0	100.0–200.0	200.0–300.0	100.0–200.0	100.0–200.0	100.0–200.0	100.0–200.0	40.0–400.0
Z0G	0.005–0.015	0.005–0.015	0.005–0.015	0.005–0.015	0.005–0.015	0.005–0.015	0.005–0.015	0.005–0.020
Z0V	2.0–2.5	0.01–0.20	2.0–2.5	0.01–0.20	0.01–0.20	2.0–2.5	0.01–0.20	0.01–2.5
TS	294.0–301.0	293.0–300.0	298.0–304.0	275.0–293.0	276.0–282.0	255.0–264.0	293.0–305.0	255.0–305.0
WET	0.7–1.0	0.7–1.0	0.7–1.0	0.5–0.8	0.7–1.0	0.7–1.0	0.3–0.6	0.0–1.0

realistic range).

3. Influence of parameter ranges and model complexities on optimal parameter estimation

Figure 1 shows the calculated root mean square errors (RMSEs) and biases between simulated and observed energy fluxes at the seven sites. These simulations were obtained by running the model with the best parameter set identified by the BSI. For the SLAM mode, use of global and realistic ranges has similar RMSEs for both sensible and latent heat fluxes simulations, although use of realistic ranges has slightly larger RMSEs at all sites (Figs. 1a and 1b). For the RS mode, use of realistic ranges results in significantly larger RMSEs for sensible and latent heat fluxes at the two tropical rainforest sites. At the other sites, however, use of global and realistic ranges leads to similar RMSEs. Therefore, use of global and realistic ranges gives similar model outputs for both modes for most sites.

The simple mode has larger RMSEs. When the complex mode is used, RMSE errors for sensible heat and latent heat fluxes are significantly reduced at the tropical forest and pasture sites as well as the midlatitude crops site when compared to the simple mode. This means that a complex model has better performance than a simple model when both are optimized for Abracof (ABF), Abracop (ABP), Amazon (AMA), and Armcart (ARM) sites. It should be noted that the same 10 parameters are calibrated for simple and complex CHASM modes so that this comparison is re-

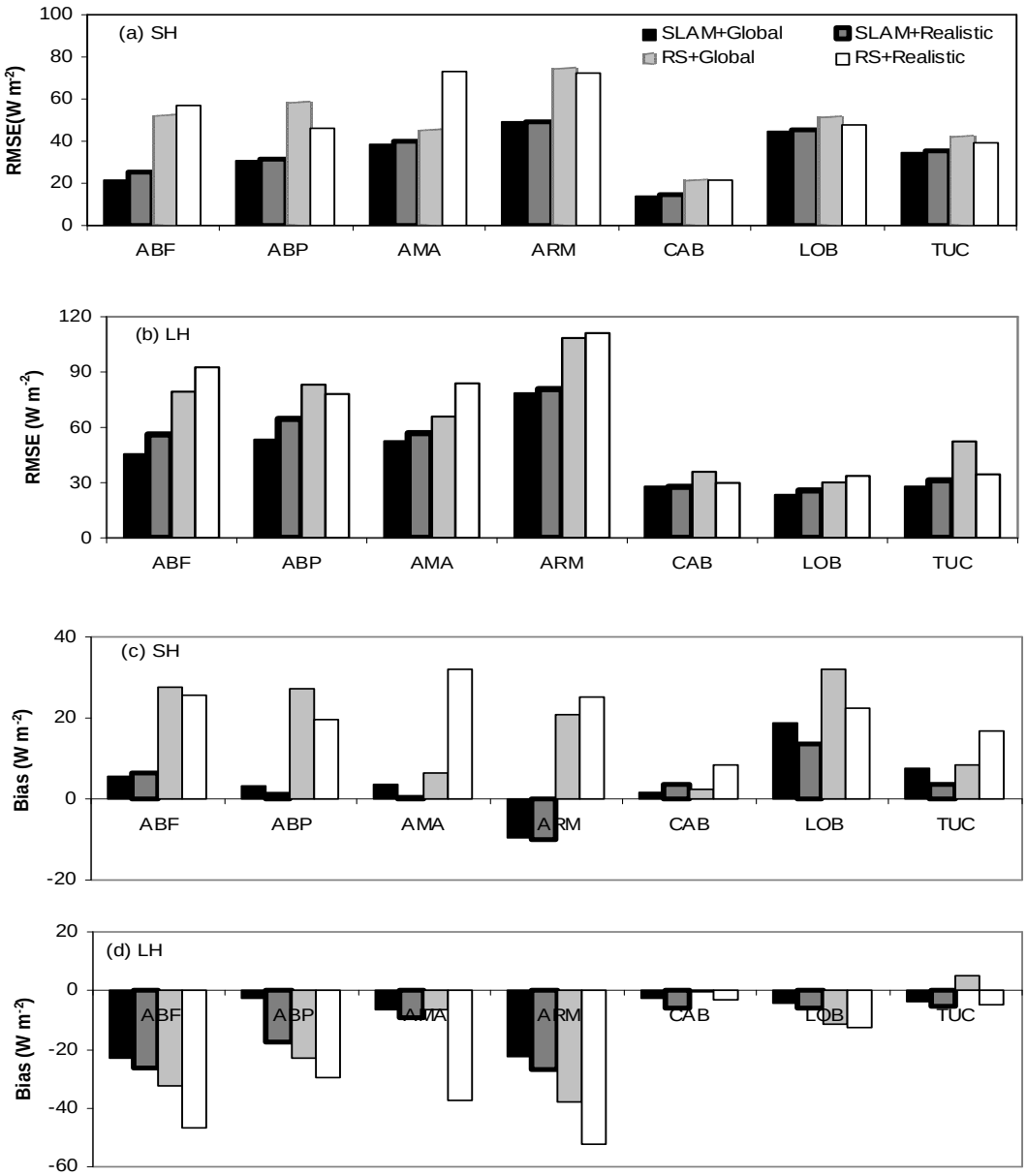

Fig. 1. Root mean square errors (RMSEs) and biases between observed and simulated fluxes at seven measurement sites. (a) RMSEs of sensible heat flux (SH), (b) RMSEs of latent heat fluxes (LH), (c) bias of sensible heat flux (SH), and (d) bias of latent heat flux (LH) when two parameter ranges and CHASM modes are used.

latively fair. However, it is still debatable if this conclusion has a general sense because RMSEs for sensible and latent heat fluxes are marginally reduced only for the Cabauw, Loobos and Tucson sites when the use of simple and complex modes is compared.

The analysis of bias for sensible and latent heat fluxes shows that CHASM overestimates latent heat flux and underestimates sensible heat flux. However, the simple mode generates larger positive biases for sensible heat fluxes, and larger negative biases for latent heat fluxes at almost all sites than the complex mode (Figs. 1c and 1d) regardless of the use of parameter ranges. Therefore, the overall assessment based on RMSE and bias shows that the complex mode performs better than the simple mode. For both the RMSE and bias analysis, the range of parameter and model complexities have the least impacts at the Cabauw site when compared to the other sites. The reason for this will be discussed in Section 6.

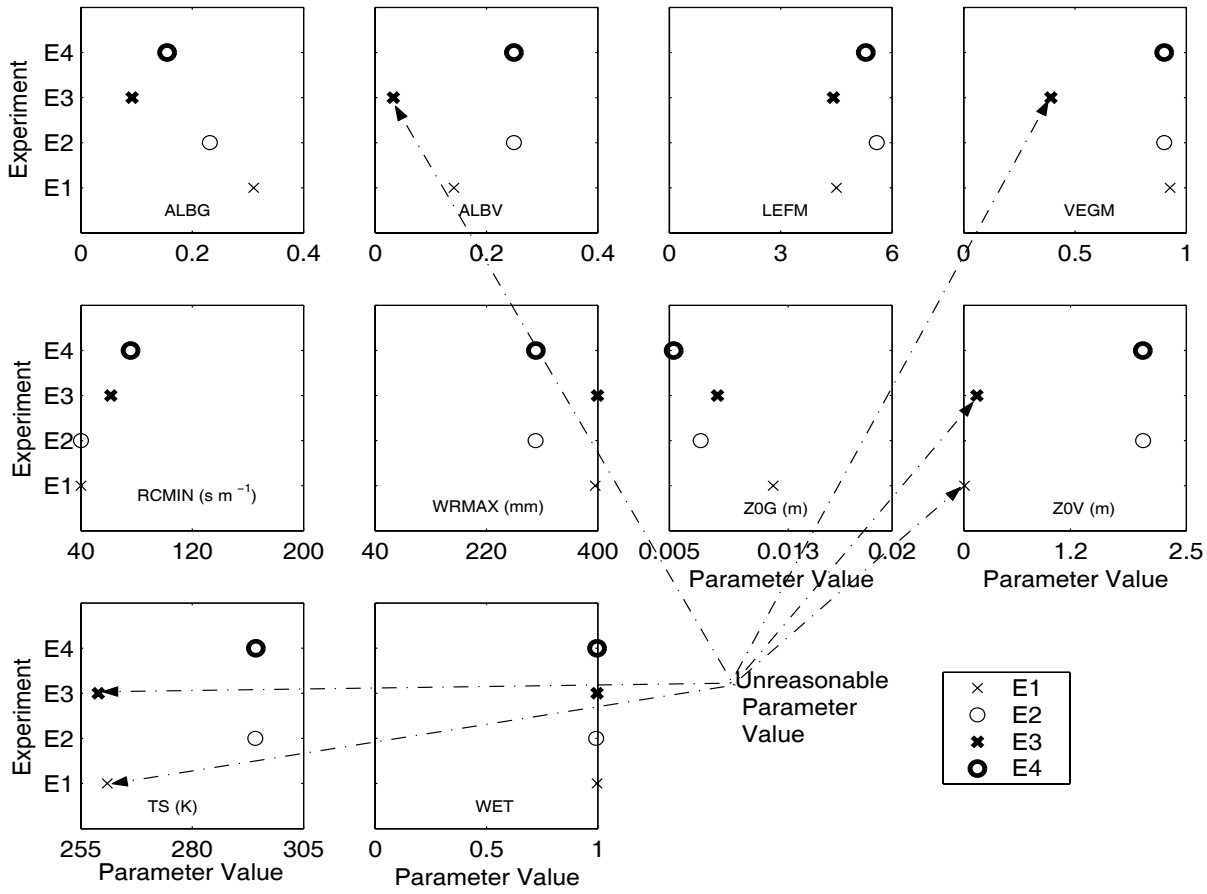

Fig. 2. Ten searched optimal parameter values for four experiments at the Abracof tropical forest site (black cross (E1)=SLAM+global; black circle (E2)=SLAM+realistic; bold cross (E3)=RS+global; bold circle (E4)=RS+realistic).

Although global and realistic ranges produced similar RMSEs for the seven sites, optimal parameters used in calculating RMSEs were relatively different. Figure 2 shows optimal parameter values searched by the BSI for four experiments at the Abracof tropical forest.

When the global ranges were used, unreasonable optimal parameter values were obtained for vegetation albedo (ALBV), vegetation cover fraction (VEGM), initial soil temperature (TS) and vegetation roughness length (Z0V) for the RS mode, and vegetation roughness length (Z0V) for the SLAM mode. These optimal values are too small when compared to observational values. However, reasonable optimal values were searched for 10 CHASM parameters when realistic ranges were used. Comparison of simple and complex modes shows that the simple mode tends to have more unreasonable optimal values. This statement holds true for the other two forest sites (Figs. 3 and 4). Unreasonable optimal values also occurred at the midlatitude grassland, midlatitude crops, tropical pasture, and semi-arid sites (see Figs. 3 and 4) when global ranges were used. However, use of realistic ranges results in reasonable and consistent optimal values for the two modes and 10 CHASM parameters at all sites.

While the selection of parameter ranges impacts estimation of optimal parameters, the use of global ranges does not change the conclusion that a complex model performs better than a simple model when optimal model parameters are used (Xia et al., 2002). However, Xia et al.'s conclusion was drawn from using unreasonable optimal parameter values. These unreasonable optimal values do not necessarily lead to unreasonable simulations of sensible and latent heat fluxes because they are a result of minimizing differences between simulated and observed fluxes. Despite the fact that the results from offline simulations are usually reasonable during the calibration period, it is not clear if the results are still reasonable from cou-

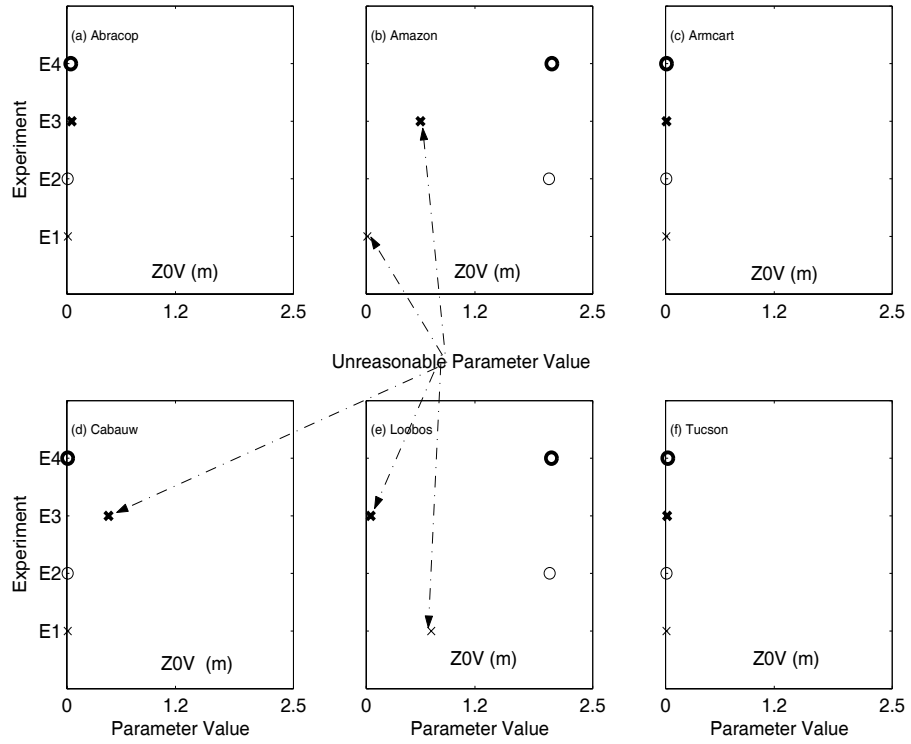

Fig. 3. Same as Fig. 2 but for vegetation roughness length at the other six measurement sites.

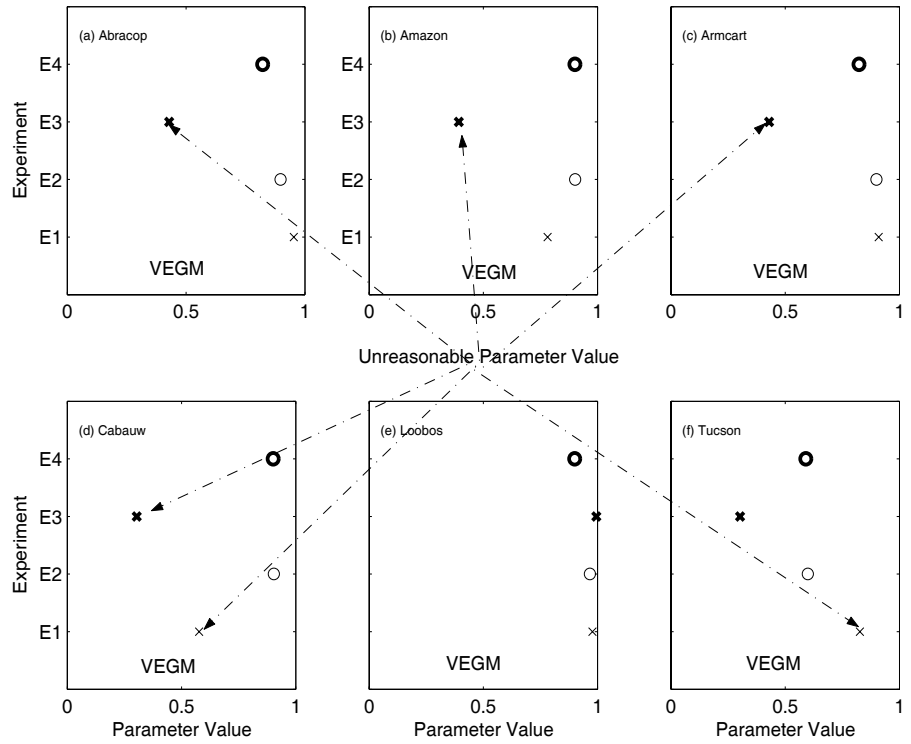

Fig. 4. Same as Fig. 2 but for vegetation cover fraction at the other six measurement sites.

pled land-atmosphere simulations. Further research is needed to examine how the use of unreasonable optimal values affects coupled land-atmosphere simulations.

4. Influence of parameter ranges and model complexities on frequency distribution of parameters

Use of global ranges also leads to unreasonable frequency distribution for some CHASM parameters, depending on sites and CHASM modes. In contrast, use of realistic ranges gives relatively consistent frequency distributions for most CHASM parameters and sites. Frequency distributions of vegetation roughness length (Z0V) and vegetation cover fraction (VEGM) at seven sites are two examples (Figs. 5–8). Comparison of the results from using global and realistic ranges shows that at all sites, the simulations with global ranges have a much larger degree of uncertainty than those with realistic ranges, which is obvious because global ranges themselves are much larger than realistic ranges. For tropical and midlatitude forests, the maximum frequency of Z0V for global range cases (Figs. 5a, 5c, and 5e) occurs between 0.0 and 0.5 m. This is not reasonable because Z0V is usually between 2.0 and 2.5 at all forest sites. This unreasonable frequency distribution is caused by the use of global ranges, because using realistic ranges results in consistent frequency distributions (see Figs. 5a–f). At sites of tropical pasture, midlatitude grassland, midlatitude crop, and semi-arid shrubs, it is reasonable that the maximum frequency of Z0V appears near zero for both global ranges and realistic ranges (see Figs. 6a–h). The simple mode has larger uncertainty estimates than the complex mode for six of the eight cases. For vegetation cover fraction, complexities of the CHASM model result in large differences in frequency distributions when global ranges are used (see Figs. 7 and 8), but they give consistent frequency distributions at five of the seven sites when realistic ranges are used. In the case of global ranges, maximum frequencies appeared on the right end of the parameter values for the complex mode and on the left end for the simple mode at six of the seven sites. Therefore, different frequency distributions indicate that, at least, the result of one mode (complex or simple) is not reasonable for this study. The complex mode gives reasonable frequency distributions at Abracof, Abracop, Amazon, and Armcart, and the simple mode gives reasonable frequency distributions at Tucson. In addition, at Cabauw when global parameter ranges are used, both modes give unreasonable frequency distributions because the maximum frequency appears in the area with small vegetation cover fraction values. In practice, however, observational vegetation cover is larger than 0.90 at the site. Therefore, the complex mode is more often likely to give reasonable frequency distributions than does the simple mode even though global parameter ranges are used. In this study, the simple mode gives reasonable frequency distribution at Tucson, but we believe that this result is less representative.

Figure 9 shows two-dimensional scatter plot distributions of vegetation roughness length and vegetation cover fraction at Abracof when global and realistic ranges are used. It should be noted that Figs. 9a–d have different scales for both x-axes and y-axes. The results when global parameter ranges are used show that, for a complex CHASM mode, the high frequency area appears on the bottom right of Fig. 9a where the roughness length is small and vegetation fraction values are large. For a simple CHASM mode, the high frequency area appears on the bottom left of Fig. 9b where both the roughness length and vegetation cover fraction are small. However, the two high frequency areas searched are unreasonable for both CHASM modes because they should be located on the top-right of Figs. 9a and 9b where both vegetation cover fraction and roughness length are large. However, when a realistic range was used, a high frequency area appears on the left of Figs. 9c and 9d where the parameter values are reasonable because realistic ranges are used.

In summary, selection of parameter ranges has a significant impact on frequency distributions of model parameters for both modes. In contrast, selection of model complexities has small effects on frequency distributions of model parameters.

5. Influence of parameter ranges and model complexities on uncertainty estimates

5.1 *Influence on uncertainty estimates of model parameters*

Uncertainties of model parameters can be represented by marginal PPD as discussed in section 2.3. As an example, marginal PPD distributions of minimum stomatal resistance for the two modes at the seven sites are shown in Fig. 10. Comparison of four experiments shows that for both modes, using global ranges leads to larger uncertainties in estimates of model parameters than using the realistic ranges at five of the seven sites. However, a simple model had much larger uncertainty than a complex model when the same parameter range is used. It is also clear that two CHASM modes generate different PPD distributions at all of the seven sites except for Cabauw. As discussed in our introduction, uncertainties of land surface models mainly come from

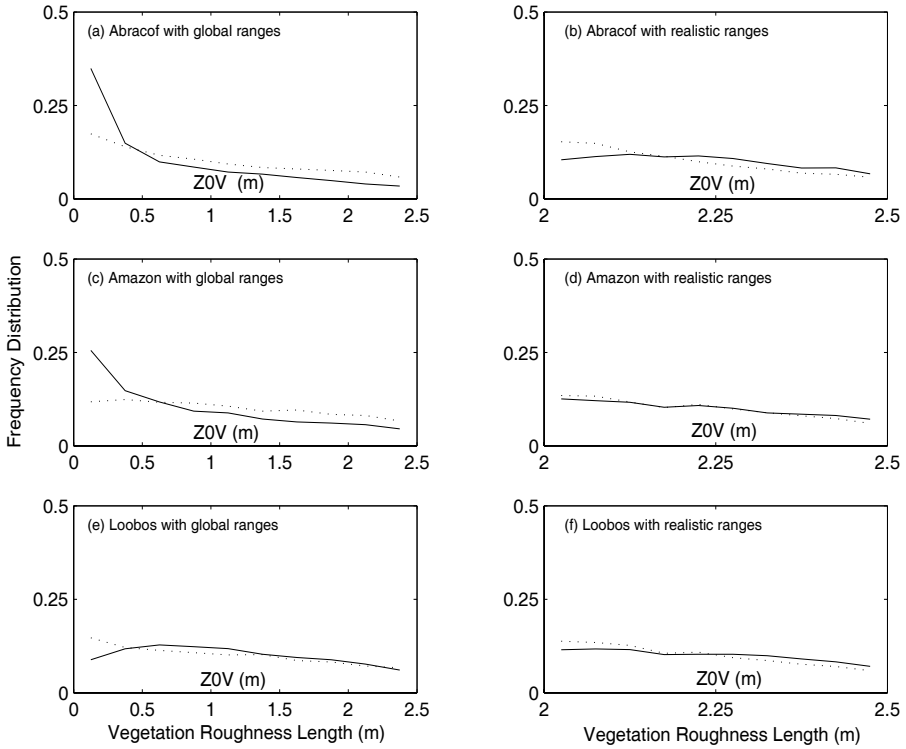

Fig. 5. Frequency distribution of vegetation roughness length at three forest sites when two ranges and two CHASM modes are used (results of realistic ranges are shown in the right column, and results of global ranges are shown in the left column. Solid line represents SLAM mode, and dotted line represents RS mode).

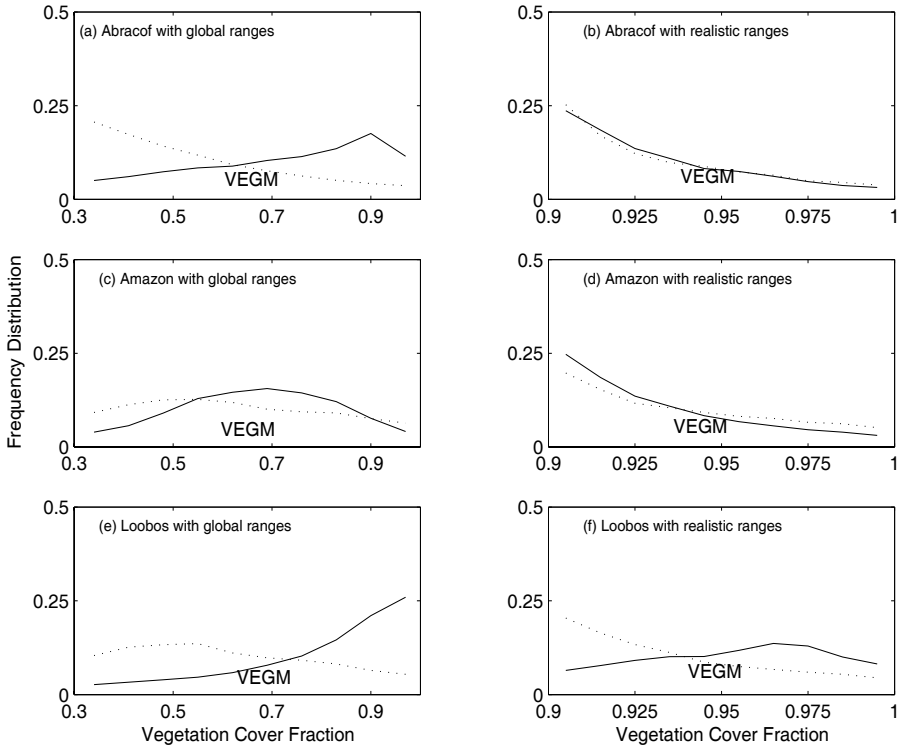

Fig. 6. Same as Fig. 5 but for vegetation cover fraction.

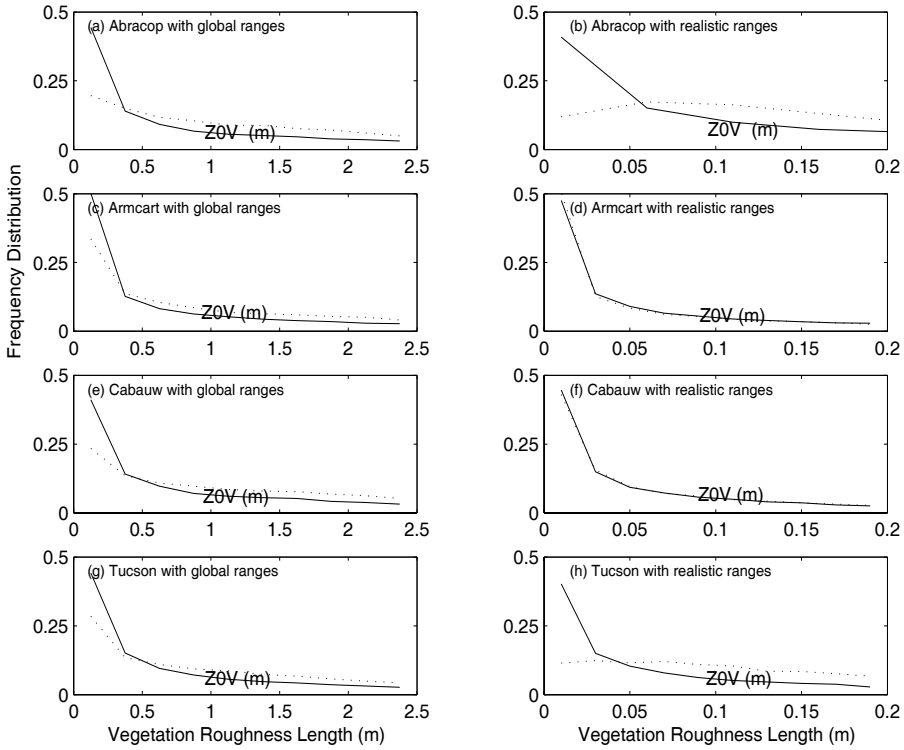

Fig. 7. Same as Fig. 5 but for tropical pasture, mid-latitude crop, mid-latitude grassland and semi-arid sites.

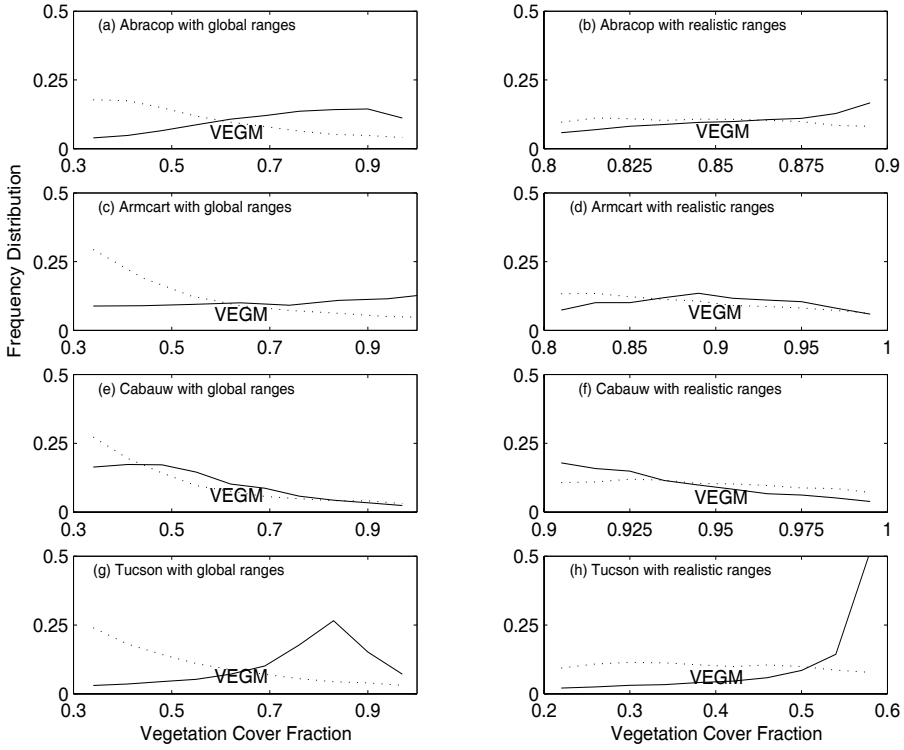

Fig. 8. Same as Fig. 7 but for vegetation cover fraction.

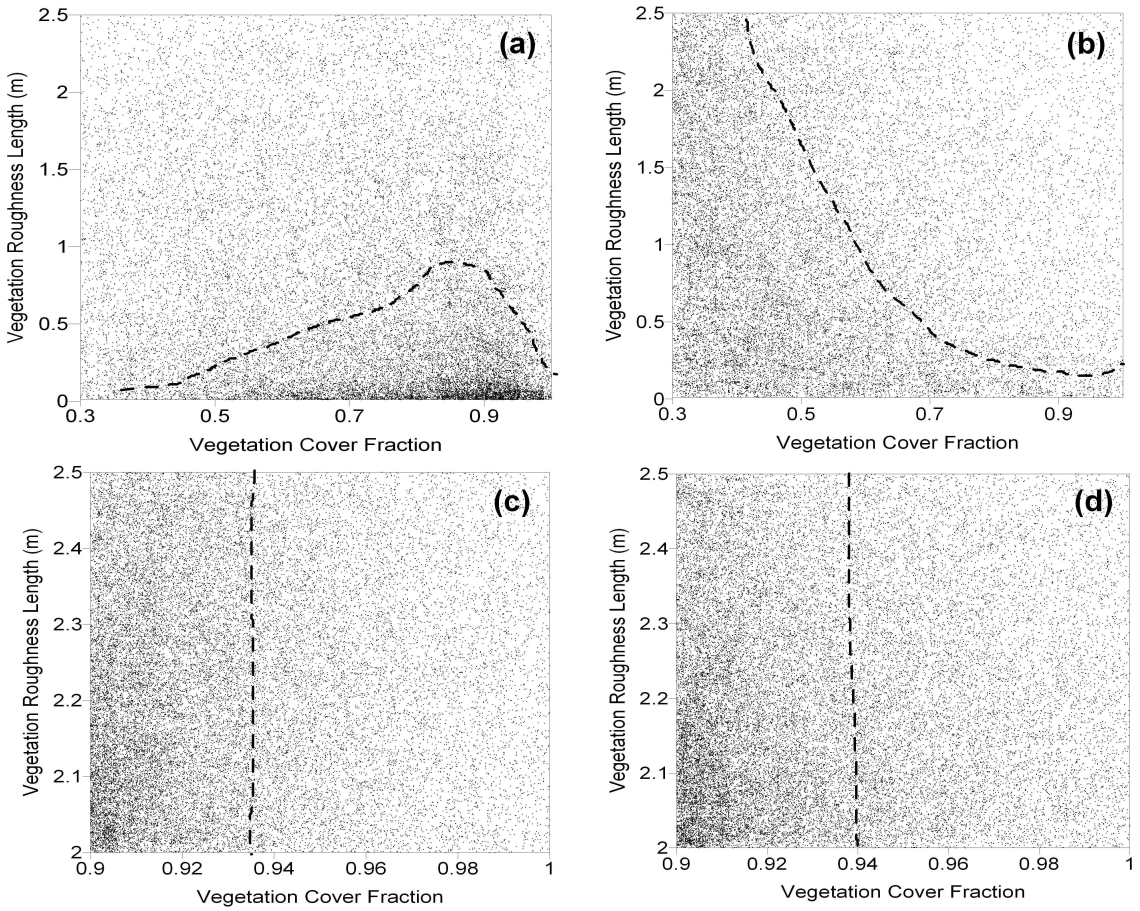

Fig. 9. Distributions of parameter points for four experiments at Abracof tropical forest site for (a) SLAM with global range, (b) RS with global range, (c) SLAM with realistic range, and (d) RS with realistic range.

insufficient model complexity and inaccurate parameter values if forcing data and calibration data are assumed to be accurate. The comparison of four experiments shows that the complexity of model structure instead of the range of parameters dictates the degree of uncertainty in the optimal values of parameters.

5.2 *Influence on uncertainty estimates of sensible and latent heat fluxes*

Uncertainty ranges of simulated sensible and latent fluxes are directly related to estimated error function values when a percentage of total parameter sets (say 10%) is given. This error function value is an arithmetic average of the error of latent heat and the error of sensible heat. The larger the error function values, the greater the uncertainty ranges of sensible and latent heat fluxes. Figure 11 shows the cumulated distribution function for the seven sites. The results show that insufficient model complexities result in large error function values at six of the seven sites if a given percentage of the best parameter sets (e.g., 10%) is used to estimate the uncertainty range of sensible and latent heat fluxes as in Frank and Beven (1997). Use of parameter ranges results in certain differences for error functions. However, these results are mixed so that it is difficult to draw a conclusion.

Overall, it is clear that the selection of parameter ranges and model complexities has significant impacts on frequencies of model parameters, distributions of error functions, marginal PPDs, and uncertainty estimates of simulated energy fluxes at some sites. Comparison of parameter ranges and model complexities shows that the latter has larger effects on model optimization and uncertainty analysis.

6. Discussion

Although most calibration studies in land surface schemes have used global parameter ranges, little study has been conducted to discuss the impacts of parameter ranges. It is widely recognized that calibration algorithms are just "blind" computer programs

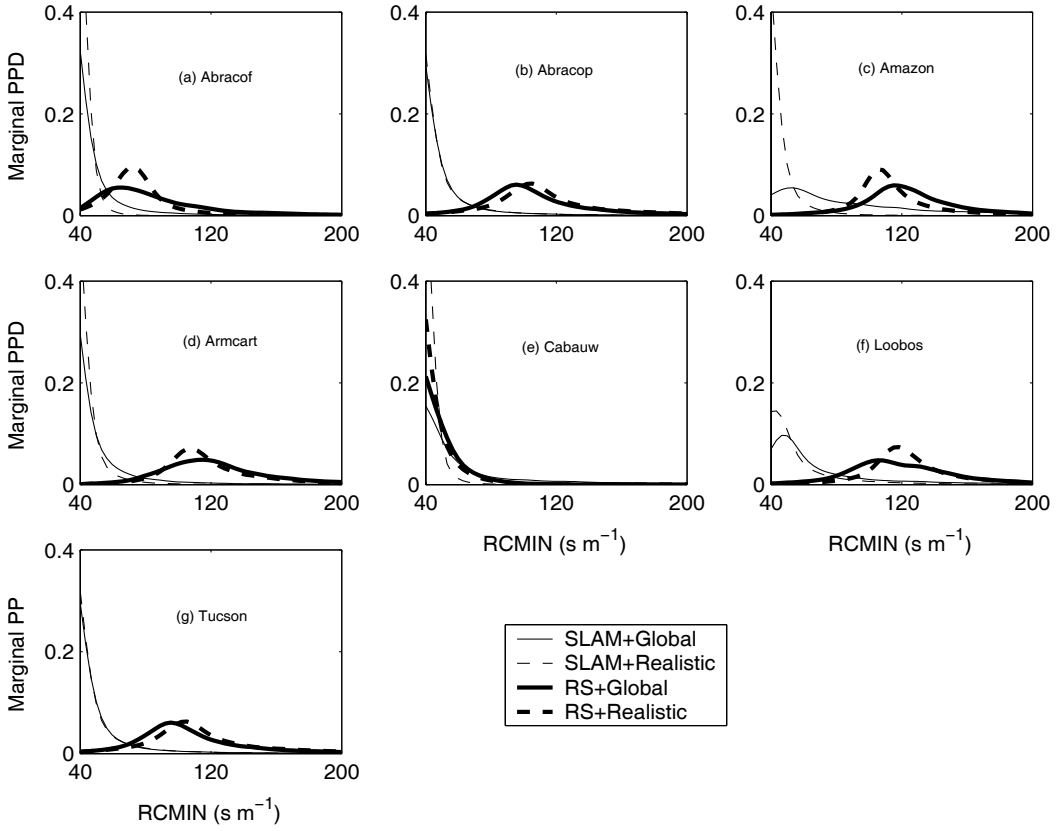

Fig. 10. Marginal posterior probability density function (PPD) at (a) Abracof tropical forest site, (b) Abracop tropical pasture site, (c) Amazon tropical forest site, (d) Armcart mid-latitude crop site, (e) Cabauw mid-latitude grassland site, (f) Loobos pine forest site, and (g) Tucson semi-arid site (thin solid line=SLAM+global range; thin dashed line=SLAM+realistic range; thick solid line=RS+global range; and thick dashed line=RS+realistic ranges. An arithmetic average of error functions for sensible and latent heat fluxes is used in this study).

and that there is a chance that an unrealistic global optimum will be found when a wide search space is specified because of multiple uncertainty sources. In the land surface research community, it is a common practice that global parameter ranges and calibration algorithms are used to search for optimal parameters. The use of global ranges can result in some unrealistic physical parameters. The reasons for choosing global ranges are as follows. First, the model parameters are usually dependent on climate, vegetation type, and soil type so that the ranges are difficult to define accurately for most model parameters. Second, researchers tend to consider that a small range, although safer, may miss a "true" global optimum that a calibration run should be targeted at. The selection process of parameter ranges is sometimes a trail-and-error process and there is a trade-off between the risk of missing a global optimum and the risk of unrealistic physical parameters. As suggested in this study, a cautious and careful selection of physical parameter ranges by reviewing a wide range of references and reports may be an important step before a land surface scheme is calibrated. The reason is that only a global optimum in a realistic model parameter space is meaningful, particularly when we want to use these calibrated model parameters to drive general circulation models as done by Sen et al. (2001).

It should be noted that the impacts of parameter ranges and model complexities on optimal parameter and uncertainty estimates are the smallest for the Cabauw site in most cases when compared to the other sites. This may be a result of the interaction between the data quality of atmospheric forcing and energy fluxes, site vegetation and climate. As suggested by Beljaars and Bosveld (1997), the atmospheric forcing data are well controlled and checked. Few missing values exist at Cabauw for energy flux observations. At the other data sites, 30%–70% of observed energy fluxes are missing, and thus only part of the energy flux observations can be used for calibration. There-

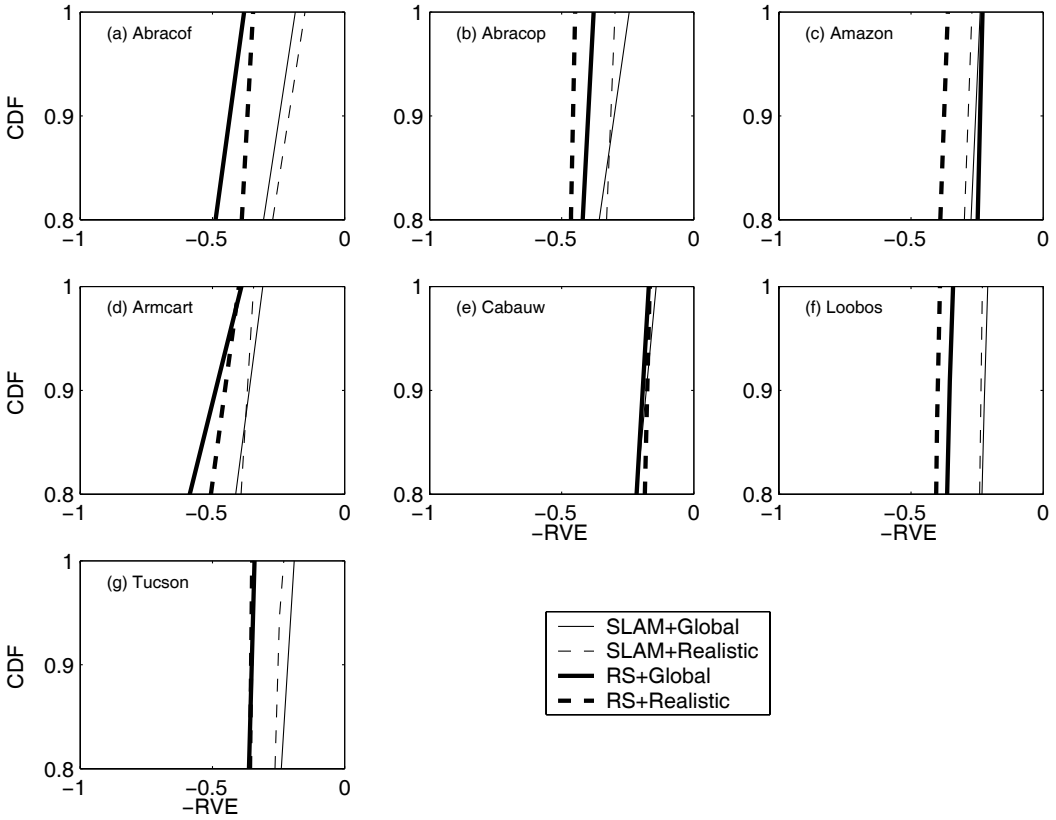

Fig. 11. Same as Fig. 10 but for empirical cumulative distribution functions (CDF) of error functions (an arithmetic average of error functions for sensible and latent heat fluxes) for the top 10%–20% of the results simulated with all parameter sets.

fore, small calibration samples are more sensitive to model parameter ranges and model complexities. In addition, the Cabauw vegetation and climate well constrain the sensitivity of model complexities. The PPD distribution of minimum stomatal resistance is a good example. As shown in Fig. 10, the PPD of minimum stomatal resistance is significantly sensitive to model complexities for all sites except for Cabauw.

7. Conclusions

The analysis presented in this paper demonstrates that the use of parameter ranges and complexities of the land surface model has a significant impact on the selection of optimal parameters, frequency distribution of the parameters, the marginal posterior probability density, and estimates of uncertainty of simulated sensible and latent heat fluxes. Care must be taken when referencing previous work where the global ranges or approximate global ranges were used because some unreasonable optimal parameter values might result. Although this analysis does not change the conclusion of Xia et al. (2002) that a complex land surface model performed better than a simple model when optimal parameters were used, it is unknown whether these unreasonable optimal values have an impact on other previous optimization studies. For uncertainty estimates, it is obvious that Jackson et al. (2003) and Xia et al. (2004a) may have amplified uncertainty estimates of the CHASM model parameters because they used an approximate global range. This study also suggests that realistic rather than global ranges should be used in the intercomparison of land surface models in the ongoing optimization studies such as PILPS San Pedro (see www.sahra.arizona.edu/pilpssanpedro), in the estimation of parameter uncertainties in offline simulations, and in the search for optimal parameters for GCMs.

Model complexities show larger effects on model optimization and uncertainty analysis, which is not surprising. It is well known that any model, regardless of its degree of complexity, is a simplification of the real world. However, insufficient model complexities would result in large uncertainties in both optimal parameters and heat fluxes, as shown in this study. Although model parameters can compensate for some

deficiencies in model complexities (Xia et al., 2002), this compensation is not enough so that simulations of energy fluxes using a simple model with insufficient complexity still has larger simulation errors than those from a complex model, even though optimal parameters are used.

Acknowledgments. The authors was supported by the G. Unger Vetlesen Foundation and the Institute for Geophysics of University of Texas at Austin. It was also supported under NASA Grant NAG5-12577 and NOAA Grant NA03OAR4310076. Special thanks go to Dr. Andy Pitman for providing the CHASM. We acknowledge the Royal Netherlands Meteorological Institute for providing the Cabauw dataset. The authors wishes to thank one reviewer whose comments greatly improved the quality of this manuscript.

REFERENCES

Bastidas, L. A., H. V. Gupta, O. L. Sen, Y. Liu, S. Sorooshian, and W. J. Shuttleworth, 2004: Comparative evaluation of land surface models using multicriteria methods. *J. Geophys. Res.* (in press).

Beljaars, A. C. M., and F. Bosveld, 1997: Cabauw data for the validation of land surface parameterization schemes. *J. Climate*, **10**, 1172–1193.

Bonan, G. B., 1996: A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide. NCAR Tech. Note NCAR/TN-417, 150 pp.

Bowling, L. C., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e), 1, Experiment description and summary intercomparisons. *Global Planet. Change*, **38**, 1–30.

Cash, J. H. C., and C. A. Nobre, 1997: Some results from ABRACOS. *Bull. Amer. Meteor. Soc.*, **78**, 823–828.

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

Dai, Y., and Coauthors, 2003: The Common Land Model. *Bull. Amer. Meteor. Soc.*, **84**, 1013–1023.

Desborough, C. E., 1999: Surface energy balance complexity in GCM land surface models. *Climate Dyn.*, **15**, 389–403.

Desborough, C. E., A. J. Pitman, and B. McAvaney, 2001: Surface energy balance complexity in GCM land surface models, Part II: Coupled simulations. *Climate Dyn.*, **17**, 615–626.

Dirmeyer, P. A., A. J. Dolman, and N. Sato, 1999: The Global Soil Wetness Project: A pilot project for global land surface modeling and validation. *Bull. Amer. Meteor. Soc.*, **80**, 851–878.

Dorman, J. L., and P. J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the simple biosphere model (SiB). *J. Appl. Meteor.*, **28**, 833–855.

Franks, S. W., and K. J. Bven, 1997: Bayesian estimation of uncertainty in land surface-atmosphere flux predictions. *J. Geophys. Res.*, **102**, 23 991–23 999.

Gulf, A. D., G. Fisch, and M. G. Hodnett, 1995: The albedo of Amazonian forest and ranchland. *J. Climate*, **8**, 1544–1554.

Gulf, A. D., J. L. Esteves, A. de O. Marques Filho, and H. R. da Rocha, 1996: Radiation, temperature and humidity over forest and pasture in Amazonia. *Amazon Deforestation and Climate*, Cash et al., Eds., John Wiley, 175–191.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, S., W. J. Shuttleworth, and Z.-L. Yang, 1999: Parameter estimation of a land surface scheme using multicriteria methods. *J. Geophys. Res.*, **104**, 19491–19503.

Henderson-Sellers, A., 1993: A factorial assessment of the sensitivity of the BATS land-surface parameterization scheme. *J. Climate*, **6**, 227–247.

Henderson-Sellers, A., 1996: Soil moisture simulation: Achievements of the RICE and PILPS intercomparison workshop and future directions. *Global Planet. Change*, **13**, 99–115.

Henderson-Sellers, A., M. F. Wilson, G. Thomas, and R. E. Dickinson, 1986: Current Global and Surface Data Sets for Use in Climate-Related Studies, NCAR Tech. Note 3 272-STR, 240pp.

Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489–503.

Henderson-Sellers, A., K. McGuffie, and A. Pitman, 1996: The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): 1992–1995. *Climate Dyn.*, **12**, 849–859.

Jackson, C., Y. Xia, M. K. Sen, and P. Stoffa, 2003: Optimal parameter estimation and uncertainty analysis of a land surface model: A case study from Cabauw, Netherlands. *J. Geophys. Res.*, **108**, 4583, doi:10.1029/2002JD002991.

Jackson, C., M. Sen, and P. Stoffa, 2004: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Climate*, **17**, 2828–2841.

Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.*, **97**, 2697–2715.

Manabe, S., 1969: Climate and the ocean circulation. 1. The atmospheric circulation and the hydrology of the earth's surface. *Mon. Wea. Rev.*, **97**, 739–805.

Oke, T. R., 1987: *Boundary Layer Climates.* 2nd ed., Methuen, New York, 435pp.

Pitman, A., 1994: Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model. *J. Climate*, **7**, 1856–1869.

Pitman, A. J., Y. Xia, M. Leplastrier, and A. Henderson-Sellers, 2003: The CHAmeleon Surface Model (CHASM): Description and use with the PILPS Phase 2e forcing data. *Global Planet. Change*, **38**, 121–135.

Schlosser, C. A., and Coauthors, 2000: Simulation of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d). *Mon. Wea. Rev.*, **128**, 301–321.

Sellers, P. J., and Coauthors, 1989: Calibrating the simple biosphere model for Amazonian tropical forest using field and remote-sensing data. 1. Average calibration with field data. *J. Appl. Meteor.*, **28**, 727–759.

Sellers, P. J., and Coauthors, 1996a: A revised land surface parameterization (SiB2) for atmospheric GCMs. 1, Model formulation. *J. Climate*, **9**, 676–705.

Sellers, P. J., and Coauthors, 1996b: A revised land surface parameterization (SiB2) for atmospheric GCMs. 2, The generation of global fields of terrestrial biophysical parameters from satellite data. *J. Climate*, **9**, 706–737.

Sellers, P. J., and Coauthors, 1997: Modeling the exchanges of energy, water and carbon between continents and the atmosphere. *Science*, **275**, 502–509.

Sen, M. K., and P. L. Stoffa, 1995: *Global Optimization Methods in Geophysical Inversion*. Elsevier, Amsterdam, 281pp.

Sen, M. K., and P. L. Stoffa, 1996: Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophysical Prospecting*, **44**, 313–350.

Sen, O. L., L. A. Bastidas, W. J. Shuttleworth, Z.-L. Yang, H. V. Gupta, and S. Sorooshian, 2001: Impact of field-calibrated vegetation parameters on GCM climate simulations. *Quart. J. Roy. Meteor. Soc.*, **127**(B), 1199–1223.

Shao, Y. P., and A. Henderson-Sellers, 1996: Validation of soil moisture simulation in landsurface parameterization schemes with HAPEX data. *Global Planet. Change*, **13**, 11–46.

Shuttleworth, W. J., 1984: Observations of radiation exchange above and below Amazonian forest. *Quart. J. Roy. Meteor. Soc.*, **110**, 1163–1169.

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

Unland, H. E., P. R. Houser, W. J. Shuttleworth, and Z. L. Yang, 1996: Surface flux measurement and modeling at a semi-arid Sonoran Desert site. *Agricultural and Forest Meteorology*, **82**, 119–153.

Wood, E. F., and Coauthors, 1998: The project for intercomparison of land-surface parameterization schemes (PILPS) phase 2(c) Red–Arkansas River basin experiment: 1. Experiment description and summary intercomparisons. *Global Planet. Change,* **19**, 115–135.

Wright, I. R. and Coauthors, 1996: Towards a GCM surface parameterization for Amazonia. *Amazon Deforestation and Climate*, Gash et al., Eds., John Wiley, 473–504.

Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas, 2002: Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set. *J. Hydrometeor.*, **3**, 181–194.

Xia, Y., M. K. Sen, C. Jackson, and P. L. Stoffa, 2004a: Multi-dataset study of optimal parameter and uncertainty estimation of a land surface model with Bayesian stochastic inversion and Multicriteria method. *J. Appl. Meteor.*, **43**, 1477–1497.

Xia, Y., P. L. Stoffa, C. Jackson, and M. K. Sen, 2004b: Effect of forcing data errors on calibration and uncertainty estimates of the CHASM model: A multi-dataset study. *Observations, Theory and Modeling of Atmospheric Variability. World Sci. Ser. Meteor. Meter. Asia*, **3**, World Scientific, Singapore, 340–355.

Yang, Z.-L., R. E. Dickinson, W. J. Shuttleworth and M. Shaikh, 1998: Treatment of soil, vegetation and snow in land-surface models: A test of the biosphere-atmosphere transfer scheme with the HAPEX-MOBILHY, ABRACOS, and Russian data. *J. Hydrol.*, **213**, 109–127.

Zhang, H., A. Henderson-Sellers, A. J. Pitman, J. L. McGregor, C. E. Desborough, and J. J. Katzfey, 2001: Limited-area model sensitivity to the complexity of representation of the land surface energy balance. *J. Climate*, **14**, 3965–3986.