An ensemble conditional nonlinear optimal perturbation approach: Formulation and applications to parameter calibration

Xiangjun Tian,[1] Zhenghui Xie,[1] and Aiguo Dai[2]

Received 9 August 2009; revised 6 May 2010; accepted 14 May 2010; published 30 September 2010.

[1] The conditional nonlinear optimal perturbation (CNOP) method proposed by Mu et al. (2003) has been a useful tool for studying predictability dynamics. Its further applications are, however, hampered by the need of an adjoint model. Aiming to solve this problem, an ensemble conditional nonlinear optimal perturbation (EnCNOP) approach is proposed in this paper by merging the Monte Carlo method and the proper orthogonal decomposition (POD) technique into the CNOP to transform an implicit optimization problem into an explicit one. This approach is further formulated for parameter calibration. Numerical experiments with a 1-D model of the soil water equation show that the EnCNOP approach outperforms a dual-pass optimization framework on the basis of the ensemble Kalman filter (EnKF) and a shuffled complex evolution approach (SCE-UA) in terms of both increasing the calibration precision and reducing the computational costs. In particular, the EnCNOP approach outperforms the SCE-UA when only an approximate value range of the input parameters is known. Our EnCNOP method is further implemented within a complicated highly nonlinear land surface model (namely, NCAR CLM3) and evaluated through a case study which illustrates its applicability to complex, highly nonlinear problems.

Citation: Tian, X., Z. Xie, and A. Dai (2010), An ensemble conditional nonlinear optimal perturbation approach: Formulation and applications to parameter calibration, *Water Resour. Res.*, *46*, W09540, doi:10.1029/2009WR008508.

1. Introduction

[2] The conditional nonlinear optimal perturbation (CNOP) method proposed by *Mu et al.* [2003] has been applied to investigate the predictability of El Niño–Southern Oscillation (ENSO) [*Duan et al.*, 2004] and the sensitivity of the ocean thermohaline circulation to finite amplitude perturbations [*Mu et al.*, 2004], as well as the passive variability of the thermohaline circulation [*Sun et al.*, 2005]. These studies show that the CNOP is a useful tool for studying predictability dynamics, mainly because it can reveal the effect of nonlinearity on climate predictability. The CNOP mathematically belongs to one kind of constraint optimization in which the control variables (or initial perturbations) are usually expressed implicitly in the cost function. The cost function is used to quantify deviations between model and observations, and the naming of control variables means the variables to be assimilated, calibrated, or optimized. To compute the gradient of the cost function with respect to the control variables, one has to integrate an adjoint model, whose development and maintenance require significant resources, especially when the forecast model is highly nonlinear and the model physics contain parameterized discontinuities [*Xu*, 1996; *Mu and Wang*, 2003]. Because of this, CNOP has been applied mainly in theoretical studies using simple models so far. The need for an adjoint model has become a major roadblock for broad applications of CNOP, and how to retain the primary advantage of the CNOP while avoiding the need of the adjoint model has become a critical issue in further applications of the CNOP. In fact, similar problems also occur in the four-dimensional data assimilation (4DVar) [*Kalnay et al.*, 2007; *Tsuyuki and Miyoshi*, 2007; *Gauthier et al.*, 2007], and ensemble-based strategies have been applied to remove the need of an adjoint model [e.g., *Tian et al.*, 2008a].

[3] Aiming to solve this problem for the CNOP, an ensemble conditional nonlinear optimal perturbation (EnCNOP) approach is proposed here using the Monte Carlo method and the proper orthogonal decomposition (POD) technique [*Ly and Tran*, 2001, 2002; *Blum et al.*, 2009; S. Volkwein, Model reduction using proper orthogonal decomposition, unpublished manuscript, 2008, available at http://www.uni-graz.at/imawww/volkwein/publist.html]. It begins with a 4-D ensemble obtained from forecast ensembles produced using the Monte Carlo method. We then apply the POD technique to this 4-D forecast ensemble, so that the orthogonal base vectors not only can capture the spatial structure of the state but also can reflect its temporal evolution. After the model status being represented by a truncated expansion of the base vectors obtained using the POD technique, the control variables in the cost function appear explicitly, so that the adjoint model is no longer needed.

[4] As pointed out by *Mu et al.* [2002], the CNOP also can potentially be used for parameter (or model) calibration, which is to adjust parameters in an already existing model using available observations [*Goldberg and Kaczka*, 1979]. Because of the importance of model parameters, many efforts

[1]ICCES/LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.
[2]National Center for Atmospheric Research, Boulder, Colorado, USA.

0043-1397/10/2009WR008508

have been devoted to developing calibration methods [e.g., *Duan et al.*, 1993, 1994; *Boyle et al.*, 2000; *Cheng et al.*, 2002; *Cooper et al.*, 2007; *Yapo et al.*, 1998]. A global optimization method known as the SCE-UA (shuffled complex evolution approach) developed by *Duan et al.* [1993, 1994] has shown promise as an effective and efficient optimization technique for calibrating watershed models. An implicit assumption in common parameter calibrations is that the state errors only result from the inaccurate model parameter values at the time when the parameter calibration is conducted, but in fact the state errors could propagate with model integration. Therefore, it can be argued that model parameter calibration has been conducted without considering the state error propagation in most previous studies, partly because of a lack of time-dependent observations. We have addressed this issue using a dual-pass optimization framework [e.g., *Tian and Xie*, 2008; *Tian et al.*, 2008b, 2009; see also *Hendricks Franssen and Kinzelbach*, 2008; *Vrugt et al.*, 2005], which usually consists of two usual optimization passes (one for the state variables and the other for the parameters) in contrast to only one pass in other common optimization algorithms or assimilation methods [*Duan et al.*, 1994; *Evensen*, 2004]. The two passes simultaneously optimize both model states and parameters using observational information, thereby leading to improved simulations. Using an ensemble constructed by perturbing model parameters and states, the EnCNOP approach can be further formulated for parameter calibration, in which the model states and parameters can be optimized simultaneously. Here our focus is on the EnCNOP method itself and its application to model parameter calibration; applications of the EnCNOP to climate predictability problems are not addressed in this study.

[5] We conducted several numerical experiments using a 1-D soil water equation with three key parameters and synthetic data to evaluate our new method for parameter calibration. Our method was compared with the ensemble Kalman filter (EnKF) [e.g., *Evensen*, 2004; *Kalnay et al.*, 2007; *Beezley and Mandel*, 2008], the SCE-UA, and a dual-pass optimization framework based on the EnKF and the SCE-UA [*Jia et al.*, 2009]. We found that our new EnCNOP method performs better than the other three methods in terms of both increasing the calibration (or assimilation) precision and reducing the computational costs. In particular, the EnCNOP approach outperforms the SCE-UA in the case when only an approximate value range of input parameters is known. Further evaluations within a comprehensive land surface model also demonstrate that the EnCNOP can be used to calibrate parameters within highly nonlinear models and produce much improved simulation results.

2. Ensemble Conditional Nonlinear Optimal Perturbation

[6] Let M_τ be the propagator of a nonlinear model from 0 to τ, and u_0 be an initial perturbation superposed on the basic state $U(t)$, which is a solution to the nonlinear model and satisfies $U(t) = M_t(U_0)$, where U_0 is the initial value of U.

[7] For a chosen norm $\|\cdot\|$, an initial perturbation $u_{0\delta}$ is called CNOP, if and only if

$$J(u_{0\delta}) = \max_{\|u_0\| \le \delta} \| M_\tau(U_0 + u_0) - M_\tau(U_0) \|, \qquad (1)$$

where $\|u_0\| \le \delta$ is the constraint condition of initial perturbation defined by the norm $\|\cdot\|$ [*Mu et al.*, 2003].

[8] CNOP is the initial perturbation whose nonlinear evolution attains the maximal value of the function J at time τ. In other words, CNOP leads to the global maximum of the objective function J. The nonlinear constrained optimization problem considered in equation (1) can be easily transformed into a minimization one as follows:

$$J(u_{0\delta}) = \min_{\|u_0\| \le \delta} G(M_\tau(U_0 + u_0) - M_\tau(U_0)), \qquad (2)$$

where G is a cost function. Then CNOP can be computed using a sequential quadratic programming (SQP) solver [*Powell*, 1982], which can solve nonlinear minimization problems with either equality or inequality constraint conditions. Unfortunately, an adjoint model is needed in such solver, and it is difficult to obtain the adjoint model for complex nonlinear systems. Here we make use of the Monte Carlo method and the POD technique to avoid this problem. The POD technique is to start with an ensemble of data, called *snapshots*, collected from an experiment or a numerical procedure of a physical system. This technique is very similar to the Karhunen-Loeve expansion [*Ly and Tran*, 2001]. The POD technique is then used to produce a set of base vectors which span the snapshot collection. The goal is to represent the ensemble of the data in terms of an optimal coordinate system. That is, the snapshots can be generated by a smallest possible set of base vectors. The POD technique has been widely used in the data assimilation community: for example, POD applications in ensembles by *Fang et al.* [2008], *Vermeulen and Heemink* [2006], and *Heemink et al.* [2001]. This technique is also used by *Tian et al.* [2008a]; here we apply it to the CNOP approach. Some formulations of *Tian et al.* [2008a] are repeated here to present a self-contained description of the methodology used here. The basic idea of our approach is to produce an ensemble space (Ω_e) using the Monte Carlo method to approximate the analysis space (Ω_a). The so-called analysis space is usually a subspace of the R^n space (n is the dimensions of the control variables) satisfying some necessary constraints. The Monte Carlo method is often used to approximate the parameter space. In our method, the Monte Carlo method is used to produce random real numbers according to some given distributions (e.g., normal distribution). The POD technique is used to determine an optimal set of base vectors for the construction of the ensemble space Ω_e. The ensemble space is constructed by the Monte Carlo method using random samples to approximate the analysis space. The following is an outline of the EnCNOP approach.

[9] Assuming there are S time steps within the time window $[0, \tau]$, generate N random perturbation fields with the constraint condition $\|u_0\| \le \delta$ using the Monte Carlo method and add each perturbation field to the initial basic field at $t = t_0$ to produce N initial fields $U_{0,n}(n = 1,2,\cdots,N)$. Integrate the forecast model (propagator) $U_n(t) = M_t(U_{0,n})$ with the initial fields $U_{0,n}(n = 1,2,\cdots,N)$ throughout the time window $[0, \tau]$ to obtain the state series $U_n(t_i)$ $(i = 1,\cdots,S)$ and then construct the perturbed 4-D fields $X_n(n = 1,2,\cdots,N)$ over the time window:

$$X_n = (U_n(0), U_n(T)),\ n = 1, 2, \cdots, N. \qquad (3)$$

All the perturbed 4-D fields $X_n(n = 1,2,\cdots,N)$ support the ensemble space Ω_e represented by the matrix $A = (X_1,\cdots,X_N)$.

As the ensemble size N increases with addition of random samples, the ensemble space Ω_e approaches the analysis space Ω_a adequately. Consequently, the analysis $X_a = (U_a(0), U_a(T))$ can be considered approximatively in the ensemble space Ω_e and expressed by the linear combinations of a set of base vectors of Ω_e. The POD technique is then used to seek the appropriate base vectors to span the linear space Ω_e [*Ly and Tran*, 2002]. We form a new ensemble by focusing on the deviations from the mean as follows:

$$\delta X_n = X_n - \overline{X}, \tag{4}$$

which form the matrix $B(M \times N)$, where $\overline{X} = \frac{1}{N}\sum_{n=1}^{N} X_n$, $M = M_g \times M_v \times 2$, and M_g and M_v are the number of the model spatial grid points and the number of the model variables, respectively. To compute the POD modes (base vectors), we solve an $M \times M$ eigenvalue problem

$$BB^T V = \lambda V, \tag{5}$$

where λ is a diagonal matrix composed of the singular values BB^T and V is the orthogonal matrices composed of the singular vectors of BB^T. The POD modes are given by $\phi_k = BV_k/\sqrt{\lambda_k}$, $1 \le k \le r$, where r(the number of the POD modes) is defined as follows:

$$r = \min\left\{ r, I(r) = \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{N}\lambda_i} : I(r) \ge \gamma \right\}, 0 < \gamma \le 1. \tag{6}$$

[10] The truncated reconstruction of analysis vector X_a in the ensemble space Ω_e is given by

$$X_a = \overline{X} + \sum_{k=1}^{r} \alpha_k \phi_k, \tag{7}$$

Substituting equation (7) into equation (2), the control variable becomes $\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_r)$ instead of $X_{a,\delta}$, so the control variable is expressed explicitly in the cost function and the adjoint model is not needed any more.

3. EnCNOP for Parameter Calibration

[11] In this section, we apply the new EnCNOP approach to parameter calibration. Let $M_\tau(U_0, W)$ be the propagator with the parameter W to be calibrated from 0 to τ. The basic state $U(t)$ is a solution to the nonlinear model and satisfies $U(t) = M_t(U_0, W)$, where U_0 is the initial value. The parameter W can be calibrated through

$$J(w_\delta) = \min_{\|w\| \le \delta} \| M_\tau(U_0, W + w) - y \|, \tag{8}$$

where $\|\cdot\|$ is the chosen norm and y is the observation at time τ. The final calibrated parameter $W_\delta = W + w_\delta$.

[12] Within the current EnCNOP framework (8), only the observations at the end of the time window $[0, \tau]$ can be utilized, which is obviously not ideal for parameter calibration. How to appropriately incorporate multisource or time-varying observations into the EnCNOP framework is an important aspect that has not yet been addressed. In addition, satellite observations are playing an increasingly important role in data assimilation and parameter calibration. The so-called observation operator, which establishes a mapping between the forecast states and the observational variables, is urgently needed in using satellite observations. However, it is not considered in the present form of the EnCNOP method. The EnCNOP objective function for parameter calibration that makes good use of multisource and/or time-varying observations and nonlinear observational operators can be further defined as follows:

$$J(w_\delta) = \min_{\|w\| \le \delta} \sum_{i=1}^{m} [H_i(M_{t_i}(U_0, W + w)) - y_i]^T \cdot [H_i(M_{t_i}(U_0, W + w)) - y_i], \tag{9}$$

where H_i is the observation operator and y_i is the observation at time t_i.

[13] Similarly, generate N random parameter perturbations with the constraint condition $\|w\| \le \delta$ using the Monte Carlo method and add each perturbation to the background parameter W to produce N parameter fields W_n. Integrate the forecast model (propagator) $U_n(t) = M_t(U_0, W)$ with the different parameter fields $W_n(n = 1,2,\cdots,N)$ throughout the time window $[0, \tau]$ to obtain the state series $U_n(t_i)(i = 1,\cdots,S)$, and then construct the ensembles $X_n(n = 1,2,\cdots,N)$ using the perturbed parameters and the perturbed states at the observational times:

$$X_n = (W_n, U_n(t_1), \cdots, U_n(t_m)) \tag{10}$$

It should be noted that the EnCNOP for parameter calibration in this section can optimize both the model states and parameters simultaneously because the ensembles in equation (10) consist of both the perturbed parameters and states.

[14] Subsequently, the transformations in equations (4), (5), and (7) are conducted to represent linearly the analysis vector $X_a = (W_a, U_a(t_1),\cdots,U_a(t_m))$ by the set of POD base vectors ϕ_k, $1 \le k \le r$. Substituting equation (7) into equation (9), the control variable becomes $\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_r)$ instead of X_a

$$J(\alpha) = \sum_{j=1}^{m} \left[y_j - \mathrm{H}_j\overline{U}(t_j) - \mathrm{H}_j\Phi_j\alpha\right]^T \left[y_j - \mathrm{H}_j\overline{U}(t_j) - \mathrm{H}_j\Phi_j\alpha\right], \tag{11}$$

where H_j is the tangent linear observation operator and $\Phi_j = (\phi_1(t_j),\phi_2(t_j),\cdots,\phi_r(t_j))$.

[15] The gradient of the cost function in equation (11) is obtained through the following simple calculations:

$$\nabla J(\alpha) = 2\sum_{j=1}^{m} -\left[\mathrm{H}_j\Phi_j\right]^T \left[y_j - \mathrm{H}_j\overline{U}(t_j) - \mathrm{H}_j\Phi_j\alpha\right]. \tag{12}$$

One can solve the optimization problem as follows

$$\nabla J(\alpha) = 0 \tag{13}$$

and

$$\left(\sum_{j=1}^{m} \left[\mathrm{H}_j\Phi_j\right]^T \left[\mathrm{H}_j\Phi_j\right]\right)\alpha = \sum_{j=1}^{m} \left[\mathrm{H}_j\Phi_j\right]^T \left[y_j - \mathrm{H}_j\overline{U}(t_j)\right]. \tag{14}$$

Equation (14) can be solved directly, and the optimized parameters are then obtained.

[16] If the observation operator H_j is highly nonlinear, the linearization of the observation operator H_j could lead to suboptimal performance of the proposed method. To avoid the use of the tangent linear observation operator H_j, we construct the ensembles $X_n(n = 1,2,\cdots,N)$ using the perturbed parameters and the perturbed observational states (instead of the perturbed state forecasts) at the observational times:

$$X_n = (W_n, H_1(U_n(t_1)), \cdots, H_m(U_n(t_m))). \tag{15}$$

The observation operator H_j is then involved in the construction of the ensembles in equation (15), and the linearization of the observation operator is also no longer needed. A similar technique has been successfully implemented in an explicit 4DVar method [*Wang et al.*, 2010] to avoid the introduction of the tangent linear observation operator. For simplicity, we use $X_n = (W_n,V_n(t_1),\cdots,V_n(t_m))$ instead of $X_n = (W_n,H_1(U_n(t_1)),\cdots,H_m(U_n(t_m)))$. Equations (11), (12), and (14) are then transformed into the following:

$$J(\alpha) = \sum_{j=1}^{m} \left[y_j - \overline{V}(t_j) - \Phi_j\alpha\right]^T \left[y_j - \overline{V}(t_j) - \Phi_j\alpha\right], \tag{16}$$

$$\nabla J(\alpha) = 2\sum_{j=1}^{m} -\left[\Phi_j\right]^T \left[y_j - \overline{V}(t_j) - \Phi_j\alpha\right], \tag{17}$$

$$\left(\sum_{j=1}^{m} \left[\Phi_j\right]^T \left[\Phi_j\right]\right)\alpha = \sum_{j=1}^{m} \left[\Phi_j\right]^T \left[y_j - \overline{V}(t_j)\right]. \tag{18}$$

[17] Actually, this kind of both state and parameter techniques has previously been implemented within the Kalman filter framework [e.g., *Hao and Ghil*, 1995; *Navon*, 1998]. Because our method originates from an explicit 4DVar [*Tian et al.*, 2008a], it shares some essential features with the 4DVar method. Therefore, the proposed method is generally a variation-based method, which is the significant difference between our method and the filter-based calibration methods. It should be noted that the parameterized discontinuities are not considered in this method. The significant difference between our method and the SCE-UA method is as follows. The main idea of the SCE-UA method is to search for the final optimized parameters in their value domains by means of complex evolution. On the contrary, in our method, we first obtain the linear base vectors of the ensemble space and then represent the final optimized parameters by the linear combination of the base vectors.

4. Calibration Experiments Using a 1-D Model

[18] In this section, the applicability of the new EnCNOP method in parameter calibration is evaluated through several numerical experiments with a 1-D soil water equation used in the NCAR Community Land Model version 3 (CLM3) [*Oleson et al.*, 2004] with three key parameters to be calibrated. In addition, we also compare calibration and assimilation results using the EnCNOP, EnKF, SCE-UA, and a dual-pass optimization framework on the basis of the EnKF and the SCE-UA (referred to as EnKF+SCE-UA [*Jia et al.*, 2009]). In the EnKF case, the EnKF method is used only in assimilating observational data, and no parameters are calibrated at all. In the SCE-UA case, the corresponding method is used only in calibrating the three parameters with the observational information, and there is no assimilation of observational data. In the EnKF+SCE-UA case, the whole optimization is divided into two passes in one time step: one pass for parameter calibration using the SCE-UA and then another pass for assimilation using the EnKF. The observational information is both utilized in the two passes.

4.1. Setup of Experiments

[19] The 1-D Richards equation used in the CLM3 is expressed as follows:

$$\frac{\partial\theta}{\partial t} = -\frac{\partial q}{\partial z} - E - R_{fm}, \tag{19}$$

where θ is the volumetric soil moisture, q is the vertical soil water flux, E is the roots' evapotranspiration rate from the bare soil or due to uptake, and R_{fm} is the melting (negative) or freezing (positive) rate, and z is the depth from the soil surface. Both q and z are positive downward.

[20] The soil water flux q is described by Darcy's law [*Darcy*, 1856]:

$$q = -k\frac{\partial(\varphi + z)}{\partial z}, \tag{20a}$$

where $k = k_s\left(\frac{\theta}{\theta_s}\right)^{2b+3}$ is the hydraulic conductivity, $\varphi = \varphi_s\left(\frac{\theta}{\theta_s}\right)^{-b}$ is the soil matrix potential, k_s is the saturated hydraulic conductivity, φ_s is the saturated water head, θ_s is the saturated volumetric moisture, and b is the Clapp and Hornberger constant. The parameters vary with local soil texture and other properties. Parameters k_s, φ_s, and b are important for soil moisture simulations. Therefore, the following experiments focus on the calibration of these three parameters. The CLM3 computes soil water content in the 10 soil layers within top 3.43 m depth through equations (19) and (20a) (see *Oleson et al.* [2004] for details). The upper boundary condition is

$$q_0(t) = -k\frac{\partial(\varphi + z)}{\partial z}\bigg|_{z=0}, \tag{20b}$$

where $q_0(t)$ is the water flux at the land surface (referred to as infiltration). The time step Δt is 30 min.

[21] A site in Northwest China (47.43°N, 126.97°E) was chosen as the simulation location. Four sets of parameter values for the site (Table 1) were used for the model parameters with no errors (Set 4, calculated by the CLM3 using the high-resolution soil texture data released with the CLM3 by NCAR) or varying degrees of errors (Sets 1–3). We then ran the CLM3 forced with observation-based 3-hourly forcing data for the site [*Qian et al.*, 2006; *Tian et al.*, 2007] from 1 January 1992, to 31 December 1992 after 10-year spinning-up period (the time period over which the model run can reach an equilibrium) to obtain a 1-year time series of simulated infiltration (i.e., the water flux q at the surface, cf. equation (20b), Figure 1a) for driving the 1-D Richards equation (19). Figure 1b shows the time series of the simulated skin (top 1.75 cm) soil moisture with the four sets of parameters. It is clear that without parameter calibration, the simulated states deviate from the true solution (black line) significantly. We generated synthetic "observations" by

Table 1. Parameters Used in the Soil Water Equation Model (Equations (14) and (15))[a]

Parameters	b	k_s (m s^{-1})	φ_s (m)	θ_s (m^3 m^{-3})
Set 1	**7.465**	2.07263E-6	−3.6779	0.46
Set 2	**7.465**	**2.34586E-6**	−3.6779	0.46
Set 3	**7.465**	**2.34586E-6**	−3.**8177**	0.46
Set 4	8.634	2.07263E-6	−3.6779	0.46
Set 5	4.320	4.**23000E**-6	−0.**1000**	0.46

[a]Set 4 represents the best estimate as it is derived from the high-resolution soil texture data, whereas Sets 1–3 are perturbed values (in bold) from Set 4. Set 5 represents a comprehensively coarse initial value.

adding 1% random errors to the time series of the true solution, and these "observations" were assimilated using the various methods in the calibration or assimilation experiments (but not in the simulation experiments). Only the skin soil moisture from the synthetic "observations" was used because satellite observations only measure soil moisture in the top few centimeters. The frequency for calibrating (assimilating) the synthetic "observations" is once per day. The ensemble size used in the EnCNOP and EnKF methods is 60 (through sensitivity experiments). The L^2 norm is chosen in this study. We used $\gamma = 0.95$ in our experiments.

4.2. Experimental Results

[22] To evaluate the performance of the three parameter calibration algorithms (EnCNOP, SCE-UA and EnKF+SCE-UA), a relative error (Ew) for the calibrated parameter (W) is used, which is defined as follows:

$$E_w = \frac{W_c - W_t}{W_t} \times 100\%, \tag{21}$$

where the index c denotes the calibrated parameter value and t represents the true parameter value. The RMS errors of the optimized soil moisture are also used to evaluate the four optimization methods (EnCNOP, EnKF, SCE-UA, and EnKF+SCE-UA).

[23] Figure 2 shows that the EnCNOP performs better than the SCE-UA and EnKF+SCE-UA methods in experiments using Set 3 parameters. The relative errors of the calibrated parameter b for the EnCNOP method are between −5% and 5%, which are substantially lower than the other two methods (SCE-UA and EnKF+SCE-UA) (Figure 2a), for which the relative errors range from −20% to −5%. For parameter k_s (Figure 2b), the relative errors are mostly within −8% to 4% for the EnCNOP, better than the −20% to −5% error range for the SCE-UA and EnKF+SCE-UA methods. The results are similar for parameter φ_s (Figure 2c). Furthermore, Figure 2 shows that the calibrated parameters by the SCE-UA and EnKF+SCE-UA fluctuate much more acutely than the EnCNOP, which suggests that there are more random noises in the calibration by the other two methods.

[24] The time series of the RMS errors of the optimized soil moisture using the four calibration or assimilation methods (calibration only: SCE-UA; assimilation only: EnKF; and calibration + assimilation: EnKF+SCE-UA, EnCNOP) during the calibration or assimilation process are shown in Figure 3a. The EnKF performs worst of the four methods, with a RMS error significantly higher than the other methods. This is expected because the EnKF can only adjust the forecast state using the observations sequentially. As a result, the systematic errors resulted from the inaccurate parameters accumulated with model integration, which introduces large errors in variable estimation and leads to suboptimal performance. Figure 3a also shows that the RMS errors for the EnCNOP are the smallest, outperforming the other methods considerably. Presumably, this superior performance results from the simultaneous optimization of the model states and parameters in the EnCNOP.

[25] Figure 3b shows the time series of RMS errors of the soil moisture simulated by the 1-D Richards model with the

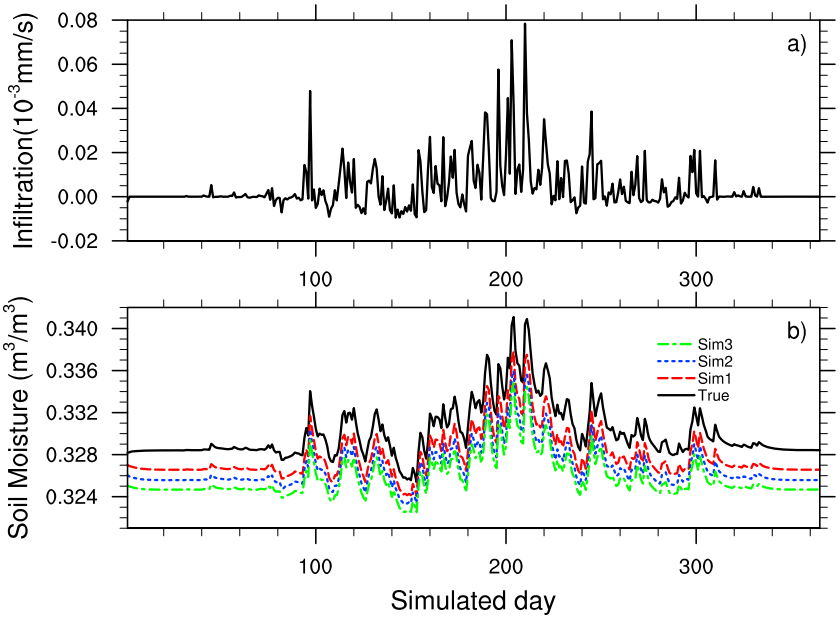

Figure 1. (a) The CLM-simulated infiltration used in the calibration experiments from 1 January 1992 (Day 1) to 31 December 1992. (b) Time series of the simulated skin soil moisture (volumetric soil moisture) with the four sets of parameters.

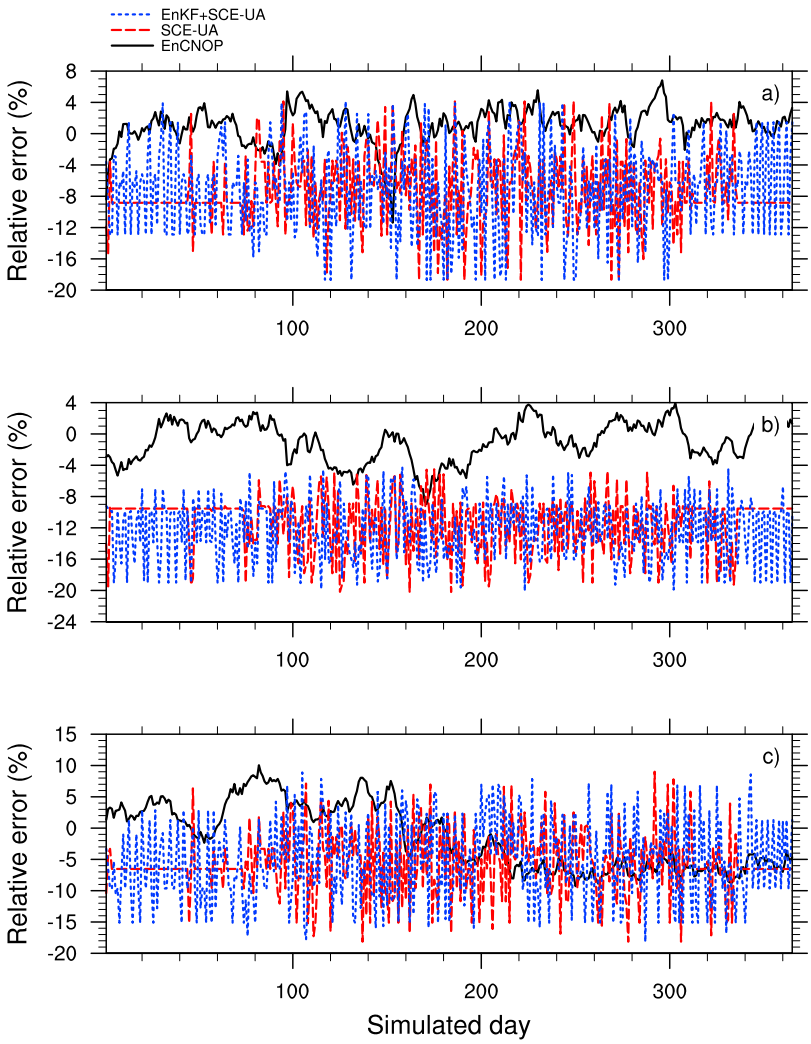

Figure 2. Relative error (E_w) for the calibrated parameters (a) b, (b) k_s, and (c) φ_s in the calibration experiments using the 1-D Richards model by the three calibration methods (EnCNOP, SCE-UA, and EnKF+SCE-UA).

time-averaged calibrated parameters derived by the three calibration methods. The RMS errors for the three methods (EnCNOP, SCE-UA, and EnKF+SCE-UA) are all very small (<0.0012 $m^3\ m^{-3}$), suggesting that these methods are all fairly efficient and effective for parameter calibration. Nevertheless, the RMS errors for the EnCNOP are still lower than the other two methods. The above results suggest that the EnCNOP method improves parameter calibration, with a performance superior to the other methods compared here. The calibration results also indicate that the EnKF+SCE-UA method outperforms the SCE-UA method, which further supports the argument to simultaneously optimize the model states and parameters.

[26] To investigate the effects of a changing number of parameters to be calibrated on the final calibration results of the EnCNOP method, we designed another group of experiments by taking the 1-D Richards equation with the former three sets of parameters in Table 1 as the forecast model, which is used to simulate soil moisture. Comparisons of the time series of RMS errors (Figure 4a) reveal that the changing number of calibration parameters has little effect on the final EnCNOP calibration results, except that the calibration results almost approach perfection with very small errors (Figure 4b) for the one parameter case. Compared with the EnCNOP, the EnKF+SCE-UA calibration results are more likely to be affected by the changing number of calibration parameters (Figure 5).

[27] Usually, the parameter value ranges can only be roughly estimated. Consequently, the sensitivity of the calibration to the constraint imposed by a chosen parameter value range becomes a very important issue. Another group of numerical experiments was conducted here to test the sensitivity of the EnCNOP to varying parameter value ranges. Two parameter value ranges (δ_S = (7.634 to 9.634, 1.32E-6 to 2.82E-6, −4.42 to −2.92) and δ_D = (6.634 to 10.634, 0.57E-6 to 3.57E-6, −5.17 to −2.17)) are adopted in this study. The value range δ_S represents a relatively accurate constraint condition of parameter perturbation, whereas the value range δ_D denotes a rough estimate of parameter perturbation. For simplicity, only the EnKF+SCE-UA method was included in this experiment for comparing with the EnCNOP approach. Figure 6 shows the relative errors for the three calibrated parameter for the EnCNOP and EnKF+SCE-UA methods. Clearly, the calibration results of the EnCNOP are not sen-

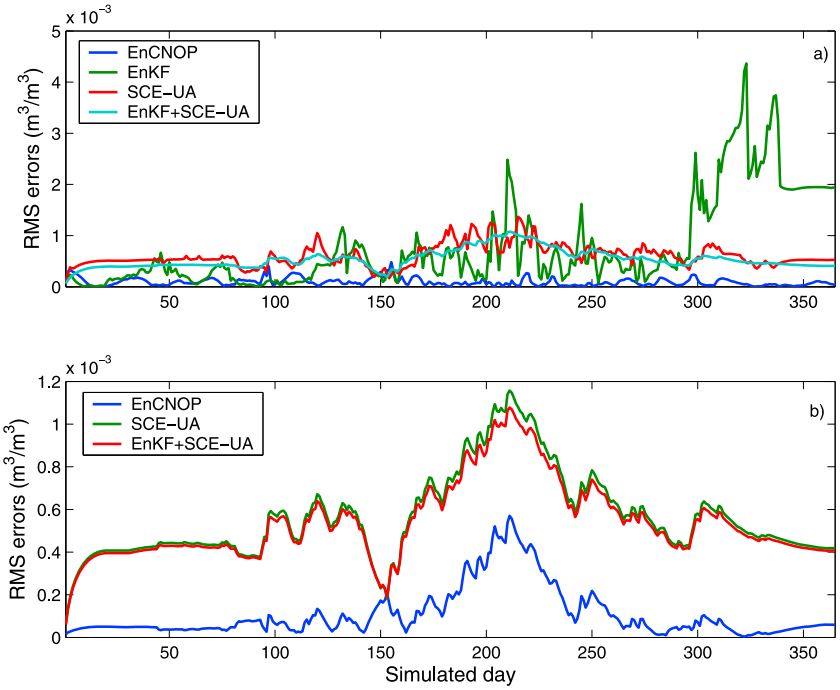

Figure 3. Time series of RMS errors of the optimized soil moisture (a) for the four methods (EnCNOP, EnKF, SCE-UA, and EnKF+SCE-UA) during the calibration (EnCNOP, SCE-UA, and EnKF+SCE-UA) or assimilation (EnKF) process. (b) Simulated by the 1-D Richards model with the time-averaged calibrated parameters derived by the three calibration methods (EnCNOP, SCE-UA, and EnKF+SCE-UA).

sitive to the imposed parameter value range, whereas results from the EnKF+SCE-UA method change substantially with the parameter value range (Figure 6). Since the EnCNOP is hardly affected by the parameter value ranges, its RMS errors change little with the different parameter value ranges (Figure 7), suggesting that the EnCNOP approach should perform well even if only a rough parameter value range can be obtained. On the other hand, Figure 7 shows that different

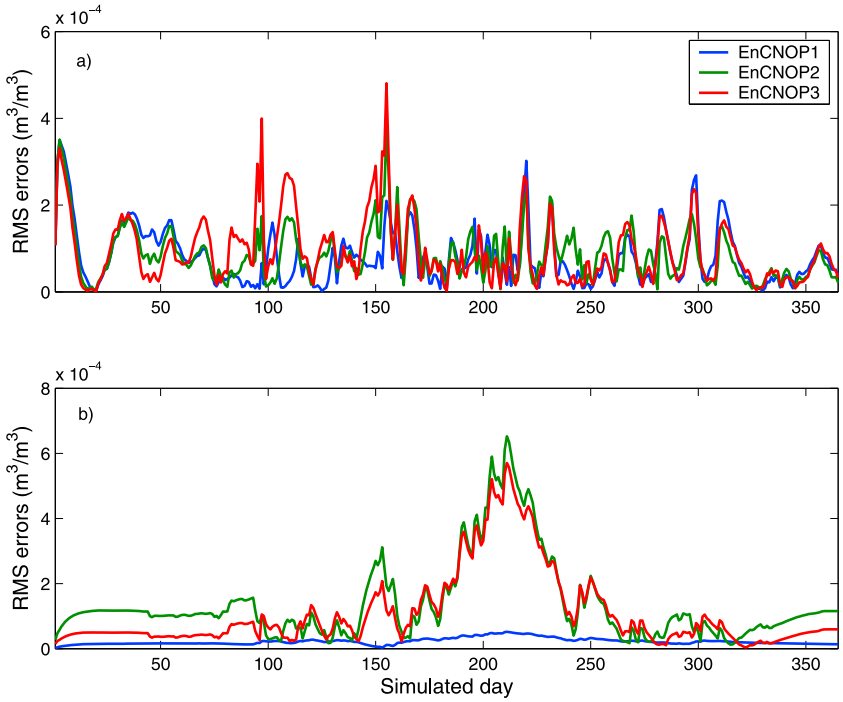

Figure 4. Time series of RMS error (m^3/m^3) of the optimized soil moisture (a) for the EnCNOP with one, two, and three parameters to be calibrated during the calibration process, respectively; (b) simulated by 1-D Richards model with the time-averaged calibrated parameters derived by the EnCNOP approach.

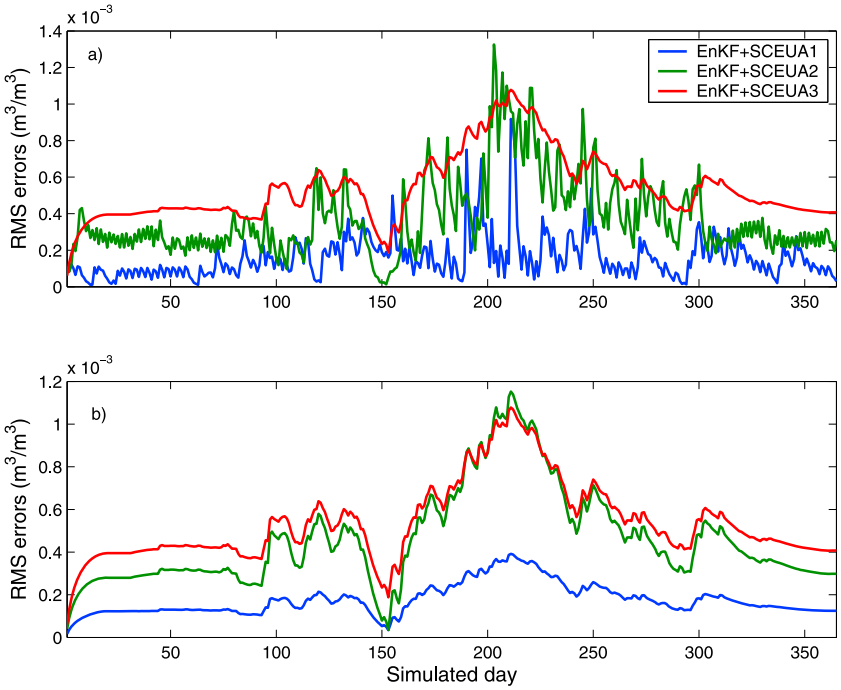

Figure 5. Same as Figure 4, but for the EnKF+SCE-UA.

parameter value ranges affect the final calibration results of the EnKF+SCE-UA method so much that any misestimate of the parameter value range can lead to large increases in the RMS errors for the optimized soil moisture. Additionally, to test its robust applications in the wider parameter ranges, we also used a comprehensively coarse initial value (Set 5 in Table 1) and a significantly rough value ranges (δ_W = 1 to 10, 1.0E-7 to 1.0E-5, −8.0 to −0.05) in this group of experiments. In this case (EnCNOP-W), the EnCNOP performs almost same as the EnCNOP-D case (Figure 7), which further illustrates that the EnCNOP suffers little from the parameter value ranges.

[28] For the experiments discussed above, the ratio of the computational costs for the four calibration (assimilation) methods (EnCNOP, EnKF, SCE-UA and EnKF+SCE-UA) is ~1:4:20:30. A large number of iterations in the SCE-UA and EnKF+SCE-UA optimization resulted in their high computational costs. These ratios may change for other case studies because the minimization of the cost function and the iteration time during each calibration vary greatly within different numerical models. Nevertheless, it at least shows that the computational cost of the EnCNOP should be comparable to, if not much lower than, the other methods.

5. Evaluation Using the CLM3

5.1. CLM3 Experiments

[29] In this section, our EnCNOP method is further implemented within the comprehensive land surface model CLM3 [*Oleson et al.*, 2004] and evaluated through a case study at a reference site in Inner Mongolia (112.5°E, 40.5°N), where in situ observations of soil moisture data and satellite Advanced Microwave Scanning Radiometer for EOS (AMSR-E) brightness temperature (Tb) data for 2006 are available (see *Tian et al.* [2009], for details). The CLM3 is a land surface model designed for use in coupled climate models. In the CLM3, a nested subgrid hierarchy in which grid cells are composed of multiple land units, snow/soil columns, and plant functional types (PFTs) is used to represent the spatial heterogeneity of land surface. Each grid cell can have a different number of columns, and each column can have multiple PFTs. The PFTs, which differ in their ecological and hydrological characteristics, are used to represent the vegetation cover. The grid averaged atmospheric forcing is used to force all subgrid units within a grid cell. Biophysical processes simulated by the CLM3 include solar and long-wave radiation interactions with vegetation canopy and soil; momentum and turbulent fluxes from canopy and soil; heat transfer in soil and snow; hydrology in the canopy, soil, and snow; and stomatal physiology and photosynthesis. These processes are simulated for each subgrid land unit, column, and PFT independently, and each subgrid land unit maintains its own state variables. As in section 4, only the same three parameters, k_s, φ_s and b, are calibrated here, even though the CLM3 has many other parameters. The CLM3 experiments represent an application to a real-world case in contrast to the theoretical 1-D model case presented in section 4.

[30] We performed two CLM3 experiments: one used the CLM3 default values for the three soil parameters (uncalibrated case: $k_s = 6.87 \times 10^{-6}$ *m/s*, $\varphi_s = -1.35$ *m*, and $b = 6.25$) and one with calibrated values for the three parameters (calibrated case). In the uncalibrated case, the CLM3 with the three uncalibrated parameters was forced with observed atmospheric forcing to simulate soil moisture content. In contrast, the calibrated case was divided into two steps: First, the CLM3 was forced with the atmospheric forcing to simulate the soil moisture content, and the satellite Tb data were used to calibrate the three parameters by the EnCNOP. In this step, we used the Q-h radiative transfer model (RTM) from *Wang and Choudhury* [1981] as the observation operator to simulate the Tb from CLM3-simulated land surface variables

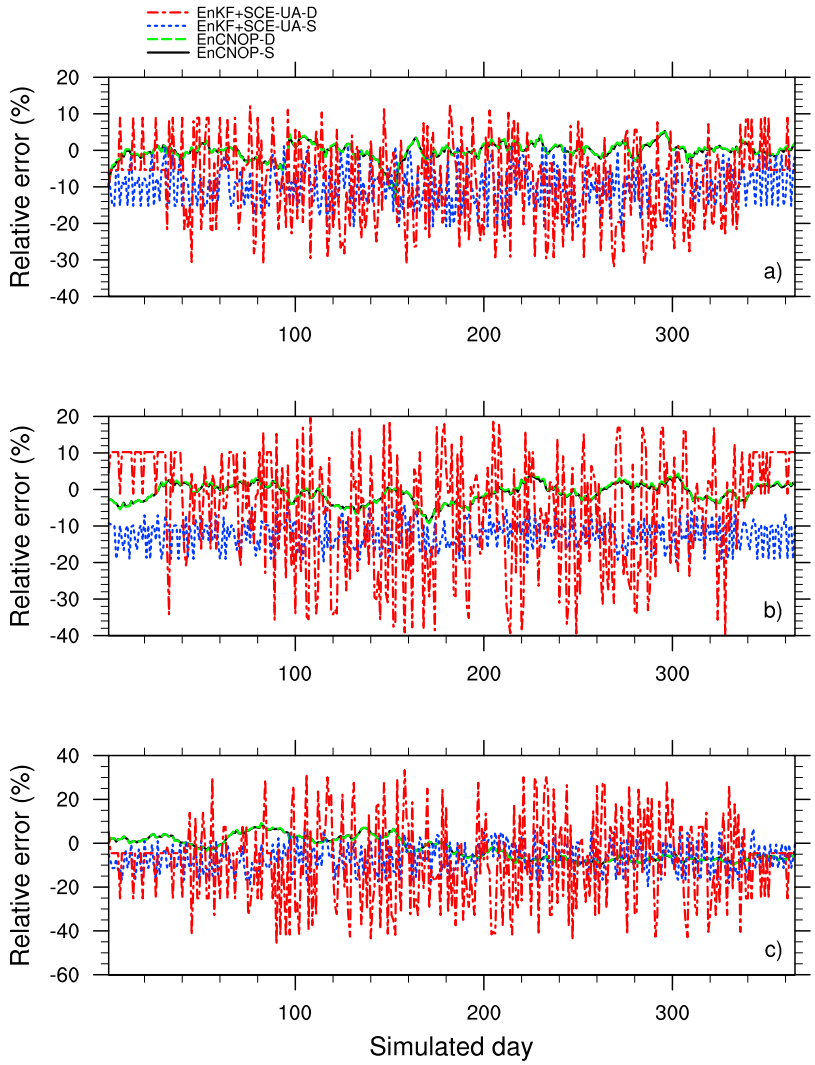

Figure 6. Relative error (E_w) for the calibrated parameters (a) b, (b) k_s, and (c) φ_s in the calibration experiments using the 1-D Richards model by the three calibration methods (EnCNOP, SCE-UA, and EnKF+SCE-UA) with the two different constraint conditions discussed in the paper.

and then compared it with the satellite observed Tb (see *Tian et al.* [2009] for details). A fairly wide initial value range (k_s: $1.0 \sim 12.0 \times 10^{-6}$ *m/s*, φ_s: $-0.5 \sim -10.0$ *m*, and b: $0.5 \sim 15.0$) was used in the experiment. Second (after the calibration), the CLM3 with the three calibrated parameters was run with the same atmospheric forcing to obtain the final simulated soil moisture content without using the satellite Tb data again. This differs from the assimilation case of *Tian et al.* [2009] in which the Tb data were used in both the calibration and assimilation phases.

5.2. Experimental Results

[31] Figure 8 shows time series of the three calibrated parameters during the parameter calibration period. The values of the calibrated parameter b generally fluctuate in magnitude between 5 and 7. This scope is greatly less than its initial guessed range. Similar results are also found for the other two parameters (Figures 8a and 8b). It indicates that the calibration narrows the parameters to a relatively small range compared with their initial guessed parameter range. This should help stabilize the solution of the proposed method. Unfortunately, there is little information about the accurate values of the three parameters. We have to use an alternative way to evaluate our method indirectly. Figure 9 shows time series of daily volumetric soil moisture from 1 January to 31 December 2006 from observations and the CLM3 simulations with the three calibrated parameters ($k_s = 3.13 \times 10^{-6}$ *m/s*, $\varphi_s = -1.26$ *m*, and $b = 6.42$.) and the uncalibrated parameters ($k_s = 6.87 \times 10^{-6}$ *m/s*, $\varphi_s = -1.35$ *m*, and $b = 6.25$), respectively. The simulated soil moisture using the calibrated parameters is capable of reproducing the temporal evolution of the observed soil moisture in terms of both its amplitude and seasonal phase. It is encouraging that, to a considerable extent, this soil moisture improvement further propagates to lower layers where satellite observations are unavailable and thus not used in the calibration. Since the Q-h RTM is highly nonlinear [*Tian et al.*, 2009], the encouraging results reported here demonstrate the EnCNOP's applicability to highly nonlinear observation operators in parameter calibration. We recognize that the improvements shown in Figure 9 are due not only to the use of the EnCNOP method but also to the use

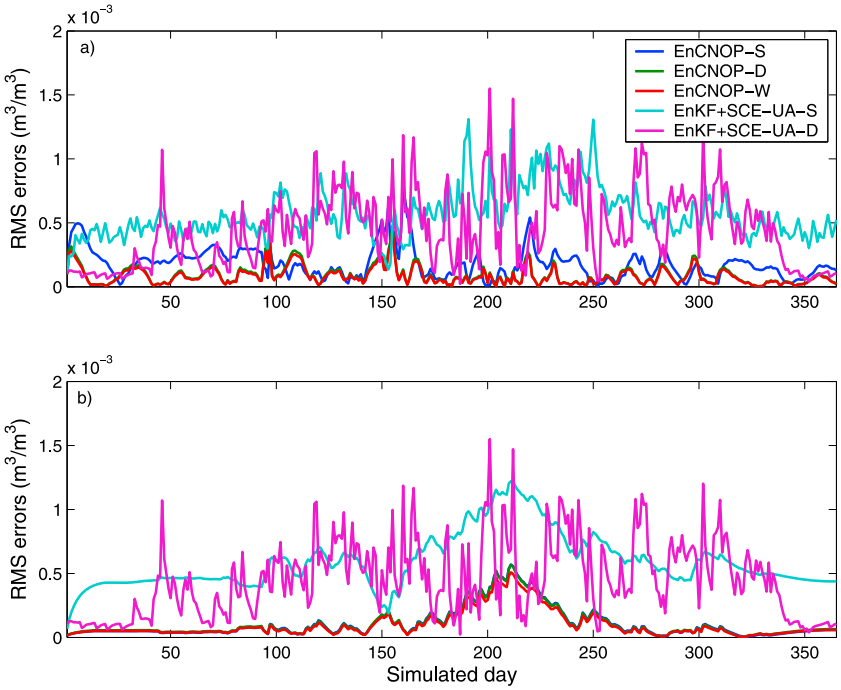

Figure 7. Time series of RMS error (10^{-3}m^3/m^3) of the optimized soil moisture (a) for the two methods (EnCNOP and EnKF+SCE-UA) with the three different constraint conditions during the calibration process; (b) simulated by the 1-D Richards model with the time-averaged calibrated parameters derived by the two same approaches with the three different constraint conditions.

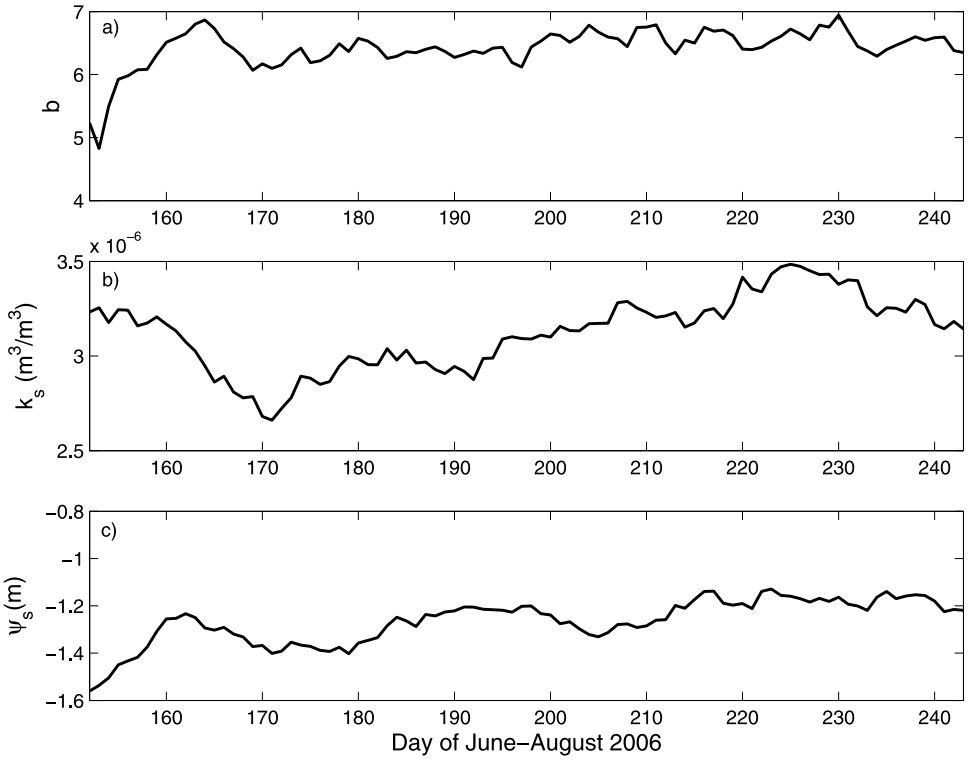

Figure 8. Time series of the three calibrated parameters (a) b; (b) k_s; (c) φ_s.

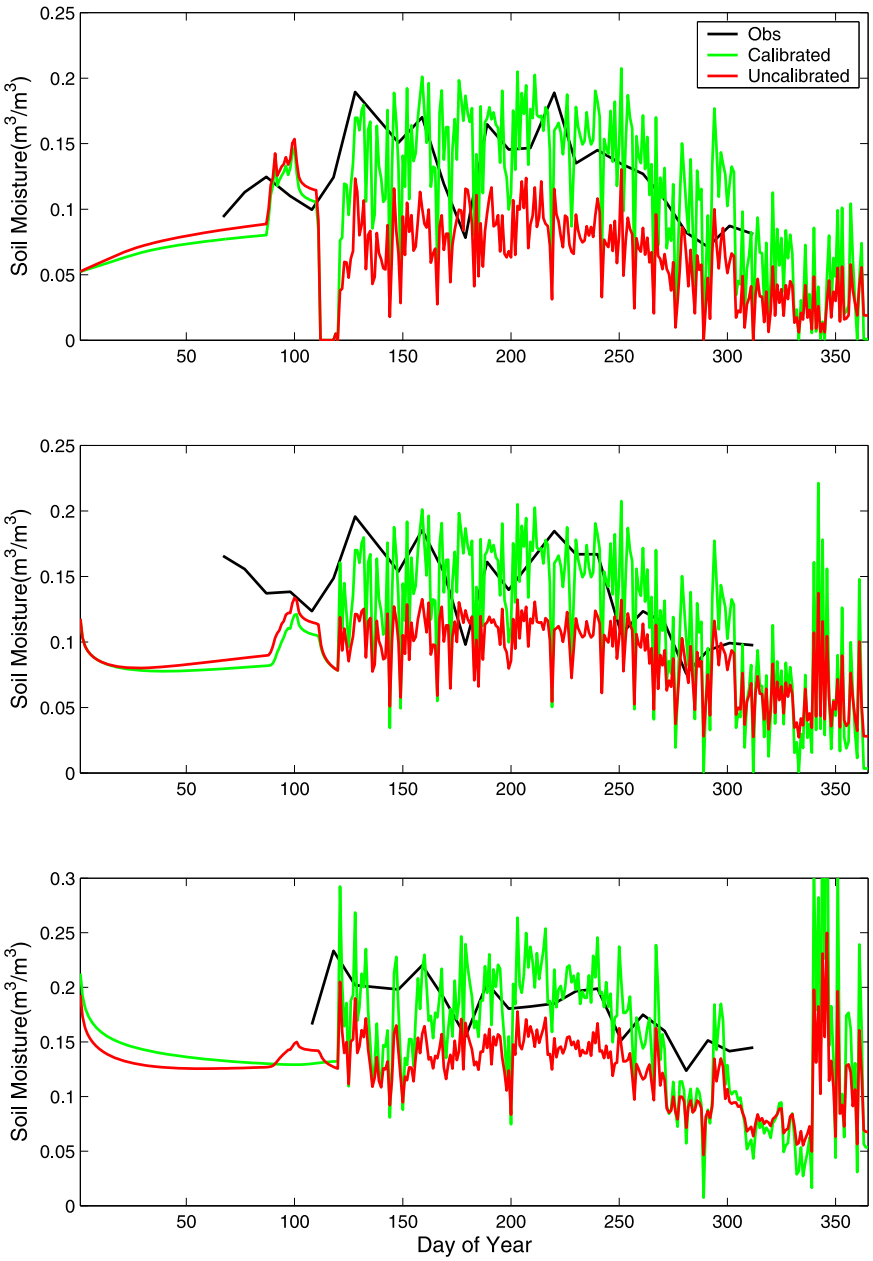

Figure 9. Time series of daily volumetric soil moisture (in meters cubed per meter cubed) from 1 January to 31 December 2006 from observations and the CLM3 simulations with the three time-averaged calibrated or the three uncalibrated parameters, respectively, for the (a) top (0–10 cm), (b) second (10–20 cm), and (c) third (40–50 cm) soil layers.

of the additional Tb data in obtaining the locally calibrated parameters for the soil water equation in the CLM3.

6. Summary and Concluding Remarks

[32] The conditional nonlinear optimal perturbation or CNOP of *Mu et al.* [2003] mathematically belongs to one kind of constraint optimization method. To compute the gradient of its cost function with respect to the control variables, one has to integrate the adjoint model, which is extremely hard to obtain for complex nonlinear models. To retain the main strength of the CNOP while removing the need of an adjoint mode of the forecast model, we have developed an ensemble CNOP method in this paper (called EnCNOP). Our new method merges the Monte Carlo method into the traditional CNOP to transform an implicit optimization problem into an explicit one. The POD method is used to efficiently approximate a forecast ensemble produced by the Monte Carlo method in a 4-D space using a set of base vectors that span this ensemble. After the analysis variables are expressed by the linear combinations of the base vectors in the linear space, the control variables in the cost function appear explicitly, so that the adjoint model, which is used to derive the gradient of the cost function, is no longer needed. Thus, the optimization process in CNOP is simplified greatly.

[33] The new EnCNOP approach was then formulated for parameter calibration by generating model perturbations

through perturbing the parameters to be calibrated and constructing the ensemble using the perturbed parameters and states, in which the model states and parameters are optimized simultaneously.

[34] Several numerical experiments performed using a 1-D model of soil water equation with three key parameters to be calibrated show that the new EnCNOP method (used for simultaneous calibration and assimilation) outperforms the EnKF (for assimilation), SCE-UA (for calibration), and EnKF+SCE-UA (for both calibration and assimilation) methods with errors being reduced to a fraction of the latter three. In our experiments, the EnKF performed worst because errors associated with the incorrect parameters used in the soil water hydrodynamic model can propagate and grow rapidly with model integration. Furthermore, the EnCNOP approach performs better than the SCE-UA if only a rough estimate of the parameter value range is available. Additionally, the experiment conducted with a comprehensively coarse initial value and a significantly rough value ranges further illustrates that the EnCNOP suffers little from the parameter value ranges. EnCNOP's applicability to complex, highly-nonlinear models was further investigated through a case study using the NCAR CLM3. Combined with satellite Tb data, the EnCNOP method was able to produce locally calibrated values for the three parameters in the CLM3 soil water equation, which resulted in much improved soil moisture in subsequent simulations. These results demonstrate that the EnCNOP provides a promising new tool for parameter calibration when certain observations (e.g., satellite Tb) are available.

[35] **Acknowledgments.** This work was supported by the National Natural Science Foundation of China (grant no. 40705035), the National Basic Research Program (grant nos. 2010CB428403), and the National High Technology Research and Development Program of China (grant nos. 2009AA12Z129 and 2007AA12Z144). We thank W. Duan for a conversation that partly motivated this study.

[36] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

Beezley, J. D., and J. Mandel (2008), Morphing ensemble Kalman filters, *Tellus, Ser. A*, *60*(1), 131–140.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*(12), 3663–3674.

Blum, J., F. X. Le Dimet, and I. M. Navon (2009), Data assimilation for geophysical fluids, in *Computational Methods for the Atmosphere and the Oceans*, *Handbk. Numer. Anal.*, vol. 14, edited by R. Temam and J. Tribbia, pp. 377–434, Elsevier, New York.

Cheng, C. T., C. P. Ou, and K. W. Chau (2002), Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration, *J. Hydrol.*, *268*, 72–86.

Cooper, V. A., V. T. V. Nguyen, and J. A. Nicell (2007), Calibration of conceptual rainfall–runoff models using global optimisation methods with hydrologic process-based parameter constraints, *J. Hydrol.*, *334*, 455–466.

Darcy, H. (1856), *Les Fontaines Publiques de la Ville de Dijon*, Dalmont, Paris.

Duan, Q., V. K. Gupta, and S. Sorooshian (1993), A shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*, 501–521.

Duan, Q., S. Sorooshian, and V. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284.

Duan, W. S., M. Mu, and B. Wang (2004), Conditional nonlinear optimal perturbation as the optimal precursors for El NinoSouthern Oscillation events, *J. Geophys. Res.*, *109*, D23105, doi:10.1029/2004JD004756.

Evensen, G. (2004), Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dyn.*, *54*, 539–560, doi:10.1007/s10236-004-0099-2.

Fang, F., C. C. Pain, I. M. Navon, M. D. Piggott, G. J. Gorman, P. E. Farrell, P. Allison, and A. J. H. Goddard (2008), A POD reduced order 4D-Var adaptive mesh ocean modelling approach. *Int. J. Numer. Meth. Fluids*, *60*(7), 709–732, doi:10.1002/fld.1911.

Gauthier, P., M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau (2007), Extension of 3DVAR to 4DVar: Implementation of 4DVar at the Meteorological Service of Canada, *Mon. Weather Rev.*, *135*(6), 2339–2354.

Goldberg, A., and E. E. Kaczka (1979), Comparison of models of individual choice in a complex setting, paper presented at 11th Conference on Winter Simulation, Oper. Res. Soc. of Am., San Diego, Calif.

Hao, Z., and M. Ghil (1995), Sequential parameter estimation for a coupled ocean-atmosphere model, in *Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, Tokyo, March 1995, Rep. WMO/TD 651*, vol. 1, pp. 181–186, World Meteorol. Organ., Geneva, Switzerland.

Heemink, A. W., M. Verlaan, and A. J. Segers (2001), Variance reduced ensemble Kalman filtering. *Mon. Weather Rev.*, *129*, 1718–1728.

Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, *44*, W09408, doi:10.1029/2007WR006505.

Jia, B., Z. Xie, X. Tian, and C. Shi (2009), A soil moisture assimilation scheme based on ensemble Kalman filter using microwave brightness temperature, *Sci. China, Ser. D*, *52*(11), 1835–1848, doi:10.1007/s11430-009-0122-z.

Kalnay, E., H. Li, T. Miyoshi, S. C. Yang, and J. Ballabrera-Poy (2007), 4-D-Var or ensemble Kalman filter?, *Tellus, Ser. A*, *59*(5), 758–773.

Ly, H. V., and H. T. Tran (2001), Modeling and control of physical processes using proper orthogonal decomposition, *Math. Comput. Model.*, *33*, 223–236.

Ly, H. V., and H. T. Tran (2002), Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor, *Q. Appl. Math.*, *60*(3), 631–656.

Mu, M., and J. Wang (2003), A method for adjoint variational data assimilation with physical "on-off" processes, *J. Atmos. Sci.*, *60*, 2010–2018.

Mu, M., W. S. Duan, and J. C. Wang (2002), The predictability problems in numerical weather and climate prediction, *Adv. Atmos. Sci.*, *19*, 191–204.

Mu, M., W. S. Duan, and B. Wang (2003), Conditional nonlinear optimal perturbation and its applications, *Nonlinear Proc. Geophys.*, *10*, 493–501.

Mu, M., L. Sun, and D. A. Henk (2004), The sensitivity and stability of the ocean's thermocline circulation to finite amplitude freshwater perturbations, *J. Phys. Oceanogr.*, *34*, 2305–2315.

Navon, I. M. (1998), Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, *Dyn. Atmos. Oceans*, *27*(1–4), 55–79.

Oleson, K. W., et al. (2004), Technical description of the Community Land Model (CLM), *Tech. Note NCAR/TN-461+STR*, 186 pp., Natl. Cent. for Atmos. Res., Boulder, Colo. (Available at http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLM_Tech_Note.pdf)

Powell, M. J. D. (1982), VMCWD: A FORTRAN subroutine for constrained optimization, *Rep. 1982/NA4*, Univ. of Cambridge, Cambridge, U. K.

Qian, T., A. Dai, K. E. Trenberth, and K. W. Oleson (2006), Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations, *J. Hydrometeorol.*, *7*, 953–975.

Sun, L., M. Mu, D. J. Sun, and X. Y. Yin (2005), Passive mechanism of decadal variation of thermohaline circulation, *J. Geophys. Res.*, *110*, C07025, doi:10.1029/2005JC002897.

Tian, X., and Z. Xie (2008), A land surface soil moisture data assimilation framework in consideration of the model subgrid-scale heterogeneity and soil water thawing and freezing, *Sci. China, Ser. D*, *51*(7), 992–1000, doi:10.1007/s11430-008-0069-5.

Tian, X., A. Dai, D. Yang, and Z. Xie (2007), Effects of precipitation-bias corrections on surface hydrology over northern latitudes, *J. Geophys. Res.*, *112*, D14101, doi:10.1029/2007JD008420.

Tian, X., Z. Xie, and A. Dai (2008a), An ensemble-based explicit four-dimensional variational assimilation method, *J. Geophys. Res.*, *113*, D21124, doi:10.1029/2008JD010358.

Tian, X., Z. Xie, and A. Dai (2008b), A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model, *J. Geophys. Res.*, *113*, D14127, doi:10.1029/2007JD009650.

Tian, X., Z. Xie, A. Dai, C. Shi, B. Jia, F. Chen, and K. Yang (2009), A dual-pass variational data assimilation framework for estimating soil

moisture profiles from AMSR-E microwave brightness temperature, *J. Geophys. Res.*, *114*, D16102, doi:10.1029/2008JD011600.

Tsuyuki, T., and T. Miyoshi (2007). Recent progress of data assimilation methods in meteorology, *J. Meteorol. Soc. Jpn.*, *85B*, 331–361.

Vermeulen, P. T. M., and A. W. Heemink (2006), Model-reduced variational data assimilation, *Mon. Weather Rev.*, *134*, 2888–2899.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Wang, B., J. Liu, S. Wang, W. Cheng, J. Liu, C. Liu, Q. Xiao, and Y.-H. Kuo (2010), An economical approach to four-dimensional variational data assimilation, *Adv. Atmos. Sci.*, *27*(4), 715–727, doi:10.1007/s00376-009-9122-3.

Wang, J. R., and B. J. Choudhury (1981), Remote sensing of soil moisture content cover bare fields at 1.4GHz frequency, *J. Geophys. Res.*, *86*, 5277–5282, doi:10.1029/JC086iC06p05277.

Xu, Q. (1996), Generalized adjoint for physical processes with parameterized discontinuities. Part I: basis issues and heuristic examples, *J. Atmos. Sci.*, *53*(8), 1123–1142.

Yapo, P., H. Gupta, and S. Sorooshian (1998). Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1), 83–97.

A. Dai, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA. (adai@ucar.edu)

X. Tian and Z. Xie, ICCES/LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China. (tianxj@mail.iap.ac.cn; zxie@lasg.iap.ac.cn)