Development of a Chinese land data assimilation system: its progress and prospects

Li Xin\textsuperscript{1}*, Huang Chunlin\textsuperscript{1}, Che Tao\textsuperscript{1}, Jin Rui\textsuperscript{1}, Wang Shugong\textsuperscript{1}, Wang Jiemin\textsuperscript{1}, Gao Feng\textsuperscript{1}, Zhang Shuwen\textsuperscript{2}, Qiu Chongjian\textsuperscript{2} and Wang Chenghai\textsuperscript{2}

(1. Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China; 2. College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China)

Accepted on January 23, 2007

Abstract The objective of land data assimilation is to merge multi-source observations into the dynamics of land surface model for improving the estimation of land surface states. We have developed a land data assimilation system for China’s land territory. In this system, the Common Land Model and Simple Biosphere Model 2 are used to simulate land surface processes. The radiative transfer models of thawed and frozen soil, snow, lake, and vegetation are used as observation operators to transfer model predictions into estimated brightness temperatures. A Monte-Carlo based sequential filter, the ensemble Kalman filter, is implemented as data assimilation method to integrate modeling and observation. The system is capable of assimilating passive microwave remotely sensed data such as special sensor microwave/imager (SSM/I), TRMM microwave imager (TMI), and advanced microwave scanning radiometer enhanced for EOS (AMSR-E) and the conventional in situ measurements of soil and snow. A spatiotemporally consistent assimilated dataset for soil moisture, soil temperature, snow and frozen soil, with a spatial resolution of 0.25 degree and temporal resolution of one hour, has been produced. This paper introduces the development of Chinese land data assimilation system and the progress made on data assimilation algorithms, land surface modeling, microwave remote sensing of land surface hydrological variables, and the preparation of atmospheric forcing data. The distinct characteristics and challenges of developing land data assimilation system and the perspectives for future development are also discussed.

Keywords: land data assimilation, land surface model, passive microwave remote sensing, Kalman filter.

Modeling and observation are two fundamental approaches to study water and energy cycles on various scales and to retrieve spatiotemporal evolution of earth surface system. Both of the methods have their advantages and disadvantages. The main advantage of modeling is that it uses knowledge of underlying physics and dynamics to provide a complete description of state evolution in time. The main advantage of observation is that it can provide direct information of the true state, whether it is taken in situ or by remote sensing.

Nevertheless, both modeling and observation are of high uncertainties. As far as modeling is concerned, many complex land surface parameterization schemes have been developed for the existing global and regional climate models and land process models, but the accuracy of model simulations is still low. For instance, in the global soil wetness project (GSWP), ten representative land surface models, including Simple Biosphere Model 2 (SiB2), Simplified Simple Biosphere Model (SSiB), Biosphere-Atmosphere Transfer Scheme (BATS), and National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction (NOAA/NCEP) ETA model have been applied to simulate the global distribution of soil moisture in 1987 and 1988. However, it has been concluded that none of these models can simulate soil moisture accurately in any place\textsuperscript{[1]}. This is because, on the one hand, the existing models are far from perfect due to the errors in physical process formulations and parameterization schemes. On the other hand, it is very difficult to determine the initial conditions of model state and the parameters of soil and vegetation in certain regions. As for the conventional in situ observation, it usually measures more accurate land surface states but has the weakness in temporal inconsistency and spatial incompleteness. Although the global observation network for land surface variables and fluxes has been enhanced, for example, the Cooperated Enhanced Observing Period

\* Supported by National Program on Key Basic Research Projects (Grant No. 2001CB309404), the Chinese Academy of Sciences (Grant No. 2003102), and CAS International Partnership Project (CXTD-Z2005-2)
\* To whom correspondence should be addressed. E-mail: lixin@lzb.ac.cn
(CEOP), which is a Global Energy and Water Cycle Experiment (GEWEX) project, coordinates a global network of 36 reference sites\(^1\), it is still very difficult to construct an image of global and regional energy and water cycle accurately by objective analysis. The major reason is that the spatial variations of land surface states are too strong. Compared with the significant heterogeneity of land surface, the \textit{in situ} observations, although being densified, are still very sparse.

Recently, a great amount of satellite remote sensing data have been used to retrieve land surface variables, greatly supplementing the insufficiency of \textit{in situ} observations. However, remote sensing itself cannot provide a continuous spatiotemporal evolution of land surface states as well. This is because:

1. The remote sensing observation is instantaneous, but land surface process is continuous in both space and time.

2. The remote sensing is an indirect measurement because the relationship between remote sensing observations and land surface variables is implicit. In general, we can establish a forward model between land surface variables and satellite observations such as brightness temperatures, but the number of observations is usually less than that of land surface variables and the forward models are often complicated and nonlinear. These make the inversion very difficult, even slight errors in observations could make the inversion an ill-posed problem. Therefore, a priori knowledge should be used to improve the possibility and accuracy of inversion. Land surface model, as a physical constraint, can provide a priori knowledge for remote sensing inversion.

3. Most of the remote sensors are not able to detect the subsurface information. Microwave can only sense the top few centimeters of the wet soil. However, for hydrological predictions, the soil moisture in root zone and deeper soil layer is just as important.

4. There exist errors in remote sensing observation, including instrument errors and representative errors caused by instrument inaccuracy and land surface heterogeneity, respectively.

Therefore, it is important to combine model simulation with instrument observation. Suppose that a remote sensing inverse model can be physically constrained by a land surface model and in the meanwhile the trajectory of land surface modeling can be adjusted by remote sensing data in order to release the accumulated errors, we can make the best use of multi-source and different spatial and temporal resolution observations and thus obtain a high-resolution and spatiotemporal consistent data sets, which can better represent the dynamic evolution of land surface. The new emerging land data assimilation method provides a practical way for us to reach the above goals.

Data assimilation, which originates from the atmospheric and oceanographic sciences, is defined as a method to produce as accurate as possible a description of the system state under observations by using all the available information and by taking into account the observation and model errors\(^2\)\(^3\)\(^4\). In the fields of land surface and hydrological sciences, data assimilation was not well established as a distinct field until the mid-1990's\(^5\). The early work was involved with retrieving soil moisture using passive microwave remote sensing data constrained by a land process model\(^6\)\(^7\), and evaluating the existing data assimilation algorithms\(^8\)\(^9\). In recent years, we see a blooming of developing and applying land data assimilation systems (LDAS), which has been becoming the frontiers of land surface and hydrological researches. The land data assimilation has also displayed some different characteristics other than atmospheric and oceanographic data assimilation and made important progress in theory and method as well as the development of operational assimilation systems\(^10\)\(^11\)\(^12\)\(^13\)\(^14\). The most representative events are that several operational LDAS have been developed, including north-American land data assimilation system (NLDAS), global land data assimilation (GLDAS)\(^15\)\(^16\), European land data assimilation system (ELDAS)\(^17\), and Chinese land data assimilation (CLDAS), which is introduced in this paper. More detailed review of LDAS can be found in Refs. [18] and [19].

The CLDAS has been developed jointly by the Cold and Arid Regions Environmental and Engineering Research Institute of Chinese Academy of Sciences and the College of Atmospheric Science of Lanzhou University. The objective of CLDAS is to develop an operational LDAS for the whole China's land territory with a spatial resolution of 0.25° and temporal resolution of one hour and in the mean time to improve the presentations of cold region process in
both land surface models and radiative transfer models. CLDAS should be capable of assimilating the passive microwave remote sensing data including special sensor microwave/imager (SSMI), TRMM microwave imager (TMI), and advanced microwave scanning radiometer enhanced for EOS (AMSR-E) into the land surface model and distributed hydrological model.

This paper is organized as follows. First, we provide a review of CLDAS development, including system design, applications and developments of data assimilation algorithms, integration of land surface models with microwave radiative transfer models, and preparation of atmospheric forcing data sets. Then, we discuss the characteristics and challenges of developing LDAS. Finally, we prospect future research highlights and propose the applicability of LDAS in land surface science.

1 Development of CLDAS

The CLDAS is composed of land models, radiative transfer models, land assimilation methods, and various data sets. The flowchart of this system is shown in Fig. 1.

Fig. 1. Flowchart of the Chinese land data assimilation system.

(1) The land surface model is driven in an offline mode. The forcing data are downscaled from the NECP re-analysis data by an atmospheric data assimilation system. They are with a spatial resolution of 0.25° and a temporal resolution of one hour. The dynamic geo-biophysical parameters are derived from remote sensing data, e.g., the MODIS leaf area index product. The vegetation morphological and physiological parameters as well as the soil hydraulic and thermal parameters are derived from vegetation and soil texture maps.

(2) The land model is then initialized with a prescribed probability distribution function (pdf), usually, a multi-dimensional Gaussian distribution is used to generate an ensemble of size N, which will be used as the first guess model state.

(3) After the initial field, parameters and forcing are given, the model integrates forward until the next observation is available.

(4) When there is an observation, use radiative transfer model or other observation operators to transfer the ensemble of model forecasts into the ensemble of observation estimates. In CLDAS, the land surface variables concerned are mainly soil moisture, soil temperature, and snow, so we used passive microwave remote sensing observation which is much sensitive to these variables.

(5) A set of random vectors are generated by a multi Gaussian distribution with zero mean vector and a prescribed covariance matrix. The brightness temperature data from passive microwave sensors are perturbed with this random field to produce an ensemble of observations.

(6) The innovation vector is calculated according to the difference between the ensemble of brightness
temperature estimates and the ensemble of true observation. The Kalman gain matrix is calculated from ensemble error statistics. Then, the increment is obtained from the multiplication of the innovation vector and the Kalman gain.

(7) The model forecast (or the initial field) is updated by adding the increment. The assimilated state vector and the analyzed error are usually calculated as the mean vector and covariance matrix of forecast ensemble, respectively. Finally, the land model is re-initialized by the updated field and the system goes into the next assimilation cycle.

1.1 Data assimilation algorithm

In CLADS, we adopted ensemble Kalman filter (EnKF) as the assimilation algorithm, which is a kind of Monte-Carlo based filter and was first proposed by Evensen[20]. Its advantage is that it uses nonlinear model operator and observation operator directly and therefore keeps all the model dynamics. On the other hand, when ensemble size is small, the computation efficiency of EnKF is comparable with the variational method. After several improvements[21-24], it has been applied widely in the field of land, atmospheric and oceanographic data assimilation[23,25].

Research progress in the aspect of developing data assimilation algorithms included:

(1) We designed and tested several variants of EnKF algorithms, compared and validated different EnKF schemes using synthetic data generated by forward model, studied the influence of initial field error, model error and observation error on inversion speed, evaluated nonlinear effect, and developed a method of generating random perturbing vectors which conform to multi-dimensional normal distribution with prescribed covariance[26]. The results showed that: (i) Though initial state of soil moisture deviates far away from true observation, the whole soil moisture profile can be retrieved quickly when soil moisture observation in the surface layer is available. (ii) The initial errors affect the inversion speed of soil temperature insignificantly while model and observation errors affect the inversion speed considerably. The latter also has a great impact on the inversion accuracy of soil temperature and moisture profiles. (iii) The nonlinearity has some negative influence on the optimal estimates of soil moisture profile but not very seriously. (iv) The dispersion of samples gradually decreases when estimation converges to the true value, which will exclude observation information and the inversion accuracy of soil temperature and moisture profile will decrease, therefore, we need to find a better way to solve the problem. (v) In comparison with different ensemble Kalman filter schemes, we draw a conclusion that the method of observation without perturbation (such as square-root series)[24] is better than that of observation with perturbation[21], and the latter is better than the early ensemble Kalman filter scheme[20].

(2) We carried out many experiments on assimilating satellite observations. The passive microwave brightness temperatures, i.e. TMI, AMSR-E and SSM/I, were successfully assimilated into land surface modeling[27-30]. The assimilating of AMSR-E observation into Sib model is illustrated in Fig. 2. The results showed that: (i) The estimation of surface hydrological variables can be improved significantly by assimilating passive microwave observation and the deviations can be corrected. When atmospheric forcing, particularly the precipitation, is significantly biased, the estimation from assimilation is much better than that from model simulation. (ii) The accuracy of land surface modeling has great effect on assimilation results. If the variation range of soil

Fig. 2. Assimilation results of soil moisture on the Qinghai-Tibetan Plateau in June and July, 2003. Land surface model: JMA (Japan Meteorological Agency) new Sib observation data; AMSR-E brightness temperatures of vertical and horizontal polarizations at 6.92GHz and 10.65 GHz.
moisture in the surface layer is very large, the assimilation results will fluctuate strongly and are worse than simulation results sometimes. The assimilated soil moisture in the root and deep layers is much improved. (iii) The quality of observation operator has essential effect on assimilation results. The sensitive parameters in radiance transfer models play an important role. For example, surface roughness is a crucial parameter which influences the inversion accuracy of soil moisture significantly. So, it is necessary to develop methods that are capable of assimilating state variable and estimating model parameters simultaneously. (iv) Compared with in situ soil moisture observation, the pixel-scale estimation by assimilating passive microwave remote sensing data is unbiased statistically. In addition, the assimilation results display the same time variation trend as in situ observations. However, the residual errors are large. This is because the assimilation results reflect average condition in large scale but observation in situ only represents local-scale characteristics.

(3) Simulated annealing based data assimilation method, which is a continuous assimilation approach, is developed independently by our group. The simulated annealing is a Monte-Carlo optimization algorithm and is independent of model operator and observation operator, so development of adjoint or linearized model is not needed and all the model dynamics can be kept. We developed an experimental data assimilation system based on simulated annealing algorithm, and successfully assimilated soil moisture observations into SiB2 model in point and regional experiments using observation data from Global Energy and Water Cycle Experiment Asian Monsoon Experiment—Tibet (GAME-Tibet) and Soil Moisture Experiment 2002 (SMEX02) projects[31–33]. The drawback of this algorithm is its low efficiency, thus it is not feasible to develop an operational land data assimilation system to obtain long time series and large area assimilated data sets. However, if we could improve the algorithm with an elaborate convergence rule, the efficiency of the algorithm could be much improved and therefore it could be potentially used as a practical continuous data assimilation method.

(4) We developed and used other assimilation methods and carried out some numerical experiments. For example, the adaptive Kalman filter was applied to estimate soil moisture profile from soil temperature observations. The thermal conductivity of soil is a function of soil moisture, therefore, we can retrieve soil moisture information from soil temperature if only the error matrix can be estimated accurately[34]. Additionally, measuring soil temperature is more convenient and less expensive than measuring soil water content, so the method has potential values in application. We also made some efforts to develop adjoints of Common Land Model (CoLM) and SiB2[35,36] since variational data assimilation is another high-efficiency data assimilation approach.

1.2 Improvement of land surface model

The Common Land Model (CoLM) version 3 was used in the operational version of CLDAS. Other land models such as SiB2 and JMA new SiB were also employed in numeric experiments.

CoLM was developed by combining the best features of three existing successful and relatively well documented and modular land models; the NCAR (National Center for Atmospheric Research) Land Surface Model (LSM), the Biosphere-Atmosphere Transfer Scheme (BATS), and the Chinese Academy of Sciences Institute of Atmospheric Physics LSM[37]. CoLM has many outstanding advantages. It considers the heterogeneity in the grid, namely, the arbitrary number of the vegetation types in one grid, so the linear ensemble of the fluxes according to the area fraction of each type can be obtained. It well represents snow and frozen/thaw process, which is very necessary for the land process simulation and assimilation in the cold and arid areas. It couples with a runoff parameterization, and introduces the two-big-leaf model to accurately describe photosynthesis and stomatal conductance of vegetation[38].

The CoLM was used to simulate the large-scale land process characteristics of the Qinghai-Tibetan Plateau and arid regions in western China. The energy and water cycles in these regions were analyzed. The freeze/thaw process in the CoLM was further validated when developing CLDAS. On the aspect of technical integration, the CoLM has been successfully coupled with ensemble Kalman filter, radiative transfer models for various land surface types, forcing data and parameters datasets.

SiB2 is another excellent land process model[39] and has been used by many operational climatic models because of its outstanding capability in simulating the surface energy and water balance. We have incor-
porated a new frozen soil parameterization scheme into the SiB2 because the soil freeze/thaw is an important land surface process on the Qinghai-Tibetan Plateau. Lack of frozen soil parameterization will result in significant underestimation of soil moisture in frozen season and therefore lead to incorrect energy and water partition. The new frozen soil parameterization included the following improvements to the original SiB2: The governing equations of water balance in SiB2 were modified to involve the soil freezing/thawing process; the unfrozen water content was expressed by a power function derived from experiments; the interlayer water flow and the hydraulic conductivity were calculated according to soil freezing. On the aspect of energy balance, the equation to calculate the surface effective thermal capacity was improved to consider the latent heat of fusion; the approximate Stefan solution was used to predict the frost/thaw depth over time and to calculate soil temperature using a simple relationship with frost depth. The modified model was validated by the soil moisture and temperature observations on the Qinghai-Tibetan Plateau. The results showed that it can realistically simulate phase change process, liquid soil water content, ice content, and frost/thaw depth.

Parameter calibration is a critical step in land surface modeling. The simulation effect of land surface model not only relies on physical mechanism of model, initialization accuracy and the quality of forcing data, but also depends on the representation and accuracy of parameters. We carried out studies on the parameter calibration for distributed hydrological model and land surface model using modern optimization algorithms. The Shuffled Complex Evolution (SCE) and the Shuffled Complex Particle Swarm Optimization (PSO) algorithms were realized in C++. A new algorithm named SCE-PSO was developed. It combines the strength of PSO and SCE and is designed as a parallel algorithm. Numerical experiments showed that the SCE-PSO has better performance than simulated annealing algorithms, genetic algorithms, and SCE algorithm. In the next work, the parameter estimation algorithms should be introduced into the assimilation system for developing a hybrid system which is capable of optimally estimating model parameters and state variables simultaneously.

1.3 Passive microwave remote sensing of land surface hydrological variables

The data being assimilated into CLDAS are mainly remote sensing observations of land surface hydrological variables, i.e., the multi-channel, multi-polarization passive microwave remote sensing data. They include SSM/I, TMI and AMSR-E.

Different microwave radiative transfer models were chosen for four types of land surfaces including soil, frozen soil, snow and vegetation, compromising the model performance and computation efficiency (Table 1). When the surface is covered by snow and the simulated snow depth > 2 cm, the radiative transfer model for snow was used, otherwise the radiative transfer model for frozen/thaw soil was employed. All of the radiative transfer models have been coupled with the CoLM and thus integrated into the land data assimilation system successfully.

<table>
<thead>
<tr>
<th>Surface type</th>
<th>Observation data being assimilated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSM/I (GHz)</td>
</tr>
<tr>
<td>Soil</td>
<td>Q/h</td>
</tr>
<tr>
<td>Snow</td>
<td>MEMLS</td>
</tr>
<tr>
<td>Vegetation</td>
<td>A zero order radiative transfer model. No assimilation if LAI &gt; 1.5</td>
</tr>
</tbody>
</table>

The radiative transfer model for soil used in CLDAS is the Advanced Integral Equation Model (AIEM) and the so-called Q/h model. The AIEM, as an improvement of IEM[43-45], is a physically based model for surface scattering. It was mainly used in one-dimensional experiments because the computation cost is huge. The Q/h model is much empirically but with high computation efficiency so it was used in the operational version of CLDAS. For Q/h model, it is of critical importance to obtain reasonable values of the "effective roughness". Although many investigations have been carried out to estimate the Q and h based on the relationship between effective roughness and measurable parameters such as the
RMS height, correlation length and correlation type of rough surface\textsuperscript{[46–48]}, there still exist a lot of uncertainties. Therefore, the “effective roughness” must be obtained by model calibration in practical applications, otherwise, the parameter error would lead to significant deviations in assimilation results.

The freeze/thaw status of soil was an important land process on the Qinghai-Tibetan Plateau and in other cold regions. It is a new challenge to assimilate the freeze/thaw status of surface soil.

For the radiative transfer of frozen soil, a land-surface process/radiobrightness model (LSP/R) with coupled heat and moisture transport for freezing soil was employed\textsuperscript{[49]}. In LSP/R, the predictions from land surface model, i.e., liquid water content, ice content, soil temperature and canopy water content were used to calculate the brightness temperature. The assimilation results showed that this model can improve the estimation of phase change in soil. The classification of soil freeze/thaw is also very important. We developed a decision tree algorithm based on scattering and thermal emission characteristics of frozen soil and many statistic analyses. By using this algorithm, a one-year dataset for the surface freeze/thaw status and the number of the frozen days in China were obtained. Compared with the soil temperature data at 4 cm beneath surface collected from the GAME-Tibet experiments, the classification accuracy can reach 92%.

The snow is a random medium with densely distributed ice particles\textsuperscript{[48]}. The microwave emission model of layered snowpack (MEMLS) was employed as the radiative transfer model for snow\textsuperscript{[50]}, which is a physical-based model that approximates volume scattering and absorption of snow using the six-flux theory and takes the boundary scattering between each layer into consideration. The scattering coefficient is parameterized by the snow density based on the measurement data, while the absorption coefficient, the effective dielectric constant, the refraction index and the reflectivity are calculated by the physical-based model and the dielectric constant of ice particle. The function to calculate the scattering coefficient was obtained by fitting the microwave scatter/radiation information and the snow density measurement, which can be difficultly applied in other areas to simulate the snow radiation. Therefore, Matzler had improved the MEMLS by using Born approximation to calculate the scattering coefficient\textsuperscript{[51]}. The prominent advantages of MEMLS are no limiting on snow layers, wider frequency range and fast computation efficiency. So, it is very suitable to be used in the CLDAS as the observation operator for snow.

By comparing various snow radiative transfer models and validating model inversion results with in situ measurements, we have developed a more appropriate algorithm to retrieve snow depth and snow water equivalent in China, particularly the Qinghai-Tibetan Plateau\textsuperscript{[52]}. Recently, a long time series of snow depth and snow water equivalent dataset (1978—2005) for China’s whole land territory since passive microwave remote sensing in operation was developed. The dataset would provide an important basis for snow hydrology and snow climate researches in China\textsuperscript{[53, 54]}.

1.4 Forcing data for land surface model

The near surface air temperature, wind speed, humidity, radiation, and precipitation data are needed to force the land surface model. The quality of forcing data is a key factor that affects the performance of land data assimilation system. In areas with sparse in situ data and complex terrain, it is impractical to obtain a reliable forcing field from observation data using objective analysis alone. There are two alternative methods. The first one is to use the reanalysis data or remote sensing products directly. However, the reanalysis data is usually with a coarse resolution, e.g. the spatial resolution of NCEP reanalysis data is one degree and the temporal resolution is 6 hours, therefore, when they are used to force the land surface model, a simple interpolation, which technically is not able to add much information to the original forcing data, is needed to be conducted. The second method is using a regional climatic model, which can generate forcing data with higher resolution but may introduce more errors.

In CLDAS, the forcing data were prepared by an atmospheric data assimilation system based on Newtonian nudging. The NCEP reanalysis data were dynamically downscaled using the meso-scale climatic model MM5. By combining the reanalysis data with model simulation, this system can produce forcing data with higher spatial and temporal resolution. In the scheme, the Newtonian nudging terms are added into the diagnosis equations of wind, temperature and water vapor, which drive simulation results approaching
the analyzed grid. Because the near surface reanalysis on the Qinghai-Tibetan Plateau usually has larger errors, proper relaxing coefficients should be carefully selected so that the high-layer simulation can approach the reanalysis data and the near surface data can be obtained mainly from model simulation.

The atmospheric analysis fields derived from the atmosphere assimilation system are used as the forcing data for CLDAS, which include the wind speed and wind direction at 10 m, air temperature and specific humidity at 2 m, surface air pressure, downward and upward solar radiations, longwave radiation, and precipitations in large scale and convective scales. The forcing data for CLDAS are in a spatial resolution of 0.25° and temporal resolution of 1 hour. Comparisons with the objective analysis of meteorological measurements and the uncontrolled modeling showed that the atmospheric data assimilation system can produce more reliable downscaling of forcing fields than objective analysis or uncontrolled modeling alone.

2 The characteristics of land data assimilation

The advances in land data assimilation research and the establishment of some operational LDAS such as NLDAS[15,16], ELDAS[17] and CLDAS (this paper) have brought this field onto a rapidly developing trajectory. It definitely will play a very important role in land and hydrological sciences. The land data assimilation also has displayed some unique characteristics which are different from atmospheric and oceanic data assimilation. In this section, we would like to analyze these characteristics and try to propose some scientific questions we need to solve.

(1) Initial-value or boundary-value problem?

The atmospheric modeling is considered as an initial value problem, so that the objective of atmospheric data assimilation is to obtain an optimized estimation of the initial field. However, the land data assimilation, although inherited from the atmosphere science, is different from atmospheric data assimilation in terms of its objective. What distinguished land data assimilation from atmospheric and oceanographic data assimilation is that it is not only an initial value problem but also a boundary value problem, additionally, model parameters also play a very important role. Therefore, in a LDAS, the optimization should not only be applied to model state but also to forcing (boundary conditions) and modeling parameters. For instance, Reichle et al.[55] reported that their variational data assimilation approach can quantify the errors of precipitation data. Moradkhani et al.[56] developed a method to estimate state variables and parameters in hydrologic model simultaneously by using the ensemble Kalman filter.

(2) What observations to be assimilated?

Due to the limited in situ observations and their poor representativeness, the remote sensing data have turned into the primary observations in a LDAS. The remote sensing data are indirect measurements so that a “forward” model is needed to transfer the state variables into the estimated remote sensing observations. In general, the microwave radiative transfer models used in LDAS, whether physically or empirically based, can be very accurate in an ideal experiment. However, these models have always shown larger bias when they are used to simulate the radio-brightness characteristics of real land surface states. These biases may come from the neglect of some important processes within the radiative transfer, the coarse resolution of passive microwave remote sensing data, the lack of heterogeneity representation to the real land surface, the disagreement between the land process model output and the radiative transfer model input (e.g. the sensible depth), and other factors. Therefore, it is very necessary to develop sophisticated but computational effective model operators (radiative transfer models for land surface states) for obtaining a reliable LDAS. In addition, several key issues should be prudently considered for different land surface variables. For instance, should the brightness temperature data or their transform (such as polarization ratio, brightness temperature gradient and so on) be assimilated? Which channel(s), which polarization(s) of brightness temperature data should be assimilated? Should remote sensing retrieval products with high quality be assimilated? Should the brightness temperature time series or, for example, its Fourier transform, be assimilated? In addition, the methods to eliminate the atmosphere and vegetation effects on microwave signals and the integrated use of microwave, infrared and visible remote sensing information may be important approaches to improve the reliability of remotely sensed data in LDAS.

(3) Stability of the system

The ensemble Kalman filter has been widely and
successfully implemented in the field of land data assimilation. However, there are two potential drawbacks within the analysis results by the ensemble Kalman filter. First, the analysis data are not restricted strongly by land surface model. Therefore, when these analyzed data are used as the initial field in a land surface model, it may be diverged. Secondly, the forcing data and land surface state variables are obtained from two independent systems, while the flux estimation depends on their gradient. Therefore, the larger difference between the analysis results and simulation outputs may result in larger bias of flux calculation or even physically unreality results. Thus, it is important to develop stable, self-adjusted, and more physically constrained LDAS. In this view, the integration of the variational and the ensemble Kalman filter assimilation methods is an important approach.

(4) Spatial autocorrelation of the errors

The land surface models are generally one-dimensional, so that the error correlation matrices usually are set in empirical. Even if the spatial correlation is taken into account, the traditional spatial correlation functions in atmosphere data analysis are used as the primary means. With the development of LDAS, especially the application of distributed hydrological models in LDAS, the spatial correlation of errors in land surface state variables and remote sensing observations will consequentially be paid more attention. The geostatistics provided more abundant variogram models and approaches to describe the spatial correlation than those in atmosphere field, therefore it is believed that the geostatistics will play a significant role in error estimation of land surface observations.

3 Conclusions and prospects

Land data assimilation, as a new field, is rapidly growing up by absorbing the advantages of atmospheric and oceanographic data assimilation. In nature, land data assimilation is within the category of methodology, however, its objectives are not only in the technological aspect, but also focus on the more effective integration of model simulation and observation which are two fundamental ways of land surface science. Land data assimilation system will play an important role with the development of the land surface system science.

This paper introduces the research progress on the development of CLDAS. Compared with the NL-DAS and ELDAS, CLDAS is still lagged behind at the aspects of data abundance, universality and profundity. However, the strong points of CLDAS are that more assimilation methods have been developed and implemented, an atmosphere data assimilation system has been used to produce the forcing data, and many passive microwave remote sensing data have been assimilated operationally in the system. The currently ongoing research priorities are:

(1) The development of new data assimilation methods. For instance, the particle filter and other generalized Bayesian filters, the singular value decomposition (SVD) method\[^{[37]}\], and the coupling of Bayesian filter and variational assimilation method.

(2) Embedding of assimilation capability into the multi-components of land surface model. In general, land surface model is composed of different modules according to their functions. The visible and near-infrared remotely sensed measurement can work on the vegetation radiative transfer, photosynthesis and the like modules, while the microwave remote sensing data can be used in the soil modules. If we can design different assimilation methods based on functions and characteristics of different modules, it will be in favor of adequately fusing the remote sensing information into land surface modeling.

(3) The next generation remote sensing satellite, e.g. soil moisture and ocean salinity (SMOS), will provide an opportunity for assimilating new passive microwave remote sensing data, that is more suitable to observe land surface hydrological variables\[^{[38-60]}\]. In addition, the upcoming Chinese satellite-borne passive microwave remote sensing data should be used. By the utility of the theoretic achievements of Chinese State Key Basic Research Project on microwave remote sensing\[^{[48,61-63]}\], we will develop more sophisticated radiative transfer models for LDAS.

(4) In the aspect of atmospheric forcing data preparation, more reanalysis data, such as the Global Precipitation Climatology Project (GPCP) precipitation data and various satellite data products should be merged to generate a higher resolution and more reliable forcing dataset.

(5) The catchment-scale land/hydrological data assimilation system should be developed, which is ca-
pable of assimilating the satellite and many kinds of automated observations in real-time for hydrological forecast and water resources management.

In conclusion, the development of LDAS helps to fill the gap between the land process simulation and observations, integrating the two important means at a uniform scheme. This kind of integration is vital to the earth surface system science where the observations are rare and the process simulations have large uncertainties. The advantages of LDAS, which also are its great challenges and opportunities, will be from remote sensing field, just like McLaughlin said that "Hydrologic data assimilation and remote sensing technology will grow together..."[64]. Anyway, we believe that, with the growing up of land data assimilation technique, remote sensing will be no longer a method standing at the edge of earth surface science, which always was given an incredulous stare and was deemed to be a tool, but a very necessary part of simulation and prediction, and a fundamental component of earth surface system science. Both of the shortages of data and the uncertainty of information will be resolved in near future. In the earth observation system (EOS) era, the LDAS will help us to obtain more accurate and consistent information of the earth surface system from massive data and information. The exploration has already been underway, and the way ahead is longer, which poses the cooperative challenge and opportunity to all of us.

Acknowledgements Prof. Jin Yasu, the chair scientist of the China State Key Basic Research Project provided many suggestions to the paper. Prof. Dai Yongjiu at the School of Geography and Remote Sensing Science of Beijing Normal University provided the latest CoLM and its guidance. Prof. Liou Yuei-an, National Central University provided the land-surface process/radiobrightness model with coupled heat and moisture transport for freezing soil.

References

17. van den Hurk B. ELDAS Final Report. KNMI, ECMWF, 2005
30 Li X. Development of a land data assimilation system: System implementation and numerical experiments at the CECOP reference sites. CAREERI/CAS, 2006, 65
Qiu C and Chou J. Four-dimensional data assimilation method based on SVD: Theoretical aspect. Theoretical and Applied Climatology, 2006, 83(1-4); 51-57


