Development of airborne remote sensing data assimilation system

Development of airborne remote sensing data assimilation system

B R Gudu[1,3], H Y Bi[2,3], H Y Wang[2], S X Qin[2,3], J W Ma[2]

[1]Institute of Remote Sensing Application, Chinese Academy of Sciences, Beijing, China.

[2]Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China.

[3]University of Chinese Academy of Sciences, Beijing, China

*sxqin@ceode.ac.cn

Abstract. In this paper, an airborne remote sensing data assimilation system for China Airborne Remote Sensing System is introduced. This data assimilation system is composed of a land surface model, data assimilation algorithms, observation data and fundamental parameters forcing the land surface model. In this data assimilation system, Variable Infiltration Capacity hydrologic model is selected as the land surface model, which also serves as the main framework of the system. Three-dimensional variation algorithm, four-dimensional variation algorithms, ensemble Kalman filter and Particle filter algorithms are integrated in this system. Observation data includes ground observations and remotely sensed data. The fundamental forcing parameters include soil parameters, vegetation parameters and the meteorological data.

1. Introduction

In recent years, data assimilation, an effective method to combine observations into land dynamic models, has drawn more and more attention in the Earth science. A land surface data assimilation system is composed of 5 parts: land surface models, data assimilation algorithms, observations, forcing data that forces the land surface models. Detailed function of each part is introduced in the following.

Land surface model is the most important part in a data assimilation system, since data assimilation is finished in the dynamic framework of land surface model. Land surface model parameterizes water and energy exchange procedure between land, atmosphere and ocean. Land surface model simulate state at next time based on the current state, meteorological data, vegetation data, soil data and other forcing data. However, simulation trajectory may gradually depart from the true trajectory because of the uncertainty of parameterization, forcing data and boundary conditions. When observations are available, observations are assimilated into land surface model to correct trajectory. Then the corrected state is used as the initial or boundary conditions for the state at next time. These steps are repeated until all the times are finished.

Observations are used to correct the trajectory of land surface model. Generally, there two kinds of observations: ground observations and remote sensing data. No matter what kind of observation is, observation operator is used to map the observations into the state domain, both in physical field and dimensional filed. When observation is the same with state variable, identity matrix is often used. When remote sensing data is assimilated, nonlinear radiation transmission model is often adopt in the assimilation procedure.

Data assimilation algorithm assimilates observations into land surface model and is another important part in data assimilation system. Many researchers are devoted to introducing newly reported achievements in mathematics into data assimilation, such as particle filter and Bayesian theory. Generally, assimilation algorithms can be divided into sequential assimilation algorithms and cost function based assimilation algorithms, with variation algorithms and Kalman filter series algorithms as typical algorithms for each strategy.

Forcing data is used to force the land surface model. Generally, forcing data is composed of meteorological data, soil data, vegetation data and other parameters. Meteorological data describes the meteorological condition of the study area during the assimilation period, including temperature, pressure, humidity, precipitation, wind and other elements. Soil data describes the soil usage, soil characteristics and other soil conditions of the study area, such as porosity, soil texture, the fraction of sand and clay, soil moisture and other parameters. Vegetation data describes the vegetation cover and vegetation characteristics of the study area, such as the count of vegetation covers in a grid, the tree height, LAI and other vegetation parameters. Other control parameters define the condition of the land surface model, such as the model mode, the start/end data, temporal and spatial resolution, output parameters, output format and other parameters.

In this work, a land surface data assimilation system for China Airborne Remote Sensing System is developed. This data assimilation system is composed of land surface Variable Infiltration Capacity (VIC, [1,2,3]) model, ground observations and remote sensing data, Ensemble Kalman Filter (EnKF) algorithm, and forcing data. In the following section, details of VIC model, observation data, EnKF algorithm and the development of this system would be introduced.

2. Land surface VIC model

As a semi-distributed macro-scale hydrological model, VIC balances both the water and surface energy within the grid cell; and its sub-grid variations are captured statistically. Distinguishing characteristics of the VIC model include: sub-grid variability in land surface vegetation classes; sub-grid variability in the soil moisture storage capacity; drainage from the lower soil moisture zone (base flow) as a nonlinear recession; and the inclusion of topography that allows for orographic precipitation and temperature lapse rates resulting in more realistic hydrology in mountainous regions. Since its existence, VIC has been well calibrated and validated in a number of large river basins over the continental US and the globe. Applications using the VIC model cover a variety of research areas.

3. Data assimilation algorithm

Generally, data assimilation algorithms can be divided into sequential data assimilation algorithms and cost function based data assimilation algorithms, such as variation algorithms, Kalman filter series algorithms and other newly developed data assimilation algorithms. In this data assimilation system, three-dimensional variation algorithm, four-dimensional variation algorithms, ensemble Kalman filter and Particle filter algorithms are integrated. In the following section, theoretical foundations of data assimilation algorithms integrated in this system are introduced.

3.1. Three-dimensional variation algorithm

Both three-dimensional variation algorithm and four-dimensional variation algorithm belong to cost function based data assimilation category. They share the same theoretical foundation and differ at the definition of cost function. In this subsection, only the theoretical foundation of three-dimensional variation algorithm is introduced.

The theoretical function of three-dimensional variation algorithm is the definition of cost function. The cost function describes the distance between observations and estimations. Values which minimize the cost function are the optimal estimation of states. Since cost function is of high dimensions, gradient function is always needed in the computation. The cost function and gradient function of three-dimensional variation algorithm are defined in equation (1) and (2), respectively.

$$J(x) = J_{\mathrm{b}} + J_0 = \frac{1}{2}(x - x_{\mathrm{b}})^{\mathrm{T}} B^{-1}(x - x_{\mathrm{b}}) + \frac{1}{2}(y_{\mathrm{o}} - H(x))^{\mathrm{T}} R^{-1}(y_0 - H(x)) \quad (1)$$

$$\nabla J = \nabla J_{\mathrm{b}} + \nabla J_{\mathrm{o}} = B^{-1}(x - x_{\mathrm{b}}) + H^{\mathrm{T}} R^{-1}(H(x) - y) \quad (2)$$

x is the state, $J(x)$ is the cost function, J_b and J_o are cost functions of background and observations, y_o is observation, $H(x)$ is observation operator, R is the observation error matrix, B is the background error matrix.

3.2. *Ensemble Kalman Filter*

EnKF was firstly proposed by Evensen in 1994 and it is now the most widely used algorithm in data assimilation field. Compared with other assimilation algorithms, such as variation algorithms, there is no need to define the tangent model and adjoint model of the dynamic model in EnKF, which means that EnKF can be used to solve non-linear and non-continuous problems. EnKF is the combination of Mont Carlo method and Kalman filter. It first defines a series of ensembles, then predicts state variables using land surface model, then updates state variables with the observations if observations are available, and finally the optimal state is computed using these updated ensembles. The main steps of EnKF is demonstrated in equations (3)~(9).

Assume N ensembles have been defined before assimilation.

(1) Prediction

$$X_{i,k+1}^{\mathrm{f}} = M_{k,k+1}(X_{i,k}^{\mathrm{a}}) + w_{i,k}, w_{i,k} \sim \mathrm{N}(0, Q_k) \quad (3)$$

In equation (1), superscript “f” and “a” mean forecast (prediction) and analysis (update) respectively. Subscript “i”, “k” and “k+1” mean the ith ensemble, the kth time step and the k+1th time step respectively. X is the state variable, M is the land surface model and here is VIC model in this work, w is the error of land surface model and it follows the Gaussian distribution.

(2) Update

If observations are available at time t+1, then observations are used to update the state of each ensemble using equations (2)~(6).

$$X_{i,k+1}^{\mathrm{a}} = X_{i,k+1}^{\mathrm{f}} + K_{k+1}[Y_{k+1}^{\mathrm{o}} - H_{k+1}(X_{i,k+1}^{\mathrm{f}}) + v_{i,k}], v_{i,k} \sim \mathrm{N}(0, Q_k) \quad (4)$$

$$K_{k+1} = P_{k+1}^{\mathrm{f}} H^{\mathrm{T}} (H P_{k+1}^{\mathrm{f}} H^{\mathrm{T}} + R_k)^{-1} \quad (5)$$

$$P_{k+1}^{\mathrm{f}} H^{\mathrm{T}} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{i,k+1}^{\mathrm{f}} - \bar{X}_{k+1}^{\mathrm{f}})(H(X_{i,k+1}^{\mathrm{f}}) - H(\bar{X}_{k+1}^{\mathrm{f}}))^{\mathrm{T}} \quad (6)$$

$$H P_{k+1}^{\mathrm{f}} H^{\mathrm{T}} = \frac{1}{N-1} \sum_{i=1}^{N} [H(X_{i,k+1}^{\mathrm{f}}) - H(\bar{X}_{k+1}^{\mathrm{f}})][H(X_{i,k+1}^{\mathrm{f}}) - H(\bar{X}_{k+1}^{\mathrm{f}})]^{\mathrm{T}} \quad (7)$$

In equation (2), K is the gain matrix and it is computed using equation (4). Y is the observation, H is he observation operator and it projects the state into the observation domain, v is the observation error and it also follows Gaussian distribution. Then the optimal estimation and variance of state at time k+1 are computed using equations (7) and (8) respectively.

$$\bar{X}_{k+1}^{\mathrm{a}} = \frac{1}{N} \sum_{i=1}^{N} X_{i,k+1}^{\mathrm{a}} \quad (8)$$

$$P_{k+1}^{\mathrm{a}} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{i,k+1}^{\mathrm{a}} - \bar{X}_{k+1}^{\mathrm{a}})(X_{i,k+1}^{\mathrm{a}} - \bar{X}_{k+1}^{\mathrm{a}}) \quad (9)$$

3.3. Particle Filter

PF originates from the Bayesian theory. It considers the uncertainty of data assimilation and represents the posterior distribution of the state using a set of particles which evolve recursively as new information becomes available.

Assume particles are generated from the initial probability density function (pdf) in equation (10). Weight associated with each particle is computed from equation (11). Then the pdf of states at time *k* can be approximated with equation (12).

$$x_k^i \sim q(x_k \mid x_{k-1}^i, z_k) \tag{10}$$

$$\tilde{w}_k^i = w_{k-1}^i \frac{p(z_k \mid x_k^i)\, p(x_k^i \mid x_{k-1}^i)}{q(x_k^i \mid x_{k-1}^i, z_k)} \tag{11}$$

$$p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i) \tag{12}$$

where x is the state, superscript i is the *ith* particle in PF, subscript k means time step, z is the observation vector, $\tilde{w}_k^i$ is the weight of particle x_k^i at time *k*, which can be iteratively computed with w_{k-1}^i, $p(z_k \mid x_k^i)$ is the likelihood function, $q(x_k^i \mid x_{k-1}^i, z_k)$ is the importance sampling function from which particles are generated.

4. Data assimilation system design and development

In this data assimilation system, VIC model is firstly forced to run by the forcing data. If observations are available, EnKF is used to assimilate observations into the VIC model to correct the trajectory of VIC model. Then optimal estimation of state at this time step is computed with these ensembles. Finally, repeat these steps until all the time steps are finished.

In order to develop this data assimilation system, the following 6 steps are needed [4].

(1) VIC model source code migration and modification. As the source code of the VIC model is developed in the UNIX operating system and this assimilation system is developed under the Windows operating system, thus code migration and necessary modification of the source code are the first steps for the VIC model.

(2) VIC model calibration and forcing data preparation. VIC model must be calibrated to determine some parameters before usage. Forcing data includes meteorological data, soil data, vegetation data and other control parameters. These data must be prepared under the same temporal and spatial resolution. In order to get rid of the effect of burn-in period of VIC model, VIC model should be forced to run ahead of the assimilation date.

(3) Data assimilation algorithms development. Although there are many applications of EnKF in data assimilation, no executable source code of these classic algorithms are available. Therefore, we first analysis the main theory of these algorithms, then summarize the technique procedure, and finally realize these algorithms with C/C++ language.

(4) Observation model definition. If remote sensing data is available, radiation transfer model is used as observation model; otherwise identity matrix is used as observation model.

(5) Experiments of the VIC model and data assimilation algorithms. In order to validate the model and assimilation algorithms, many experiments are carried out based on the fundamental parameters we prepared.

(6) Integration of data assimilation algorithm, VIC model, database and other data process algorithms.

(7) Development of the interface. After step (1)-(4), we can integrate the VIC model, data assimilation algorithms, the database, data pre-processing and interface of the system.

IOP Conf. Series: Earth and Environmental Science **17** (2014) 012189 doi:10.1088/1755-1315/17/1/012189

5. Conclusion and discussion

Data assimilation, an effective way of combining observations and dynamic models, can greatly improve the accuracy of state variable and has drawn more and more attention in recent years. In this work, a land surface data assimilation system for China Airborne Remote Sensing System is developed. This data assimilation system is composed of land surface VIC model, EnKF algorithm, ground observations and remote sensing data, and interface of the system. This work firstly introduces the function of each part in this system. Then the technique procedure of this system is introduced. Finally, the development of this system is introduced.

References

[1] Huilin G, Qiuhong T, Xiaogang S, Chunmei Z, Bohn T, Fengge S, Sheffield J, Ming P, Lettenmaier D P and Eric F W 2009 Water Budget Record from Variable Infiltration Capacity (VIC) Model Algorithm Theoretical Basis Document. University of Washington.

[2] Liang X and Zhenghui X 2003 A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity land surface model *Journal of geophysical research* D16.

[3] Liang X, Wood E F, Lettenmaier D P 1996 Surface soil moisture parameterization of the VIC-2L model: Evaulation and modification *Global and Planetary Change* 195-206

[4] Haoyu W, Sixian Q, Jianwen M and Hongjian Y 2012 Assimilation algorithm integrated computing and visualization for airborne remote sensing data *Computer Engineering and Applications* 124-127.