

A 3DVAR Land Data Assimilation Scheme: Part 1, Mathematical Design

Lanjun Zou^{1*a,b,c}, Wei Gao^{a,d}, Tongwen Wu^b, Xiaofeng Xu^b,
Bingyu Du^a, and James Slusser^d

^aSino-US Cooperative Center for Remote Sensing, Nanjing University of Information Science & Technology, Nanjing, PRC

^bLaboratory for Climate Studies, China Meteorological Administration, Beijing, PRC

^cJiangsu Key Laboratory of Meteorological Disaster, Nanjing University of Information Science & Technology, Nanjing, PRC

^dUSDA UV-B Monitoring and Research Program, Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO

ABSTRACT

Land surface states have significant control to the water and energy exchanges between land surface and the atmosphere. Thus land surface information is crucial to the global and regional weather and climate predictions. China has built abundant meteorological stations that collect land surface data with good quality for many years. But applications of these data in their numerical weather and climate prediction models are quite low efficient. To take the advantages of land surface data in numerical weather and climate models, we have developed a three dimension variational (3DVar) Land Data Assimilation Scheme (LDAS). In Part 1 of this paper, we present the mathematical design of the 3DVar LDAS. By assimilating a single point observational datum into a background setup, the LDAS is tested to demonstrate its capability and usage. In the other part of this paper, we will demonstrate the results and error analysis of assimilating China's air temperature observational data of the meteorological stations into ECMWF's model background using the 3DVar Land Data Assimilation Scheme.

Key words: land surface, data assimilation, 3DVAR

1. INTRODUCTION

Land surface is an important interface that exchanges energy with atmosphere as a main part of the earth. The most key factors of land surface that impact atmospheric circulation are albedo, soil moisture and roughness of surface. Radiation energy budget varies over different lands due to different albedos, and soil temperature may directly affect the sensible heat in land-atmosphere exchange. Despite the thermal difference between the land and the ocean, the heterogeneity of the soil, vegetation and terrain leads to the heterogeneity of sensible and latent heat flux, these thermal fluxes with turbulence form will impact the energy transfer of meso-scale and macroscale circulation. Soil moisture is not only related to albedo, thermal capacity, radiation and vegetation growth, but also related to precipitation, evaporation and runoff, influencing the water and heat distribution and redistribution in land, thus it plays an important role to weather and climate^[1,2].

It is clear that land surface is crucial to weather and climate prediction, however, uncertainties may exist in land data and

* Address correspondence to: wgao@uvb.nrel.colostate.edu, phone (970) 491-3609, fax (970) 491-3601, lanjunzou@gmail.com; phone: +86 10 6840-0082

land surface variables are depicted coarsely in GCMs, so it is not enough for multi-model consideration of climate system. Because of lack of observation data with high quality and unsatisfied model results, how to reduce the land data uncertainties and how to use much more data into the models are considered, it is urgent to develop a good land data assimilation scheme for data acquirement and model applications^[3,4].

Land data assimilation effectively combines land surface data of different types with different spatial-temporal distributions and different error characteristics under the comprehensive considerations of observation error and model background error. Comparing to atmosphere and ocean data assimilation, land data assimilation is a new field to take effort in, however, the relatively mature concepts and methods in ADAS (Atmosphere Data Assimilation Scheme) and ODAS (Ocean Data Assimilation Scheme) can be applied in LDAS. However, LDAS is different with ADAS and ODAS because of the characteristics of land surface processes and land surface parameters. As this point is concerned, we develop a 3DVAR Land Data Assimilation Scheme. Its design framework will be introduced in this paper. A single point observation test and ECMWF ERA-40 air temperature data are used to demonstrate its validity and usability. Iteration steps and correlation length of background in assimilation procedure are determined depending on cost function descending and correlation analysis of background^[5,6,7,8,9].

2. SCHEME OF 3DVAR LAND DATA ASSIMILATION

2.1 3DVAR land data assimilation system scheme

2.1.1 Cost function

The basic concept of 3DVAR assimilation is to find the minimum of a quadratic functional (named as cost function) that presents the difference of analysis to observation and background. Cost function in this land data assimilation scheme is taken as

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[y - R(x)]^T O^{-1}[y - R(x)] \quad (1)$$

where x and x_b are analysis and background vectors with length N respectively, while y is an observation vector with length m . N is the freedom degree of analysis field, m is the number of observations. B is an $N \times N$ background error covariance matrix, while O is an $m \times m$ observation error covariance matrix. R is an observation operator, which maps analysis variables into observation variables and observation positions. Under the condition that observation variables coincide with model variables, R is a simple linear interpolation operator, otherwise R is a complex nonlinear operator.

2.1.2 Model of background and observation error

Cost function has two parts, in essential, which is a weighted linear combination with analysis to background and observation. And the inverses of B and O (B^{-1} and O^{-1}) are regarded as weightings of background and observation contributed in analysis. Clearly, the result of assimilation is up to background error covariance B and observation error covariance O (Daley, 1991). Vertical correlation of background error is not considered in this assimilation scheme, so we only need to process horizontal correlation and define background error covariance B as

$B = (b_{i,j})_{N \times N}$, where $b_{i,j}$ is error covariance between the i -th and the j -th grid. Background error is regarded

Gaussian, so the form of background error covariance can be given as

$$B = a \exp(-r^2 / c^2 \cos \phi) \quad (2)$$

where a is background error variance and c is horizontal correlation length taken as Constant. r is the distance between arbitrary two background grids and ϕ is the latitude of midpoint between the two grids.

Meanwhile, observation error is regarded as uncorrelated between different observation points, which is reasonable because the measurements are independent for conventional observations, so the observation error covariance matrix is diagonal and easy to inverse. This will greatly bring convenience and cut computation cost in assimilation procedure.

2.1.3 Principle of 3DVAR and minimization algorithm of cost function

To obtain the optimal analysis, the aim is to minimize the cost function with respect to analysis variable x , i.e., to make the gradient of cost function equal to zero. Thus, it makes the minimum problem of functional converted to the problem of finding the solution of $\nabla J(x) = 0$.

Concerning the tremendous dimension of analysis variable x , increment analysis method is adopted in this scheme. Also, B is decomposed by Cholesky method to reduce its condition number.

A variable W is set to meet $W = C^{-1}(x - x_b)$, where $B = CC^T$. Substituting it into the cost function, it follows

$$J = \frac{1}{2} W^T W + \frac{1}{2} [y - R(x_b + CW)]^T O^{-1} [y - R(x_b + CW)] \quad (3)$$

To find the minimum of cost function, it requires $\partial J / \partial W = 0$, that is

$$\frac{\partial J}{\partial W} = W - C^T L^T O^{-1} [y - R(x_b + CW)] = g \quad (4)$$

in which L is the tangent linear operator of observation operator R , i.e. $L = \left. \frac{\partial R}{\partial x} \right|_{x=x_b}$.

The conjugate-gradient method was found to represent a good compromise in convergence rates and computer memory requirements between simpler and more complex methods of nonlinear optimization (Fletcher and Reeves, 1964; Hestenes and Stieffel, 1952; Polak and Ribiere, 1969; Beale, 1972; Shanno, 1970 and 1978; Perry, 1976, 1977 and 1978). Based on previous work of Hestenes and Steiffel (1952), Fletcher and Reeves (1964) proposed a conjugate-gradient method applied to general nonlinear functions. Navon and Legler (1987) assessed use of different available conjugate-gradient algorithms in large-scale typical minimization problems in meteorology considering computational efficiency and accuracy as principal criteria. This conjugate-gradient method is used to calculate the gradient of J in this scheme. Two preliminary notes to be used in the iteration algorithm are given as followings:

i) If the cost function J is defined as

$$J(W) = \frac{1}{2} W^T W + \frac{1}{2} [y - R(x_b + CW)]^T O^{-1} [y - R(x_b + CW)] \quad (5)$$

then $\nabla^2 J = I + C^T L^T O^{-1} LC$, where $L = \left. \frac{\partial R}{\partial x} \right|_{x=x_b}$.

ii) The Euclidean norm is defined as

$$\|g\| = (g_1^2 + g_2^2 + \dots + g_n^2)^{1/2} \quad (6)$$

where $g = (g_1, g_2, \dots, g_n) \in R^n$.

The minimization algorithm is given in the followings:

1) Set $W = 0, \beta = 0, d = 0, EPS=1.0E-10$. (initial condition)

2) Do iter=1,steps

3) Compute the gradient of J

$$\nabla J = grad_1 = W - C^T L^T O^{-1} [y - R(x_b + CW)] \quad (7)$$

4) Set

$$d = -grad_1 + \beta d \quad (\text{descent direction}) \quad (8)$$

5) Stepsize is obtained as

$$\alpha = \frac{(\nabla J)^T (\nabla J)}{d^T (\nabla^2 J) d} = \frac{\|grad_1\|^2}{d^T (\nabla^2 J) d} \quad (9)$$

6) Generate a new W by

$$W = W + \alpha d \quad (10)$$

7) Recompute the gradient of J

$$\nabla J = grad_2 = W - C^T L^T O^{-1} [y - R(x_b + CW)] \quad (11)$$

8) Update β as

$$\beta = \frac{\|grad_2\|^2}{\|grad_1\|^2} \quad (12)$$

9) Enddo

A convergence criterion for stopping the iterations is tested ($\|\nabla J\| < EPS$ or iteration number is larger than a given value, for example, 10), and, if it is satisfied, the iterations must stop.

We modify the criteria of iteration stop by controlling the norm of cost function and iteration step number in the procedure. Practical applications in this scheme illustrate that this algorithm quickly converges in most circumstances, which only needs to deal with multiplication and addition of matrixes and vectors.

3. SINGLE POINT OBSERVATION TEST

To verify the validity and usability of our assimilation system, one typical single point observation test is carried out. This test supposes the background value set as zero and the single point observation set as one taking on the central grid

of the background.

China zone of longitude as 75,105,135E and latitude as 15, 35, 55N is taken as background (all zeroes) and single point observation (one) is take in central grid (105E, 35N). Horizontal correlation length is taken as 500km. Thus a list of

results with different background error covariance σ_b^2 and observation error covariance σ_o^2 (σ_o^2 is 0.01, 0.1, 1, 10,

100 times as σ_b^2) are given in the following table(Table 1):

Background error variance σ_b^2	Observation error variance σ_o^2	Norm of cost function J after 3 iterations
0.01	0.0001	2.827086658996336E-006
	0.001	2.751396918299065E-006
	0.01	0.000000000000000E+000
	0.1	1.235978874802868E-009
	1	2.704572676925920E-012
0.1	0.001	9.910650158317935E-007
	0.01	1.263638001791634E-006
	0.1	0.000000000000000E+000
	1.0	1.794940374111320E-008
	10.0	3.595401754097338E-012
1	0.01	2.921509891784524E-007
	0.1	3.497462633950121E-007
	1	0.000000000000000E+000
	10	9.516144816590355E-011
	100	6.362566723483454E-013

Table 1. The norm of cost function descends with different background and observation error variances

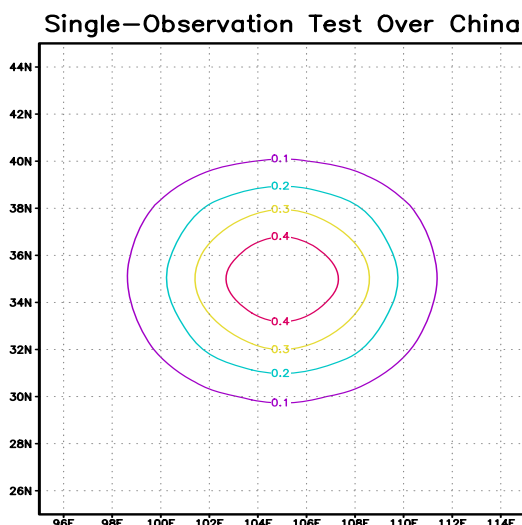


Fig.1 Single Point Observation Test Over China

This figure again verifies the validity of our land data assimilation scheme, because the central point of assimilation analysis is 0.5 and analysis belongs to Gaussian.

In most circumstances, the gradient of cost function descends almost to zero in 3 steps, that is to say the

conjugate-gradient method is quite effective. In order to analyze the assimilation result concerning about background

error variance and observation error variance, r is defined as the ratio of σ_b^2 and σ_o^2 , i.e. $r = \frac{\sigma_b^2}{\sigma_o^2}$. When

$r > 1$ (or $r < 1$), that is the effect of observation (background) is larger than that of background (observation), the assimilation is closer to observation (background) than background (observation). When $r = 1$, that is the effect of observation equals to that of background, the assimilation equals to half of background and observation. In fact, error variances reflect relative magnitudes of errors, while their reciprocals reflect relative weights in the assimilation.

Assuming the background and observation error covariance is equivalent, the single point observation test comes to the following result with Figure 1.

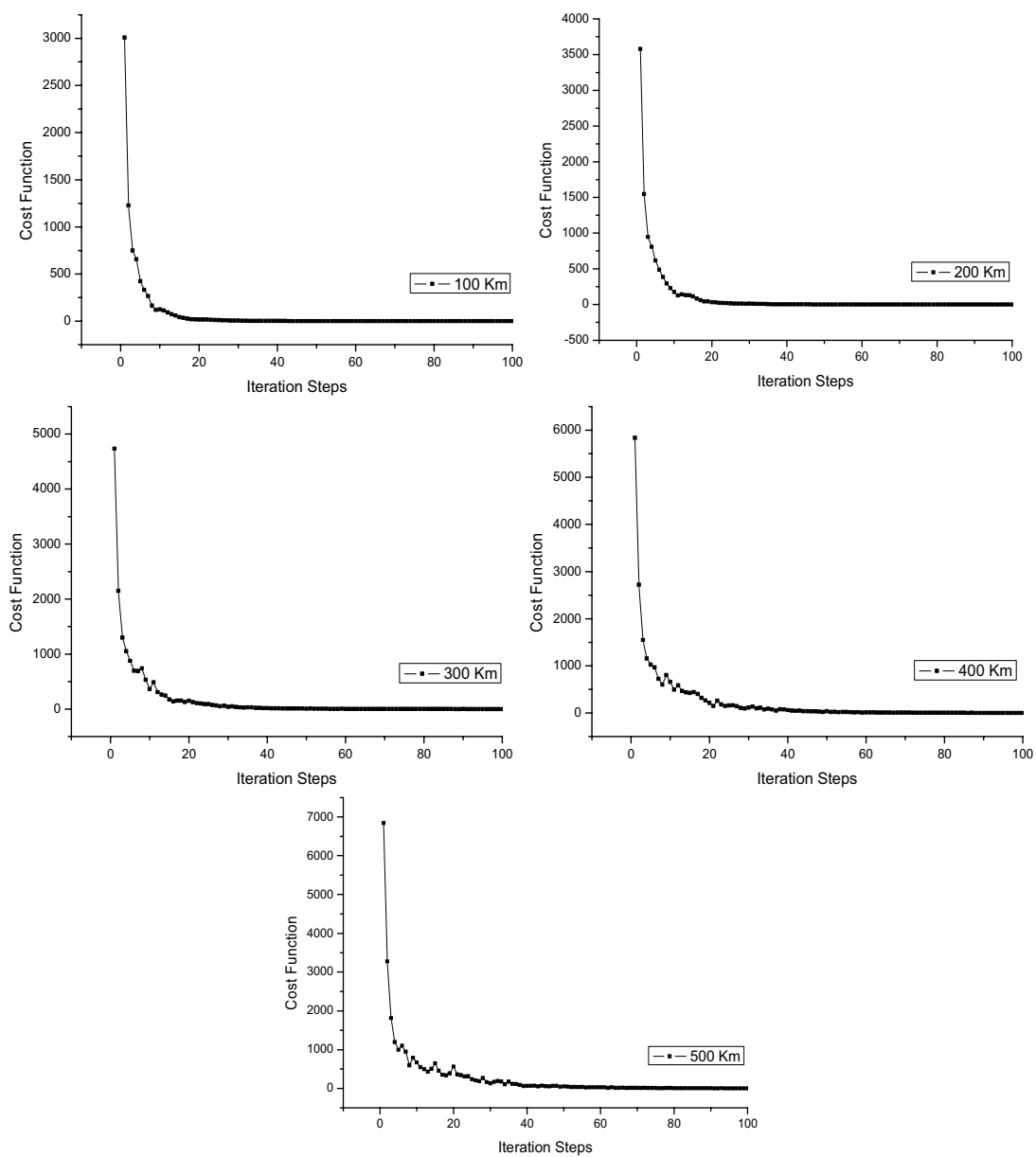


Figure 2 Cost function vs. iteration steps with different correlation length

4. DETERMINATION OF ITERATION STEPS AND CORRELATION LENGTH FROM ECMWF ERA-40 AIR TEMPERATURE DATA

ECMWF ERA-40 has good and the most time-consistent product quality for the globe as a whole, and it is a common-used data set for climate research, but it also has uncertainties, we choose air temperature from ERA-40 to assimilate Chinese station data as a test to determine the iteration steps and correlation length in assimilation procedure. ECMWF ERA-40 has a horizontal resolution of 2.5×2.5 degree.

4.1 Determination of iteration steps

It must take assimilation accuracy and efficiency into consideration at the choice of iteration steps. Iteration steps depend on the iteration algorithm, observation longitude and latitude, background grid distance and selected assimilation region. Assimilation system always iterates 100 steps with background horizontal correlation distance taken as 100, 200, 300, 400 and 500 km respectively, and then the descent velocity of cost function is examined to find out the iteration step satisfying accuracy and low-cost time requirement (See Fig. 2).

The value of cost function rapidly descends with iteration. Cost function is very small at the tenth step comparing to that at the first step when background horizontal correlation length is taken as 100, 200, 300, 400 and 500 km respectively, so the step number of this assimilation scheme is taken as 10.

4.2 Determination of background horizontal correlation length

Background error covariance matrix B , controlling the way that information spreads out at observation position, presenting uniform analysis increment at adjacent model grids and vertical levels, and ensuring model variables dynamically coincide with each other, is crucial to analysis results. Suppose that the distribution of background error is isotropic and uniform, then background error covariance can be modeled by Gaussian function, i.e.,

$B = a \exp(-r^2 / c^2 \cos \phi)$. Correlation length must be derived from statistics of numerous background data.

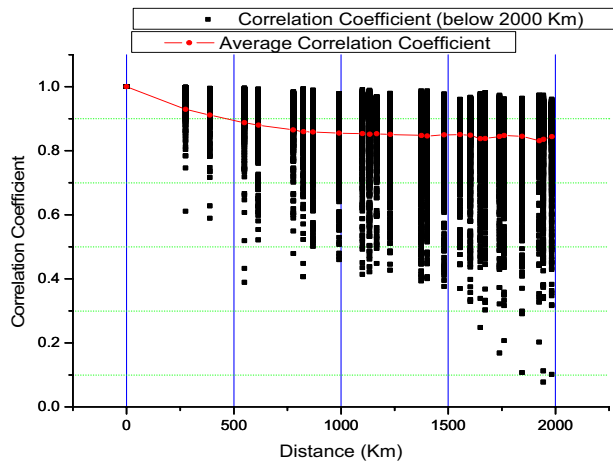


Fig. 3 Correlation coefficients vs. distance

Black dots: correlation coefficient; Red line: Average correlation coefficient

After statistics about correlation coefficients of arbitrary two of all the grids and average correlation coefficients at every distance, the distance with correlation coefficients above 0.9 is taken as correlation length, that is 500 km (See Fig. 3).

5. SUMMARY AND CONCLUDING REMARKS

Concerning about the low efficient application of Chinese station climatic data into the model, a 3DVAR Land Data Assimilation Scheme is developed and its single point observation test is carried out in the first part of this paper to verify the validity and usability of this 3DVAR LDAS. Iteration steps and correlation length of background are cautiously determined as 10 and 500 km in assimilation procedure by using ECMWF ERA-40 air temperature data as background and China station data as observation. This 3DVAR LDAS can be used to assimilate station observation efficiently and will greatly help the research of climate.

Acknowledgements

ECMWF ERA-40 data used in this study have been provided by ECMWF. Authors would like to express their thanks to Dr. Zaizhi Wang for discussions about the 3DVAR scheme and Dr. Xiwu Zhan for his helpful suggestions about LDAS and comments about this paper. This work was supported by ,USDA UV-B Monitoring and Research Program under a grant from USDA CSREES(2005-34263-14270), Project 973 (2004CB418303, 2006CB403605), National Natural Science Foundation of China (40471097).

REFERENCES

1. Beale, E. M., 1972: A derivation of conjugate-gradients. Numerical Methods for Non-linear Optimization. F. A. Lootsma, Ed., Academic Press, 39-43.
2. Fletcher R. and C. M. Reeves, 1964: Function minimization by conjugate-gradients. The Computer Journal, 7,149-153.
3. Hestenes M. R. and E. Stieffel, 1952: Methods of conjugate-gradients for solving linear systems. J. Res. Natl. Bur. Stand., 48,409-436.
4. Navon I. M. and D. M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. Monthly Weather Review, 115, 1479-1502.
5. Perry, A., 1976: A modified conjugate-gradient algorithm. Discussion Pap. No. 229, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University.
6. Polak, E. and G. Ribiere, 1969: Note sur la convergence de methods de directions conjuguées. Rev. Franc. Informat. Rech. Operationnelle, 16,35-43.
7. R. Deley, 1991, Atmospheric Data Analysis, Cambridge University Press.
8. Shanno, D. F., 1970: Conditioning of quasi-Newton methods for function minimization. Math. Comp., 24, 647-657.
9. X. L. Zou, F. Vandenberghe, M. Pondeva and Y. H. Kuo,1997, Introduction to adjoint techniques and the MM5 adjoint modeling system, NCAR Technical Note, NCAR/TN435STR.