Using Ensemble Kalman Filter to Assimilate Land Surface Temperature and Evapotranspiration

Yang Wang
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
wangyang@lzb.ac.cn

Yaonan Zhang
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
yaonan@lzb.ac.cn

Guohui Zhao
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
zhgh@lzb.ac.cn

***Abstract*— Ensemble Kalman filter (EnKF) is an efficient algorithm in dealing with nonlinear and discontinuous data assimilation problems. We designed a scheme that integrated the EnKF and Simplified Simple Biosphere model (SSiB) to improve the estimate of land surface temperature and evapotranspiration (ET) using Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature (LST) products. This scheme can make a judgment whether there are MODSI LST products available to assimilate at every time step. Then we compared the assimilation results with SSiB open loop simulation and station observations. The results showed that the EnKF algorithm could improve the land surface temperature and evapotranspiration estimate. Then we discussed five challenges during the experiment. In a word, this scheme provides a practical way for improving land surface models estimates with assimilating remote sensing observations.**

I. Introduction

Water and energy circulation are two important processes. Accurately simulation of them will help us to research land-atmosphere interaction, land surface processes and climate change processes. Land surface models convert real world into physical equations and dynamic processes. With that, we can obtain variables that we concerned over time and space. J. K. Entin et al. had used many land surface process models, such as Simple Biosphere model (SiB), Simplified Simple Biosphere model (SSiB), Biosphere-Atmosphere Transfer Scheme (BATS) and Eta model, to simulate the global soil moisture [1]. The results showed that none of the models could accurately simulate the soil moisture in any regions. Consequently, low simulation accuracy usually restricts the use of land surface model. On the other hand, traditional observations can obtain relatively true value of land surface characters; however, spatial heterogeneity and time discontinuity are two disadvantages of observation data [2]. How to integrate model simulation and observation data has been a focus in geosciences for decades. In recent years, data assimilation techniques realize the integration perfectly.

Data assimilation is method, which can dynamically merge together observations with a numerical model in order to determine the model state variables as accurately as possible considering the observation errors and model errors. In 1960s, data assimilation technique had been proposed. This method was first used in Numerical Weather Prediction [3], and then applied into oceanographic models. In 1990s, data assimilation method was used in land surface models and hydrological models. The essential idea of data assimilation is that using appropriate observation operator to integrate different spatial resolution data into land surface model; adjusting the model outcome by assimilating the observation data and decreasing the error of the simulations [2].

Generally, from algorithmic point of view, data assimilation exists at present under two forms, variational assimilation and sequential assimilation [3]. Variational assimilation aims at globally adjusting a model solution to all the observations available over the assimilation period. 3-dimensional variational (3D-Var) and 4-dimensional variational (4D-Var) are two main algorithm in variational assimilation which need an adjoint model when dealing with the nonlinear models, and it is difficult to derive an adjoint model from a land surface model [4]. On the other hand, Kalman filter and its family are typical algorithms in sequential assimilation. Unfortunately, traditional Kalman filter [5] cannot handle the nonlinear and discontinuous problems. Kalman filter is first proposed by Kalman in 1960, and widely used in linear data assimilation. R. N. Miller, M. Ghil and F. Gauthiez [6] developed extended Kalman filter (EKF) to assimilate nonlinear problems in 1994. However, this method performance poorly when the problems were complex and highly nonlinear. G. Burgers, P. J. van Leeuwen and G. Evensen [7] first proposed ensemble Kalman filter that applies an ensemble of model state variable to represent the error statistics of the model estimate and to predict the error statistics continuously [8]. It has been used in various fields, such as numerical weather report, ocean prediction, land surface models and hydrological models [7, 9, 10]. The assimilation results showed that EnKF is an effective method to deal with nonlinear assimilation problems, and performed well in non-Gaussian error statistics in some case [11]. C. L. Huang [8] used ensemble Kalman filter (EnKF) to assimilate soil moisture in Revised Simple Biosphere model (SiB2) with microwave remote sensing data, and the results showed that surface soil moisture were significantly improved. C. L. Huang [12] used Moderate Resolution Imaging Spectroradiometer (MODIS) Product to assimilate soil surface temperature and deep soil temperature in CoLM with EnKF. The simulation accuracy was improved by 1K. X. J. Han [11] had reviewed the modern nonlinear filters, and compared the performance of EnKF, unscented Kalman

CHINACOM 2010, August 25-27, Beijing, China

DOI 10.4108/adim.2010.4

filter (UKF), sampling importance resampling particle filter (SIR-PF) and unscented particle filter (UPF) in Lorentz system, and used them in VIC-3L model to assimilate soil moisture with microwave remote sensing data. W.T. Crow [13] used remote sensing data to retrieve soil moisture and assimilated the runoff prediction in Sacramento hydrologic model. J.D. Bolten [14] used Advanced Microwave Scanning Radiometer (AMSR-E) data to assimilate soil moisture for Operational Agricultural Drought Monitoring. R. Reichle, J. Walker, R. Koster and P. Houser [15] used the EnKF to estimate soil moisture profile and found that the EnKF had a better performance than EKF. R.C. Pipunic, J.P. Walker and A. Western [16] have attempted to use indirect data assimilation technique to improve evapotranspiration (ET) estimation accuracy.

In this study, we integrated ensemble Kalman filter algorithm and SSiB [17] to improve the accuracy of land surface temperature and evapotranspiration (ET) simultaneously with assimilating the MODIS Land Surface Temperature (LST) products.

II. METHOD

Generally, a data assimilation system needs model, data assimilation algorithm, observation operator and data sets. In our experiment, we choose SSiB as the model, and ensemble Kalman filter as the data assimilation algorithm. We use MODIS LST products to assimilate land surface temperature, which is the model state variable that we defined. Therefore, the observation operator is the formula that transforms radiant temperature into observation temperature. The flowchart of our data assimilation scheme which assimilating MODIS LST products with EnKF is showed in Fig. 1.

A. Land surface model

The Land surface model used in this study is the Simplified Simple Biosphere model (SSiB) modified by Y.K. Xue from the Simple Biosphere model (SiB) which was developed by P.J. Sellers, Y. Mintz, Y.C. Sud and A. Dalcher [18]. In SSiB, the calculation of radiation fluxes, aerodynamic resistance and surface resistance are simplified and it has been proven that the simplification had little effect to the results. D. Q. Zhu has validated SSiB model in arid region and the sensitivity results showed that this model are suitable for arid regions if only the parameters were correctly defined [19]. In SSiB, the land surface temperature, T_{gs}, is calculated as [17],

$$C_{gs}\frac{\partial T_{gs}}{\partial t}=R_{ngs}-H_{gs}-\lambda E_{gs}-\frac{2\pi C_{gs}}{\tau}(T_{gs}-T_d) \tag{1}$$

$$H_{gs}=\frac{T_{gs}-T_a}{r_d}\rho C_p \tag{2}$$

$$\lambda E_{gs}=[f_h e_{*(gs)}-e_a]\frac{\rho c_p}{\lambda}\frac{1}{r_{surf}+r_d} \tag{3}$$

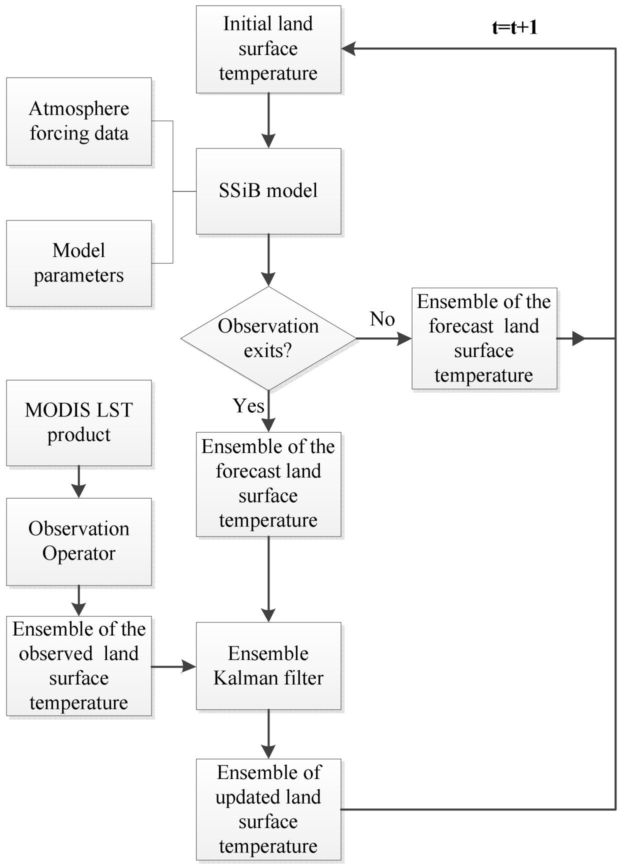

Fig 1. The flowchart of data assimilation scheme

where C_{gs} is the effective heat capacity of soil, R_{ngs} is the net radiation flux at the ground, H_{gs} is the sensible heat flux from the ground, E_{gs} is the latent heat flux from the ground, λ is the latent heat of vaporization, τ is the day length, T_d is the temperature for deep soil, r_d is the aerodynamic resistance between ground and canopy air space, ρ is the density of air, c_p is the specific heat of air,f_h is the relative humidity of the air at the soil surface, $e_{*(gs)}$ is the saturation vapor pressure at temperature T_{gs}, T_a is the temperature of air, e_a is the vapor pressure in canopy air space, r_{surf} is the surface resistance.

B. Data assimilation algorithm

As we mentioned in the first part, the EnKF is an efficient method for nonlinear data assimilation. In this study, random perturbations with the appropriate expectation and variance were added to the observations, which generate an ensemble of observations. Researches have proven that this process can bring about lower variance in the ensemble of updated model state variables. The error covariance for the forecast and analysis estimate, P^f and P^a of model state variable X are defined as [9],

$$P^f \cong P_e^f=\overline{(X^f-\overline{X}^f)(X^f-\overline{X}^f)^T} \tag{4}$$

$$P^a \cong P_e^a=\overline{(X^a-\overline{X}^a)(X^a-\overline{X}^a)^T} \tag{5}$$

where overbar denotes an average over the ensemble, and the superscripts a and f refer to analysis and forecast, respectively.

Generally, the EnKF has two steps, forecast step and analysis step. In the forecast step, the model state X is calculated by [9],

$$X^f_{i,t+1} = M(X^a_{i,t}) + u_i \tag{6}$$

where $X^f_{i,t+1}$ is the forecasted state variable of the ith ensemble member at time t+1, $X^a_{i,t}$ is the analyzed state variable of the *i*th ensemble member at time t, M is the model operator, u_i is the model error which assumed to satisfy Gaussian distribution with zero mean and covariance matrix Q.

In the analysis step, the forecast of each ensemble member is updated as follows [9],

$$X^a_{i,t} = X^f_{i,t} + K_t[(Y_t + \varepsilon_i) - H(X^f_{i,t})] \tag{7}$$

$$K_t = P^f_t H^T [H P^f_t H^T + R]^{-1} \tag{8}$$

where H is the observation operator that relates the model state variables to the observation, Y_t is the observation at time t. ε_i is the Gaussian distributed observation error with zero mean and covariance matrix R, K_t is the Kalman gain at time t, P^f_t is the forecast background covariance matrix at time t.

In this data assimilation scheme, we define land surface temperature as the model state variable, X_t; MODIS LST products as the observation variable Y_t; and the number of ensemble size as 50. In the forecast step, the ensemble of land surface temperature is generated by SSiB estimate of land surface temperature adding appropriate random error. At every time step, the scheme will check if there are MODIS LST products available; if not, the scheme moves on as usual; if yes, the scheme begins to launch data assimilation as follows,

1) Generate the ensemble the MODIS LST products with adding appropriate random error using (6).
2) Calculate the forecast error covariance with forecast land surface temperature using (4).
3) Calculate the Kalman gains with observation operator and MODIS LST products covariance using (8).
4) Calculate the analysis land surface temperature using (7).
5) Update the forecast land surface temperature at time t+1 with analysis land surface temperature at time t using (6).

In this process, the land surface temperature was quantificationally adjusted by MODIS LST observation, which is considered more accurate than the model simulation. And this process illustrated above will repeat every time step. Therefore, at the end of the scheme, the observations will greatly affect the model estimates, if the observations are accurate enough, the model estimates will improve the simulation accuracy.

C. Experiment Area

The one-dimensional assimilation experiments were conducted at Arou station (E100°27′, N38°02′; 3030m) which is located in upstream of the Heihe river basin in northwestern China. Arou station lies on the southern slope of Qilian Mountain northeast of Qinghai province. The annual average air temperature is about 1℃, while the annual precipitation is about 270 - 600mm. The underlying surface around the station is predominated by meadows with 20 – 30 cm height. In Arou station, there are a set of automatic meteorological station and a set of eddy correlation system.

III. Results and Discussion

The EnKF assimilation scheme is used to assimilate MODIS LST products from June 1st to June 30th (Julian Day 152 to 181), with a time step of one hour. The value of land surface temperature from observation, SSiB open loop run and data assimilation method are shown in Fig. 2 (a); and the value of ET are shown in Fig. 2 (b).

In land surface temperature estimate, there are significant improvements when the MODIS LST products are available. However, in the last a few days, the improvements are not very obvious, due to frequently precipitation and absence of MODIS LST products. In addition, in precipitation days, the quality of MODIS LST products will decline owing to the reflection and scattering effects of the clouds. Sometimes, the MODIS LST products will be a null value because of clouds. Another important reason for the indistinctive improvement is that the SSiB open loop simulation of land surface temperature is comparative close to the observations. The root mean square

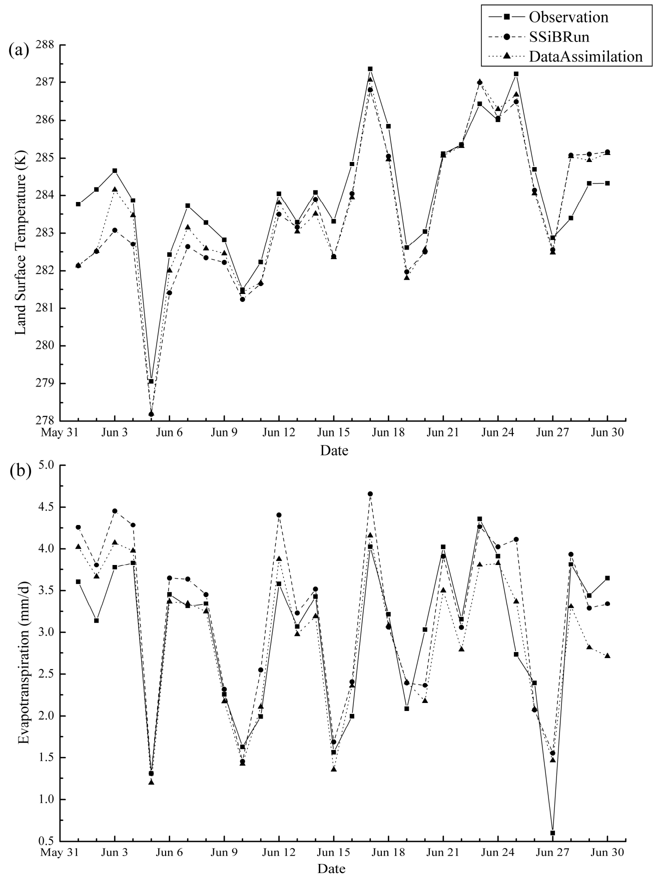

Fig 2. The results of observation, SSiB open loop run, data assimilation in land surface temperature and evapotranspiration

error (RMSE) of SSiB estimate of land surface temperature is 0.89K, while the RMSE of assimilation with MODIS LST is 0.76K, which has been improved by 13%. The error between SSiB open loop simulated and observation is about 0.74K average and 1.68K maximum, while the error between assimilated results and observation is about 0.63K average and 1.92K maximum.

On the other hand, in ET estimation, the assimilated results also significantly improved the accuracy of ET estimate. However, in the last three days, ET assimilation also has obviously underestimated. The main causes of this situation are the absence and low quality of MODIS LST products. There is only one scene of MODIS LST product available in the last five days. Moreover, if the quality of MODIS LST product is low, the data assimilation algorithm will also lead the model to the opposite direction. Therefore, a data set quality control scheme is very important to a data assimilation system. Fortunately, the quality of MODIS LST product has an acceptable accuracy for our study area. The RMSE of SSiB open loop simulation is 0.49, while the RMSE of data assimilation simulation is 0.43, which has been improved by 12%. The average daily ET of SSiB simulation is overestimated by 0.23mm, while the ET estimate of EnKF algorithm is 2.93mm, underestimated by 0.06mm. The statistics shows that the EnKF algorithm has respectable achievements in ET estimate.

However, we have faced five challenges or problems in simulation and assimilation. First, owing to the influence of terrain and heterogeneity of spatial scale, there are differences between the value retrieved from MODIS LST products and observed from the station. Moreover, these differences are not belonging to systematic error. Therefore, more experiments and researches are needed to resolve this problem. Second, although the time step of the SSiB model is one hour, the MODIS LST products collected by the satellites are instantaneous values. Apparently, an instantaneous value cannot match with a one-hour average value. Accordingly, it is crucial to build a reliable relationship between the MODIS LST products and station observation. Third, the error covariance of the model state variable and the observation, Q and R, which mentioned in (6) and (7) respectively, are difficult to define. Usually, there are systematic biases of the model estimate, therefore, the error distribution cannot be Gaussian distributed with zero mean. In our experiment, we can infer that the EnKF has a limit to deal with systematic bias problems. Consequently, model parameters and model structures are also needed to optimize in order to reduce the systematic biases. Besides, ET observations are usually converted from latent heat flux observed by eddy correlation system, which also have great indeterminacy in data quality. In our experiment, frequency response [20] corrections and Webb, Pearman & Leuning corrections [21] has been done for the eddy correlation data. Finally, the performance of EnKF algorithm also depends on the frequency of available observations. Unfortunately, the maximum frequency of MODIS LST products is one scene per day. Due to the influence of the clouds, there are only 15 scenes of the MODIS LST products available in 30 days, that is to say there are only 15 observations can be assimilated in 720 time steps. Consequently, in order to obtain better assimilation results, we need more satellites and sensors to provide us more remote sensing data.

From this paper, we can infer that data assimilation method can improve the model estimates with accurate observations when the models cannot describe the real process analyticly and observations are accurately enough. Therefore, this algorithm can be imported into other fields, such as chemical engineering, marine navigation and military fields.

IV. Conclusion

This paper shows how the EnKF works and provides us an executable way to improve model estimates, such as land surface temperature and evapotranspiration. We have designed a one-dimensional land surface temperature assimilation scheme with EnKF algorithm and SSiB. This scheme can decide whether there are MODIS LST products to assimilate. Moreover, from the results, we can draw a conclusion that the EnKF algorithm can improve the accuracy of land surface temperature and ET estimate with MODIS LST products. Though the improvement are not evident enough, we found some other way to make the estimate better, which is also important, such as optimizing the model parameter, studying the relationships between the MODIS LST products and station observations in mountain area and using other kind of satellites to get more frequent and reliable data. On the other hand, we can also use other kind of data, such as soil moisture, LAI and precipitation, to assimilate corresponding model state variables in different models, such as Common Land Model (CLM), Dynamic Global Vegetation Model (DGVM) and Soil Water Assessment Tool (SWAT).

Acknowledgment

This work is supported by Informationization Foundation of Chinese Academy of Sciences (CAS), "The E-Science Environment for Ecological and Hydrological Model Research in Heihe River Basin" (Grant number: 29O920C61); Project for Incubation of Specialists in Glaciology and Geocryology of National Natural Science Foundation of China (Grant number: J0930003/ J0109); and Second Phase of the CAS Action-Plan for West Development (Grant number: KZCX2-XB2-09-03).

References

[1] J. K. Entin, A. Robock, K. Y. Vinnikov, V. Zabelin, S. X. Liu, A. Namkhai, and T. Adyasuren, "Evaluation of Global Soil Wetness Project soil moisture simulations," *Journal of the Meteorological Society of Japan,* vol. 77, pp. 183-198, Mar 1999.

[2] X. Li, C.L. Huang, et al., "Development of a Chinese Land Data Assimilation System: its Progress and Prospects," *Progress in Natural Science*, vol. 17, no. 8, pp. 881-892, Feb. 2007.

[3] O. Talagrand, "Assimilation of Observations, an Introduction," *Journal of the Meteorological Society of Japan*, vol. 75, no. 1B , pp. 191-209, Mar. 1997.

[4] D. McLaughlin, "Recent Developments in Hydrological Data Assimilation," *Reviews of Geophysics,* vol. 33, pp. 977-984, 1995.

[5] R.E.Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans ASME J Basic Eng,* vol. 82, pp. 35-45, 1960.

[6] R. N. Miller, M. Ghil, and F. Gauthiez, "Advanced Data Assimilation in Strongly Nonlinear Dynamical-Systems," *Journal of the Atmospheric Sciences,* vol. 51, pp. 1037-1056, Apr 1994.

[7] G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review,* vol. 126, pp. 1719-1724, Jun 1998.

[8] C. Huang, X. Li, L. Lu, and J. Gu, "Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter " *Remote Sensing of Environment,* vol. 112, pp. 888-900, 2007.

[9] J. L. Anderson, "An ensemble adjustment Kalman filter for data assimilation," *Monthly Weather Review,* vol. 129, pp. 2884-2903, 2001.

[10] G. Evensen, "Sequential Data Assimilation with a Nonlinear Quasi—Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistic," *Journal of Geophysical Research,* vol. 99, pp. 10143- 10162, 1994.

[11] X. Han and X. Li, "An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation," *Remote Sensing of Environment,* vol. 112, pp. 1434-1449, Apr 2008.

[12] C. L. Huang, X. Li, and L. Lu, "Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter," *Remote Sensing of Environment,* vol. 112, pp. 1320-1336, 2008.

[13] W. T. Crow and D. Ryu, "A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals," *Hydrology and Earth System Sciences,* vol. 13, pp. 1-16, 2009.

[14] J. D. Bolten, W. T. Crow, X. W. Zhan, T. J. Jackson, and C. A. Reynolds, "Evaluating the Utility of Remotely Sensed Soil Moisture Retrievals for Operational Agricultural Drought Monitoring," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 3, pp. 57-66, Mar. 2010.

[15] R. Reichle, J. Walker, R. Koster, and P. Houser, "Extended versus ensemble filtering for land data assimilation," *J. Hydrometeor,* vol. 3, pp. 728–740, 2002.

[16] R.C. Pipunic, J.P. Walker and A. Western, "Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1295-1305, Apr. 2008.

[17] Y.K. Xue, P.J. Sellers, et al., "A Simplified Boisphere Model for Global Climate Studies," *Journal of Climate*, vol. 4, no. 3, pp. 345-364, Mar. 1991.

[18] P.J. Sellers, Y. Mintz, Y.C. Sud and A. Dalcher, "A Simple Biosphere Mode (SIB) for Use within General Circulation Models," *J. Atmos. Sci.*, vol. 43, no. 6, pp. 505-531, Mar. 1986.

[19] D.Q. Zhu, X.Q. Gao, W. Chen, "alidation of SSiB M odel over Gobi in Dunhuang and its Sensitivity to Vegetation Parameters," *Journal of Desert Research*, vol. 26, no. 3, pp. 466-472, May. 2006.

[20] C. J. Moore, "Frequency response corrections for eddy correlation systems," *Boundary-Layer Meteorology*, Vol. 37, pp. 17-35, Apr. 2004.

[21] E. K. Webb , G. I. Pearman , R. Leuning." Correction of flux measurements for density effects due to heat and water vapour transfer," *Quarterly Journal of the Royal Meteorological Society*, vol. 106, pp. 85-100, Jan. 1980.