ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **10**: 127–131 (2009)
Published online 20 April 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/asl.220

Royal Meteorological Society

Soil model parameter estimation with ensemble data assimilation

Biljana Orescanin,[1] Borivoj Rajkovic,[1] Milija Zupanski[2]* and Dusanka Zupanski[2]
[1] *Department of Meteorology, Faculty of Physics, University of Belgrade, Belgrade, Serbia*
[2] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, U.S.A*

**Correspondence to: Milija Zupanski, Cooperative Institute for Research in the Atmosphere, Colorado State University, 1375 Campus Delivery, Fort Collins, CO 80523-1375, U.S.A. E-mail: zupanskim@cira.colostate.edu*

Received: 12 January 2009
Revised: 5 March 2009
Accepted: 6 March 2009

Abstract

A parameter estimation problem in context of ensemble data assimilation is addressed. In an example using a one-point soil temperature model, the parameters corresponding to the emissivity and to the effective depth between the surface and the lowest atmospheric model level are estimated together with the initial conditions for temperature. The nonlinear synthetic observations representing various fluxes are assimilated using the Maximum Likelihood Ensemble Filter (MLEF). The results indicate a benefit of simultaneous assimilation of initial conditions and parameters. The estimated uncertainties are in general agreement with actual uncertainties. Copyright © 2009 Royal Meteorological Society

Keywords: **data assimilation; soil temperature model; ensemble**

1. Introduction

Data assimilation is typically understood as a process of combining a model with observations as a way of estimating the optimal state of the system. Data assimilation algorithmic framework provides other very important applications beyond initial state determination, such as the estimation of (1) the model bias, (2) the empirical model parameters, and (3) the information content of observations. Improving on the definition of the initial state of a system is an essential step toward predicting the evolution of the state and thus essential to numerical weather prediction (NWP). Improvement of model bias and empirical model parameters has an impact on weather prediction, but even more so on climate prediction, when the impact of the initial conditions is less relevant. As such, assimilation emerges as an important tool in evaluating models and improving model parameterization methods, as well as for evaluating observing systems and maximizing the value of observations.

Advanced data assimilation methods can also quantify the uncertainty of the optimal estimate, thus essentially providing the information about the posterior probability density function (PDF). Among most promising advanced data assimilation methodologies is the ensemble data assimilation (EnsDA), which includes the algorithms such as the ensemble Kalman filter (EnKF) (Evensen, 1994, 2003; Houtekamer and Mitchell, 1996, 1998), reduced-order Kalman filters (Verron *et al*., 1999; Hemmink and Verlaan, 2001), square-root filters (Bishop *et al*., 2001; Anderson, 2001, 2003; Whitaker and Hamill, 2002; Ott *et al*., 2004; Zupanski, 2005).

Among most challenging problems of data assimilation in general is the empirical model parameter estimation. There are several reasons for the difficulty of parameter estimation – key model parameters are not directly observable, and parameters are often nonlinearly related to observations and to model state. Being important for improving model parameterization, and thus for NWP and climate prediction, there has been a considerable interest in developing feasible and reliable methods for parameter estimation (Forest *et al*., 2000; Andronova and Schlesinger, 2001; Knutti *et al*., 2002; Gregory *et al*., 2002; Annan *et al*., 2005; Zupanski *et al*., 2007).

In this work, we address the issue of empirical parameter estimation in a simple, yet challenging problem of a one-point soil temperature model with assimilation of nonlinear observations. We evaluate the capability of ensemble data assimilation, in our case the Maximum Likelihood Ensemble Filter (MLEF) (Zupanski, 2005; Zupanski and Zupanski, 2006) to simultaneously estimate the initial conditions and model parameters. The paper is organized as follows. In section 2 we describe the experiment, results are discussed in section 3, and in section 4 conclusions are drawn.

2. Experimental design

In this section, we briefly describe the EnsDA algorithm, the prediction model, and the experiments.

2.1. Ensemble data assimilation

In this paper we employ the MLEF algorithm, an EnsDA algorithm based on control theory (Zupanski, 2005; Zupanski and Zupanski, 2006). The MLEF is chosen because it is well suited for use with highly nonlinear observation operators, since it employs an iterative nonlinear minimization algorithm to solve the nonlinear and possibly non-differentiable analysis problem (Zupanski *et al.*, 2008). Model parameters can be estimated within the MLEF using a state augmentation method (Zupanski *et al.*, 2007). An advantage of this approach is that it automatically allows cross covariances between the state vector and parameters, otherwise very difficult to represent. As other EnsDA algorithms, the MLEF estimates the uncertainty of the analysis, in addition to estimating the optimal solution. In this paper, we use a common assumption that all errors are Gaussian distributed although it may not be always preferable for parameter estimation.

2.2. Model

The employed model is a one-dimensional soil model for wet bare ground. The model is based on the scheme for parameterization of surface processes developed by Miyakoda and Sirutis (1977), Miyakoda *et al.*, (1986) 1984) and Janjic (1990). The model calculates equation of thermal balance (temperature θ_s) in a thin surface layer of finite depth. Equation consists of contribution from seven fluxes: net incoming short wave solar radiation (${R_S}^{sw}$), net incoming long wave atmospheric radiation (R_A^{lw}), outgoing long wave radiation of the Earth surface (F^{lw}), heat transfer between the surface ground layer and shallow layers (F_g^{shall}), heat transfer between the surface ground layer and deeper layers (${F_g}^{deep}$), turbulent flux of the latent heat trough the air/ground interface (F^{LH}), and turbulent flux of the sensible heat trough the air/ground interface (F^{SH}).

The equation representing the thermal balance is:

$$\left(\frac{\rho_s}{c_s}\right)_E d_s \frac{\partial \Theta_s}{\partial t} = {R_S}^{sw} + R_A^{lw} + F^{lw} - F_g^{shall} - {F_g}^{deep} + F^{SH} + F^{LH} \quad (1)$$

where ${R_S}^{sw}$ and R_A^{lw} have prescribed values, ρ is density, c is thermal capacity, and d_s is depth of surface ground layer. The subscripts s and g denote surface and ground, respectively. The fluxes are calculated as:

$$\begin{aligned} F^{lw} &= a_1 \varepsilon (\theta_s)^4 \\ F_g^{shall} &= a_2 [R_S^{sw} + R_A^{lw} - a_1 \varepsilon (\theta_s)^4] \\ {F_g}^{deep} &= a_3 (\theta_s - \theta_g) \\ F^{SH} &= a_4 \frac{\theta_{LM} - \theta_s}{\Delta z_e} \\ F^{LH} &= \frac{a_5}{\Delta z_e} \end{aligned} \quad (2)$$

In Equation (2) the constants a_1, a_2, a_3, a_4, a_5 are defined in the model, and subscript LM defines the lowest level of atmospheric model.

The HAPEX-MOBILHY data was used as external forcing. Data used include: short wave solar radiation, long wave atmospheric radiation, precipitation, temperature and wind speed at 2 m, humidity, and surface pressure. Details of this data set are given in a series of publications documenting the HAPEX-MOBILHY experiment (e.g. Goutrobe *et al.*,1989; Goutorbe and Tarrieuu, 1991). The HAPEX-MOBILHY data used for this paper are for the period from January to April.

2.3. Observations

The true state is defined as one realization of the model with specified values for parameters $\varepsilon = 0.9$, $\Delta z_e = 2$ m and $\theta_s = 320$ K. Observations are obtained by adding Gaussian perturbations to the true state. The observations include five fluxes: F^{lw}, F_g^{shall}, ${F_g}^{deep}$, F^{LH}, F^{SH} with prescribed observation errors (standard deviations): stdev(F^{lw}) = 20, stdev(F_g^{shall}) = 5, stdev(${F_g}^{deep}$) = 0.7, stdev(F^{LH}) = 100, and stdev(F^{SH}) = 50. Note from Equation (2) that the fluxes are nonlinearly related to soil temperature and to the emissivity and effective depth parameters, implying nonlinear observation operators.

2.4. Experiments

The experiments are designed to test the value added due to data assimilation with parameter estimation. Therefore, we define two data assimilation experiments: (1) with simultaneous adjustment of the initial conditions and model parameters, and (2) with adjustment of the initial conditions only, while using the *incorrectly specified* model parameters. The first experiment illustrates the impact of augmented data assimilation that includes simultaneous estimation of both the initial conditions and model parameters. The second experiment mimics a common situation in data assimilation: state vector is estimated, while the model parameters have a predefined, often an incorrect value. In addition, we define a third, no-DA experiment, in which no data assimilation is performed. This experiment is used to define an upper limit to the soil temperature error.

In our example, the initial condition is defined as the soil temperature. Two parameters are estimated: the emissivity, denoted ε, and the effective depth between the surface and the lowest atmospheric model level, denoted Δz_e. The initial values used for the 'truth' are $\theta_s = 320$ K, $\varepsilon = 0.9$, and $\Delta z_e = 2$ m. In all other experiments the incorrect initial setup is $\theta_s = 293.3238$ K, $\varepsilon = 0.1$, and $\Delta z_e = 3$ m. Note that the chosen 'truth' values can be changed according to real situation. The initial ensemble perturbations are assumed Gaussian, with zero mean and standard deviations of 1 K for soil temperature, and 0.2 for both parameters. In the experiment (1) the augmented control variable is $x = (\theta_s, \Delta z_e, \varepsilon)$, and in the experiment (2) the control variable is $x = \theta_s$. The data assimilation cycle is 15 min, and there are 48 cycles, corresponding to a 12-h period. In all experiments we

DOI: 10.1002/asl

use four ensembles, sufficient for complete representation for the three-dimensional augmented variable of experiment (1).

3. Results

First we examine how the MLEF performs when simultaneously estimating the initial conditions and the model parameters (Experiment (1)). Although the assimilation problem is of small dimension (augmented control vector is only three-dimensional), the observation operators are nonlinear with respect to soil temperature and to chosen parameters, making the assimilation problem more challenging. As a verification measure, we used root mean square (RMS) error, in our case equal to absolute error since we have a one-point model. The following formula for RMS error is used for all variables θ_s, Δz_e, ε.

$$RMS = \sqrt{\frac{1}{N}\sum_i (x_i - x_t)^2} = |x - x_{true}|$$

In Figure 1, the estimation of the augmented control vector is shown during the full assimilation period of 12 h. One can see that all components of the control vector converge to true values. The emissivity parameter needs about 30 analysis cycles before it converges, while the effective depth parameter needs only three analysis cycles. The soil temperature converges after about ten cycles.

The uncertainty of the augmented control vector in the experiment (1) is shown in Figure 2. We show both the estimated uncertainty, defined as the analysis error variance, and the actual uncertainty. In our case the actual uncertainty, defined as $|x - x_{true}|$, is equal to the RMS error.

The effective depth parameter uncertainty (Figure 2(a)) shows a consistently overestimated uncertainty, still of the order of the actual uncertainty. For the emissivity parameter (Figure 2(b)) and of the soil temperature (Figure 2(c)) there is an underestimation of uncertainty during the initial adjustment of about ten cycles. After that, the estimated uncertainties are of the same order as the actual uncertainties. This is an encouraging result, given that most ensemble data assimilation algorithms have problem with underestimated uncertainty, prompting the need for error covariance inflation. The MLEF achieves satisfactory results without error covariance inflation, effectively avoiding the need for prescribing this value. For the soil temperature uncertainty the values are too small to be seen, but they are in general agreement, with a slight overestimation after the 18th cycle, and an underestimation after the 36th cycle.

We now examine the impact of data assimilation experiments on the soil temperature error (Figure 3). Since all experiments start from incorrect values for soil temperature and for parameters, we can examine the impact of data assimilation. As expected, the best result is obtained in the case when both the initial conditions and parameters are adjusted (Experiment (1)). The no-DA experiment shows the largest soil temperature error that exists in the system, given the incorrect initial values of soil temperature and parameters. After the initial adjustment, the temperature error is approximately constant, about 5–6 K. The experiment (2) produces somewhat better results than the no-DA experiment during the initial adjustment (first ten cycles), but then it becomes almost identical to the no-DA experiment. In other words, the state estimation

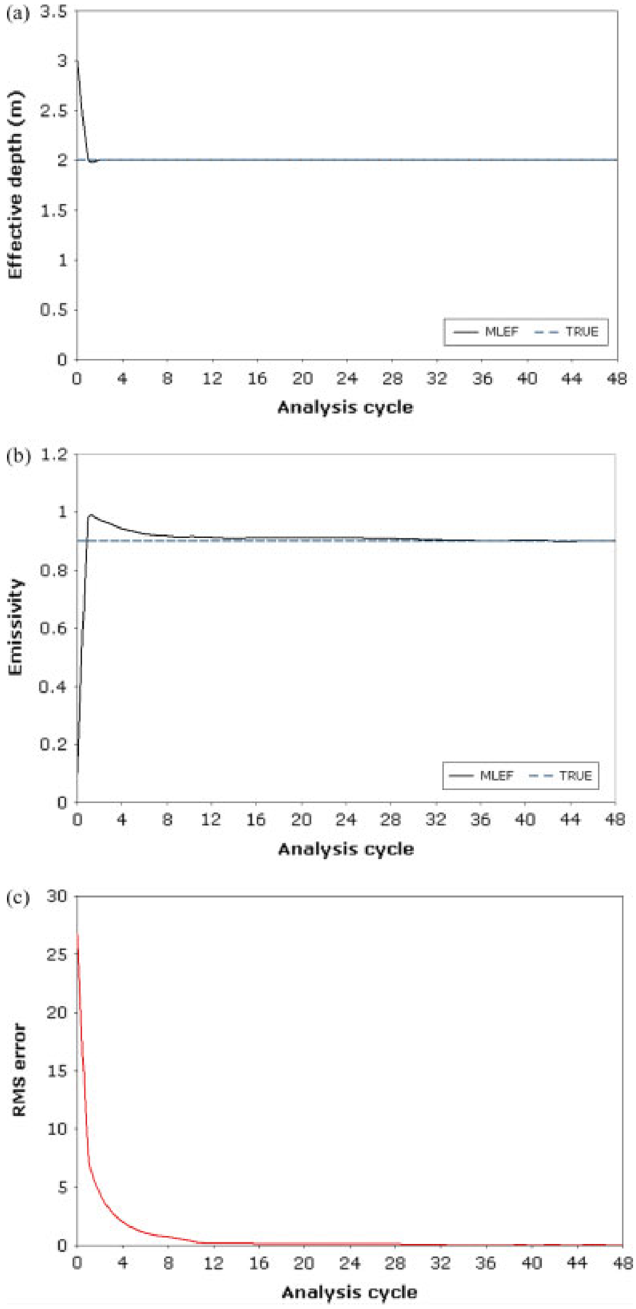

Figure 1. (a) Parameter Δz_e -true – represents the true value of parameter, MLEF – value of parameter obtained by EnsDA; (b) Parameter ε -true – represents the true value of parameter, MLEF – value of parameter obtained by EnsDA; (c) soil temperature error.

DOI: 10.1002/asl

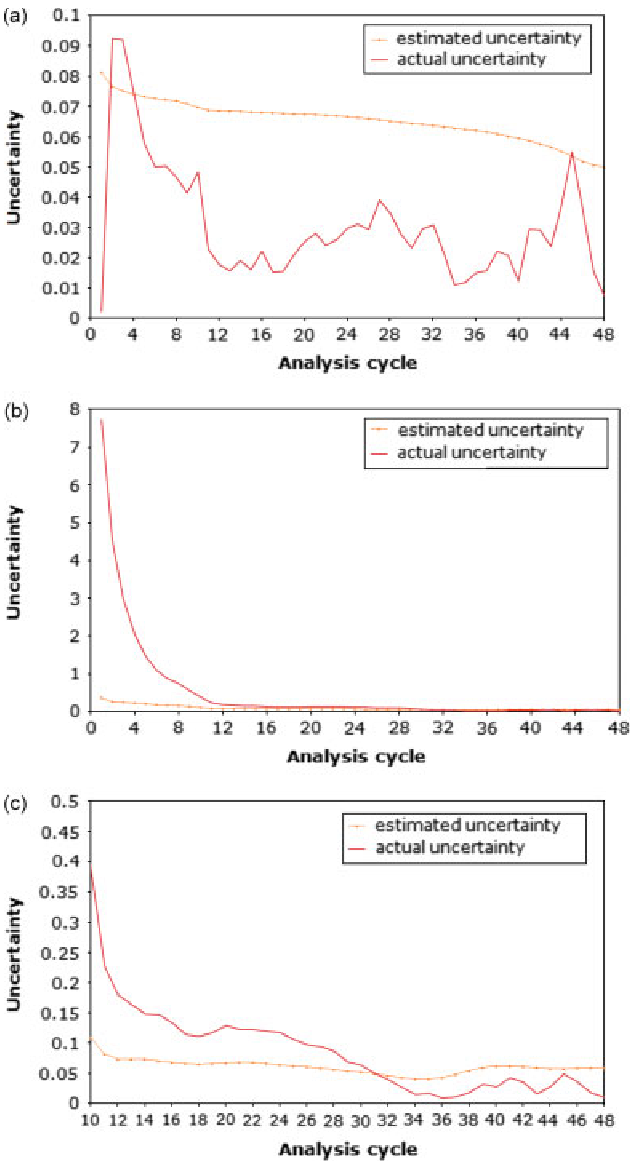

Figure 2. Uncertainty for (a) effective depth parameter, (b) emissivity parameter, and (c) soil temperature. *Solid red line* denotes the actual uncertainty, and the *dashed red line* the estimated uncertainty.

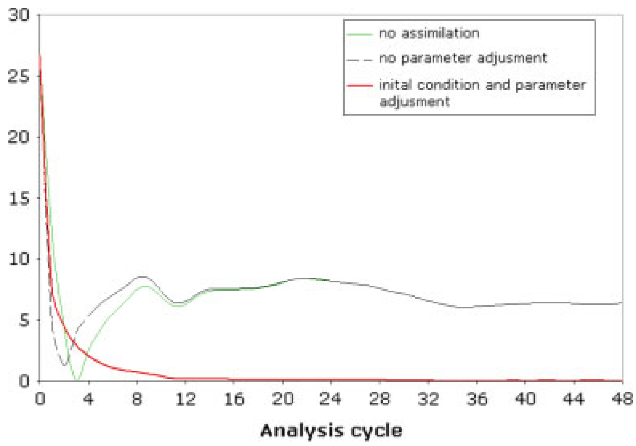

Figure 3. Impact of parameters and initial conditions on soil temperature error in the experiments without data assimilation (*green thick solid line*), with simultaneous assimilation of initial conditions and parameters (*red thin solid line*), and with assimilation of temperature only, but no parameter adjustment (*black dashed line*).

could not produce an improvement: there is almost no sensitivity to initial conditions. Apparently, the parameters are controlling the system, eventually leading to the same incorrect solution as without data assimilation. This indicates a potential impact of incorrect parameters in realistic data assimilation – even if the initial conditions are adjusted, data assimilation can lead to large forecast errors. Simultaneous adjustment of the initial conditions and several key model parameters is likely a better choice.

4. Conclusions

A simple one-dimensional soil temperature model was used to examine the capability of the MLEF to simultaneously estimate the initial conditions and model parameters. Synthetic observations represent fluxes, nonlinearly related to the soil temperature and to the chosen emissivity and effective depth parameters. The results confirm the MLEF potential to produce optimal values and reliable uncertainty estimates.

We also show that typical data assimilation that adjusts only initial conditions may not be appropriate if the model has incorrectly specified key parameters. As indicated by our experiments, this may lead to a large error of the prediction model. A simultaneous estimation of the initial conditions and model parameters seem to be a preferred choice.

In future, we plan to investigate the parameter estimation in more realistic environment using complex models and real observations, with applications to climate prediction and carbon data assimilation.

Acknowledgements

This paper is an extension of the diploma work done by the first author at undergraduate studies at University of Belgrade, Serbia. We greatly appreciate the help from Vladimir Djurdjevic. We are also indebted to Ana Vukovic, Milos Jovanovic, Jovanka Nikolic, and Veljko Petkovic for useful comments.

References

Anderson JL. 2001. An ensemble adjustment filter for data assimilation. *Monthly Weather Review* **129**: 2884–2903.

Anderson JL. 2003. A local least squares framework for ensemble filtering. *Monthly Weather Review* **131**: 634–642.

Annan JD, Hargreaves JC, Edwards NR, Marsh R. 2005. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling* **8**(1–2): 135–154. DOI: 10.1016/j.ocemod.2003.12.004.

Andronova NG, Schlesinger ME. 2001. Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research* **108**(D8): 22 605–22 611.

Bishop CH, Etherton BJ, Majumdar S. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Monthly Weather Review* **129**: 420–436.

Evensen G. 1994. *Sampling Strategies and Square Root Analysis Schemes for the EnKF*. Springer: Berlin. DOI: 10.1007/s10236-004-0099-2.

Evensen G. 2003. *The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation*. Springer: Berlin. DOI:10.1007/s10236-003-0036-9.

Forest CE, Allen MR, Stone PH, Sokolov AP. 2000. Constraining uncertainties in climate models using climate change detection techniques. *Geophysical Research Letters* **27**: 569–572.

Goutorbe JP, Noilhan J, Cuenca R, Valancogne C. 1989. Soil moisture variations during HAPEX-MOBILHY. *Annals of Geophysics* **7**: 415–426.

Goutorbe JP, Tarrieu C. 1991. HAPEX-MOBILHY data base. In *Land Surface Evaporation*,Schmugge, Andre (eds). Springer: Berlin; pp 403–410.

Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA. 2002. An observationally based estimate of the climate sensitivity. *Journal of Climate* **15**(22): 3117–3121.

Houtekamer PL, Mitchell HL. 1996. A system simulation approach to ensemble prediction. *Monthly Weather Review* **124**: 1225–1242.

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review* **126**: 796–811.

Heemink AW, Verlaan M, Segers AJ. 2001. Variance reduced ensemble Kalman filtering. *Monthly Weather Review* **129**: 1718–1728.

Janjic ZI. 1990. The step-mountain coordinate: Physical package. *Monthly Weather Review* **118**: 1429–1443.

Miyakoda K, Sirutis J. 1977. Comparative integrations of global models with various parameterized processes of sub grid-scale vertical transports: Description of the parameterizations. *Beitraege zur Physik der Atmosphaere* **50**: 445–487.

Miyakoda K, Sirutis J, Ploshay J. 1986. One-month forecast experiments without anomaly boundary forcings. *Monthly Weather Review* **114**: 2363–2401.

Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**: 273.

Knutti R, Stocker TF, Joos F, Plattner GK. 2002. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**: 719–723.

Verron J, Gourdeau L, Pham DT, Murtugudde R, Busalacchi AJ. 1999. An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation. *Journal of Geophysical Research* **104**: 5441–5458.

Whitaker J, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Monthly Weather Review* **130**: 1913–1924.

Zupanski M. 2005. Maximum likelihood ensemble filter. Part I: Theoretical aspects. *Monthly Weather Review* **133**: 1710–1726.

Zupanski D, Zupanski M. 2006. Model error estimation employing ensemble data assimilation approach. *Monthly Weather Review* **134**: 1337–1354.

Zupanski D, Denning AS, Uliasz M, Zupanski M, Schuh AE, Rayner PJ, Peters W, Corbin KD. 2007. Carbon flux bias estimation employing maximum likelihood ensemble filter (MLEF). *Journal of Geophysical Research* **112**: D17107. DOI: 10.1029/2006JD008371.

Zupanski M, Navon IM, Zupanski D. 2008. The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society* **134**: 1039–1050.