UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ENSEMBLE KALMAN FILTER DATA ASSIMILATION IN THE PRESENCE OF LARGE MODEL ERROR

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ANDREW TAYLOR
Norman, Oklahoma
2010

ENSEMBLE KALMAN FILTER DATA ASSIMILATION IN THE PRESENCE OF LARGE MODEL ERROR

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Lance Leslie, Chair

_____
Dr. David Stensrud

_____
Dr. Kelvin Droegemeier

_____
Dr. Petra Klein

_____
Dr. May Yuan

The heavens declare the glory of God;
the skies proclaim the work of His hands.

Day after day they pour forth speech;
night after night they display knowledge.

There is no speech or language
where their voice is not heard.

- Psalms 19:1-3, NIV

ACKNOWLEDGMENTS

I am thankful for a supportive committee. Funding from a grant acquired by Lance Leslie and Dave Stensrud enabled me to focus on my Ph.D. program without the pressures of an external job through the first four years. Lance's encouraging words have motivated me to press on and finish. I am very appreciative of the time Dave has spent with me over the last several years, watching seminar rehearsals, reading drafts, and offering valuable suggestions when I've met roadblocks. Many helpful tips from Petra Klein regarding sensible heat flux estimation were incorporated into this research. Kelvin Droegemeier went above and beyond what was necessary in reviewing my prospectus, and his recommendations have greatly improved this dissertation. May Yuan went out of her way to be available for meetings, despite her very busy schedule.

Though several substantial modifications have been made to the computer code used in this project, the basic code was inherited. Thanks are due to Lou Wicker, David Dowell, Tadashi Fujita and Dave Stensrud for their work on the EnKF code I used. Also, Evgeni Fedorovich generously provided code for the profile flux estimation method.

The management and staff at the Norman WFO have been in full support of my degree completion, permitting me to work on my research during quiet shifts. My wife, Jessica, refused to let me give up despite my repeated threats to quit. Countless words of encouragement have come from family and friends at just the right time. Though I threatened to quit, in my mind I knew I would not and could not. The Lord has taught me to dream big, and to follow through. He has fulfilled innumerable desires of my heart, and I owe Him everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Though assimilation of synthetically generated surface flux "observations" into numerical forecast models has been attempted, the topic of observed flux data assimilation over land has received little attention. This may be partially due to the lack of available flux observations in most areas. This study examines model analyses and observed data from the summer of 2004. Sensible heat flux estimates are calculated from Oklahoma Mesonet observations of temperature and wind using an iterative profile method reliant on boundary layer similarity theory. These sensible heat flux estimates are assimilated into the MM5 model, coupled with the Noah LSM, using an ensemble Kalman filter (EnKF) data assimilation scheme. A forecast is run over a 48 h period, with EnKF updates to the model grids occurring every hour. A control forecast with no assimilation is also run over the same 48 h period, as well as a forecast including the assimilation of standard surface data from the Oklahoma Mesonet. Results from these three forecasts are compared and contrasted.

It is shown that the EnKF scheme correctly updates the model low level temperature and moisture fields according to the values of the various terms in the governing equations. However, in the case of sensible heat flux data assimilation, model fields are sometimes updated in a way that creates analyses further from observed data rather than closer. Several factors are noted that make assimilation of sensible heat flux a challenge. Large differences between Mesonet estimates of sensible heat flux and horizontally interpolated values from the model can result in changes to the model forecast fields that are far too great. When combined with covariances generated by the model that are not

physically based, the problem is amplified. Consistency checks within the EnKF scheme show that the sensible heat flux observation error selected at the beginning of the study may be far too low, and that the ensemble spread is likely not nearly large enough to provide a reasonable estimate of the model error. These issues will need to be addressed before sensible heat flux assimilation can be expected to consistently perform well.

Chapter 1

Introduction

The importance to human activities of the state of the atmosphere and its effects on the land surface below cannot be overestimated. Spatial and temporal variations in temperature, atmospheric moisture content, wind and precipitation influence both business and recreational enterprises. Accurate forecasts of near-surface atmospheric variables such as temperature and wind direction/speed do more than just inform the general public of the best days for having a picnic outdoors. They also can mitigate large economic losses in various industries. For example, when the temperature is near a critical value, even a small forecast error could result in severe crop losses or unnecessary time and expense being used to protect crops in the case of orchards (Katz et al. 1982, Stewart et al. 1984). Errors in forecast wind direction/speed impact both aviation and electric power generation by wind turbines. Although many of the impacts resulting from variations in atmospheric moisture content and precipitation are longer-term, these variables have important short-term impacts as well. Taken together, accurate forecasts of near-surface variables also are crucial to the ability to correctly forecast episodes of destructive thunderstorms.

Numerical weather prediction (NWP) models simulate the evolution of the atmospheric state by starting with a set of initial conditions and numerically solving

the governing equations subject to boundary conditions. Due to the approximations involved and to uncertainty in the initial conditions, forecasts produced by such models tend to diverge from reality with time. Solutions to systems of non-linear differential equations, such as those that govern atmospheric evolution, can be extremely sensitive to small differences in initial conditions (Lorenz 1963). In order to bring the trajectory of the model back toward the "truth" periodically, assimilation schemes have been devised which use observations of atmospheric, ocean and land surface variables to adjust the model state according to the estimated errors in the observations and in the model forecasts. By improving the initial analyzed state of the atmosphere and the systems to which the atmosphere is coupled, such as the oceans and the land surface, the subsequent forecasts also should improve.

Surface observations are among the data available for assimilation into numerical models. These include observations of both the near-surface atmosphere and characteristics of the land surface itself. Assimilation of these observations can improve analyses of the low-level atmosphere and land surface. Standard atmospheric surface observations include observations of variables such as 2 m air temperature, 10 m wind speed, 2 m dew point/mixing ratio/specific humidity and station air pressure. Observations of land surface characteristics are far less common and include soil temperature and soil moisture at various depths, soil type and vegetation fraction. Special surface observations may include components of the surface energy balance such as received shortwave and longwave radiation, reflected shortwave radiation, emitted longwave radiation and surface heat fluxes.

Improvements in both analyses and forecasts have been realized by assimilating near surface atmospheric and land observations into NWP models (Crawford et al. 2001; Kurkowski et al. 2003). It would stand to reason that assimilation of additional surface observations not currently included in assimilation cycles could improve analyses and forecasts even further. The Oklahoma Mesonet (Brock et al. 1995; McPherson et al. 2007) provides several observation types that are not collected by the standard national surface observing network operated by the National Weather Service (NWS). These types include soil temperature and moisture observations from various depths, atmospheric radiation measurements, temperature and wind observations from a second level and surface heat flux measurements.

Soil temperature, and even more often soil moisture, observations have been assimilated into both simple and complex models of land surface processes for many years (Walker and Houser 2001; Crow and Wood 2003; Zhang and Fredericksen 2003). However, surface heat flux observations have not been assimilated directly into a coupled land-atmosphere model. Also, aside from small networks of surface stations and field programs, most surface flux observations have been estimated via remote sensing of surface temperatures and reflectances (Xinmei et al. 1993; McNider et al. 1994; Friedl 1996; Kustas et al. 1999). Despite the fact that satellite data have become available at higher and higher resolution, flux estimates are still often better using area-averaged data as the high-resolution data contain too many small-scale observation errors (Friedl 1996). In contrast, the Oklahoma Mesonet estimates sensible heat fluxes using a gradient method (Brotzge and Crawford 2000) at around 100 sites. Ground heat flux is estimated using two conducting plates

installed below ground at the same sites. At approximately 10 sites, extra instrumentation is added allowing approximate closure of the surface energy budget. The closure is approximate because latent heat flux is estimated as a residual.

Estimates of surface fluxes obtained from in situ measurements at Oklahoma Mesonet stations could add valuable information to land surface and low-level atmospheric analyses which has not been utilized thus far. This information could greatly improve the NWP model planetary boundary layer (PBL) initial state, which would improve modeling of atmospheric processes near the land surface. One method by which these data may be assimilated into land surface models (LSMs) and NWP models is the ensemble Kalman filter (EnKF) (Evensen 1994). The EnKF approach has many advantages over other data assimilation methods, including the ability to estimate the background error covariance matrix from the ensemble spread and easy parallel implementation on a supercomputer. Estimation of the background error covariance matrix makes the EnKF less costly than the extended Kalman filter. The ensemble created for use by the EnKF can be used for more than just improving data assimilation; the ensemble mean is, on average, a more accurate forecast than any of the individual members (Leith 1974).

Using the ensemble Kalman filter to assimilate Oklahoma Mesonet observations offers a novel way to test the value of new observations for improving numerical model initial conditions and forecasts of near-surface variables. In particular, this project will assess the assimilation of estimated sensible heat fluxes, in addition to standard surface observations.

Chapter 2

Review of related work

2.1 Ensemble forecasting

Deterministic forecasts of the evolution of the atmosphere over time inform the user of just one possible solution to the forecast problem. Numerous characteristics of the atmosphere suggest that a stochastic approach to numerical weather prediction (NWP) is more appropriate than a deterministic one (Epstein 1969). First, the initial state of the atmosphere can never be known exactly, even if observations are perfect. Whether NWP models represent the state of the atmosphere by a finite grid or using a superposition of waves, gaps exist both spatially and temporally in which the state of the atmosphere is unknown. The chaotic nature of the atmosphere (Lorenz 1963) means that an infinitesimal change in the initial conditions used to solve the governing equations may result in an entirely different solution. Therefore, accurate knowledge regarding the error distribution in the specified initial conditions is helpful for determining how much confidence may be placed in the initial analysis. The model itself can be the source of further errors in the forecast, as no NWP model is perfect. Sub-grid scale parameterizations of physical processes, as well as many

model parameters (constants), are approximations to reality and can cause the model to diverge from the true atmospheric state.

The idea of making stochastic weather forecasts has been around for decades. Epstein (1969) explored various methods of calculating uncertainty estimates as part of weather forecasts using the minimum hydrodynamic equations of Lorenz (1960). Since the prospect of integrating the continuity equation of probability density in full was not feasible, Epstein neglected third and higher moments to form a closed system of equations for the first and second moments. He measured the performance of this system against the performance of a group of forecasts made from a Monte Carlo distribution of analyses with a specified mean and variance. The validity of neglecting third moments as a closure assumption was investigated by Fleming (1971), who found that when third moments are important, neglecting them can eventually cause the estimate of the mean to drift away from the truth. An analysis method was developed using uncertainties modeled with first and second moments (Epstein and Pitcher 1972), and this method was eventually used to analyze real data (Pitcher 1977). Unfortunately, the stochastic dynamic forecast technique of directly solving an approximation to the continuity equation of probability was too computationally expensive for all but the very smallest model state vectors. Another approach was needed if stochastic forecasting was to become feasible in an operational context.

The focus changed from solving the continuity equation of probability density to forming forecast ensembles by making Monte Carlo selections from initial conditions with a specified mean and variance and integrating these analyses forward with the

model equations. This was the benchmark to which Epstein (1969) compared his stochastic dynamic forecasts. Gleeson (1970) envisioned a group of parallel forecasts as being a cloud of points in a phase space with the number of dimensions of the model. The density of the cloud of points at any given location and time was proportional to the probability of the atmospheric state being at that point in phase space. Later, this idea was applied to a turbulence model, and the probability density function (PDF) of the forecast was approximated by the ensemble mean and the accompanying covariance matrix (Leith 1974). Because the approximate PDF in such a system is defined by a mean and a variance, the PDF is assumed to be Gaussian.

Ensemble forecasts have many characteristics that make them an attractive alternative to deterministic or other stochastic forecasts. For instance, ensemble forecasts provide information about the range of possible outcomes through an estimate of the forecast variance, but without the necessity of linearizations and closure assumptions that are inherent in solving stochastic dynamic equations. The forecast PDF progresses forward in time according to the model equations, which are usually nonlinear. Ensemble forecasts also do not use as much computer power as would be necessary to make stochastic dynamic forecasts for systems with large state vectors (Leith 1974).

The ensemble approach to forecasting offers improved skill relative to that of deterministic forecasts. The ensemble mean forecast should have less error than the control (deterministic) forecast as long as the uncertainty in the initial analysis is properly represented (Epstein 1969; Leith 1974, Murphy 1988; Toth and Kalnay

1993). Using ensembles of only two members each, the ensemble mean forecast of 500 hPa streamfunction was shown to have less error than the control forecast, as long as the initial analysis errors were large enough (Houtekamer and Derome 1994). Many also relate the skill of the ensemble mean to the amount of spread among the ensemble members (Gleeson 1970; Murphy 1988; Houtekamer and Derome 1994; Buizza 1997; Hamill et al. 2000), but there is not always a straightforward relationship between these two elements of an ensemble forecast (Anderson 1996; Whitaker and Loughe 1998). Additionally, Whitaker and Loughe (1998) found the ensemble spread to be most useful as a predictor of the skill of the mean when the spread was very small or very large, or in regions of large day-to-day variability of spread.

Though much of the testing of the ensemble technique has been conducted under the perfect model assumption (i.e., the model is an exact representation of physical processes), model error may also be accounted for in an ensemble (Houtekamer et al. 1996; Houtekamer and Lefaivre 1997; Stensrud et al. 2000). Zhu et al. (2002) discussed the benefits of the improved skill of ensemble forecasts from an economic viewpoint, performing cost-loss analyses on ensemble forecasts and finding their economic value to users to be much greater than that of deterministic forecasts. Others have also investigated the economic value of ensemble forecasts by varying ensemble size and creating user decision models (Richardson 2000; Palmer 2002).

Though much debate continues about the best way to construct perturbations to ensemble members, many assert that random perturbations do not result in a large enough ensemble spread when the number of ensemble members is limited (Toth and

Kalnay 1993; Houtekamer and Derome 1995; Buizza and Palmer 1995; Molteni et al. 1996; Toth and Kalnay 1997). One possible way to determine fast-growing perturbations to add to ensemble members is the so-called singular vector method (Buizza and Palmer 1995; Molteni et al. 1996; Palmer et al. 1998; Barkmeijer et al. 1998). In this method, perturbations are dynamically constrained by using a linear tangent and adjoint of the forecast model to find the fastest-growing perturbations over a specified time interval. These perturbations are added to the model fields to form the ensemble members. Another dynamically constrained method, which does not require the use of a linear tangent or adjoint model, is the breeding of growing modes (BGM) (Toth and Kalnay 1993; Tracton and Kalnay 1993; Toth and Kalnay 1997). The BGM technique involves adding an initial arbitrary perturbation to an analysis, integrating both the control and the perturbed analyses forward for a few hours, subtracting the control analysis from the perturbed analysis, then taking the result and scaling it back so that it has the same root-mean-square (rms) size as the initial arbitrary perturbation. The rescaled difference is added back to the control as the new perturbation, and the process can be continued indefinitely. Toth and Kalnay (1993) found that the growth rate of their BGM perturbations was not dependent on the initial perturbation after about three or four days. By identifying the fastest-growing perturbations in a model, both the singular vector and BGM methods aim to create ensembles which sample as much of the tails of the forecast PDF as possible.

Recent research indicates that Monte Carlo methods of generating perturbations may sometimes be preferable to dynamically constrained methods. Hamill et al. (2000) discussed experiments for which the perturbed observation (PO) method,

where perturbed observations centered on the analysis PDF are added to first guess fields in parallel data assimilation cycles to initialize ensemble members for a forecast, performed better than either the BGM or singular vector methods despite the fact that the PO perturbations did not grow as fast. The PO probabilistic forecasts were improved and the resultant spread-skill relationships were more meaningful. However, the experiments were run using a quasi-geostrophic model, and the PO method has yet to be proven operationally. As an alternative, Hamill et al. (2000) suggested using different model configurations as in Stensrud et al. (1999), perturbing land surface variables, or adding stochastic forcing to the model equations. The first two of these suggestions, along with the use of the BGM perturbation method, make up a relatively cost-effective way to generate ensemble members and make ensemble forecasts accounting for variability of initial conditions.

Ensemble forecasting techniques are currently being utilized for both short-range and small-scale forecasts. A Short-Range Ensemble Forecasting (SREF) system was implemented operationally in May 2001 by the National Centers for Environmental Prediction (NCEP) (Du et al. 2006). The SREF system is made up of ensemble members derived from varying atmospheric forecast models with different parameterization schemes, horizontal resolutions, and initial conditions. Forecasts from the SREF system provide probabilistic guidance, which has been used with some success in the case of severe convective storms (Weiss et al. 2007; Guyer and Bright 2008).

Storm-scale ensembles have also been used in the deterministic prediction of thunderstorms (Kong et al. 2009; Xue et al. 2008). During a NOAA Hazardous

Weather Testbed (HWT) Spring Experiment, Kong et al. (2009) produced a remarkably accurate 30 h forecast of composite reflectivity across the central and eastern continental United States with a multi-model, multi-parameterization scheme ensemble forecast system. Because the ensemble mean tends to forecast too little precipitation, a technique known as probability matching was used to increase precipitation intensity while keeping the spatial distribution forecast by the ensemble mean. Xue et al. (2008) examined two cases from an earlier HWT experiment to assess the forecast improvements resulting from the assimilation of radar data. It was found that radar data assimilation had a positive effect on the forecast past 24 hours with weakly forced convection, but radar data had very little effect in strongly forced events.

2.2 Ensemble Kalman filter scheme

In addition to the useful variance information ensembles provide along with their forecasts, the use of ensembles can also help to improve data assimilation. Data assimilation systems are intended to solve the problem of finding the most likely state of the atmosphere given a background (first-guess) field and observations, both of which contain errors. One data assimilation method that yields the optimal linear estimate of the initial conditions, as long as the first- and second-order statistics of the observations and background fields are correctly specified (Blanchet et al. 1997), is the Kalman filter (Kalman 1960). The Kalman filter updates model fields according to the standard optimal interpolation (OI) analysis equation,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y}^o - H(\mathbf{x}^f)) \quad (1)$$

where $\mathbf{x}^a$ is the analysis vector, $\mathbf{x}^f$ is the background vector, $\mathbf{P}^f$ is the background error covariance matrix, $\mathbf{H}^{\mathrm{T}}$ is the linearized observation operator, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y}^o$ is the vector of observations, and $H()$ denotes the full observation operator. This notation follows that of Ide et al. (1997). The background error covariance matrix within the Kalman filter contains covariances between each model grid point and all of the other model grid points. Thus, the influence of an observation may be spread non-isotropically across the model grid. Equation (1) can be simplified by defining the Kalman gain $\mathbf{K}$,

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \quad (2)$$

The background error covariance matrix, which partially determines the Kalman gain, may be updated as

$$\mathbf{P}^f_{k+1} = \mathbf{F}\mathbf{P}^f_k\mathbf{F}^{\mathrm{T}} + \mathbf{Q} \quad (3)$$

if the forecast model is linear (Evensen 2003), where the subscripts are time indices, the linear forecast model is denoted $\mathbf{F}$, and $\mathbf{Q}$ is the model error covariance matrix. If the forecast model being used is nonlinear, the extended Kalman filter (EKF) may be used in lieu of the traditional Kalman filter (Reichle et al. 2002a), though the EKF tends to be unstable when the nonlinearity is strong (Miller et al. 1994). The

background error covariance matrix is still updated using eq. (3), but with **F** being the linear tangent to the nonlinear model operator $f()$.

An alternative to the EKF for which linear tangent estimates to the forecast model are unnecessary is known as the ensemble Kalman filter (EnKF), first discussed by Evensen (1994). Rather than calculating computationally expensive updates to $\mathbf{P}^f$ directly using eq. (3), the elements of $\mathbf{P}^f$ are estimated from the covariance information given by an ensemble of forecast states. The true value of $\mathbf{P}^f$ is defined in the Kalman filter as

$$\mathbf{P}^f = \overline{(\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^{\mathrm{T}}}, \tag{4}$$

where the superscript t denotes the true atmospheric state. The EnKF estimate of $\mathbf{P}^f$ is

$$\mathbf{P}^f \approx \mathbf{P}_e^f = \overline{(\mathbf{x}_i^f - \overline{\mathbf{x}^f})(\mathbf{x}_i^f - \overline{\mathbf{x}^f})^{\mathrm{T}}}, \tag{5}$$

where the subscript e denotes an ensemble estimate, $\overline{\mathbf{x}^f}$ is the vector of ensemble mean values, and the bar over the entire right-hand side of eq. (5) signifies the mean over all ensemble members i. The estimated $\mathbf{P}_e^f$ is substituted for $\mathbf{P}^f$ in the calculation of the Kalman gain defined in eq. (2).

The EnKF data assimilation technique has many advantages over other commonly used techniques such as three-dimensional variational (3D-Var) algorithms (Lorenc 1986; Parrish and Derber 1992) or four-dimensional variational (4D-Var) algorithms (Bennett et al. 1997; Rabier et al. 1998). First of all, as discussed earlier, ensemble forecasts provide variance information that is not available from deterministic

forecasts. The use of an ensemble to generate covariance information means that the covariances are flow-dependent, in contrast to the isotropic covariance structures typical of 3D-Var techniques. No tangent linear or adjoint models are necessary for the EnKF. These models can be costly to develop, and are typically used by both EKF and 4D-Var schemes. Depending on the number of ensemble members used, the computational costs of EnKF and 4D-Var schemes can be similar (Lorenc 2003). The covariances are updated in the EnKF by the fully nonlinear model, in contrast to the tangent linear model used in the EKF (eq. (3)). Solution of eq. (3) is not practical for high-dimensional state vectors, and the EnKF represents an approximation to the EKF that may improve operational data assimilation.

While EnKF schemes offer benefits that other schemes do not, their use also presents some unique challenges. In theory, an infinite ensemble should exactly determine the background error covariances, but because ensembles are finite they only approximate the covariances. Noisy or inaccurate spatial covariances can be calculated from finite ensembles, leading to strong correlations between distant locations. To alleviate this problem, Houtekamer and Mitchell (1998) used an influence radius, ignoring background covariances between positions greater than a specified distance from each other. Another approach has been to apply a Schur product (Houtekamer and Mitchell 2001; Hamill et al. 2001) by multiplying covariance estimates with a distance-dependent correlation function. Such solutions reduce the effects of distant correlations, but they may still result in a choppy, discontinuous analysis (Evensen 2003).

Lorenc (2003) pointed out some difficulties with the EnKF technique in the process of comparing it to 4D-Var. The error probability distribution function (PDF) calculated with EnKF methods will be imprecise in situations when the actual error PDF is strongly non-Gaussian. Quality control of observations is not as simple with a sequential algorithm used in the EnKF since all of the observations are not processed together. Lorenc (2003) suggested that a hybrid EnKF-4D-Var scheme modeling the evolution of covariances within assimilation windows using a variational fit and between windows using an EnKF approach could lessen the effects of these difficulties. Bennett et al. (1996) and Bennett et al. (1997) developed a four-dimensional variational scheme known as the method of representers that can account for model error which is in principle equivalent to Kalman filtering (Kalnay 2003). It is also faster than the traditional 4D-Var scheme because the work is done in observation space rather than model space. This approach avoids model linearization, solving the nonlinear problem by iteration. However, the method also carries many disadvantages that make use of the EnKF more practical (L. Leslie, personal communication).

Many types of EnKF algorithms exist that are suited for a wide range of data assimilation problems. One algorithm involves the use of a pair of ensembles, with the covariances developed with one ensemble used to spread the influence of observations in the other ensemble (Houtekamer and Mitchell 1998; Mitchell and Houtekamer 2000; Houtekamer and Mitchell 2001; Houtekamer et al. 2005). This algorithm was developed to avoid an inbreeding problem arising from updating an ensemble with covariances determined using the same ensemble. Such a practice

could underestimate the uncertainty in a forecast (Houtekamer and Mitchell 1998). Evensen and van Leeuwen (2000) developed an ensemble Kalman smoother which propagates an EnKF solution backward in time using ensemble covariances; it is useful for computing climatologies, but not so much for forecasting. The ensemble adjustment Kalman filter (EAKF; Anderson 2001) and the ensemble square root filter (EnSRF; Whitaker and Hamill 2002, Tippett et al. 2003, Xue et al. 2006) each avoid the use of perturbed observations in generating analyses. This is an advantage since for finite ensembles, perturbed observations constitute an additional source of sampling error. The ensemble transform Kalman filter (ETKF; Bishop et al. 2001, Wang and Bishop 2003) allows faster calculation of background covariances than the traditional EnKF and has been used for finding the spatial arrangement of observations that most reduces the forecast error variance. Finally, hybrid schemes (Hamill and Snyder 2000; Bengtsson et al. 2003; Lorenc 2003) attempt to overcome some of the limitations of the EnKF while retaining its strengths.

Though a relatively recent innovation, the EnKF has already been used in a wide variety of applications. The first general area to which the EnKF was applied was ocean data assimilation (Evensen 1994; Evensen and van Leeuwen 1996; Blanchet et al. 1997; Keppenne 2000; Keppenne and Rienecker 2002). More recently, the EnKF has been the method of choice for many experiments in radar data assimiliation (Snyder and Zhang 2003; Dowell et al. 2004; Zhang et al. 2004; Caya et al. 2005; Tong and Xue 2005; Xue et al. 2006). Most documented EnKF studies thus far have involved the assimilation of synthetically generated data, but some (e.g., Crow 2003, Dowell et al. 2004, Houtekamer et al. 2005, Fujita et al. 2007) have used real

observations as well. The EnKF also has been applied to the assimilation of mesoscale and land surface data (e.g., Reichle et al. 2002a, Reichle et al. 2002b, Crow 2003, Zhang et al. 2006, Fujita et al. 2007). Often, satellite microwave brightness temperature has been assimilated as a proxy for soil moisture or rainfall (Reichle et al. 2002a,b; Crow 2003). Studies documented in the literature have only used synthetic and satellite observations of soil properties. However, such measurements are not necessarily directly linked to soil moisture in heavily vegetated regions, and an in situ system would provide observations that would likely improve upon those from satellites. The ability of EnKF-based schemes to account for various kinds of model error is a big advantage for soil moisture assimilation (Crow 2003), and the use of an EnKF scheme to assimilate real land surface observations from an in situ network has great potential for improvement over current operational methods.

2.3 Land-atmosphere interactions

2.3.1 *Surface energy budget*

To adequately simulate atmospheric processes such as planetary boundary layer (PBL) evolution, the growth and decay of convective storms, and mesoscale circulations, the effects due to the coupling between the surface of the earth and the atmosphere must be taken into account. In addition, the states of atmospheric variables such as temperature and mixing ratio near the ground are closely related to

the state of the land surface below. Small-scale variability in atmospheric conditions can develop due to small-scale variability in the condition of the land surface.

The link between the land surface and the atmosphere can be partially understood in terms of the surface energy budget, which describes the partitioning of net radiative energy into various components. Shortwave radiation from the sun enters Earth's atmosphere, after which it may be reflected back to space or transmitted to and absorbed by Earth's surface. Earth emits longwave radiation to the atmosphere, and a portion of the emitted longwave radiation can be absorbed by clouds and re-emitted back to Earth's surface. The sum of these emissions and absorptions gives the net radiation absorbed by Earth's surface. The surface energy budget is balanced by three fluxes of energy, namely the sensible heat flux, latent heat flux, and ground (or soil) heat flux. Sensible heat flux refers to a transfer of enthalpy from the land surface to the air, thus heating the air. Latent heat flux refers to the transfer of energy from the land surface to the air due to the evaporation of water from lakes, oceans, vegetation, soil, etc. The ground heat flux is that portion of the net radiation that results in a heating (or cooling) of sub-surface soil. Each of these fluxes varies according to annual, intraseasonal, synoptic, and diurnal time scales (Smith et al. 1992). Changes in the thermodynamic profile of the atmosphere due to surface energy fluxes, as well as local-scale variability in the Bowen ratio (the ratio between sensible and latent heat flux), can affect the near-surface temperature, humidity and wind, in addition to cloudiness and precipitation.

The magnitudes and directions of the surface energy fluxes are affected by a multitude of factors, some atmospheric and some surface-based. The surface albedo

determines the amount of solar radiation that is reflected to space, and therefore the amount of solar radiation that is absorbed by the earth and available to be partitioned into the various fluxes. Cloudiness also affects the amount of radiation received by reflecting solar radiation to space as well as absorbing some of the earth's radiation and re-emitting it back to the surface. Soil type and texture help to determine how much water may be held by the soil (field capacity), the wilting point of vegetation, and the thermal and hydraulic conductivity. The amount of evaporation from bare soil, and therefore the latent heat flux from such surfaces, is a nonlinear function of soil moisture and is often very sensitive to small changes in its magnitude (Entekhabi and Eagleson 1989). The spatial and seasonal extent of vegetation also affects the components of the surface energy budget in many ways. Different vegetation types have different albedos, the effect of which has been discussed previously. Plants also regulate the amount of evaporation from the surface through their stomatal resistances, interception of precipitation by leaves, and according to the depth of their roots. Subsequently, the Bowen ratio varies spatially and seasonally as well, affecting surface sensible heating and near-surface atmospheric moisture content.

2.3.2 *Importance of surface-atmosphere coupling to forecasting*

Land-atmosphere interactions are important on many different spatial and temporal scales, and modeling those interactions takes different forms depending on the scales of interest. Land surface processes have been included in climate models for many years because parameters such as albedo, vegetation cover, land/water

distribution, and soil moisture vary widely on such long time scales (Entekhabi and Eagleson 1989; Sato et al. 1989; Koster and Suarez 1996). Due to computer power limitations and the long time integrations necessary to model the earth's climate, global climate models (GCMs) tend to have a relatively low resolution in space and time. Land surface characteristics are highly variable spatially, so initial grid point values of these characteristics must be generated via some type of spatial averaging (Koster and Suarez 1992). Variable results often are produced by different methods for calculating composite grid point values of land surface parameters (Abramopoulos et al. 1988; Koster and Suarez 1992; Bonan et al. 1993).

Using composite land surface parameter values in a hydrology model, Abramopoulos et al. (1988) found that as the size of their grid boxes decreased from 8ºx10º to 1ºx1º, the composite values were less representative and higher resolution averaging methods were needed. Koster and Suarez (1992) outlined two major approaches to sub-grid scale parameterization of surface vegetation: the "mixture" approach and the "mosaic" approach. When the "mixture" approach is taken, a homogeneous mix of vegetation types is assumed over the entire grid box, while the "mosaic" approach allows the atmosphere to interact with each vegetation type separately. The authors found that the "mixture" approach produced larger latent heat flux and total turbulent flux values when the resistances of the vegetation types differed by more than a factor of 2, but otherwise the choice of method was somewhat arbitrary. The "mosaic" method may be preferable as it is more computationally efficient, and it also may more clearly model mesoscale circulations such as sea breezes (Avissar and Pielke 1989). Bonan et al. (1993) measured the effects of

accounting for sub-grid scale heterogeneity in leaf area index, stomatal resistance, and soil moisture in terms of grid-scale fluxes.

Climate forecasts display a strong sensitivity to soil moisture variations, partially due to their long duration. Soil moisture changes on scales of months (Paegle et al. 1996; Beljaars et al. 1996) to years affect climate variability. In addition, care must be taken when spatially averaging soil moisture over a grid box, since due to the nonlinear relationship between evaporation and soil moisture the evaporation for the average soil moisture condition is not necessarily equivalent to the average evaporation for the distribution of soil moisture conditions (Entekhabi and Eagleson 1989). Dirmeyer (1994) ran simulations with a model earth having one rectangular continent covered with prairie vegetation, finding that low initial soil moisture resulted in a decrease of precipitation. The long-term persistence of soil moisture anomalies and their effects on atmospheric circulations and precipitation was the driving force behind the assimilation of remotely sensed soil moisture observations by Walker and Houser (2001). Zhang and Fredericksen (2003), using a general circulation model, studied the effects on modeled precipitation patterns in China during the summer of 1998 due to initializing soil moisture with NCEP-NCAR reanalysis data. They discovered that soil moisture anomalies had a direct impact on surface temperatures, but an indirect impact on precipitation in that the soil moisture anomalies altered the atmospheric circulation, bringing in additional moisture from other areas. Soil moisture initialization provided more accurate temperature and precipitation forecasts than did sea surface temperature (SST) updating, even with a simple bucket soil model.

Basic land surface parameters such as albedo and soil moisture (e.g., moisture availability, defined by Manabe (1969)) have been included in climate models for some time. However, as computing capacity has increased, it has become possible to add further complexity to land surface models while still completing forecasts in a reasonable amount of time. One ground hydrology model simulating major hydrologic processes for a GCM was most sensitive to fractional vegetation cover and canopy resistance, two variables that were neglected in most previous models (Abramopoulos et al. 1988). Sato et al. (1989) used the Simple Biosphere (SiB) model (Sellers et al. 1986) with a GCM, comparing it to the use of a more standard land surface model for the time. SiB produced a more realistic partitioning of the surface energy balance and modeled the interrelationships between albedo, surface roughness, and soil moisture instead of treating them as independent of each other. Later, SiB was improved (Sellers et al. 1996) by adding parameters for CO_2 fluxes from the canopy and fractional snow cover. Additionally, in order to accurately model the surface energy balance, mesoscale fluxes must be considered as well as turbulent fluxes (Pielke et al. 1991). The inclusion of extra complexity in land surface models for use with GCMs has increased the accuracy of land surface process representations in many cases.

Though many land surface variables fluctuate on a long time scale, land-atmosphere interactions are also important considerations in short-term mesoscale forecasts. First of all, land surface conditions affect the initiation and the evolution of convective storms. Benjamin and Carlson (1986) showed that the surface can influence convective initiation through sensible heating and evaporation by affecting

cloud and stability patterns. They stated that an accurate representation of PBL/surface processes and topography can "provide significant improvement in 12-36 h forecasts of an environment favorable for severe storms." In idealized two-dimensional experiments, Benjamin (1986) found that the state of the elevated mixed layer (EML) over the U.S. Southern Plains depends on horizontal gradients of soil moisture, surface heating, and elevation. The EML allows the air below to warm, delaying and focusing storm initiation. Experiments carried out using four different soil moisture distributions in a mesoscale model showed that the distribution of soil moisture influences the EML, the dryline, precipitation amount, ageostrophic flow, and PBL instability (Lanicci et al. 1987). With abundant atmospheric moisture, convective clouds tend to form first over areas where the Bowen ratio is the smallest, and vice versa for dry days (Rabin et al. 1990), in apparent contradiction to the Charney hypothesis (Charney 1975). However, the clouds which formed on dry days were non-precipitating and increased the albedo. Rabin et al. (1990) also pointed out that land surface effects tend to be most apparent on days with weak atmospheric forcing, and that the role of the land surface coupled with strong atmospheric forcing had not been determined.

Further research has continued to provide insight into the influence of the land surface on the development of convective storms. Chang and Wetzel (1991) estimated soil moisture from antecedent precipitation and vegetation cover from NDVI, with the best simulation of a storm complex from 3-4 June 1980 resulting from the inclusion of soil moisture and vegetation variations. These variations produced differential heating leading to thermally driven circulations which

subsequently influenced storm initiation. Crook (1996) examined the sensitivity of convective storms to surface processes under weak atmospheric forcing, finding that these storms were sensitive to surface temperature and moisture dropoff (the difference between the surface potential temperature and moisture and those quantities in the well-mixed boundary layer). This sensitivity indicates a benefit to improving analyses of surface temperature and moisture, and therefore fluxes of the same.

In situ land surface observational data have also been assimilated into models in an effort to improve convective forecasting. Trier et al. (2004) used the MM5 model with the Grell cumulus parameterization and the Noah LSM to model development of convection over the Southern Plains in June 1998, assimilating soil moisture data with a high resolution land data assimilation system (LDAS). Soil moisture variations result in differences in the partitioning of surface fluxes, which create differential heating leading to small-scale solenoidal circulations. These circulations can help to initiate deep moist convection. A comparative run with coarser soil moisture data drawn from the Eta Data Assimilation System (EDAS) failed to capture storm initiation at the correct locations despite forecasting a similar dryline structure to the high-resolution LDAS run. Using various levels of complexity in their model as well as different initial soil moisture fields, Holt et al. (2006) demonstrated that more accurate land surface modeling could improve convective forecasts even in the presence of strong synoptic forcing. However, even the most complex LSM with the best soil moisture initialization initiated convection about 2-3 h late. The COAMPS model with the Weather Research and Forecasting (WRF) model version of the Noah

LSM was used to produce the convective forecasts for a case from the International H_2O Project (IHOP) in May 2002. Variations included using soil moisture fields from EDAS and from a higher-resolution source, adding a photosynthesis model to the LSM, and using a simple slab soil model.

The modeling of land surface characteristics such as soil moisture content and vegetation amount/type is also important to resolving mesoscale circulations and diagnosing mesoscale boundaries. Inferring soil moisture from infrared satellite temperature measurements and assuming weak synoptic-scale forcing, Ookouchi et al. (1984) used the model described in Pielke (1974) to study the effects of mesoscale variability in soil moisture. In their model, a substantial soil moisture contrast on flat terrain produced a circulation similar in strength to a sea breeze, showing that variations in the Bowen ratio may likely produce atmospheric circulations. Other studies regarding the initiation of atmospheric circulations due to soil moisture contrasts have been done (Yan and Anthes 1988). Work with a two-dimensional mesoscale model including detailed PBL, soil and vegetation parameterizations has shown that mesoscale circulations can be generated due to variations in soil texture alone because of their impact on moisture availability (Mahfouf et al. 1987). The authors also simulated a boundary between bare soil and vegetation, finding that the boundary was a favored location for the initiation of moist convection due to the atmospheric circulations arising from the surface flux differences. A similar study found that the effects of the land surface on atmospheric circulations were highly dependent on the extent of vegetated areas and on environmental conditions (Segal et al. 1988). Pinty et al. (1989) pointed out that the response of the PBL to soil moisture

and vegetation gradients within a model is sensitive to the model parameters, such as whether the soil is moist or dry in general. Many models neglect the changes in mixed-layer depth due to sub-grid scale sensible heat flux variability, but it has been shown that these changes can be significant (Doran and Zhong 1995). Accounting for the variability in mixed-layer depth with a simple average value is dangerous, as the distribution of depths tends to be bimodal.

Many studies have attempted to find mesoscale circulations appearing in model simulations present in actual observed data. Instead of using prescribed vegetation grids as in much past modeling work, Segal et al. (1989) obtained a real-world vegetation-irrigation "map" of an area in northeastern Colorado from GOES satellite, aircraft, and surface observations. The expected thermally driven circulations between dry and irrigated areas were only weakly reflected in the observations, but model simulations using the "map" showed that this was likely due to terrain effects and strong synoptic-scale flow. Segal and Arritt (1992) defined nonclassical mesoscale circulations (NCMCs) as circulations caused by gradients in sensible heat flux that are not sea or lake breezes. They explained that additional factors that can mask the presence of NCMCs in observations are the spatial and temporal non-uniformity of sensible heat flux and the small-scale spatial variability in soil moisture and land use, both of which are difficult to account for in numerical models. Perturbed areas (PAs) with differing sensible heat flux from surrounding areas generate mesoscale circulations, with the most common PAs appearing due to cloudiness.

Mesoscale boundaries and circulations observed in the atmosphere usually are not isolated, but instead multiple circulations and boundaries combine with each other to generate the actual observations. Using the RAMS model to examine the characteristics of drylines, Shaw et al. (1997) derived soil moisture from antecedent precipitation and vegetation from the NDVI to examine how these parameters affect the dryline. They concluded that soil moisture initialization is important to dryline formation and evolution in a model, but it is also true that the position of the dryline affects the soil moisture. Grasso (2000) also used RAMS to simulate a dryline, and he found that the dryline did not even form under a uniform soil moisture distribution. Yet another circulation that falls under the category of NCMCs is the so-called "salt breeze" (Rife et al. 2002), which results from differential heating due to the contrasting land surface characteristics between salt flats and the surrounding sand. Rife et al. (2002) conducted experiments using the MM5 model coupled to an LSM to study the interactions between salt breezes and other mesoscale circulations in the area due to variable terrain and lakes, which also result from sensible heat flux gradients.

Examinations of the influence of the land surface on atmospheric structure and evolution show that the sensible, latent and ground heat fluxes are very important to the determination of near-surface weather variables such as temperature, water vapor mixing ratio and wind. The partitioning of surface energy between sensible and latent heat fluxes is determined by factors such as the amount and type of vegetation, the soil type and the soil moisture. Therefore, the quality of the initialization of such

fields in numerical models helps to determine the quality of forecasts of near-surface variables (Kurkowski et al. 2003; Marshall et al. 2003; Segele et al. 2005).

It has been stated that soil moisture is the single most important characteristic in determining the influence of the surface on the atmosphere (McCumber and Pielke 1981). Not only does soil moisture affect the near-surface mixing ratio, but it also affects the albedo and the soil thermal conductivity which help to determine the amount of surface energy available to be partitioned into fluxes. Carlson and Boland (1978) found that the most important factors affecting near-surface temperature variation were thermal inertia and moisture availability, with moisture availability having a larger influence during the day and thermal inertia having a larger influence at night. Surface heating affects PBL structure and wind circulations. Mesoscale circulations simulated between irrigated land and dry areas, though sometimes masked by terrain effects and synoptic conditions in the actual observations, occur due to sensible heat flux gradients (Segal et al. 1989). A connection was also found between evapotranspiration and precipitation through comparison of latent heat flux with rainfall amounts at the ARM-CART site in northern Oklahoma, though the relationship varied strongly with atmospheric circulation (Sud et al. 2001). A much stronger relationship was seen for moist cases than for drought circulations. Oncley and Dudhia (1995) performed a comparison of MM5 model output surface fluxes to observations, discovering that the model fluxes were most sensitive to the choices of surface roughness and moisture availability. In order to more accurately model near-surface weather conditions, therefore, modeling of these parameters was improved.

2.3.3 *Modeling of land surface processes*

All of the parameters necessary to calculate the values of many land surface variables are not known exactly, so approximations to these variables must be made. Examples of approximations include various methods for estimating the evaporation from the land surface. Penman (1948) combined turbulent transport and surface energy balance approaches to modeling evaporation, developing an equation that makes theoretical estimates of evaporation rates from standard meteorological data. This was a step on the way to the development of the Penman-Monteith evaporation model, first stated by Monteith (1965) as

$$\lambda E = \frac{\Delta(R_{net} - G) + \rho c_p (\delta e) r_a^{-1}}{\Delta + \gamma(1 + \frac{r_c}{r_a})} \quad . \tag{6}$$

In eq. (6), Δ is the slope of the saturation vapor pressure curve with temperature, R_{net} is the net radiation at the surface, G is the ground heat flux, r_a and r_c are the aerodynamic and canopy resistances, respectively, ρ is the air density, c_p is the specific heat of air at constant pressure, δe is the air vapor pressure deficit, γ is the psychrometric constant, λ is the latent heat of vaporization, and λE is the latent heat flux. Priestley and Taylor (1972) formulated a similar equation for the evaporation rate,

$$\lambda E = \alpha \frac{s}{s + \gamma}(R_{net} - G), \tag{7}$$

where α is the empirical Priestley-Taylor parameter, *s* is the slope of the saturation specific humidity curve with temperature, and the other symbols have the same meaning as in eq. (6). Both eq. (6) and eq. (7) contain numerous assumptions, and the usefulness of each method varies depending on the application. Comparison of these methods with respect to parameterization of sensible and latent heat fluxes was undertaken by de Bruin and Holtslag (1982), who used a Priestley-Taylor based formulation which showed similar skill to the Penman-Montieth method for a grass surface with an unstable PBL. Only the air temperature, surface moisture, and net radiation were necessary inputs to run the parameterization.

The first LSMs were very basic with crude approximations to land surface processes, but many of those approximations are still used in current modeling efforts. Pan (1990) pointed out that the National Meteorological Center (NMC) global spectral model was using the bucket model of Manabe (1969) to parameterize evapotranspiration up until the late 1980s and showed that improvements were made in the near surface moisture field by using an alternate parameterization based on the Penman-Monteith model. An early effort at numerically modeling soil temperature and moisture was made by Philip (1957), who simulated the three stages of soil evaporation that he identified. Nappo (1975) investigated two methods for the parameterization of evapotranspiration, one based on moisture availability and the other based on the relative humidity just above the surface. Nickerson and Smiley (1975) developed an equation for the Monin-Obukhov length, parameterizing the latent heat flux, the ground heat flux and radiation terms to form the required inputs.

A parameterization of soil moisture content now known as the "force-restore" model (Deardorff 1977; Deardorff 1978) separated the soil into two layers so that the surface soil moisture content could be used as opposed to one bulk estimate, allowing evaporation to vary more realistically. The moisture content of the surface layer was estimated using the soil surface evaporation rate minus the precipitation rate. While the force-restore model accounted for variations in soil characteristics such as texture and type through the use of dimensionless empirical constants, McCumber and Pielke (1981) examined the sensitivity of their model to soil characteristics explicitly, finding soil moisture to be the most important factor in determining the influence of the land surface on the atmosphere. Noilhan and Planton (1989) used Deardorff's (1977) force-restore model to calculate soil properties as part of a broader parameterization which computed five prognostic land surface variables over time scales of a few days. The basic effects of vegetation were included.

Studies have also been done in which the effects on atmospheric models due to variations in the parameterization of land surface processes are assessed. Holtslag et al. (1990) designed a model initialized with radiosonde data to forecast lower level temperature and humidity profiles, PBL height and clouds. The one-dimensional profiles generated by the model could be advected along trajectories calculated by a three-dimensional limited area model. Clouds were allowed to affect the model's surface flux calculations. Scientists at the French Weather Service developed a new land surface parameterization with parameters derived from variables such as soil texture, soil type, albedo and dominant vegetation type (Bougeault et al. 1991a) and linked it to an atmospheric prediction model, comparing the output to data from a

field experiment conducted earlier (Bougeault et al. 1991b). A forecast was run for a cloud-free day with "reasonable agreement" between the temperature, humidity and wind forecasts, and observations. Argentini et al. (1992) used the one-dimensional PBL model of Zhang and Anthes (1982) to validate their new land surface parameterization prior to including it in a three-dimensional model. They were able to eliminate some input parameters derived from poorly measured biophysical variables, finding that surface flux predictions were not sensitive to variations in the parameters. Lakhtakia and Warner (1994) coupled different surface schemes to a one-dimensional PBL model in an effort to discover which situations could be modeled by a simple scheme and which situations required a more complex scheme. A time-independent slab model, a moisture availability parameter tied to precipitation, and the Biosphere-Atmosphere Transfer Scheme (BATS; Dickinson et al. 1986) were compared. The results among the three schemes varied the most for large vegetation cover, high net surface radiation, and unsaturated soil.

Land surface models (LSMs), which may either run as a stand alone scheme with appropriate atmospheric boundary conditions or coupled to an atmospheric model, constitute another category of surface process parameterization. The Simple Biosphere Model (SiB; Sellers et al. 1986, Sellers and Dorman 1987, Sellers et al. 1996) includes vegetation canopy properties such as resistance to evaporation, radiative properties and interception of precipitation by leaves, in addition to multiple soil layers. SiB is designed to be used by meteorologists and biophysicists alike. Since SiB is configured to be connected with GCMs, issues arise regarding sub-grid scale variability, for instance as it pertains to convective precipitation within a grid

cell (Sellers et al. 1986). Surface fluxes measured by SiB during early testing (Sellers and Dorman 1987) compared favorably with micrometeorological observations, but large uncertainties appeared in evapotranspiration estimates. Model flux calculations were found to be very sensitive to soil moisture variations, especially for dry soil. To make it more helpful as a model coupled to GCMs, SiB was revised (SiB2; Sellers et al. 1996) to include CO_2 flux from the canopy, satellite vegetation information, more accurate hydrological modeling and gradual albedo change due to snowmelt.

Other LSMs have been created to provide appropriate surface boundary conditions for atmospheric models. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS; Henderson-Sellers et al. 1993) compared and contrasted various characteristics of twenty-five different LSMs, with completion of model sensitivity studies as a primary goal. Chen et al. (1997) reported on results of comparing the PILPS models with observed meteorological variables, fluxes and net radiation at Cabauw, the Netherlands. Accounting for stomatal resistance to evaporation was found to be an important factor, as the range in flux calculations was halved when the three schemes not accounting for stomatal resistance were dropped. Net radiation between the schemes did not vary by any greater amount than the observational error while phase shifts in the diurnal surface flux existed between schemes, possibly due to differing treatments of thermal inertia. Wilson et al. (1987) conducted sensitivity tests on the BATS scheme, finding that it was highly sensitive to soil texture in contrast to the model run by Wetzel and Chang (1988). The land surface and PBL schemes connected to the Eta model were improved between 1995 and 1996, and Betts et al. (1997) compared forecasts from the two model versions.

They concluded that the new land surface scheme improved summer precipitation forecasts, but the flawed radiation scheme was unchanged between the two versions and flux calculations still had large inaccuracies. The authors stated that modeling fluxes depends on three crucial factors: modeled radiation, green vegetation fraction and soil moisture.

Many of the early LSMs were appropriately designed for producing temporally long forecasts with low resolution, but the need was still present for LSMs providing boundary conditions for short-term mesoscale atmospheric modeling. One model designed to be coupled with MM5 is the scheme created by Pleim and Xiu (1995), which includes a force-restore soil model and a hybrid PBL model combining nonlocal closure and eddy diffusion. Like SiB, the Pleim-Xiu (PX) scheme includes a prognostic treatment of soil moisture and considers the effects of vegetation on transpiration, but the schemes differ in that SiB was intended for use with GCMs. Testing of the PX scheme took place later (Xiu and Pleim 2001) in the form of comparison with observations from the Wangara (Clarke et al. 1971) and FIFE (Sellers et al. 1988) experiments. The PX scheme adjusted the near-surface output of the standard MM5 model toward the observed fluxes and surface temperature. Soil moisture nudging was included as a technique in the PX scheme to improve analyzed fields (Pleim and Xiu 2003) based on biases in 2 m temperature and relative humidity calculated from model output and analyzed observations. Nudging drastically decreased the spin-up time necessary to develop reasonable soil moisture fields, in addition to being computationally less expensive than other data assimilation schemes such as 4D-Var.

Another approach to modeling the land surface is embodied by the Noah LSM, so named in recognition of the entities developing it (National Centers for Environmental Prediction (NCEP), Oregon State University (OSU), the Air Force, and the Hydrologic Research Lab (HRL)). The modified OSU LSM (Pan and Mahrt 1987), or Noah, was coupled to MM5 by Chen and Dudhia (2001a,b). It included soil and vegetation information at 1 km resolution, allowing the representation of circulations unresolved by observing networks. As part of the same study, the modified OSU LSM coupled to MM5 was tested under clear sky conditions, modeling surface fluxes more accurately than a slab model and therefore improving near-surface temperature and humidity predictions. Noah was formed starting with the OSU LSM and has benefited from revisions and regular updates by the groups mentioned above (e.g., Chen et al. 1996, Schaake et al. 1996, Betts et al. 1997, Chen et al. 1997). Improvements have included an increase from two to four soil layers, the use of Eta Data Assimilation System (EDAS) soil moisture and temperature data making nudging unnecessary, and vegetation changes. Upgrades intended to more accurately model cold season processes were added by Ek et al. (2003) which improved near-surface temperature and relative humidity forecasts while not adversely affecting, and sometimes improving, upper air and precipitation forecasts. Cloudy conditions present challenges in modeling land surface processes such as fluxes, making the ability to accurately parameterize radiation, clouds, and stable PBL and surface-layer conditions very important.

2.3.4 *Observation and assimilation of land surface variables*

Initialization and verification of land surface variable forecasts requires representative observational data, and these data are obtained through a variety of methods. Remote sensing techniques allow the collection of observations worldwide. Radiative surface temperatures have been measured by satellite to infer surface parameters such as soil moisture (e.g., Schmugge 1978, Carlson et al. 1981, Price 1982, Wetzel et al. 1984, Diak et al. 1986, Wetzel and Woodward 1987). Schmugge (1978) measured microwave brightness temperature and radar backscatter as a proxy for soil moisture, with the microwave technique providing useful data on clear or cloudy days. Carlson et al. (1981) asserted that warm nights over urban areas are due to low moisture availability rather than high thermal inertia using such a method. Price (1982) used thermal infrared (IR) images to calculate grid-scale surface fluxes for a GCM as the spatial integral of estimated fluxes at 600 m resolution across a grid box. Wetzel et al. (1984) took advantage of GOES data whereas most other investigators had used data from polar-orbiting satellites, and they determined that the best predictor of soil moisture was mid-morning surface temperature differential with respect to absorbed solar radiation. Subsequently, linear regressions of morning IR surface temperature change against soil moisture and other variables were performed by Wetzel and Woodward (1987) in an effort to isolate that portion of the temperature change that could be explained by soil moisture. Diurnal surface temperature range forecast by the model used in Diak et al. (1986) was measured against satellite temperatures, resulting in a close match when the moisture availability was moderate

to high but in a poor match for most of the arid western U.S. The collaborators also pointed out that remotely sensed data should ideally be evaluated against surface-based measurements at some point to demonstrate the accuracy of using remote sensing techniques to infer land surface parameters. Jones et al. (1998a) retrieved soil moisture by assimilating satellite-derived infrared heating rates. These heating rates had a direct relationship with several model parameterization variables. In a case study (Jones et al. 1998b), this method produced realistic soil moisture profiles. The authors contended that more in situ soil moisture observations would be helpful for model validation.

Remotely sensed surface temperatures and reflectances have also been used to estimate various elements of the surface energy balance (Xinmei et al. 1993; Friedl 1996; Kustas et al. 1999). The estimation of surface fluxes using satellite data was conducted by Xinmei et al. (1993), who determined the sensible heat flux directly from the radiative temperature and calculated the latent heat flux as a residual by also utilizing red and near IR reflectances. Friedl (1996) discussed the noisiness of high-resolution surface flux proxy data from satellites and the fact that modeled and observed surface fluxes agreed well with each other when averaged over regional scales. The author asserted that high-resolution flux values determined from remote sensing techniques may be calculated only approximately. The estimation of surface fluxes using the values of surface layer or mixed-layer variables combined with radiometric temperature was explored by Kustas et al. (1999), who found that the mixed-layer method yielded reasonable flux estimates using data that were easier to obtain. However, in cases for which a regional roughness length value was used with

local radiometric temperature, the mixed-layer method resulted in serious flux estimation errors.

Remotely sensed surface temperatures and reflectances are helpful in detecting soil moisture variability and estimating fluxes in a wide variety of locations, but in situ measurements of surface variables are essential in demonstrating the validity of LSMs and model parameterizations. The measurement of surface variables such as fluxes can be difficult, as the fetch needed to make a reasonable surface flux estimate under stable PBL conditions is more than 100 times greater than the measurement height (Horst and Weil 1994). Businger et al. (1990) described the measurement of surface fluxes of sensible heat, water vapor, and trace chemical species at the NCAR Atmosphere-Surface Turbulent Exchange Research (ASTER) facility. The facility was birthed partially out of a need for ground truth in support of remote sensing from aircraft and satellites. Xinmei and Lyons (1995) compared results generated by a one-dimensional soil-canopy-PBL model to observed flux data from locations with varying amounts of vegetation cover in Canada and France. Measurement of surface fluxes has shown that flux observations are only representative over distances of 1 km or less, such as at the Southern Great Plains Atmospheric Radiation Measurement/Cloud and Radiation Testbed (ARM/CART) site (Gao et al. 1998).

Land surface data such as soil temperature, soil moisture and surface flux observations also have been collected as a result of various field experiments. The Wangara experiment in southeastern Australia (Clarke et al. 1971) yielded quality data from the surface layer and the PBL used to test many numerical boundary layer simulations (Hess et al. 1981). A team of scientists in France conducted the

Hydrologic Atmospheric Pilot Experiment-Modelisation du Bilan Hydrique (HAPEX-MOBILHY; Andrè et al. 1986) program to collect land surface and atmospheric data using both remote and in situ methods with a focus on developing land surface parameterizations for GCMs. Since interactions between the atmosphere and the land surface require interdisciplinary cooperation to understand, the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE; Sellers et al. 1988) was developed. Scientists from many different fields worked with the land surface data collected during FIFE to better characterize the role of biology in land-atmosphere interactions and to evaluate the use of satellite observations to infer land surface parameters. A temporary surface observing network was constructed on the steppe and irrigated farmland of northeastern Oregon in order to provide flux data for the testing of PBL forecasting models (Doran et al. 1992). Atmospheric soundings showed that the surface energy balance played a large role in determining the PBL depth, but advection of sensible heat flux from the steppe to the irrigated farmland and thermally driven mesoscale fluxes were also important (Doran et al. 1995). Data including wind, radiation, soil temperature and surface fluxes were collected at Cabauw, the Netherlands in 1987 with the purpose of building a dataset by which land surface schemes could be tested (Beljaars and Bosveld 1997). To form a continuous time series, missing observations were filled in by model output. The Cabauw data were used to assess the spread and skill of predictions generated by the land surface schemes included in the PILPS study (Chen et al. 1997). The Walnut River watershed in southeastern Kansas was heavily instrumented with land surface data collection equipment, including a soil moisture

array, for the Cooperative Atmosphere Surface Exchange Study and Atmosphere Boundary Layer Experiments (CASES and ABLE; LeMone et al. 2000) with the network intended for long-term use in addition to short, intense field programs. The CASES project provided a unique opportunity for land surface scheme validation with the ARM/CART site, along with several radars and profilers, located nearby.

Not only has the obtaining of land surface field measurements provided data to which forecast model output may be compared, but these same measurements can be used to improve model analyses of the state of the land surface via assimilation. In lieu of the assimilation of actual land surface observations, however, many surface parameters are often estimated using known quantities. Holtslag and van Ulden (1983) described the estimation of surface fluxes from wind speed at one level, the screen temperature and cloud cover using similarity theory. Though their net radiation and flux estimates greatly improved by using measured rather than calculated solar radiation, their scheme agreed well with observations on average. Spatially and temporally averaging both surface fluxes and gradients can produce counterintuitive results, so Mahrt (1987) used a small-scale model in individual grid boxes of a larger-scale numerical model to discover relationships between grid-averaged surface fluxes and grid-averaged flux gradients. A new formulation relating area-averaged quantities was developed, but at the time no datasets existed by which the formulation could be adequately tested. Peters-Lidard et al. (1998) showed that errors in soil thermal conductivity caused errors in the ground heat flux, which primarily fed back into the sensible heat flux and the surface temperature. They used a different thermal conductivity parameterization that reduced the ground heat flux

errors, as compared to data from FIFE (Sellers et al. 1988), resulting in better estimation of the components of the surface energy balance.

Soil moisture has also been initialized indirectly using conventional observations. Smith et al. (1994) indirectly determined soil moisture values by generating soil wetness fields using a hydrologic model. Comparison with soil moisture observations showed that the fields derived by the hydrologic model were more realistic than climatological fields, but no verification of affected atmospheric model variables was conducted. Calvet et al. (1998) attempted to retrieve the total soil moisture content and the field capacity from surface soil moisture and surface temperature, comparing their results to observations from a single site in France in 1995. They pointed out that during wet conditions, this method would suffer because surface temperature is not very sensitive to total soil moisture content. Soil moisture profiles obtained by Jones et al. (1998b) by assimilating satellite-derived heating rates into a three-dimensional mesoscale model were found to be realistic when compared with microwave, precipitation, and radar data. Xu and Zhou (2003) developed a simple linear regression method for retrieving soil water content from Oklahoma Mesonet soil temperature measurements at varying depths. Comparisons against measured soil water content showed that root mean square (rms) differences between the measurements and the soil water content calculated using regression were smaller than the soil water content observation error.

In addition to estimating the values of land surface parameters using known quantities, indirect and direct assimilation of soil moisture estimates and measurements into land surface parameterizations and LSMs has been carried out in

an attempt to improve the simulation of land-atmosphere interactions. Assimilation of soil moisture began with indirect methods. Mahfouf (1991) assimilated soil moisture values, which were estimated from near-surface atmospheric parameters, using variational and sequential techniques into a one-dimensional model with an attached land surface parameterization. Both methods produced soil moisture fields that converged to realistic values for clear-sky periods when compared to observations taken during the HAPEX-MOBILHY experiment (Andrè et al. 1986), favoring the use of the sequential method due to the lower computational cost. Exploration of the feasibility of the sequential method, or optimal interpolation (OI), for assimilating soil moisture was continued by Bouttier et al. (1993a), who found coefficients arising from the relationship between soil moisture and atmospheric parameters. They discussed the sensitivity of low-level atmospheric parameters to soil and vegetation properties, finding that the OI coefficients were extremely sensitive to vegetation cover. Bouttier et al. (1993b) also tested their assimilation method with a three-dimensional mesoscale model and found that their soil moisture analyses converged to reference values within around 48 h, despite the fact that the relationship between soil moisture and the atmosphere is nonlinear. Convergence was the fastest for the deep soil reservoir over vegetated areas and for the surface soil reservoir over bare ground.

The development of indirect assimilation techniques for soil moisture and surface fluxes has continued as an alternative to initialization from climatological databases or previous forecasts. Satellite-derived evaporation measurements were used to update soil moisture on a daily basis by van den Hurk et al. (1997), and model runs

incorporating these updates reduced low-level temperature and humidity biases when compared with forecasts from climatological soil moisture fields for a 7-day period on the Iberian Peninsula. However, their model only included one soil layer, neglecting the long-term persistence of deep soil moisture. Jones et al. (1998b) assimilated satellite-derived infrared heating rates into the Regional Atmospheric Modeling System (RAMS), which included eleven soil layers, in an effort to retrieve soil moisture fields. Realistic soil moisture profiles were obtained that demonstrated the effectiveness of the surface parameterization, but a call for more case studies with a variety of conditions was made.

Land surface description in French global and limited area atmospheric models was upgraded from a simple scheme to the Interaction Soil Biosphere Atmosphere (ISBA) parameterization (Giard and Bazile 2000). Soil moisture analyses under the ISBA parameterization are determined through an OI procedure using 2 m temperature and relative humidity observations, and the incorporation of ISBA resulted in substantial surface modeling improvements for short-term forecasts. Margulis and Entekhabi (2003) used a variational method to assimilate air temperature, humidity and radiometric surface temperature measurements into a one-dimensional PBL model, finding that much improved estimates of surface fluxes were made by using both reference-level and surface data as opposed to either type by itself. "Accurate and robust" surface flux values were calculated in tests with synthetic and FIFE data, including when factors not parameterized such as advection were important. Soil moisture data from the NCEP-NCAR reanalysis were used by Zhang and Fredericksen (2003) in a Bureau of Meteorology Research Centre

(BMRC) GCM to examine the effects of replacing climatological soil fields on modeling of climate anomalies in surface temperature and rainfall. They found that soil moisture anomalies had indirect effects on rainfall due to changed circulations in addition to the direct effects on localized surface temperature. Rodell et al. (2004) described an offline land data assimilation system (LDAS) compiling high-resolution global land surface analyses from mostly satellite observations. The LSMs used to generate analyses through assimilation techniques were forced with observation-based atmospheric fields to avoid incorporating biases from atmospheric models.

Direct assimilation of land surface observations, as well as land data assimilation using Kalman filter-type approaches, has been shown to be feasible. A 4D-Var method was used by Ruggiero et al. (1996) to assimilate surface observations at frequent intervals, and the influence of the changes in the background surface field were spread upward via the eddy diffusivity in a PBL adjustment scheme. The more frequent PBL adjustment resulted in better analyses and forecasts from both 3 h and 12 h assimilation cycles, indicating that such frequent surface assimilation could provide more accurate short-term predictions. Alapaty et al. (2001) continuously assimilated analyses of surface temperature and mixing ratio into a one-dimensional model, correcting model sensible and latent heat fluxes using differences between observations and model results. Since surface boundary conditions have a pronounced effect on the PBL, modifying surface flux values often according to frequent assimilation of surface temperature and mixing ratio resulted in improved lower atmospheric virtual potential temperature and mixing ratio profiles. Three simulations with varying soil moisture over North America were run by Walker and

Houser (2001) to test the length of time needed for the assimilation of soil moisture data to reduce error in the model fields of a GCM. In situ soil moisture data are not appropriate for a GCM due to their spatial variability, so remotely sensed data from the ISLSCP experiment were used. Synthetic observations generated from the ISLSCP soil moisture data set were assimilated into a model run with arbitrarily high soil moisture continent-wide using a one-dimensional Kalman filter, and the resulting model fields matched closely with the ISLSCP fields within one month. Crow and Wood (2003) assimilated remotely sensed brightness temperature observations, which can help to increase the vertical depth and horizontal resolution of soil moisture information, using an EnKF (Evensen 1994) method. Though the EnKF method sometimes spuriously produced non-Gaussian error structures and biased forecasts of root-zone soil moisture, it was viewed by the investigators as an "effective and computationally competitive" land data assimilation method. The assimilation of brightness temperature produced better soil moisture analyses than direct assimilation of satellite soil moisture or low-resolution precipitation data.

More recent experiments involving the assimilation of various types of land surface and near-surface observations have shown continued potential to improve both analyses and forecasts of near-surface atmospheric variables. Hacker and Snyder (2005) assimilated simulated surface observations of screen-level potential temperature, wind speed, and mixing ratio into a one-dimensional PBL model using an EnKF approach to test the usefulness of such observations for improving the model estimate of the PBL state. In their idealized experiment, they showed that the error in key PBL variables relative to their truth simulation was reduced when the

PBL was constrained by the EnKF assimilation. Through observing system simulation experiments during a winter cyclogenesis event, Zhang et al. (2006) demonstrated the effectiveness of the EnKF in keeping analyses close to the truth simulation. Simulated sounding and surface observations were assimilated over a 24 h period, after which the domain-averaged root mean square error (rmse) of the wind and temperature fields was comparable to typical observation errors. Godfrey (2006) replaced climatological soil and vegetation fields in the MM5 with analyzed fields created from soil observations from the Oklahoma Mesonet and observed vegetation information from the Advanced Very-High Resolution Radiometer (AVHRR), reporting the most substantial forecast improvements from runs using the new soil data. Actual surface observations from across the United States were assimilated using an EnKF approach by Fujita et al. (2007) with the goal of improving short-range forecasts of mesoscale features. Three different ensembles were formed using different combinations of varying initial and boundary conditions and varying the physical parameterizations included. The largest ensemble spread and the most accurate forecasts were achieved when varying both, with improvements in the forecasting of mesoscale boundaries and PBL structure especially important to making more useful short-term forecasts of severe weather.

Chapter 3

Experiment setup

In an effort to meet the goal of improving initial analyses of near-surface variables derived from NWP models by including various observations obtained at the surface, data from the Oklahoma Mesonet (Brock et al. 1995; McPherson et al. 2007) are assimilated into the Fifth-Generation Pennsylvania State University-National Center for Atmospheric Research Mesoscale Model (MM5; Dudhia 1993) using an EnKF scheme. Initial analyses are created on a domain covering the continental United States (CONUS) following the methodology of Fujita et al. (2007). The ensemble members are initialized from Global Forecast System (GFS) analyses on a large grid covering much of the Northern Hemisphere including North America, northern South America, and parts of both the Atlantic and Pacific Ocean. Bred modes are created via forecasts on this grid to develop perturbations among the members. The model then is re-initialized with GFS data on a domain consisting of primarily the CONUS and the perturbations from the bred modes are added to the members. Once this is done, 48 h forecasts are run on the CONUS domain. One-way nesting is applied twice with additional 48 h forecasts integrated on two higher-resolution domains. On the domain with the highest resolution, which encompasses the body of the state of Oklahoma, three 48 h forecasts are made: one with no assimilation, one with standard

Mesonet data assimilated every hour using the EnKF scheme, and a third with hourly assimilation of sensible heat flux estimates. Each ensemble member uses perturbed boundary conditions as well as a unique set of initial conditions. Analyses and forecasts generated with and without Mesonet data assimilation are then verified and compared with both NWS and Oklahoma Mesonet observations to test the effectiveness of the EnKF assimilation in reducing analysis error in low level temperature and mixing ratio.

3.1 Data

3.1.1 *Oklahoma Mesonet data*

The Oklahoma Mesonet (Brock et al. 1995; McPherson et al. 2007) and the Oklahoma Atmospheric Surface-layer Instrumentation System (OASIS; Brotzge 2000) provide a unique long-term data set of standard meteorological measurements as well as surface data such as soil temperature, soil moisture, net radiation and surface heat fluxes. Standard measurements used in this study include temperature and relative humidity at 1.5 m, wind direction and wind speed at 10 m, and station pressure. These measurements are taken from approximately 110 Mesonet sites. In addition, soil temperature is measured at depths of 5 cm, 10 cm and 30 cm. Soil moisture estimates at 90 OASIS Mesonet sites are determinable from the measurement of the calibrated change in temperature of the soil over time after a heat pulse is introduced (Oklahoma Climatological Survey 2005). Sensors performing this

function are placed at 5 cm, 25 cm, 60 cm, and 75 cm depths. Ground heat flux and skin temperature are measured at 84 OASIS sites. Nine Mesonet sites are designated as OASIS super sites and measure sensible and latent heat flux using an eddy covariance technique. However, directly estimated latent heat flux data are not released by the Oklahoma Climatological Survey (OCS), so the residual of the surface energy balance is used for the latent heat flux. Four-component net radiation is also measured at the super sites. A domeless net radiometer, the NR-Lite (Brotzge and Duchon 2000), measures total net radiation at the majority of the sites in the OASIS network. Sensible heat flux can be estimated at around 100 sites using a gradient method, since temperature and wind sensors are placed at two heights (Brotzge and Crawford 2000).

These additional data observed by the OASIS network have been used in a number of applications. Brotzge and Crawford (2000) confronted issues with estimating sensible heat flux using the gradient method including sensor errors and fetch variations. They also compared profile sensible heat flux estimates to eddy correlation estimates, finding that conditions conducive to large radiational heating errors resulted in the largest differences between the flux estimates. OASIS data were used to measure the effectiveness of the Interaction Soil Biosphere Atmosphere (ISBA) land surface parameterization (Noilhan and Planton 1989) within the Advanced Regional Prediction System (ARPS) run in 1-D vertical column mode by Brotzge and Weber (2002). They were able to conclude that the sensitivity of ISBA to input parameters was detrimental to its use with a mesoscale/microscale model such as the ARPS. OASIS data have been used to compare flux estimation methods

(Brotzge and Crawford 2003), to calculate spatial and temporal correlations between measurements at various Mesonet sites (Brotzge and Richardson 2003), to test the performance of the land surface parameterization in the operational Eta Model (Marshall et al. 2003), and to evaluate NCEP-NCAR reanalyses of the surface water and energy budgets (Brotzge 2004).

3.1.2 *Sensible heat flux estimation method*

With only nine Mesonet sites measuring sensible heat flux by way of an eddy covariance technique, an additional method is needed if enough sensible heat flux estimates to influence the entire model domain upon assimilation are to be generated. Brotzge and Crawford (2000) demonstrated that sensible heat flux estimates may be calculated from temperature and wind speed measured at two different levels. For this study, sensible heat flux estimates are calculated using a variant of an iterative profile method. The iterative profile method is applied to each five-minute observation within a 30 m time window, and these seven sensible heat flux estimates are averaged to come up with the estimate for inclusion in the EnKF assimilation process.

The underlying theory behind iterative profile methods for estimating surface fluxes is covered thoroughly in boundary layer meteorology textbooks (Stull 1988; Sorbjan 1989; Garratt 1994) and will not be repeated in full here. However, one important assumption made in order to use this profile method that may affect the resultant flux estimates must be discussed. Standard OASIS Mesonet sites do not

measure humidity at two levels, so the standard formulation of the iterative profile method must be simplified. The equation for the Monin-Obukhov length may be written as

$$L = \frac{u_*^2}{\kappa\left(\beta\theta_* + 0.61 g q_*\right)} \tag{8}$$

where L is the Monin-Obukhov length, u_* is the friction velocity, κ is the von Karman constant, g is the acceleration of gravity, $\beta = \frac{g}{\theta_{v0}}$ is the buoyancy parameter where θ_{v0} is a reference value for virtual potential temperature, θ_* is the temperature turbulence scale, and q_* is the humidity turbulence scale. Since standard OASIS Mesonet sites do not measure humidity at two levels, it is necessary to neglect the humidity terms in eq. (8). The new equation for Monin-Obukhov length thus reads

$$L = \frac{u_*^2}{\kappa b \theta_*} \tag{9}$$

with

$$b = \frac{g}{T_0} \tag{10}$$

where T_0 is a reference value of temperature, in this study taken to be the mean of the temperature measurements at both levels. All other symbols have the same meaning as in eq. (8).

The humidity turbulence scale (q_*) tends to be negative during the day in summer, as atmospheric moisture decreases with height near the surface. Therefore, neglecting the moisture term in eq. (8) will make the denominator of smaller magnitude, leading to larger negative values of L during the day when the surface layer is unstable. This results in the effects of wind shear having incrementally more weight than they should versus the effects of buoyancy in determining the sensible heat flux. The expressions for u_* and θ_* are

$$u_* = \frac{\kappa \Delta u}{\left[\ln\left(\frac{z_{u2}}{z_{u1}} \right) - \Psi_m\left(\frac{z_{u2}}{L} \right) + \Psi_m\left(\frac{z_{u1}}{L} \right) \right]} \tag{11}$$

and

$$\theta_* = \frac{\kappa \Delta \theta}{\left[\ln\left(\frac{z_{\theta 2}}{z_{\theta 1}} \right) - \Psi_h\left(\frac{z_{\theta 2}}{L} \right) + \Psi_h\left(\frac{z_{\theta 1}}{L} \right) \right]} \tag{12}$$

where $\Delta u = u_2 - u_1$ is the difference between the two wind speed measurements, $\Delta\theta = \theta_2 - \theta_1$ is the difference between the two temperature measurements, z_{ux} signifies the height of a wind measurement, $z_{\theta x}$ signifies the height of a temperature measurement, and Ψ_m and Ψ_h are deviations of the velocity and temperature profiles, respectively, from a logarithmic profile (Dyer and Hicks 1970; Dyer 1974). These deviations are due to differences in stability.

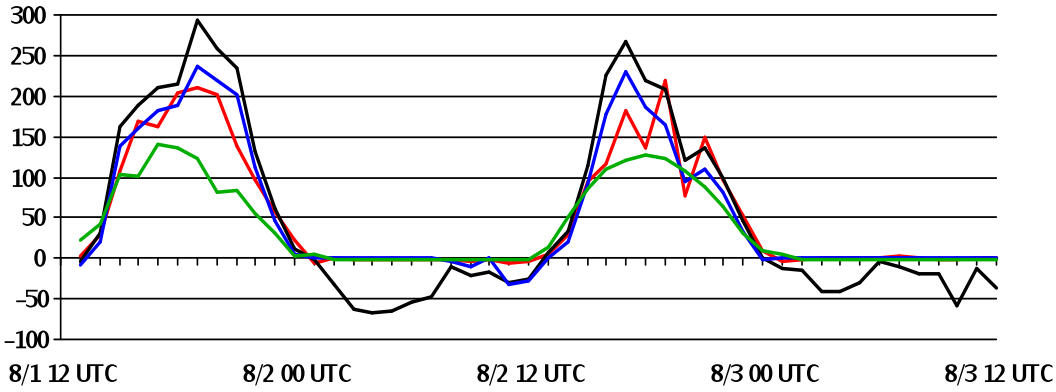


FIG. 3.1: Mesonet sensible heat flux estimates calculated using a simple gradient method (black), the profile method explained in Section 3.1.2 (blue), and eddy correlation instrumentation (red), as well as sensible heat flux estimates from the Noah LSM (green). Values are for the location of the Mesonet station at Idabel in southeastern Oklahoma from 12 UTC on 1 August 2004 to 12 UTC on 3 August. Sensible heat flux values along the *y*-axis are in W m^{-2}, while time is designated along the *x*-axis.

This iterative profile algorithm yields lower values of sensible heat flux than a simple gradient method during peak boundary layer heating. In addition, the values are acceptable at most sites when compared with eddy correlation estimates as well as flux values calculated by the Noah land surface scheme within MM5. For example, sensible heat flux estimates from Idabel, OK in early August of 2004 (Fig. 3.1) are lower during the day calculated via the profile method or eddy correlation method compared with a simple gradient method. Sensible heat flux estimates from the Noah LSM tend to be lower than any estimate using Mesonet data during the daylight hours. At night, sensible heat flux estimates are near zero except when the simple gradient method is used.

3.2 Model

The atmospheric model chosen for this study, the MM5, is a three-dimensional nonhydrostatic sigma-vertical coordinate model that may be initialized by interpolating output from a coarser-grid model to a smaller grid area with higher spatial resolution (Dudhia 1993). The MM5 modeling system includes both pre- and post-processing programs in addition to the actual code comprising the model itself. Terrain elevation of varying resolution, vegetation data and soil type data from the United States Geological Survey (USGS) are used to initialize the land surface characteristics in each domain. Analyses at 1° X 1° horizontal resolution from the GFS, a model which is part of the NCEP operational model suite, are used in these experiments to initialize MM5 atmospheric variables. Forecasts are generated on four nested domains; the horizontal resolutions of the four domains are 81 km, 27 km, 9 km and 3 km, respectively. The coarsest-domain forecasts include 24 sigma levels, with the number of levels increased in the boundary layer as the horizontal resolution increases such that domain 4 includes 26 sigma levels.

The MM5 modeling system is flexible in the number of parameterization schemes available to choose from. The schemes cover convection, radiation, land surface processes, moisture physics and boundary layer processes. For the largest domain encompassing all of North America and most of the north Atlantic and north Pacific oceans, the Betts-Miller-Janjic convective parameterization is chosen due to its suitability for coarse resolution forecasts (Dudhia et al. 2005). Other parameterization schemes used include the MRF PBL scheme (Hong and Pan 1996),

a simple ice microphysics parameterization which explicitly predicts cloud and rain water but melts snow immediately below the freezing level (Dudhia 1989), and the NCAR CCM2 radiation scheme (Hack et al. 1993). The CCM2 scheme is also chosen due to its suitability for larger grid scales. For the next two domains, the Kain-Fritsch convective parameterization (Kain and Fritsch 1993) and the Dudhia cloud-radiation scheme (Dudhia 1989) are substituted for the Betts-Miller-Janjic convective parameterization and the CCM2 radiation scheme. No convective parameterization is used for the fourth domain because the 3 km resolution makes the use of a convection scheme unnecessary.

In addition to atmospheric parameterization schemes designed to model processes occurring at the sub-grid-scale, a land surface parameterization is necessary to account for processes occurring on and below Earth's surface. The Noah LSM (Pan and Mahrt 1987; Chen et al. 1996; Schaake et al. 1996; Betts et al. 1997; Chen et al. 1997; Chen and Dudhia 2001a,b; Ek et al. 2003) is used for all four domains in this study. Additional input fields beyond those necessary for the initialization of MM5 which Noah requires are annual mean deep-soil temperature, vegetation fraction, dominant soil and vegetation types, soil moisture and soil temperature at various depths, snow depth and sea ice. Annual mean deep-soil temperature, vegetation fraction and soil/vegetation type data are acquired from the MM5 modeling system climatological data, while soil moisture and soil temperature data at 10 cm and 200 cm are obtained from the GFS analyses. A maximum snow albedo is used in the Noah LSM to prevent the albedo value from becoming too high in the presence of snow cover. Noah outputs soil temperature and moisture at four levels (5, 25, 70, and

150 cm), canopy moisture and snow amount information, runoff accumulation both at the surface and below ground, and albedo. The MRF PBL scheme passes surface-layer exchange coefficients, radiative forcing and precipitation rate to Noah, and Noah outputs the values of the surface fluxes. These data are then available to the MM5 for use in calculating the values of near-surface atmospheric variables.

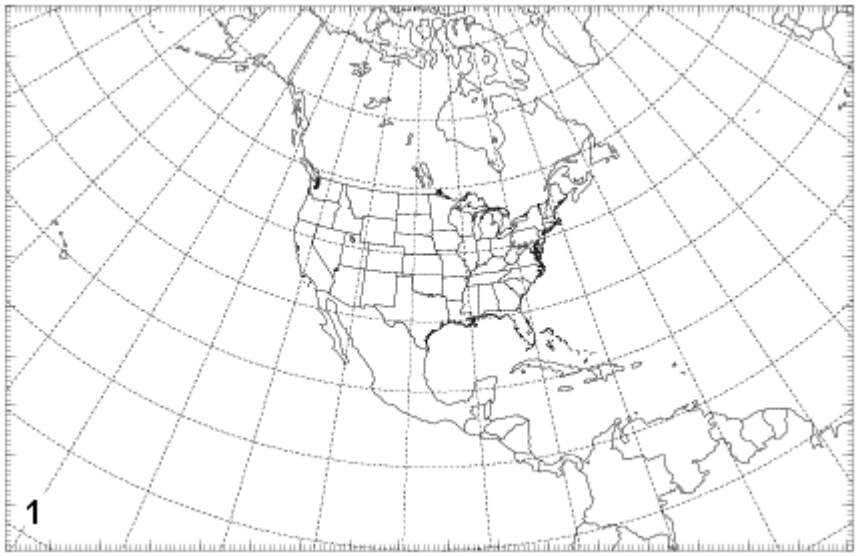


FIG. 3.2: Domain 1, used for the purpose of creating bred modes.

3.3 Creation of bred modes

The coarsest model domain is a 110x171 point grid with an 81 km horizontal resolution and 24 sigma levels in the vertical covering much of the Western Hemisphere north of the equator (Fig. 3.2). The model terrain and land use parameters are initialized from global data provided by the USGS. The initial states of both atmospheric and land surface variables are created by interpolating data from GFS global analyses to the MM5 grid. Boundary conditions also are created from

GFS analyses for the whole length of the forecast. After the development of this initial control analysis, a Monte Carlo technique (Errico and Baumhefner 1987) is used to perturb temperature, U and V wind, height, specific humidity and sea level pressure. Twenty-five perturbed initial pressure-level analyses are created via this technique. Sigma-level analyses and perturbed boundary conditions for introduction into MM5 are formed from the perturbed analyses. Once the analyses are created, a 12 h MM5 forecast is run starting from the control analysis and each perturbed

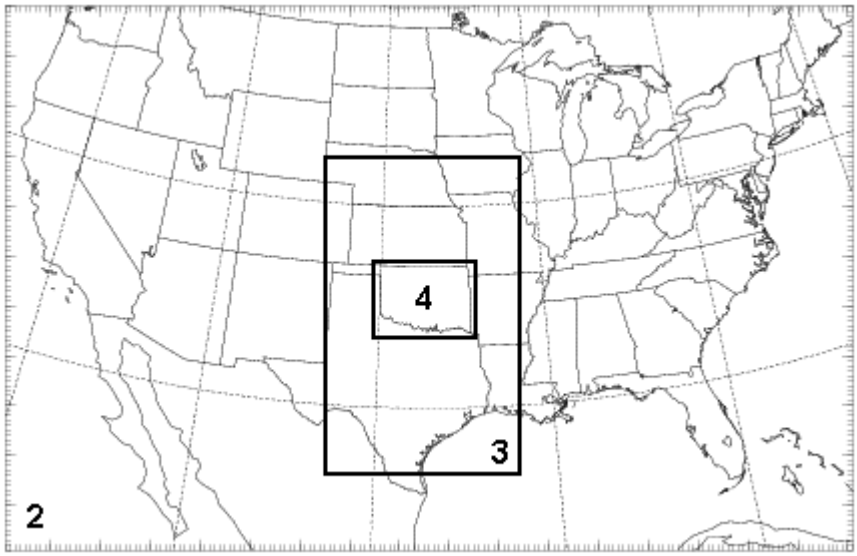


FIG. 3.3: Ensemble Kalman filter forecast one-way nested domains.

analysis. The values of the variables in the control forecast are then subtracted from each perturbation forecast, after which the differences are rescaled to the sizes of the initial perturbations and added back to the control forecast to create new perturbed analyses. A new 12 h MM5 forecast is run following the re-scaling of perturbations,

and the cycle is repeated until stopped. This methodology follows that of Fujita et al. (2007) and was developed from the earlier work of Toth and Kalnay (1993).

Once the BGM process ending at 84 h (3.5 days) has been run and the perturbations have been rescaled, the initial conditions and boundary conditions for the EnKF data assimilation period are created. The initial number of perturbations is doubled by subtracting the existing perturbations from as well as adding them to the control analysis, enabling a run with 51 ensemble members (50 perturbations plus control). A final forecast on the 81 km grid is run for the length of the EnKF data assimilation period for all 51 ensemble members. Forecast atmospheric and land surface data are output and archived every 3 h. Following their creation, the boundary condition files for each of the ensemble members are vertically interpolated back to pressure coordinates from sigma coordinates. The pressure coordinate files are then horizontally interpolated to domain 2, a 110x171 grid with a 27 km

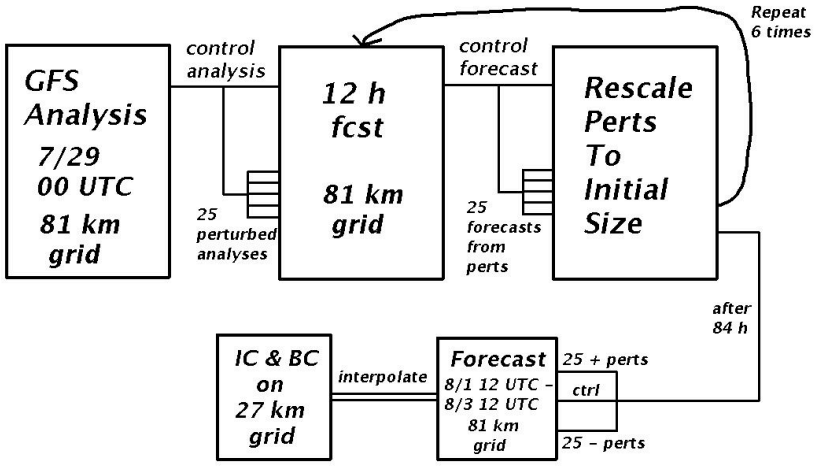


FIG. 3.4: Schematic diagram showing how ensemble members are developed.

horizontal resolution covering the CONUS (Fig. 3.3). After the horizontal interpolation to the finer grid, the pressure level data are interpolated back to sigma levels, of which there are 24 in the new domain. A schematic diagramming the process of initializing the ensemble members is included in Figure 3.4.

Having the files necessary to initialize all of the ensemble members on the CONUS domain, the EnKF assimilation can begin.

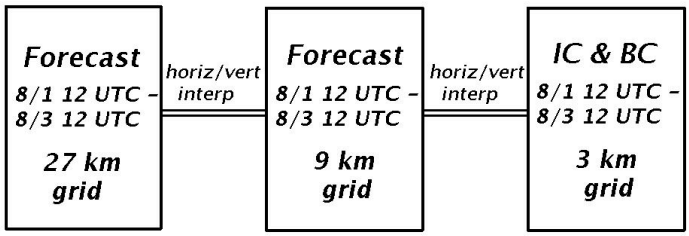


FIG. 3.5: Schematic diagram depicting process of initializing the 3 km (smallest) domain.

3.4 Ensemble Kalman filter assimilation

Surface observations from the Oklahoma Mesonet are the only data assimilated in these experiments, so a domain encompassing the body of the state of Oklahoma is chosen for the model runs including assimilation. In order to obtain the desired initial conditions for a run on such a domain, a one-way nesting procedure is used. Atmospheric variables for the control run over domain 2, which has a 27 km

horizontal resolution and covers primarily the CONUS, are initialized from GFS analyses using the same approach as for the atmospheric variables on domain 1. The other ensemble members are initialized similarly from the bred modes data. Forecasts spanning the entire length of the EnKF assimilation period are generated from the domain 2 initialization, after which initial conditions and boundary conditions are formed on a 9 km horizontal resolution (domain 3; Fig. 3.3) grid by horizontally interpolating from the coarse grid to the fine grid (Fig. 3.5). Once again, forecasts over the EnKF assimilation period are run and horizontal interpolation takes place, this time to a 3 km horizontal resolution grid (domain 4; Fig. 3.3). An additional sigma level is also added during both interpolations to increase the resolution in the PBL, and vertical interpolation from the coarse grid to the fine grid is used to create the new vertical profiles. A 1 h forecast is run from the domain 4 initial conditions, after which the assimilation of Mesonet data begins.

To prepare for comparison of the model output with observations, two-dimensional model fields corresponding to Mesonet observations of 1.5 m temperature and 1.5 m relative humidity are calculated. Other calculated fields which are needed prior to computing variable types corresponding to observations include total pressure on sigma levels, geopotential height and Monin-Obukhov length. A background file is written with both three-dimensional and two-dimensional fields. Three-dimensional horizontal and vertical wind, temperature, water vapor mixing ratio, pressure perturbation and geopotential height are saved. Additionally, the two-dimensional p* field (an MM5 pressure variable), two layer soil temperature and three layer volumetric water content fields; sensible, latent and ground heat flux

fields; friction velocity and terrain height fields are retained. With the exception of terrain height, these three- and two- dimensional fields are updated using the EnKF scheme.

Following the formation of the background file, model forecasts of each of the Mesonet variables chosen for assimilation are bilinearly interpolated from the grid to the observation locations. Surface pressure observations are adjusted hypsometrically using a standard environmental lapse rate of 6.5 K/km to account for the difference between the terrain height interpolated from the grid and the actual terrain height at the observation location. If the absolute value of the height difference is greater than 300 m, the pressure observation is discarded. The interpolated background values not in the units of the Mesonet observations are converted to allow direct comparison. Finally, a second file containing the Mesonet observations along with the converted, interpolated background values from the model grids is written.

Once the background and observations files are created, observations are assimilated sequentially to update the forecast fields. Each observation alters the analysis, so variables such as Monin-Obukhov length, surface pressure, temperature, wind and mixing ratio must be re-calculated at grid points near the observation for each ensemble member before the observation is run through the Kalman filter. The model background value of the observed variable at the observation location for each ensemble member is bilinearly interpolated from the nearby grid points to the observation point. The ensemble average background value is computed once all of the background values have been calculated for each of the individual members.

Five three-dimensional fields and ten two-dimensional fields are then updated using the EnKF scheme. These fields include all of those contained in the background file with the exception of terrain height. In particular, soil temperature values from the top two soil layers and soil moisture values from the top three soil layers in the Noah LSM are updated. Soil temperature values from the lowest two soil layers, as well as soil moisture values from the lowest soil layer, change so little over a 48 h period that updating of these variables is omitted for the sake of saving computer time. Each of the Mesonet observations is processed serially according to the following EnSRF (Whitaker and Hamill 2002) equations (Snyder and Zhang 2003; Dowell et al. 2004; Fujita et al. 2007):

$$\overline{x^a} = \overline{x^f} + WK[y^o - \overline{H(\mathbf{x}^f)}] \tag{13}$$

$$x_n^a = x_n^f + WK[y^o - \beta H(\mathbf{x}_n^f) - (1.0 - \beta)\overline{H(\mathbf{x}^f)}] \tag{14}$$

where an overbar denotes an ensemble mean value, n is an index pertaining to a particular ensemble member, x^a is an analysis value from a model field, x^f is a forecast (background) value from a model field, W is a smoothly-varying localization weight, y^o is an observation at a single point, $\mathbf{x}^f$ is a vector containing the values of all of the model background fields at all of the grid points, and H is an operator that both converts the units of model background variables into those used by the Mesonet and bilinearly interpolates the background values from grid points to observation locations.

The weight W in eq. (13) and eq. (14) is proportional to the inverse of the exponential function and is used to decrease the magnitude of the change each observation makes to the background value as the distance from the observation location to the grid point increases. A horizontal localization radius of 150 km is chosen based on the examination of spatial correlation structures plotted at 18 UTC 1 August 2004; due to the use of a range factor of 1.5 grid points as far as 225 km from the observation location can still be slightly modified. A vertical localization radius of ten grid points is selected to ensure that the near-surface observations can influence the analysis throughout the depth of the PBL. Pressure observations are allowed to influence the analysis throughout the entire vertical column.

The Kalman gain K in eq. (13) and eq. (14) and the factor β in eq. (14) are given by

$$K = \frac{\frac{1}{N-1}\sum_{n=1}^{N}(x_n^f - \bar{x}^f)[H(\mathbf{x}_n^f) - \overline{H(\mathbf{x}^f)}]}{\sigma_o^2 + \frac{1}{N-1}\sum_{n=1}^{N}[H(\mathbf{x}_n^f) - \overline{H(\mathbf{x}^f)}]^2} \tag{15}$$

$$\beta = \left\{1 + \sqrt{\frac{\sigma_o^2}{\sigma_o^2 + \frac{1}{N-1}\sum_{n=1}^{N}[H(\mathbf{x}_n^f) - \overline{H(\mathbf{x}^f)}]^2}}\right\}^{-1} \tag{16}$$

where N represents the total number of ensemble members, σ_o^2 is the observation error variance, and the other symbols have the same meaning as in eq. (13) and eq. (14). The first term in the summation in the numerator of eq. (15) is the difference

between the value for ensemble member *n* and the ensemble mean value of a model background field at a specific grid point. The second term is the difference between the value for ensemble member *n* and the ensemble mean value of the model background estimate of the variable being assimilated at the observation location. Therefore, the Kalman gain *K* at a point may be interpreted as the error covariance of the model background grid point value of the field being updated and the model background estimate of the observation being assimilated divided by the total (observation + background) error variance at the observation location. If the covariance is large, the analysis increment will be larger, and vice versa. The β term defined in eq. (16) appears in eq. (14) because EnSRF schemes do not use perturbed observations (Whitaker and Hamill 2002). When σ_o^2 is large relative to the background error variance, β is small and eq. (14) tends toward using $\overline{H(\mathbf{x}^f)}$ to calculate the innovation determined by the observation for each ensemble member. However, when σ_o^2 is small relative to the background error variance, β approaches 1 and eq. (14) tends toward using $H(\mathbf{x}_n^f)$ to calculate each innovation.

Though model grid point sensible heat flux values will be re-calculated by the Noah land surface model during the first time step following assimilation, effectively eliminating the changes to model sensible heat fluxes resulting from the EnKF update, each Mesonet estimate is used to modify model fluxes in preparation for the assimilation of the next Mesonet estimate. The EnKF scheme described in the preceding paragraphs updates the model fluxes in accordance with the covariance between the model sensible heat flux at the grid point being updated and the model estimate of the observed sensible heat flux. In nature, sensible heat fluxes vary on a

much smaller scale than a model with 3 km horizontal grid spacing may reasonably resolve.

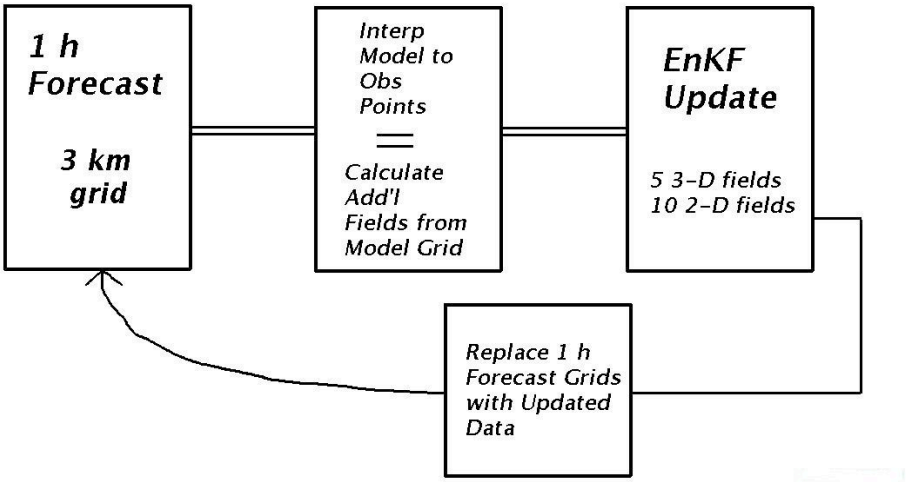


FIG. 3.6: Schematic diagramming the EnKF data assimilation process.

As an alternative and for comparison with the "original" EnKF scheme, a physically-based "new" scheme provides another method by which to update the model sensible heat flux values with each assimilated Mesonet heat flux estimate. Through an analogy with Ohm's law, the sensible heat flux at a point on Earth's surface may be represented by the following equation:

$$H = \frac{\rho c_p (T_s - T_a)}{\Omega_H} \tag{17}$$

where ρ is the density of air, c_p is the specific heat of air at constant pressure, T_s is the temperature of the topmost soil layer, T_a is the near-surface air temperature, and Ω_H is

a “resistance” dependent on the properties of the medium in which the flux is considered (Berkowicz and Prahm 1982). The variables T_s and T_a are updated using the EnKF scheme. Making the assumption that ρ, c_p and Ω_H are constants, and denoting T_s - T_a by ΔT, model grid point sensible heat flux values are updated via the following relation:

$$\frac{\Delta T_{new}}{\Delta T_{orig}} = \frac{H_{new}}{H_{orig}} \tag{18}$$

which, solving for the new sensible heat flux value, yields

$$H_{new} = H_{orig} \frac{\Delta T_{new}}{\Delta T_{orig}}. \tag{19}$$

After all of the valid Mesonet observations pass through the EnKF scheme, the 15 model output fields from the 1 h forecast that were updated through assimilation are replaced with the newly analyzed fields. Once a model restart file has been formed from the newly generated analyses and adjustments, MM5 may be restarted and run for another 1 h time period. An overview of the EnKF assimilation process followed in this study can be referred to in Figure 3.6. Mesonet observations are assimilated every hour during a 48 h period (12 UTC 1 August 2004 – 12 UTC 3 August 2004) to show their effect during two complete diurnal cycles.

Chapter 4

Results from model simulations

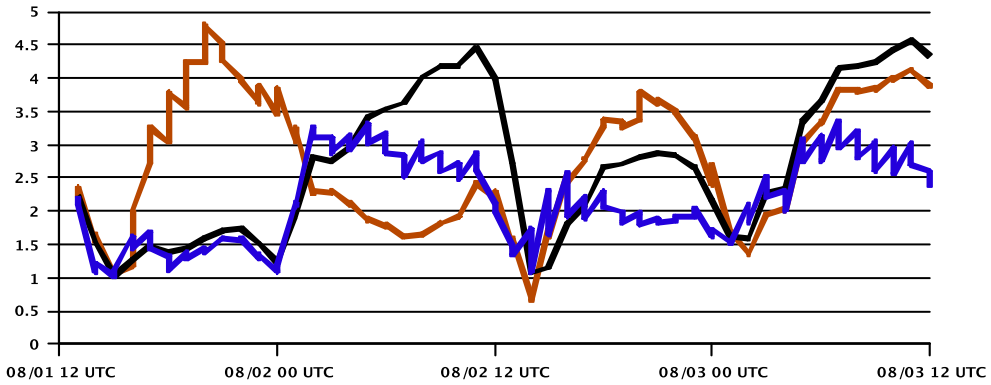
The time period over which the EnKF assimilation scheme is tested extends from 12 UTC 1 August – 12 UTC 3 August 2004. The 500 mb geopotential height analysis from 12 UTC 1 August shows a broad ridge with an axis just east of the Rocky Mountains spreading over the central and southern Plains. Some cloudiness was observed in far eastern Oklahoma during the afternoon of 1 August and limited high clouds were observed in northern and eastern Oklahoma on 2 August, but otherwise the sky was clear over the entire body of the state for the duration of the period. Winds in Oklahoma were affected by the presence of a surface ridge centered in southern Missouri and Arkansas and the presence of a surface low pressure system centered in western Kansas. Afternoon winds in most locations in eastern Oklahoma were between 0 and 6 kt with varying direction, with winds in western Oklahoma being more in the 10-20 kt range and mostly southerly to southwesterly. Radiosonde observations from Norman, Oklahoma (KOUN) at 12 UTC on 1 and 2 August indicate that warming and drying of the already dry atmospheric column above 750 mb took place during the period. Below 750 mb, the atmosphere moistened in association with southerly winds. At 12 UTC 2 August, winds veered from southerly

at 750 mb to northwesterly at 200 mb. The PBL top at 00 UTC 2 August was at 850 mb, with the atmosphere well-mixed below that level.

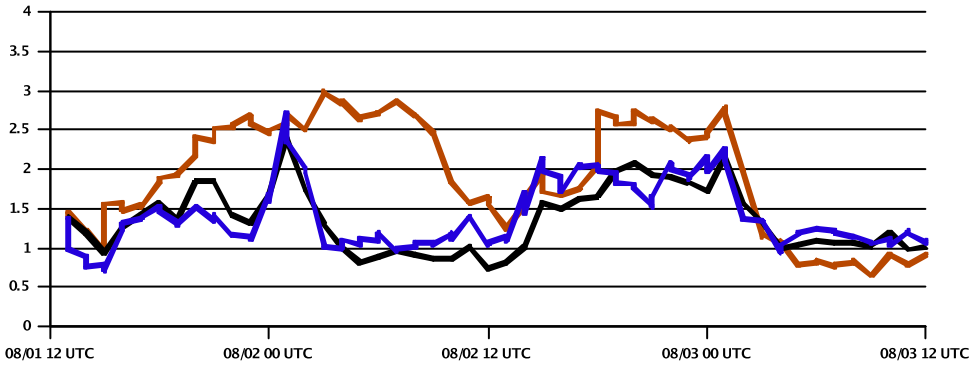
4.1 Spatially averaged error statistics

In an effort to evaluate the changes made to analyses of near-surface atmospheric variables by assimilating observations, three different MM5 ensemble forecast runs are compared. One run (CTRL) is a simple forecast over the period with no data assimilation, with the second run (STD) including hourly assimilation of over 100 Mesonet observations of 1.5 m temperature, 1.5 m relative humidity, 10 m wind, and surface pressure. The third run (FLUX) assimilates only sensible heat flux estimates calculated from observations of wind and temperature at two levels from roughly 90 Oklahoma Mesonet sites. Ensemble mean fields from the model at each hour both before and after data assimilation are compared to measurements from the NWS observing network by calculating bias and root mean square error (rmse) statistics from the observations and analysis values of 2 m temperature and 2 m mixing ratio.

Patterns emerge from the 2 m temperature rmse plot (Fig. 4.1a). Sunrise in Oklahoma City occurred at 1139 UTC 1 August, just prior to the beginning of the model simulations. Sunset was at 0133 UTC 2 August, and the next sunrise was at 1140 UTC 2 August. Sunset on the evening of 2 August and sunrise on the morning of 3 August were within a minute of these times. Error minima are located just after sunset and just before sunrise, a characteristic that should not be surprising given the tendency toward a warm bias at night and a cool bias during the day (Fig. 4.2a).



(a)

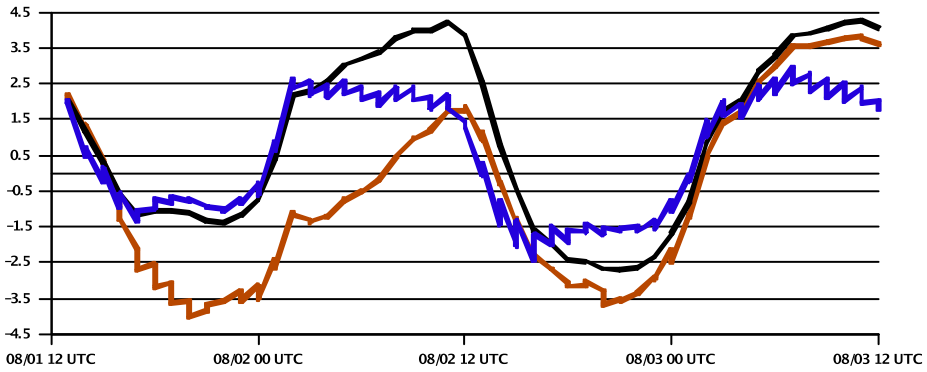


(b)

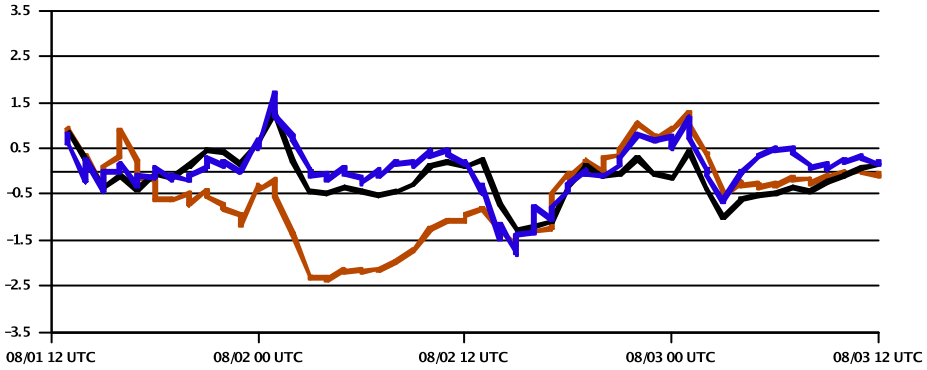
FIG. 4.1: Root mean square error comparing MM5 forecast values from the CTRL (black), STD (blue), and FLUX (red) runs to NWS observations of (a) 2 m temperature in °C and (b) 2 m mixing ratio in g kg^{-1}.

Additionally, substantial error increases often are noted in all three runs following sunrise and sunset. These changes may be attributable to transitions between a stable and an unstable PBL. As mentioned in Section 3.1.2, separate sets of equations are used by PBL parameterizations for stable and unstable conditions, and this can lead to abrupt changes in model near-surface variables.

Differences between the 2 m temperature rmse values from the CTRL and STD runs begin to appear around 03 UTC 2 August. The rmse from the CTRL and STD runs follow a similar path until this time, after which the assimilation process keeps



(a)



(b)

FIG. 4.2: Bias in MM5 forecast values from the CTRL (black), STD (blue), and FLUX (red) runs when compared to NWS observations of (a) 2 m temperature in °C and (b) 2 m mixing ratio in g kg^{-1}.

the rmse from the STD run on a general decreasing trend while the rmse from the CTRL run increases to 4.5ºC by 10 UTC 2 August. The STD run rmse is about 0.2º - 0.5ºC greater than the CTRL run rmse just after sunset and just before sunrise, but otherwise the STD run rmse is below that of the CTRL run through the end of the 48 h assimilation cycle.

The rmse from the FLUX run exhibits different behavior than the rmse from the other two runs. The FLUX run rmse is less than the STD run rmse until 16 UTC 1 August, when the FLUX rmse increases rapidly to a value near 5ºC at 20 UTC 1 August. Between 16 and 20 UTC, the assimilation of sensible heat flux estimates adds error to the temperature analysis, with the forecasts between each hour acting to reduce the error. The FLUX rmse then follows a decreasing trend and reaches a value of about 1.6ºC at 06 UTC 2 August, then increases slightly. The point in time where the trend reverses coincides with the point where the bias changes from negative (cool) to positive (warm) (Fig. 4.2a). The FLUX run again has a greater rmse than either the CTRL or STD runs during the day on 2 August, but with the maximum rmse reaching only about 3.75ºC. The FLUX run rmse reaches a minimum around sunset similar to the CTRL and STD rmse values, then follows just under the CTRL rmse through the overnight and early morning hours on 3 August.

A pattern may also be distinguished in the plots of 2 m mixing ratio rmse (Fig. 4.1b). Apart from a large jump at 01 UTC 2 August, the 2 m mixing ratio rmse from the CTRL and STD runs is similar during both the daytime and nighttime hours on the first day. On the second day, the rmse is greater during the day than at night. Maxima in rmse occur near 00 UTC on both 2 and 3 August, the first around 2.3 g kg^{-1} and the second about 2.1 g kg^{-1}. The 2 m mixing ratio rmse of both the CTRL and STD runs drop by around 1-1.5 g kg^{-1} in the following two to three hours after 00 UTC on each day. On the first day, the assimilation of standard Mesonet observations prevents an increase in 2 m mixing ratio rmse in the STD run that is noted in the CTRL run between 20 and 21 UTC. Additionally, though the rmse in the

STD run is greater than the rmse in the CTRL run by about 0.4 g kg^{-1} prior to assimilation at 01 UTC on 2 August, running the filter does result in an rmse reduction of approximately the same magnitude. The 2 m mixing ratio rmse from the STD run remains relatively close to the rmse from the CTRL run throughout the majority of the 48 h assimilation period.

The 2 m mixing ratio rmse plot from the FLUX run is markedly different from the rmse plots from the other two runs during the first twenty-four hour period, but shows similar tendencies during the second twenty-four hour period. The FLUX rmse is very close to the CTRL rmse through mid-morning on 1 August, at which point it begins a steady increase. The 2 m mixing ratio rmse from the FLUX run reaches a maximum value of around 3 g kg^{-1} at 03 UTC 2 August, while the rmse values from the CTRL and STD runs are near 1 g kg^{-1} at the same time. The FLUX rmse does not begin to decrease consistently until after 06 UTC 2 August, and it remains well above the rmse values from the CTRL and STD runs through the vast majority of the overnight and early morning hours. The 2 m mixing ratio rmse from the FLUX run is lower during the daylight hours on the second day than during the daylight hours on the first day, and decreases noticeably at around sunset on the second day in line with the rmse values from the CTRL and STD runs.

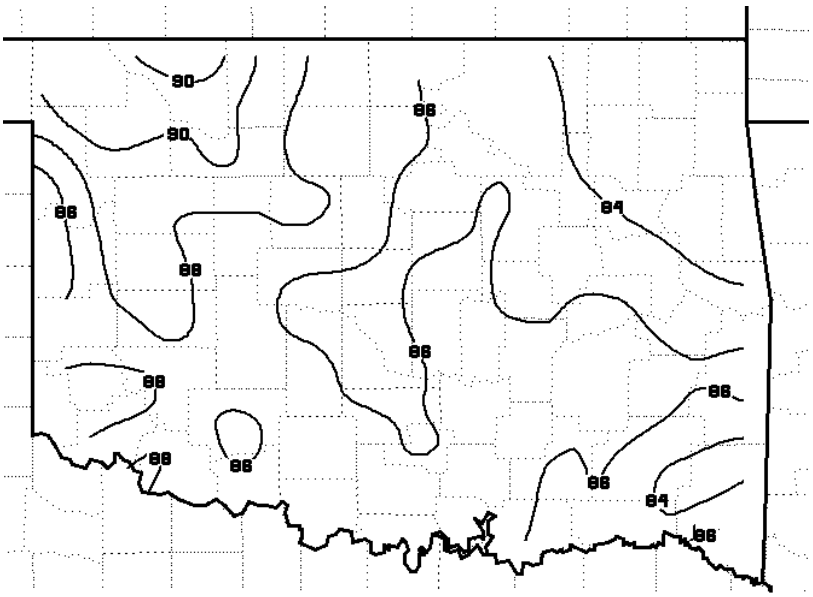
The 2 m mixing ratio bias from the FLUX run (Fig. 4.2b) also shows characteristics not displayed in the biases of the STD and CTRL runs. Forecasts of the 2 m mixing ratio from all runs have a local maximum bias at 01 UTC 2 August, and all have a dry bias of about 1.5 g kg^{-1} during the morning of 2 August. They are fairly close to one another during the last twenty-four hours of the assimilation

period. However, the FLUX run has a dry bias between 0.5 and 2.5 g kg^{-1} greater than the biases from the STD and CTRL runs from mid-day on 1 August until early morning on 2 August. The bias values from the CTRL and STD runs are predominately near zero through this period.

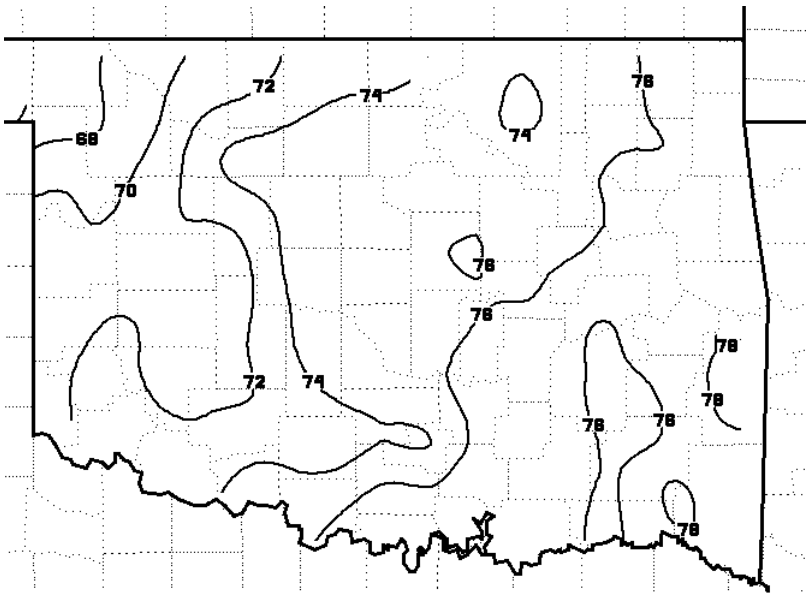
4.2 Two-dimensional surface analyses

As noted earlier, error statistics for the 2 m temperature and mixing ratio from the FLUX run are much different than those from the CTRL and STD runs, especially during the first twenty-four hours of the assimilation period. Assimilation of sensible heat flux estimates usually increases the root mean square error in 2 m temperature during the day on 1 August, sometimes by a large margin (Fig. 4.1a). These error increases result from the filter cooling temperatures by too much on average (Fig. 4.2a). Spatially averaged error in 2 m mixing ratio is also typically increased during this time (Fig. 4.1b), sometimes (but not always) due to the filter excessively decreasing the mixing ratio (Fig. 4.2b). An upward change in 2 m mixing ratio of greater than 1.5 g kg^{-1} is noted at 15 UTC on 1 August, followed by several hours of comparatively small changes, mostly downward.

To look in greater detail at a case when the assimilation of sensible heat flux estimates increases the spatially averaged error in near-surface temperature and dew point, Oklahoma Mesonet analyses, 1 hr model forecasts and model analyses produced following assimilation from 16 UTC on 1 August are compared. An



(a)

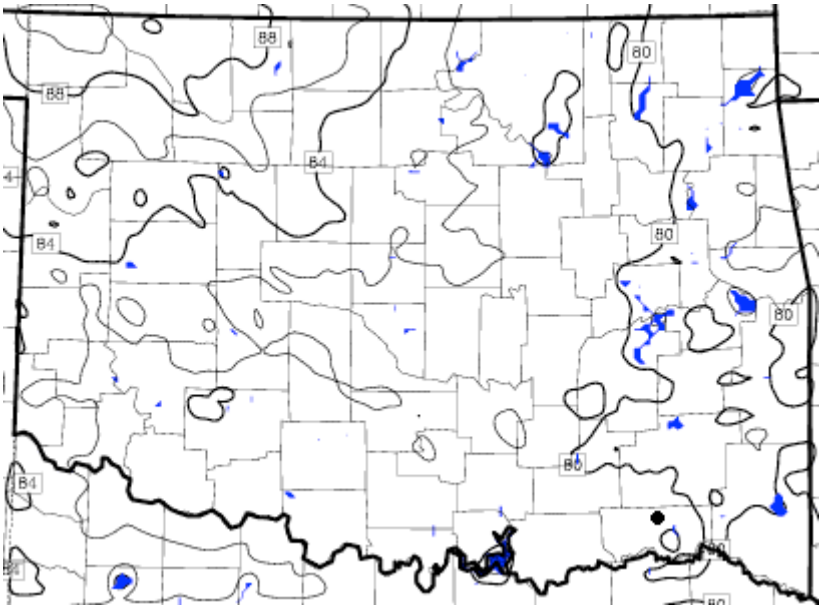


(b)

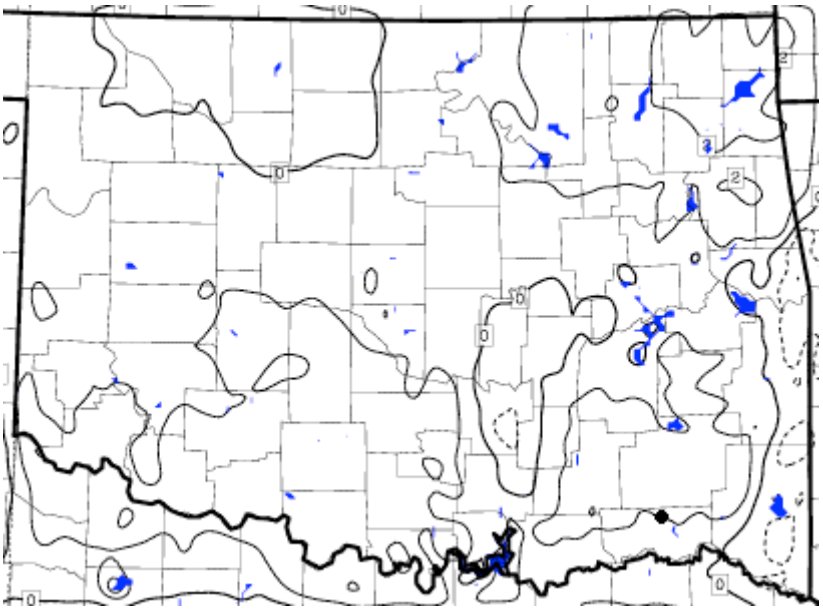
FIG. 4.3: Objective analysis of 16 UTC 1 August Mesonet observations of (a) 1.5 m temperature and (b) 1.5 m dew point. Contours are drawn every 2°F.

analysis of 1.5 m air temperature measurements from the Oklahoma Mesonet at 16 UTC on 1 August places the warmest temperatures of just above 90°F in the northwestern corner of Oklahoma, decreasing to between 82°F and 84°F in the northeastern and southeastern corners of the state (Fig. 4.3a). The 1.5 m dew point temperature analysis identifies a near-surface moisture gradient oriented from northwest to southeast across Oklahoma, with dew points measured at Mesonet sites ranging from around 66°F in far northwestern Oklahoma to around 78°F in far southeastern Oklahoma (Fig. 4.3b).

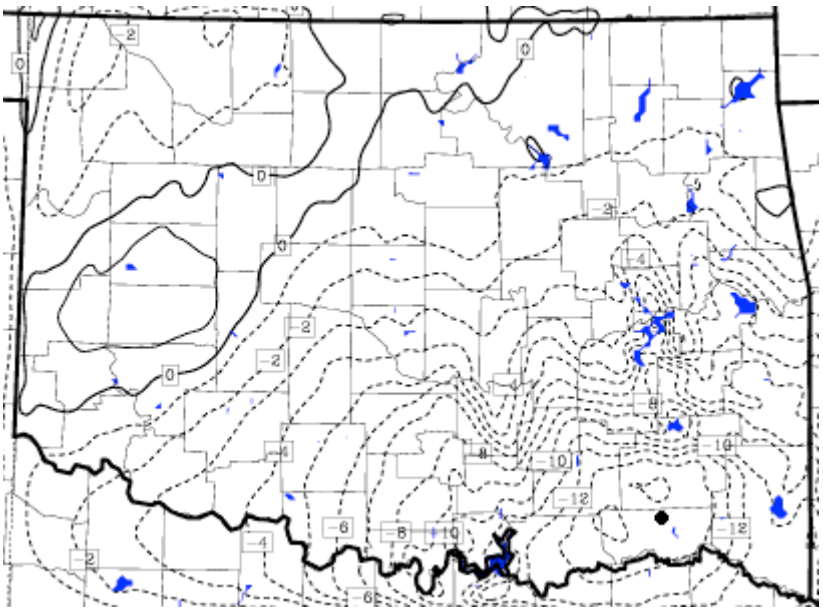
Prior to examining results from the model, a few caveats must be noted for clarity. For the sake of convenience, Mesonet analyses as well as model analyses and forecasts of temperature and dew point are displayed in degrees Fahrenheit. This does make direct comparison with the area-averaged statistics in Section 4.1 more difficult; however, the purpose of those statistics is mainly to display relative trends with less concern for exact magnitudes. Also, Mesonet temperature and moisture measurements are made at a height of 1.5 m while the analogous model output near-surface values are at 2 m. Speaking of temperature, if density were to be constant with height near the ground everywhere resulting in an autoconvective (maximum) lapse rate of approximately 34°C km^{-1}, this discrepancy would result in the Mesonet 1.5 m temperature analyses being uniformly about 0.3°F warmer than concurrent 2 m analyses. Potential discrepancies in moisture at the two levels are somewhat less clear. Finally, two-dimensional model plots are from the twenty-second member of the ensemble and are not from the ensemble mean. It will be shown later that this decision has very little impact on the interpretation of results.



(a)



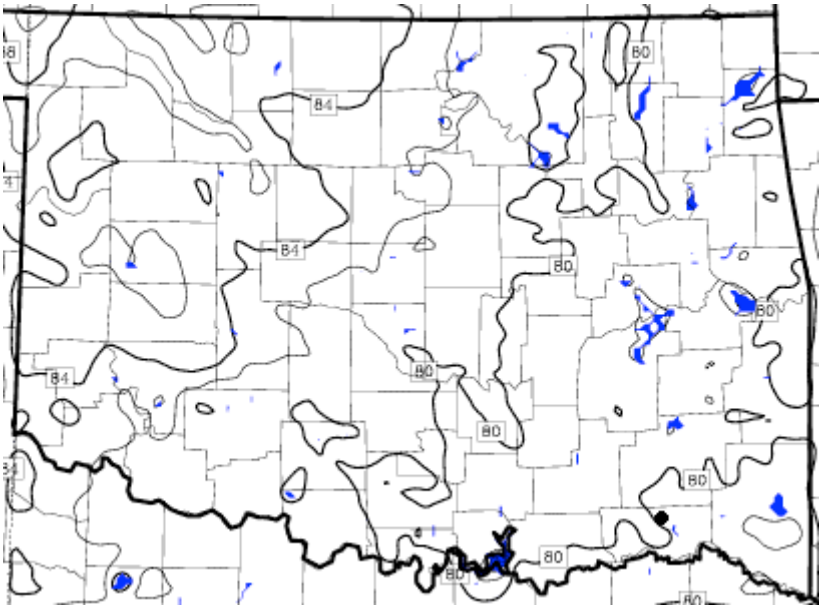
(b)



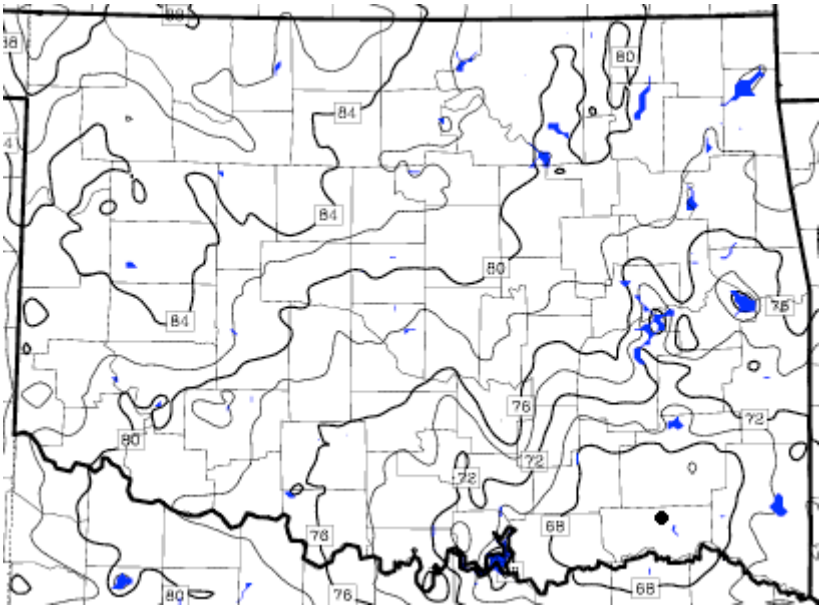
(c)

FIG. 4.4: MM5 output 2 m temperature at 16 UTC 1 August (a) from the CTRL run; (b) as a STD-CTRL run difference; (c) as a FLUX-CTRL run difference. Contours are drawn every 2°F. The black dot in SE Oklahoma denotes the grid point chosen for further discussion in Section 4.3.

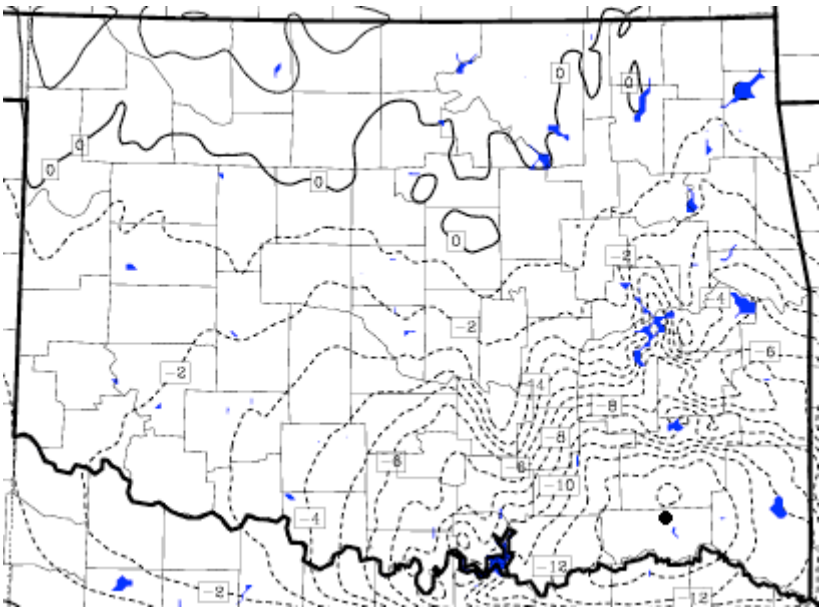
The 2 m temperature bias plot (Fig. 4.2a) from Section 4.1 shows a negative temperature bias from all three runs at 16 UTC 1 August, and this bias is clear from examination of two-dimensional model output. Aside from far northwestern Oklahoma where temperatures are near to only slightly below those from the Mesonet analysis (Fig. 4.3a), lowest model level temperatures from the CTRL forecast are from 3°F to 5°F too cool (Fig. 4.4a). One would expect hourly updates of model forecast variables using the EnKF scheme to result in a model temperature field that more closely resembles the Mesonet temperature analysis. Except for in small portions of east central and far eastern Oklahoma, lowest model level temperatures from the STD run at 16 UTC on 1 August are near to about 2°F above those from the



(a)



(b)

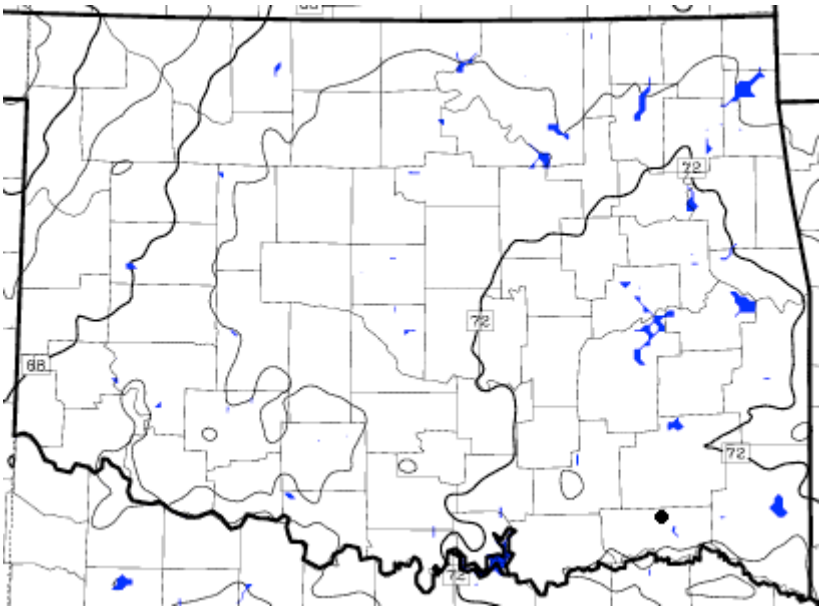


(c)

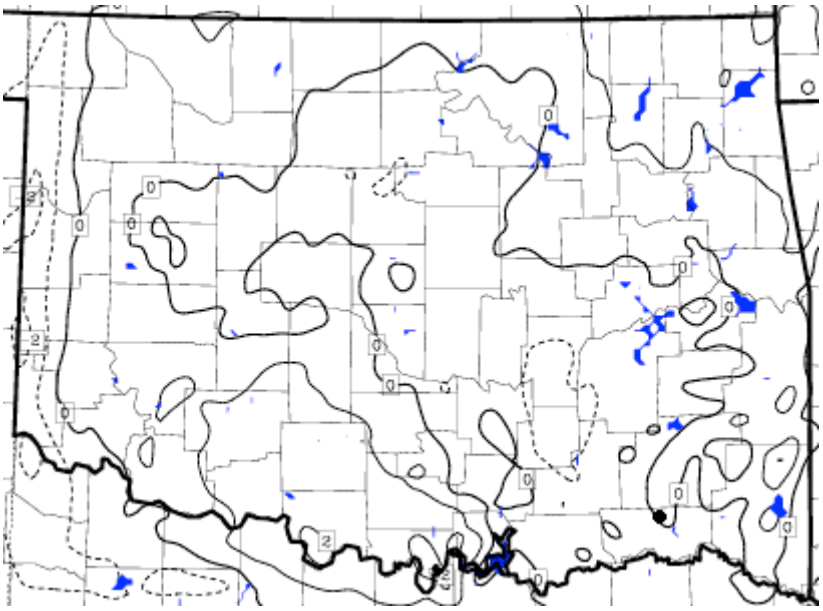
FIG. 4.5: FLUX run 2 m temperature at 16 UTC 1 August (a) prior to assimilation of fluxes; (b) after assimilation of fluxes; (c) the difference between the two. Contours are drawn every 2°F. The black dot in SE Oklahoma denotes the grid point chosen for further discussion in Section 4.3.

CTRL forecast (Fig. 4.4b), meaning that the EnKF scheme produces a temperature change in the correct direction. However, assimilation of sensible heat flux estimates fails to yield the desired improvement in the low level temperature field. As in the STD run, only modest differences between the CTRL forecast and the FLUX run are noted across the northwestern half of Oklahoma (Fig. 4.4c). Contrary to expectations, temperatures from the FLUX run are as much as *14°F cooler* than the CTRL forecast in southeastern Oklahoma, despite the fact that temperatures from the CTRL forecast are already too cool there (Fig. 4.4a).

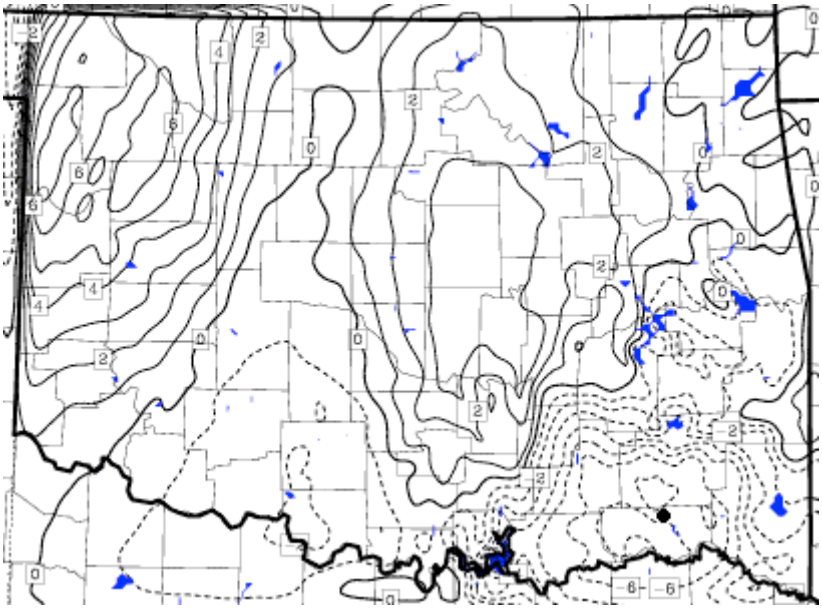
Large magnitude errors, locally greater than 15°F relative to the Mesonet analysis of near-surface temperature, initially develop in the FLUX run at 16 UTC on 1



(a)



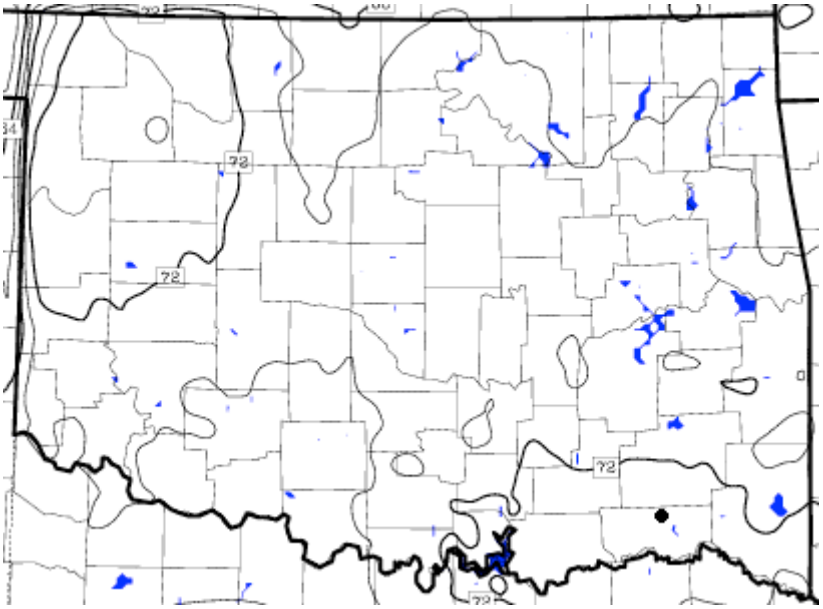
(b)



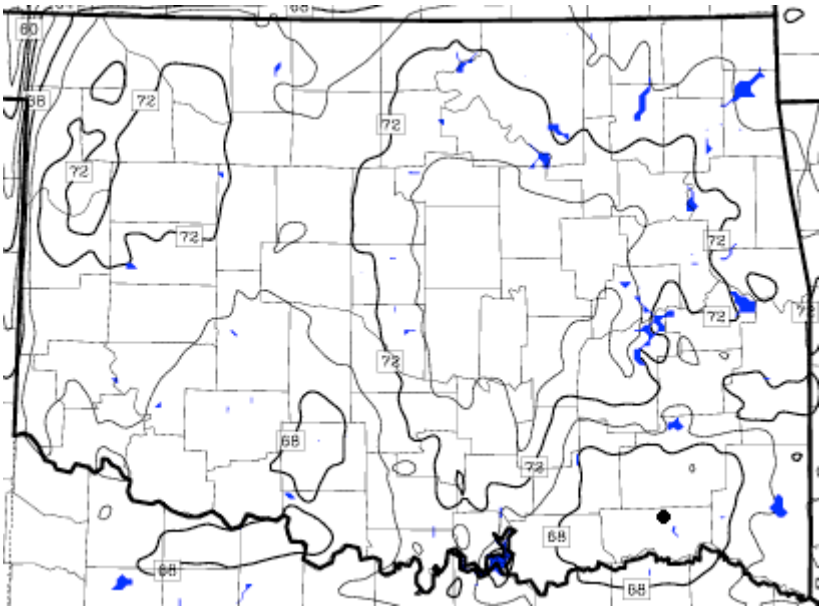
(c)

FIG. 4.6: MM5 output 2 m dew point at 16 UTC 1 August (a) from the CTRL run; (b) as a STD-CTRL run difference; (c) as a FLUX-CTRL run difference. Contours are drawn every 2°F. The black dot in SE Oklahoma denotes the grid point chosen for further discussion in Section 4.3.

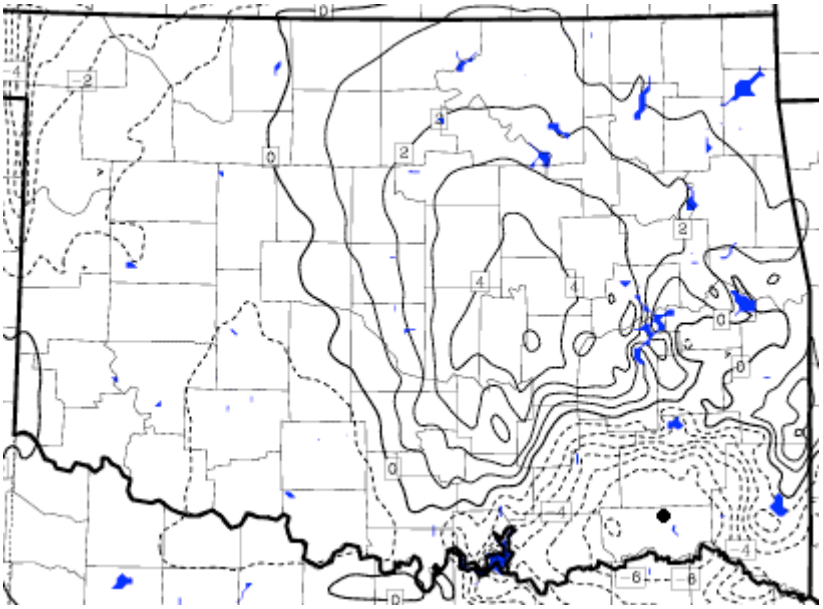
August (Fig. 4.1a; Fig. 4.5). The 16 UTC forecast from the FLUX run prior to applying the filter is in reasonable agreement with the CTRL forecast valid at the same time, with similar error characteristics (Fig. 4.5a; Fig. 4.4a). The analysis resulting from application of the EnKF scheme contains temperature values much lower than those observed in southeastern Oklahoma, with a minimum value of 66°F in central Pittsburg County (Fig. 4.5b). Temperature reductions of greater than 4°F are isolated to roughly the southeastern quarter of Oklahoma, with much smaller reductions and even some appropriate modest temperature increases occurring in the remainder of the state (Fig. 4.5c).



(a)



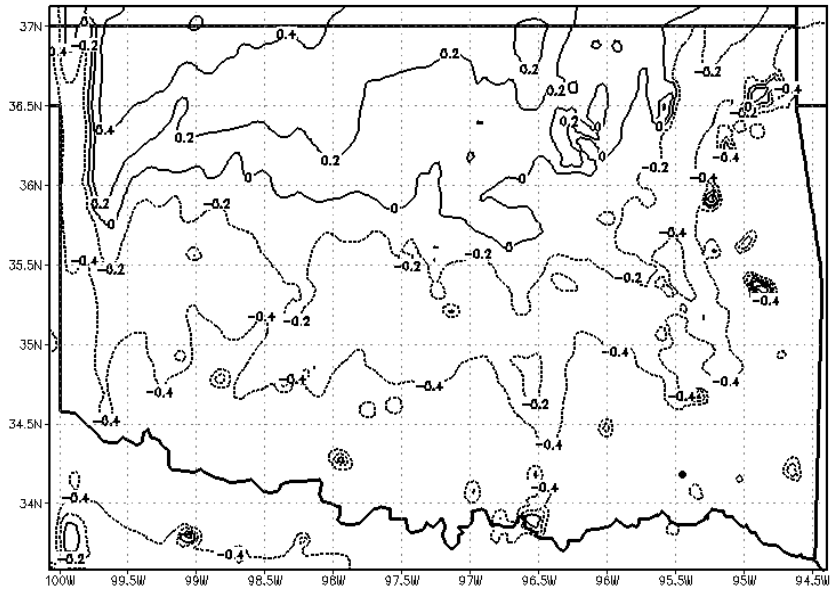
(b)



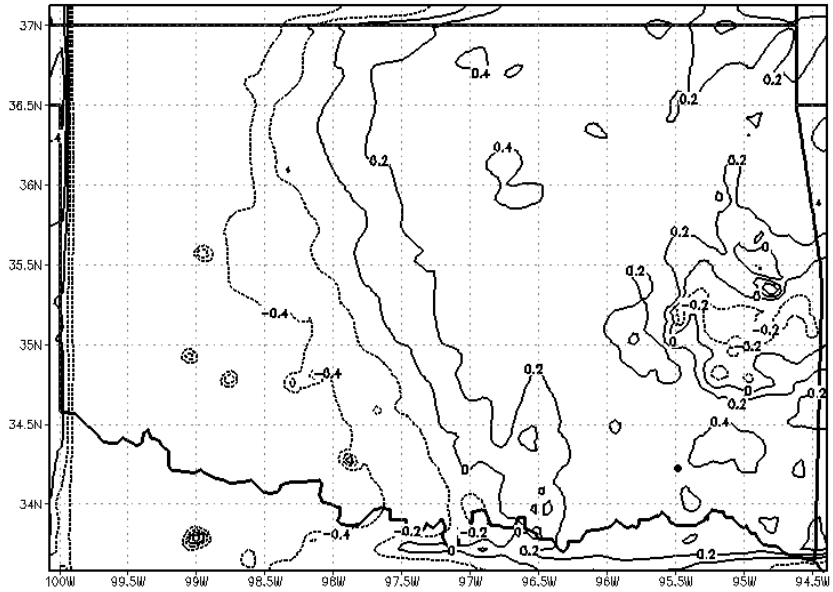
(c)

FIG. 4.7: FLUX run 2 m dew point at 16 UTC 1 August (a) prior to assimilation of fluxes; (b) after assimilation of fluxes; (c) the difference between the two. Contours are drawn every 2°F. The black dot in SE Oklahoma denotes the grid point chosen for further discussion in Section 4.3.

Similar to low-level temperature, additional details regarding changes in low-level moisture bias and rmse values at 16 UTC 1 August are revealed through examination of two-dimensional fields. Moisture bias values from the CTRL and STD runs are near zero at this time, while a pronounced moist bias develops in the FLUX run (Fig. 4.2b). Minimal rmse changes are noted in all three runs, with the rmse from the FLUX run dropping slightly (Fig. 4.1b). The four hour forecast of lowest model level dew point from the CTRL run (Fig. 4.6a) is from 4°F to 6°F lower than the Mesonet analysis (Fig. 4.3b) over much of the eastern half of Oklahoma, and from 2°F to 4°F lower over much of the western half. The 16 UTC analysis from the STD run is barely different from the CTRL forecast (Fig. 4.6b). In contrast, the 2 m dew points



(a)



(b)

FIG. 4.8: Correlation between the sensible heat flux at the designated point in southeastern Oklahoma and (a) 2 m temperature; (b) 2 m mixing ratio at all grid points. Contours are drawn at intervals of 0.2. Data for the calculation are taken from the FLUX run ensemble at 16 UTC on 1 August.

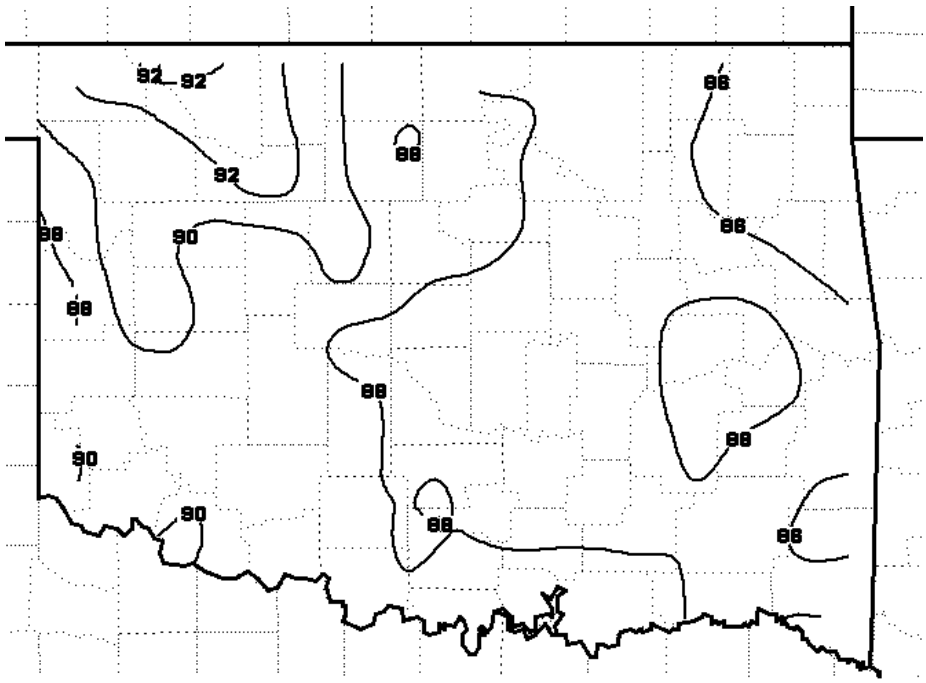


FIG. 4.9: Objective analysis of 16 UTC 2 August Mesonet observations of 1.5 m temperature. Contours are drawn every 2°F.

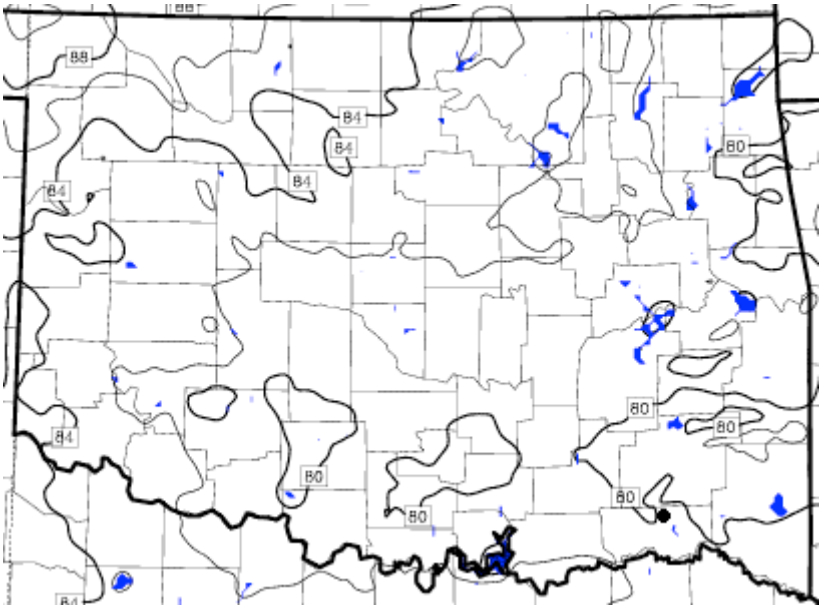
from the FLUX run analysis (Fig. 4.6c) are 3°F to 7°F higher than the CTRL forecast in the northwestern corner of Oklahoma, 4°F to 8°F lower in a portion of southeastern Oklahoma, and within 2°F of the CTRL forecast over most of the rest of the state.

The bias in the low level mixing ratio from the FLUX run increased by around 0.5 g kg^{-1} at 16 UTC on 1 August (Fig. 4.2b), while there was a modest decrease in the rmse (Fig. 4.1b). Forecast dew points from the FLUX run at 16 UTC before updating the moisture field with the Kalman filter are between 70°F and 72°F across nearly the entire domain, with a tight eastward-pointing moisture gradient in far northwestern Oklahoma (Fig. 4.7a). Following the EnKF update, dew points are reduced to the upper 60s Fahrenheit across far northwestern and much of southern Oklahoma, while dew points are increased to the middle 70s Fahrenheit in east central Oklahoma (Fig.

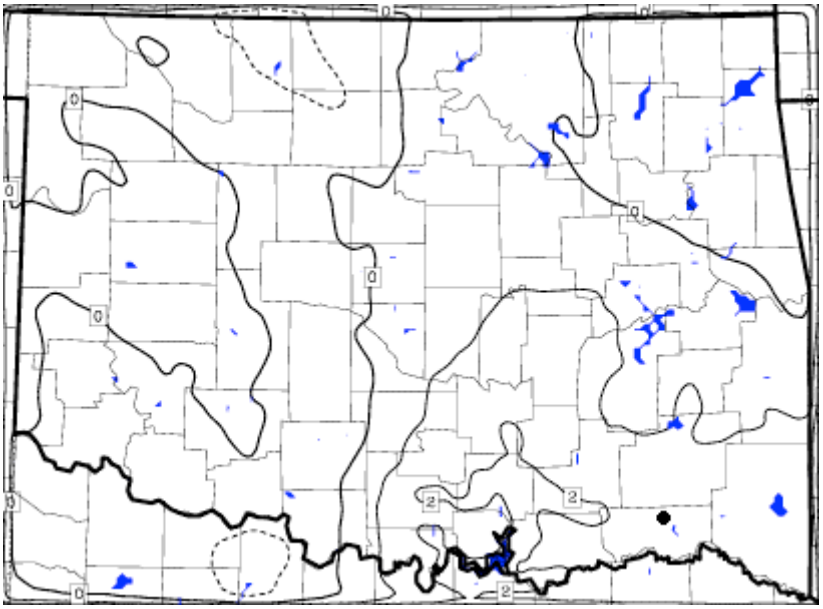
4.7b). Drying the model near-surface air in far northwestern Oklahoma and moistening the air in east central Oklahoma (Fig. 4.7c) bring the model moisture field closer to the values in the Mesonet analysis (Fig. 4.3b), while drying over southern portions of Oklahoma pushes the model farther from the Mesonet analysis.

Covariance diagrams help us to assess the effect of the EnKF scheme on low level temperature and moisture fields in the model due to the Kalman gain (eq. (15) from Section 3.4). The covariances between the model sensible heat flux at a point in southeastern Oklahoma and the (a) model 2 m temperature and (b) model 2 m mixing ratio at all other grid points are plotted in Figure 4.8. Recall from Section 3.1.2 that the iterative method for estimating sensible heat flux has a high bias. For temperature, negative covariance values of magnitude 0.4 or greater stretch across southern Oklahoma and into far eastern Oklahoma (Fig. 4.8a), suggesting that one or more nearby Mesonet estimates of sensible heat flux much greater than the horizontally interpolated model value could at least partially be responsible for the drastic temperature decrease in southeastern Oklahoma (Fig. 4.5c). Positive moisture covariances mostly from 0.2 - 0.4 prevail in the region of substantial moisture decrease (Fig. 4.8b; Fig. 4.7c). However, again assuming nearby Mesonet sensible heat flux estimates are larger than model values, a positive covariance should lead to a positive Kalman gain and an *increase* in 2 m dew point.

In some cases (e.g., Fig. 4.6c, Fig. 4.8b) when physically inconsistent changes are made to model background fields as a result of the assimilation of sensible heat flux estimates, tight gradients develop near the boundaries of the model domain. These gradients result from the large discrepancy between the GFS-based boundary

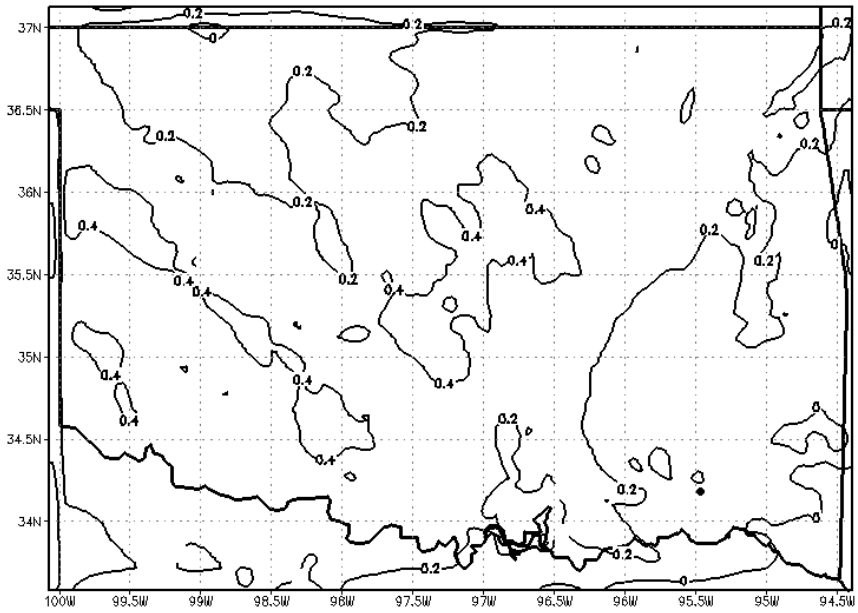


(a)

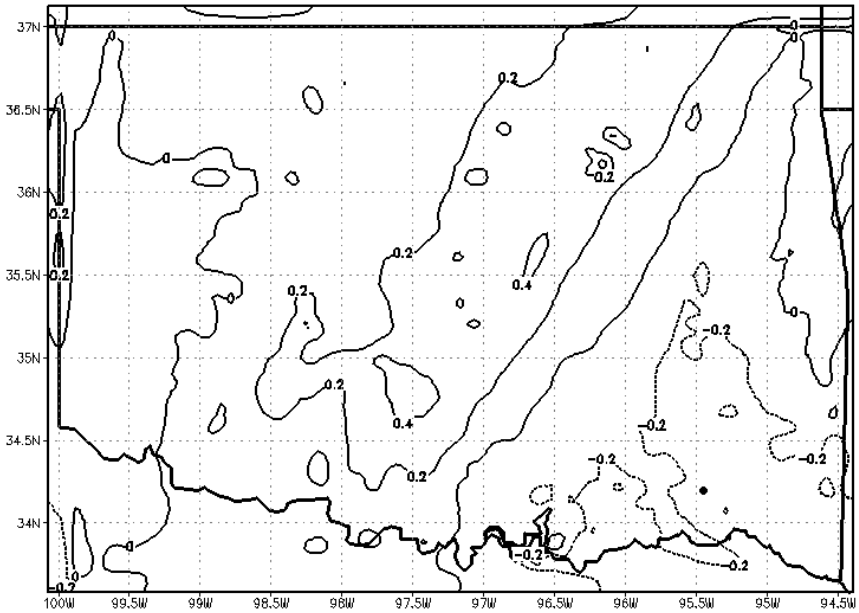


(b)

FIG. 4.10: FLUX run 2 m temperature at 16 UTC 2 August (a) prior to assimilation of fluxes; (b) change attributed to flux assimilation. Contours are drawn every 2°F. The black dot in SE Oklahoma denotes the grid point chosen for further discussion in Section 4.3.



(a)



(b)

FIG. 4.11: Correlation between the sensible heat flux at the designated point in southeastern Oklahoma and (a) 2 m temperature; (b) 2 m mixing ratio at all grid points. Contours are drawn at intervals of 0.2. Data for the calculation are taken from the FLUX run ensemble at 16 UTC on 2 August.

conditions and the updated model fields. The tight gradients tend to remain near the boundary with time, not propagating into the model domain. For instance, the boundary effects do not have any bearing on the large temperature decrease in southeastern Oklahoma.

Though large-magnitude degradations of low level temperature and moisture fields occasionally arise following EnKF assimilation of sensible heat flux estimates during the daylight hours on 1 August, such changes infrequently occur during the same period on 2 August (Fig. 4.1, 4.2). At 16 UTC on 2 August, the forecast of lowest model level temperature from the FLUX run is cooler than the Mesonet temperature analysis by roughly 4°F - 8°F across the entire domain (Fig. 4.9; Fig. 4.10a). Excepting parts of south central and southeast Oklahoma where temperatures were raised by up to 2.5°F and a relatively small portion of northern Oklahoma where temperatures were lowered by 1°F - 2°F, temperatures were changed by less than 1°F as a result of sensible heat flux assimilation (Fig. 4.10b). The spatial distribution of covariances between the sensible heat flux at the same southeastern Oklahoma point as on 1 August and the 2 m temperature at all other grid points suggests that the filter should not make large changes to the temperature field at this time (Fig. 4.11a). Application of the EnKF scheme at 16 UTC on 2 August produced 2 m dew point changes of less than 1°F everywhere (not shown) concurrent with small spatial covariance values between sensible heat flux and low level moisture (Fig. 4.11b). These results build confidence that the EnKF scheme is operating correctly.

Number	Site ID	Number	Site ID	Number	Site ID
1	ANTL	12	TISH	23	HECT
2	HUGO	13	CALV	24	BRIS
3	CLOU	14	WIST	25	RING
4	TALI	15	ADAX	26	BIXB
5	IDAB	16	OKEM	27	CHAN
6	MCAL	17	BYAR	28	TAHL
7	WILB	18	OKMU	29	WASH
8	DURA	19	PAUL	30	NRMN
9	BROK	20	HASK	31	WEST
10	CENT	21	COOK	32	OILT
11	STUA	22	SHAW	33	SPEN

TABLE 4.1: Oklahoma Mesonet observing sites contributing to the flux estimates assimilated by the EnKF scheme at the grid point of focus in southeastern Oklahoma at 16 UTC on 1 August. Observing sites are ordered according to increasing distance from the grid point.

Number	Site ID	Number	Site ID	Number	Site ID
1	ADAX	12	HASK	23	STUA
2	ANTL	13	HUGO	24	TAHL
3	BIXB	14	IDAB	25	TALI
4	BRIS	15	MCAL	26	TISH
5	BYAR	16	OILT	27	WASH
6	CALV	17	OKEM	28	WEST
7	CENT	18	OKMU	29	WILB
8	CHAN	19	PAUL	30	WIST
9	CLOU	20	RING	31	HECT
10	COOK	21	SHAW	32	NRMN
11	DURA	22	SPEN	33	BROK

TABLE 4.2: Oklahoma Mesonet observing sites contributing to the flux estimates assimilated by the EnKF scheme at the grid point of focus in southeastern Oklahoma at 16 UTC on 1 August. Observing sites are arranged sequentially.

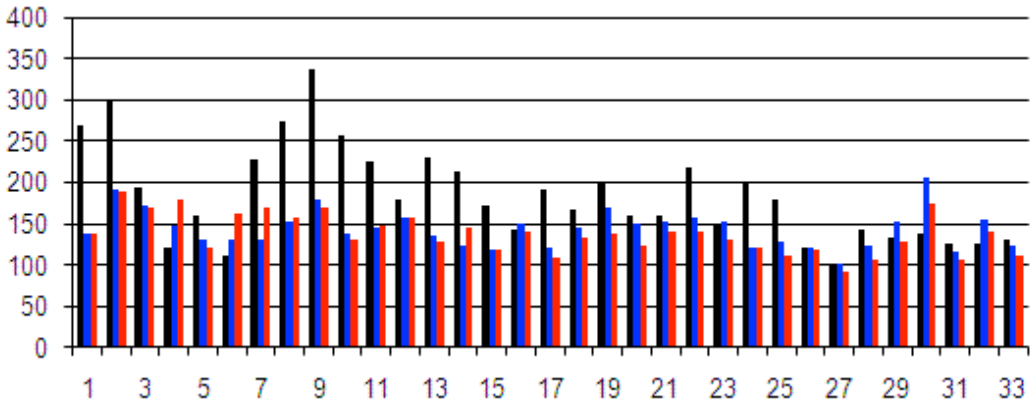


FIG. 4.12: Mesonet sensible heat flux estimates (black), model interpolated values using the original assimilation scheme (blue), and model interpolated values using the new scheme (red) at 16 UTC on 1 August. Values are from the grid point of focus in southeastern Oklahoma, and stations are ordered according to increasing distance from the grid point. Sensible heat flux values along the *y*-axis are in W m^{-2}, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

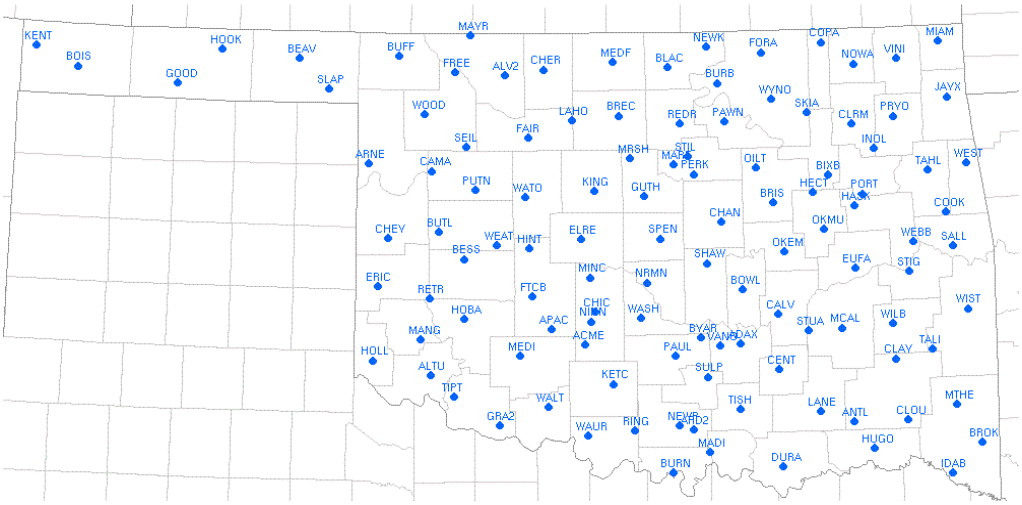


FIG. 4.13: Locations of Oklahoma Mesonet observing sites denoted by four-letter identifiers.

4.3 Point-specific analysis of EnKF equation terms

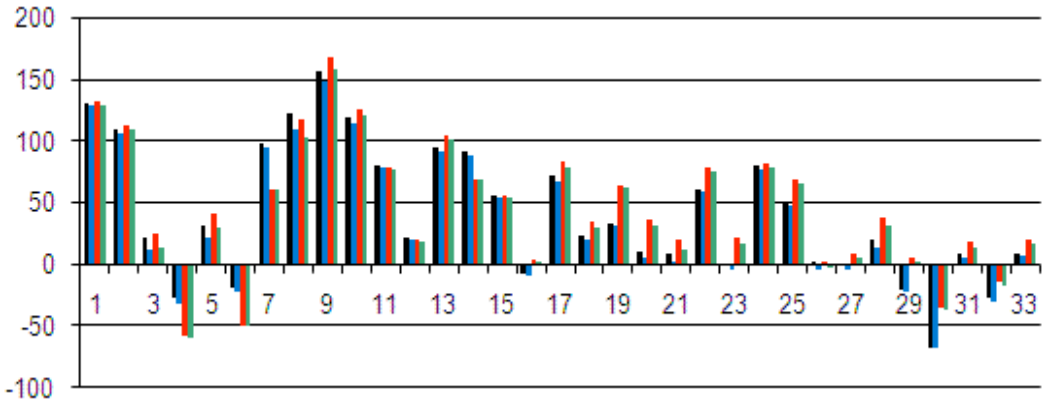
Though updates to model temperature and moisture fields from the FLUX run appear to be consistent with spatial covariances for the most part, the fact remains that use of the EnKF procedure to assimilate sensible heat flux estimates occasionally produces spurious and undesirable results. At 16 UTC on 1 August, the EnKF scheme reduces the lowest model level temperature from the FLUX run 1 h forecast in much of southeastern Oklahoma from 5°F - 15°F despite Mesonet temperatures being *warmer* than the model temperatures (Fig. 4.3a). Similarly, FLUX run lowest model level dew points are reduced in much of southeastern Oklahoma even though an analysis of Mesonet data shows dew points higher than the model forecast (Fig. 4.3b). Spatial covariances (Fig. 4.8) would assert that Mesonet sensible heat flux estimates much higher than model grid point values could lead to a large temperature decrease in southeastern Oklahoma, but should result in a dew point increase at the same time. An examination of terms from the EnKF equations for individual observations will demonstrate that sensible heat flux estimates much larger than model background values have a profound effect on the resulting analyses of low level temperature and moisture fields.

Necessarily, a numerical model must check to make sure that the dew point is less than or equal to the temperature at all grid points prior to initiating a forecast run. Since two-dimensional plots of temperature and dew point after running the EnKF scheme are drawn from the first time step of the subsequent one-hour forecast, this check would have already been run. This line of reasoning explains the dew point

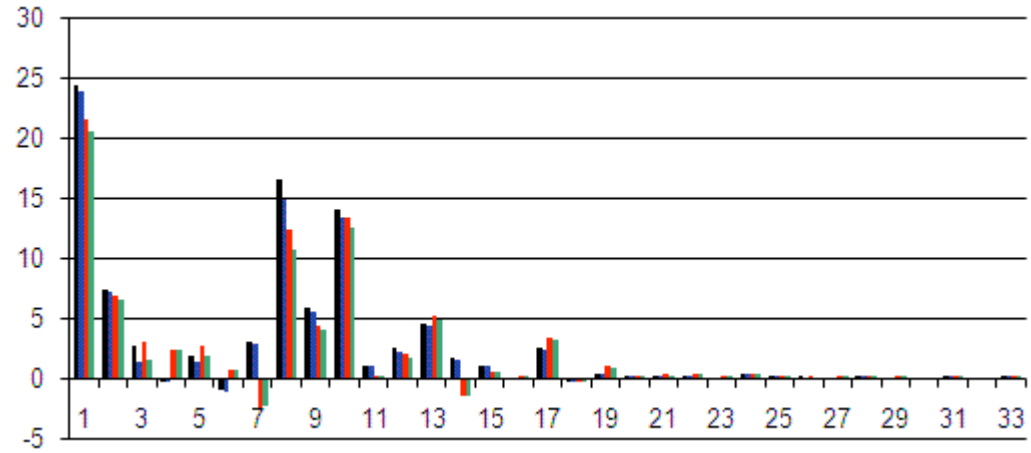
decrease in southeastern Oklahoma at 16 UTC when by all other accounts there should be an increase. Consequently, a look at specific terms from the Kalman filter equations for Mesonet observations both individually and collectively should show an increase in near-surface moisture in southeastern Oklahoma.

Hereafter, both the original and new EnKF update processes from Chapter 3 are closely examined at the same grid point in southeastern Oklahoma chosen for the generation of the covariance maps in Section 4.3. Thirty-three Mesonet sensible heat flux estimates are used to update the model fields at 16 UTC on 1 August. These Mesonet estimates, along with ensemble mean model values interpolated to Mesonet locations using the original and the new update schemes, are plotted in Figure 4.12. Mesonet sites are shown on a map in Figure 4.13, and Mesonet site identifiers from this map may be matched with numbers along the *y*-axis of Figure 4.12 using Table 4.1. Mesonet estimates of sensible heat flux regularly exceed model interpolated values, sometimes by more than 100 W m^{-2}. Twenty-four Mesonet estimates are above model interpolated values, while only nine estimates are below model values. Model fluxes vary mainly from 100 to 200 W m^{-2}, while Mesonet estimates vary from 100 to almost 350 W m^{-2}.

Innovations (differences between Mesonet sensible heat flux estimates and model interpolated values) and modifications to model grid point sensible heat flux values vary more due to changes in the update scheme rather than between the ensemble mean and an individual member (Fig. 4.14a,b). Innovation differences between the ensemble mean and the twenty-second member typically are less than 5 W m^{-2}, but differences between the original and the new scheme run up to around 30 W m^{-2},



(a)



(b)

FIG. 4.14: Graphs of (a) innovations and (b) changes to sensible heat flux during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are numbered according to increasing distance from the grid point. Sensible heat flux values along the *y*-axis are in W m^{-2}, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

especially in the latter portion of the assimilation cycle. Only five out of thirty-three Mesonet estimates (ANTL, CENT, DURA, HUGO, BROK) produce an innovation of greater than 100 W m^{-2} (Fig. 4.14a), and of those five two (ANTL and HUGO) are within 20 km of the grid point being modified (Fig. 4.13). Increases greater than 10 W m^{-2} in the grid point sensible heat flux resulting from a single observation are limited to ANTL, CENT and DURA, with most other observations changing the grid point flux by 5 W m^{-2} or less (Fig. 4.14b). In accordance with the fact that most Mesonet estimates are greater than the corresponding model interpolated values, the model grid point sensible heat flux finishes the assimilation process between 70 and 90 W m^{-2} larger than at the beginning (Fig. 4.15). The new scheme increases the grid point flux value by around 12 W m^{-2} less than the original scheme.

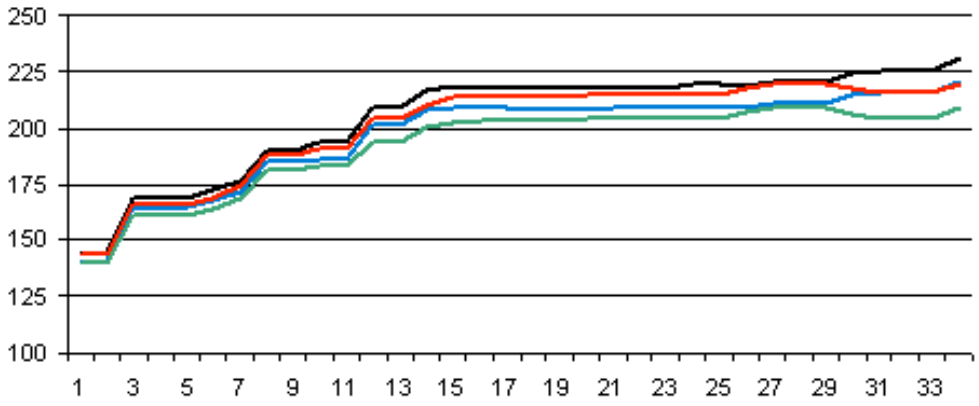


FIG. 4.15: Sensible heat flux values during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are arranged sequentially. Sensible heat flux values along the *y*-axis are in W m^{-2}, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.2.

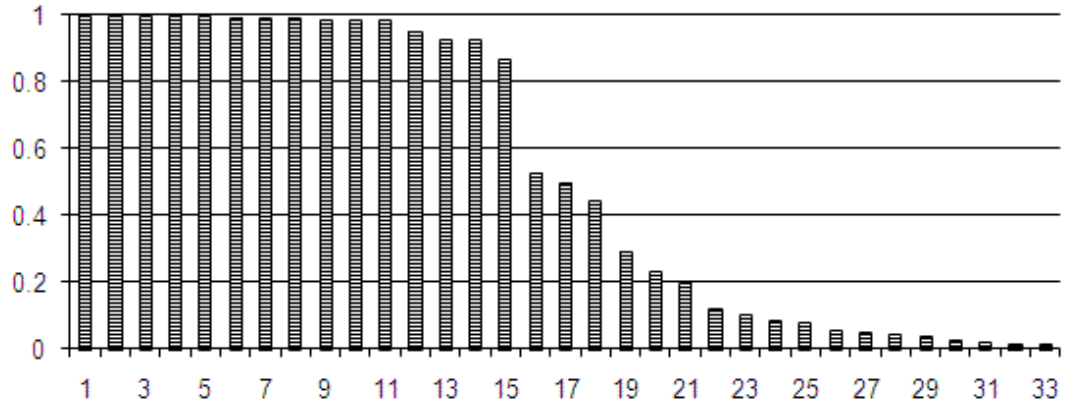


FIG. 4.16: Weight (*W*) from eq. (13) and (14) for individual Mesonet sites relative to the grid point of focus in southeastern Oklahoma, arranged according to increasing distance from the grid point. Weights are on the *y*-axis, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

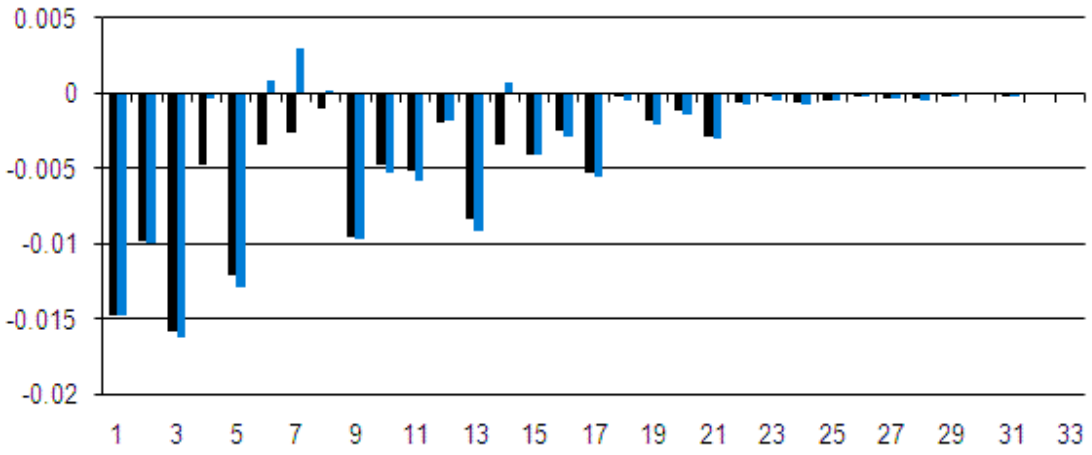


FIG. 4.17: Low level temperature Kalman gain multiplied by the weight (*WK*) from eq. (13) and (14) for individual Mesonet sites relative to the grid point of focus in southeastern Oklahoma, arranged according to increasing distance from the grid point. Values for the original EnKF scheme are in black, and values for the new EnKF scheme are in blue. Unitless *WK*-values are on the *y*-axis, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

Equations (13) - (16) from Chapter 3 define the EnKF scheme used for this case. For each observation, the corresponding sensible heat flux innovation goes into eq. (13) as $y^o - \overline{H(\mathbf{x}^f)}$ for the ensemble mean or into eq. (14) as $y^o - \beta H(\mathbf{x}_n^f) - (1.0 - \beta)\overline{H(\mathbf{x}^f)}$ for an individual member. The larger the innovation, the more the grid point variable being updated may be changed. At 16 UTC on 1 August, the value of β is less than or equal to 0.55 for every Mesonet sensible heat flux estimate such that the ensemble mean has a weight nearly equal to the weight of the individual ensemble member in calculating observation increments according to eq. (14). The weights *W* (Fig. 4.16), proportional to the inverse of the exponential function, are merely based on distance from the Mesonet site to the grid point and are the same for each model variable updated. Values of *W* vary from 0.0 to 1.0, but only four of the observations (HASK, OKEM, OKMU, PAUL) have a weight between 0.2 and 0.8.

All of the temperature Kalman gain values from the original scheme are negative (Fig. 4.17). For four of the Mesonet estimates (DURA, MCAL, WILB, WIST), the new assimilation scheme actually changes the sign of the Kalman gain from negative to positive. Kalman gains in Figure 4.17 include the localization weight, representing the *WK* term from eq. (13) and eq. (14). The number of negative Kalman gains should not be surprising given the propensity for Mesonet sensible heat flux estimates to be greater than model interpolated values (Fig. 4.12) and given the temperature covariance map for the grid point of focus (Fig. 4.8a).

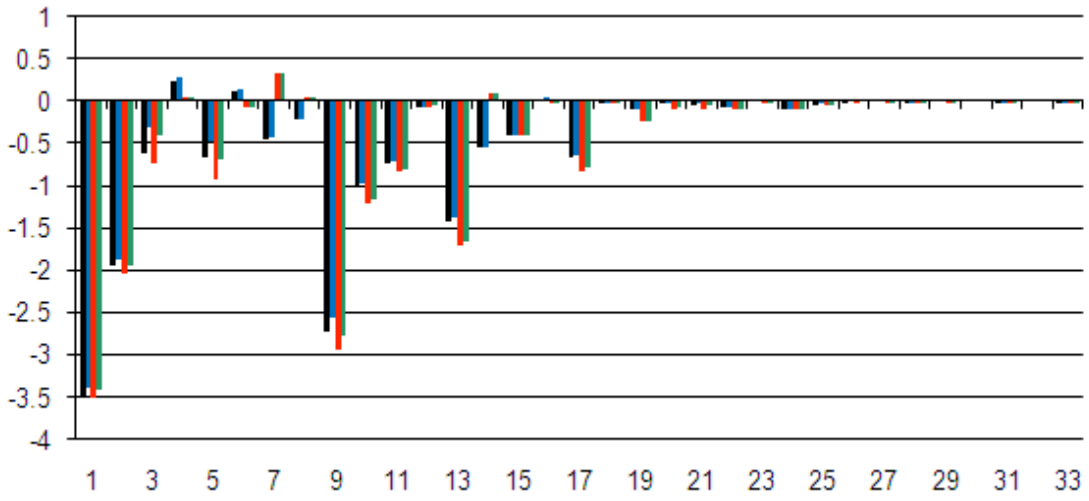


FIG. 4.18: Changes to lowest model level temperature during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are arranged in order of increasing distance from the grid point. Temperature values along the *y*-axis are in °F, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

As is the case for sensible heat flux, a very small subset of observations is responsible for the vast majority of the change in temperature by the assimilation schemes at the grid point of focus. Four Mesonet sensible heat flux estimates decrease the low level temperature at the grid point of focus by more than 1°F: ANTL, CALV, HUGO and BROK (Fig. 4.18). These four estimates contribute a staggering 65% of the temperature decrease, with the remaining twenty-nine estimates contributing the other 35%. At each of these four Mesonet sites, an innovation of 100 W m^{-2} or greater combines with a "*WK*" term that is comparatively large to produce a substantial temperature drop. The most negative "*WK*" term arises at CLOU, but with an observation increment less than 25 W m^{-2}, the influence of this

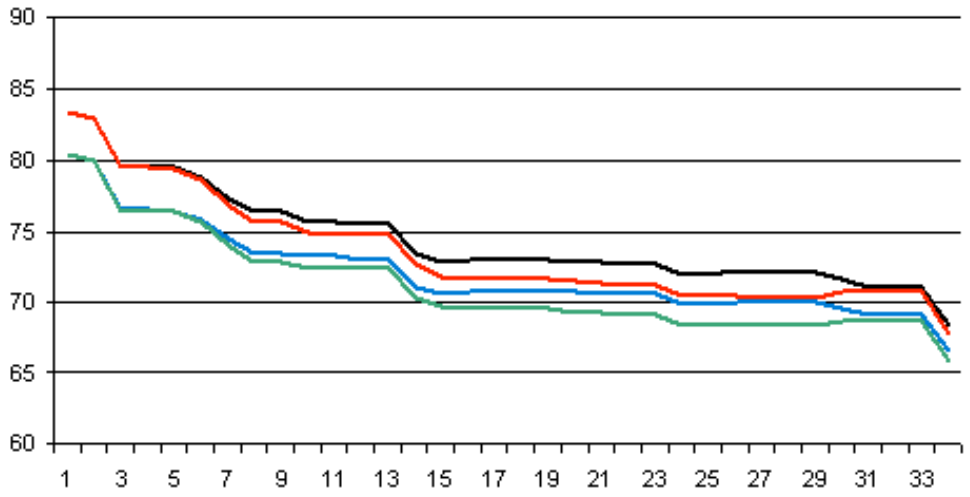


FIG. 4.19: Actual lowest model temperature values during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are arranged sequentially. Temperature values along the *y*-axis are in °F, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.2.

observation is limited. Innovations at DURA exceed 100 W m^{-2}; however, a "*WK*" term near zero means that this estimate modifies the temperature little.

At the grid point of focus, more variation in the 16 UTC total temperature change occurs between individual ensemble members and the ensemble mean than results from using the new scheme as opposed to the old scheme. Figure 4.19 shows the path that the lowest model level temperature takes during the assimilation process. The EnKF scheme makes only very small changes to the model forecast ground temperature while making much greater changes to the low level air temperature. This fact largely explains the minimal differences in temperature change between the old and new schemes. Depending on the particular member or scheme chosen,

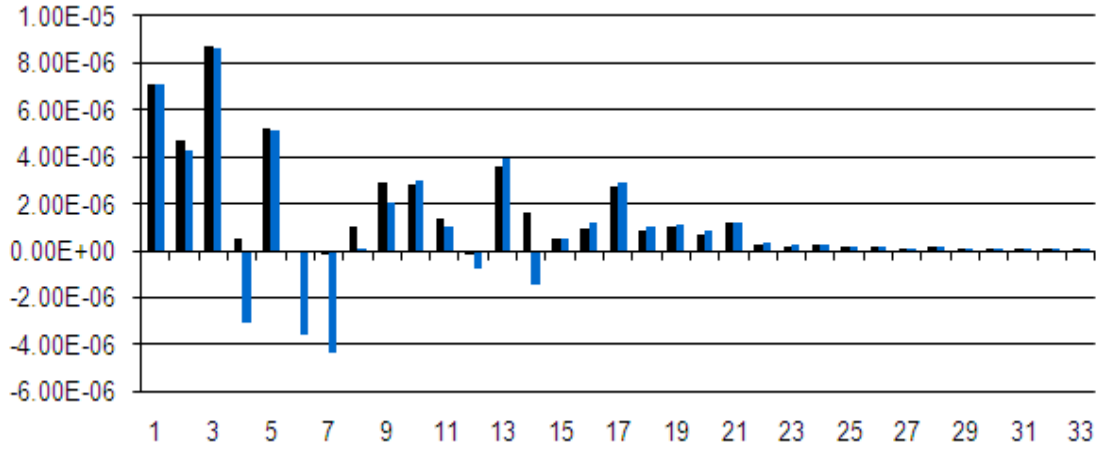


FIG. 4.20: Low level mixing ratio Kalman gain multiplied by the weight (*WK*) from eq. (8) and (9) for individual Mesonet sites relative to the grid point of focus in southeastern Oklahoma, arranged according to increasing distance from the grid point. Values for the original EnKF scheme are in black, and values for the new EnKF scheme are in blue. Unitless *WK*-values are on the *y*-axis, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

assimilation of Mesonet sensible heat flux estimates drops the low level temperature by around 15°F despite the fact that the EnKF scheme is working correctly!

Mixing ratio Kalman gain values in Figure 4.20 include localization weights in the style of Figure 4.17. Opposite to the Kalman gain for temperature, values for mixing ratio at the grid point of focus are predominantly positive when the original EnKF scheme is applied. Some of the original Kalman gain values are near zero, with four of these (MCAL, TALI, WILB, WIST) becoming negative when the new assimilation scheme is used. The signs of the Kalman gain values from the original EnKF scheme match well with the positive covariances between low level mixing ratio and the grid point sensible heat flux (Fig. 4.8b).

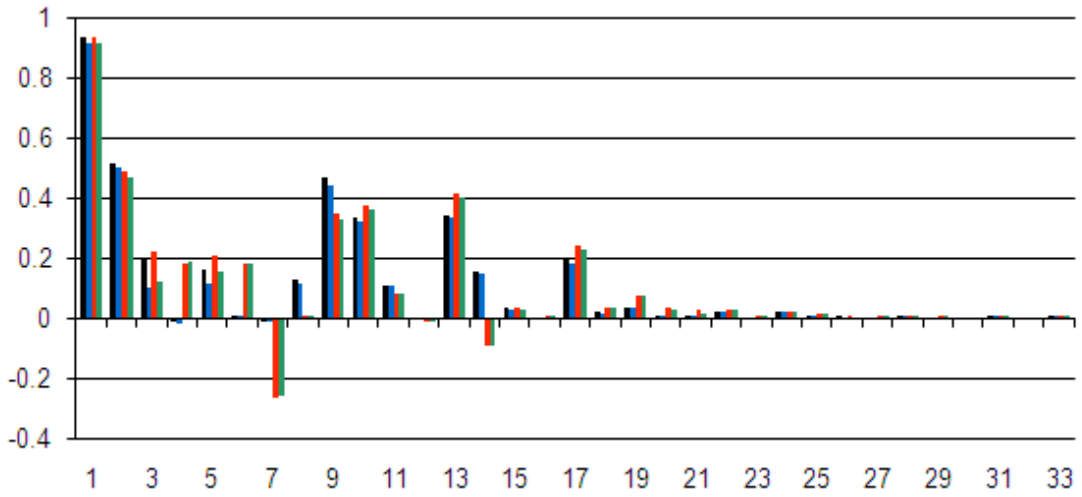


FIG. 4.21: Changes to lowest model level mixing ratio during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are arranged in order of increasing distance from the grid point. Mixing ratio values along the *y*-axis are in g kg^{-1}, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.1.

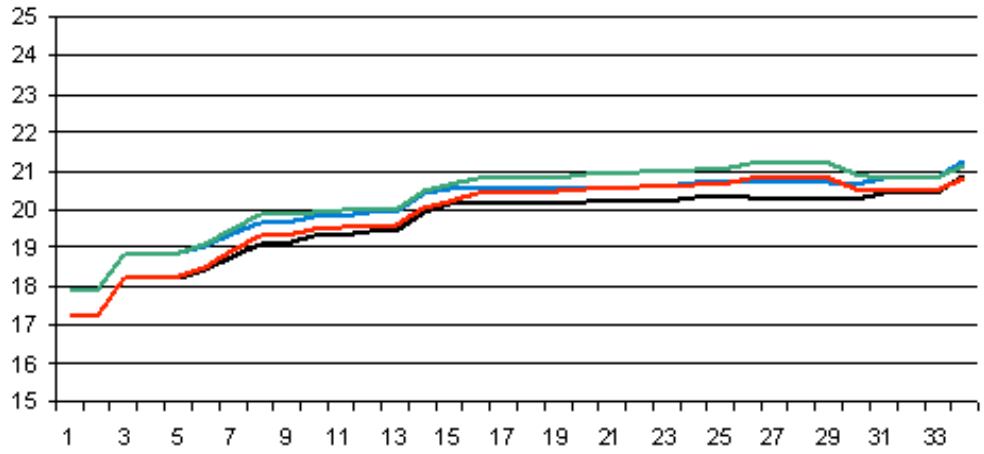


FIG. 4.22: Actual lowest model level mixing ratio values during the assimilation process from the ensemble mean using the original EnKF scheme (black), an individual ensemble member using the original scheme (blue), the ensemble mean using the new EnKF scheme (red), and the individual ensemble member using the new scheme (green) at 16 UTC on 1 August. All values are for the grid point of focus in southeastern Oklahoma, and stations are ordered sequentially. Mixing ratio values along the *y*-axis are in g kg^{-1}, while numbers along the *x*-axis refer to the Mesonet sites listed in Table 4.2.

By examining changes made to the model low level mixing ratio field as a result of the assimilation of individual Mesonet sensible heat flux estimates, it becomes plain that erroneously dry model conditions in southeastern Oklahoma (Fig. 4.7b) should be blamed on the massive temperature decrease (Fig. 4.5b,c) due to application of the EnKF scheme and not on the EnKF modifications to the moisture field. Addition of the mixing ratio changes at the grid point of focus resulting from the assimilation of all accepted Mesonet estimates (Fig. 4.21) totals between 3.25 and 3.75 g kg^{-1}. Moisture is greater from the twenty-second ensemble member than from the ensemble mean at the grid point in question prior to assimilation, with a larger moisture increase in the ensemble mean narrowing the gap between the two by the end of the assimilation cycle (Fig. 4.22). There is a difference of less than 0.1 g kg^{-1} between the moisture changes resulting from the original and the new assimilation scheme. The Mesonet flux estimate from ANTL is responsible for nearly 1.0 g kg^{-1} of the total increase in mixing ratio. At ANTL, the second largest "*WK*" value from Fig. 4.20 is paired with an innovation of approximately 130 W m^{-2}. Similar to temperature, five Mesonet estimates (ANTL, CALV, CENT, HUGO, BROK) are responsible for over 70% of the total change to the grid point mixing ratio.

Chapter 5

Conclusions

Assimilation of sensible heat flux estimates calculated from observed data into coupled models of the land-atmosphere system using an ensemble Kalman filter presents some unique problems. As evidenced by the example from 16 UTC on 1 August 2004, very large differences between sensible heat flux estimates from observations and model values interpolated to the observation location can occur. Even if these differences are legitimate, model analyses of low level temperature and moisture produced by an EnKF scheme may still represent a departure from rather than a correction toward observations. High covariance values between model sensible heat flux at the observation location and the model variable being updated at a grid point amplify the error due to extremely large innovations. A judicious choice of the observation error variance (σ_0^2 from eq. (15) and eq. (16) in Chapter 3) could at least partially address these issues, but consistency ratio values from the EnKF scheme indicate that the optimal value of σ_0^2 varies substantially and that this choice is difficult.

Multiple factors contribute to the existence of large sensible heat flux innovations. First of all, the iterative profile method used to calculate sensible heat flux estimates yields acceptable values when compared with eddy correlation estimates, but the

iterative method does tend to slightly overestimate fluxes. However, a substantial part of the magnitude of large innovations likely stems from poor model forecasts of sensible heat flux. The sensible heat flux is dependent on the gradient in low level wind speed, vegetation types and amounts, as well as soil types and moisture, all of which vary on small scales. Therefore, the sensible heat flux itself exhibits spatial variability on a much smaller scale than the 3 km horizontal resolution of the model can adequately simulate. Even if only one or two innovations within the influence radius of a grid point are unusually large, this number may be all it takes to cause the EnKF scheme to make undesirable changes in the model low level temperature and moisture fields (Section 4.3).

Updates to model fields according to EnKF-generated covariances are often superior to those from an isotropic scheme such as 3D-Var as they tend to be more flow-dependent; however, covariances are purely statistical quantities that may or may not have a physical basis. Correlations between sensible heat flux and other low level variables (Fig. 4.8a,b) show some consistent structure at a considerable distance from the grid point in southeastern Oklahoma. Within the vicinity of this grid point, the correlation with low level moisture tends to be positive (Fig. 4.8b) while the correlation with low level temperature tends to be negative (Fig. 4.8a). Combined with large positive innovations, the result is a large decrease in temperature concurrent with a large increase in moisture that is not physically possible. With such large model error at just a few Mesonet sites, reasonably accurate Mesonet estimates of sensible heat flux can combine with large covariances (and, therefore, large Kalman gain values) to make unphysical changes to the model background fields.

Choosing what number to use for the error variance (σ_o^2) of Mesonet sensible heat flux estimates is not straightforward. One way to assess the appropriateness of the pre-determined value for σ_o^2 uses the fact that the variance in innovation equals the sum of the variance in observation error and the variance in model error

$$\sigma_d^2 = \sigma_o^2 + \sigma_f^2 \tag{20}$$

where *d* denotes the innovation, *o* denotes the observation, and *f* denotes the model forecast. If assumptions regarding the observation and forecast errors are correct, the ratio of the right and left sides of eq. (20) should approximately equal unity. Averaging over all observations updating the model forecast fields (Dowell et al. 2004) at 16 UTC on 1 August, this ratio equals roughly 0.08 with both the original and the new update schemes. The ratio exhibits a diurnal trend, equaling near 1.0 close to sunrise and sunset but bottoming out at around 0.04 in the early afternoon (not shown). A very low ratio indicates that the observation error variance selected is much too low, the ensemble spread is much too low to be a good estimate of the model error, or both.

A number of tunable elements in this EnKF scheme could be modified in an effort to prevent unphysical and extreme updates to model low level temperature and moisture fields. The inverse exponential weight from eq. (13) and eq. (14) is a function of both the distance from the observation location to the grid point being updated and a pre-selected localization radius. The influence of distant high correlations between sensible heat flux and other low level variables having

questionable physical basis (Fig. 4.8a,b) could be tempered by decreasing the localization radius. The adjustment factor (weight function) from this EnKF scheme produces nearly bi-modal weights at 16 UTC 1 August (Fig. 4.16). Combining the choice of an alternate, more smoothly decreasing weight function curve with a reduction in the localization radius could limit the influence of observations relatively near a grid point while also limiting the influence of distant observations.

A more robust method for the quality control of flux observations as they enter the filter may also decrease the number of very large innovations and subsequent large changes to model fields. In this study, a simple method accepts only those Mesonet flux estimates that simultaneously equal greater than half and less than double the background model interpolated flux. This method may reject too many estimates when the magnitude of the flux is small, and too few when the magnitude is great. The high variability of sensible heat flux on a small spatial scale makes the use of a "buddy check" system somewhat problematic. For sites where eddy correlation estimates of sensible heat flux exist, a quality control method based on comparison between the eddy correlation and profile estimates could be devised. Unfortunately, in 2004 this included less than 10 sites.

Consistency ratios of less than 0.1 indicate that the prescribed observation error variance and the model forecast error variance estimated by the ensemble spread collectively are much too small. Just as the consistency ratio follows a diurnal curve, the observation error could be prescribed to follow a diurnal curve as well. In this case, the observation error would be relatively low around sunrise and sunset, and relatively high toward mid-day or early afternoon. Comparisons between eddy

correlation and profile sensible heat flux estimates from the Oklahoma Mesonet in 2004 do in fact show a maximum difference during the portion of the day when the sun angle is highest (not shown). Ensemble spread shows an increasing trend from the first day of this study into the second day, with the rms error in low level temperature and moisture analyses resulting from the EnKF scheme decreasing on 2 August when compared with 1 August (Fig. 4.1a,b). The two days were meteorologically very similar to one another across the model domain. Allowing the ensemble and the EnKF assimilation scheme to run for some time to allow the ensemble spread and therefore the estimated model error to increase would likely improve analyses, and would be preferable prior to initiating forecasts from these analyses.

Finally, increasing the number of ensemble members beyond the total of fifty-one used in this study could be beneficial. More ensemble members may produce a larger spread and, consequently, a larger estimate of model error. Additionally, a larger ensemble may be able to provide more accurate estimates of the magnitudes of relatively small covariances. With only fifty-one ensemble members, the uncertainty especially in small covariance values is great, increasing the chances of ending up with a random high value that produces an unusually large innovation.

In spite of the numerous complicating factors associated with using an EnKF scheme to assimilate Mesonet estimates of sensible heat flux, the idea may still become a useful tool in the future. It has been shown that the EnKF scheme is capable of making physically meaningful and observationally correct changes to model forecast fields. Observing systems such as the Oklahoma Mesonet and the

nearby West Texas Mesonet (Schroeder et al. 2005), both measuring wind and temperature at two levels, are multiplying and expanding yielding a corresponding increase in available sensible heat flux estimates. Finding workable solutions to the problems discussed here may permit improvements to analyses resulting from the use of the EnKF scheme to assimilate sensible heat flux. Once the negative effects of the EnKF scheme are reduced and it consistently produces useful analyses, then producing forecasts from these analyses will become a viable option.

REFERENCES

Abramopoulos, F., C. Rosenzweig, and B. Choudhury, 1988: Improved ground hydrology calculations for global climate models (GCMs): Soil water movement and evapotranspiration. *J. Climate,* **1,** 921-941.

Alapaty, K., N.L. Seaman, D.S. Niyogi, and A.F. Hanna, 2001: Assimilating surface data to improve the accuracy of atmospheric boundary layer simulations. *J. Appl. Meteor.,* **40,** 2068-2082.

Anderson, J.L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518-1530.

——, 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884-2903.

Andrè, J.-C., J.-P. Goutorbe, and A. Perrier, 1986: HAPEX-MOBILHY: A hydrologic atmospheric experiment for the study of water budget and evaporation flux at the climatic scale. *Bull. Amer. Meteor. Soc.,* **67,** 138-144.

Argentini, S., P.J. Wetzel, and V.M. Karyampudi, 1992: Testing a detailed biophysical parameterization for land-air exchange in a high-resolution boundary-layer model. *J. Appl. Meteor.,* **31,** 142-156.

Avissar, R., and R.A. Pielke, 1989: A parameterization of heterogeneous land surfaces for atmospheric models and its impact on regional meteorology. *Mon. Wea. Rev.,* **117,** 2113-2136.

Barkmeijer, J., M. van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Roy. Meteor. Soc.,* **124,** 1695-1713.

Beljaars, A.C.M., and F.C. Bosveld, 1997: Cabauw data for the validation of land surface parameterization schemes. *J. Climate,* **10,** 1172-1193.

——, P. Viterbo, M.J. Miller, and A.K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil moisture anomalies. *Mon. Wea. Rev.,* **124,** 362-383.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.,* **108(D24),** 2 1-2 10.

Benjamin, S.G., 1986: Some effects of surface heating and topography on the regional severe storm environment. Part II: Two-dimensional idealized experiments. *Mon. Wea. Rev.,* **114,** 330-343.

——, and T.N. Carlson, 1986: Some effects of surface heating and topography on the regional severe storm environment. Part I: Three-dimensional simulations. *Mon. Wea. Rev.,* **114,** 307-329.

Bennett, A.F., B.S. Chua, and L.M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys,* **60,** 165-178.

——, ——, and ——, 1997: Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation. *Meteor. Atmos. Phys.,* **62,** 129-140.

Berkowicz, R., and L.P. Prahm, 1982: Sensible heat flux estimated from routine meteorological data by the resistance method. *J. Appl. Meteor.,* **21,** 1845-1864.

Betts, A.K., F. Chen, K.E. Mitchell, and Z.I. Janjić, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data. *Mon. Wea. Rev.,* **125,** 2896-2916.

Bishop, C.H., B.J. Etherton, and S.J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420-436.

Blanchet, I., C. Frankignoul, and M.A. Cane, 1997: A comparison of adaptive Kalman filters for a tropical Pacific ocean model. *Mon. Wea. Rev.,* **125,** 40-58.

Bonan, G.B., D. Pollard, and S.L. Thompson, 1993: Influence of subgrid-scale heterogeneity in leaf area index, stomatal resistance, and soil moisture on grid-scale land-atmosphere interactions. *J. Climate,* **6,** 1882-1897.

Bougeault, P., J. Noilhan, P. Lacarrère, and P. Mascart, 1991a: An experiment with an advanced surface parameterization in a mesobeta-scale model. Part I: Implementation. *Mon. Wea. Rev.,* **119,** 2358-2373.

——, B. Bret, P. Lacarrère, and J. Noilhan, 1991b: An experiment with an advanced surface parameterization in a mesobeta-scale model. Part II: The 16 June 1986 simulation. *Mon. Wea. Rev.,* **119,** 2374-2392.

Bouttier, F., J.-F. Mahfouf, and J. Noilhan, 1993a: Sequential assimilation of soil moisture from atmospheric low-level parameters. Part I: Sensitivity and calibration studies. *J. Appl. Meteor.,* **32,** 1335-1351.

——, ——, and ——, 1993b: Sequential assimilation of soil moisture from atmospheric low-level parameters. Part II: Implementation in a mesoscale model. *J. Appl. Meteor.,* **32,** 1352-1364.

Brock, F.V., K.C. Crawford, R.L. Elliott, G.W. Cuperus, S.J. Stadler, H.L. Johnson, and M.D. Eilts, 1995: The Oklahoma Mesonet: A technical overview. *J. Atmos. Oceanic Technol.,* **12,** 5-19.

Brotzge, J.A., 2004: A two-year comparison of the surface water and energy budgets between two OASIS sites and NCEP-NCAR reanalysis data. *J. Hydrometeor.,* **5,** 311-326.

——, and D. Weber, 2002: Land-surface scheme validation using the Oklahoma Atmospheric Surface-layer Instrumentation System (OASIS) and Oklahoma Mesonet data: Preliminary results. *Meteor. Atmos. Phys.,* **80,** 189-206.

——, and K.C. Crawford, 2000: Estimating sensible heat flux from the Oklahoma Mesonet. *J. Appl. Meteor.,* **39,** 102-116.

——, and ——, 2003: Examination of the surface energy budget: A comparison of eddy correlation and Bowen ratio measurement systems. *J. Hydrometeor.,* **4,** 160-178.

——, and S.J. Richardson, 2003: Spatial and temporal correlation among Oklahoma Mesonet and OASIS surface-layer measurements. *J. Appl. Meteor.,* **42,** 5-19.

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* **125,** 99-119.

——, and T.N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.,* **52,** 1434-1456.

Businger, J.A., W.F. Dabberdt, A.C. Delany, T.W. Horst, C.L. Martin, S.P. Oncley, and S.R. Semmer, 1990: The NCAR atmosphere-surface turbulent exchange research (ASTER) facility. *Bull. Amer. Meteor. Soc.,* **71,** 1006-1011.

Calvet, J.-C., J. Noilhan, and P. Bessemoulin, 1998: Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements. *J. Appl. Meteor.,* **37,** 371-386.

Carlson, T.N., J.K. Dodd, S.G. Benjamin, and J.N. Cooper, 1981: Satellite estimation of the surface energy balance, moisture availability and thermal inertia. *J. Appl. Meteor.,* **20,** 67-87.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.,* **133,** 3081-3094.

Chang, J.-T., and P.J. Wetzel, 1991: Effects of spatial variations of soil moisture and

vegetation on the evolution of a prestorm environment: A numerical case study. *Mon. Wea. Rev.,* **119,** 1368-1390.

Charney, J.G., 1975: Dynamics of deserts and drought in the Sahel. *Quart. J. Roy. Meteor. Soc.,* **101,** 193-202.

Chen, F., and coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.,* **101,** 7251-7268.

——, and J. Dudhia, 2001a: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129,** 569-585.

——, and ——, 2001b: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part II: Preliminary model validation. *Mon. Wea. Rev.,* **129,** 587-604.

Chen, T.H., and coauthors, 1997: Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes. *J. Climate,* **10,** 1194-1215.

Clarke, R.H., A.J. Dyer, R.R. Brook, D.G. Reid, and A.J. Troup, 1971: The Wangara experiment: Boundary layer data. Tech. Paper No. 19, CSIRO, Division of Meteorological Physics, Aspendale, Australia, 362 pp.

Crawford, T.M., D.J. Stensrud, F. Mora, J.W. Merchant, and P.J. Wetzel, 2001: Value of incorporating satellite-derived land cover data in MM5/PLACE for simulating surface temperatures. *J. Hydrometeor.,* **2,** 453-468.

Crook, N.A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.,* **124,** 1767-1785.

Crow, W.T., 2003: Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble Kalman filter and surface brightness temperature observations. *J. Hydrometeor.,* **4,** 960-973.

——, and E.F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.,* **26,** 137-149.

de Bruin, H.A.R., and A.A.M. Holtslag, 1982: A simple parameterization of the surface fluxes of sensible and latent heat during daytime compared with the Penman-Monteith concept. *J. Appl. Meteor.,* **21,** 1610-1621.

Deardorff, J.W., 1977: A parameterization of ground-surface moisture content for use

in atmospheric prediction models. *J. Appl. Meteor.,* **16,** 1182-1185.

Deardorff, J., 1978: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *J. Geophys. Res.,* **83,** 1889-1903.

Diak, G., S. Heikkinen, and J. Bates, 1986: The influence of variations in surface treatment on 24-hour forecasts with a limited area model, including a comparison of modeled and satellite-measured surface temperatures. *Mon. Wea. Rev.,* **114,** 215-232.

Dickinson, R.E., A. Henderson-Sellers, P.J. Kennedy, and M.F. Wilson, 1986: Biosphere Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. NCAR/TN-275 + STR, National Center for Atmospheric Research, Boulder, CO, 69 pp.

Dirmeyer, P.A., 1994: Vegetation stress as a feedback mechanism in midlatitude drought. *J. Climate,* **7,** 1463-1483.

Doran, J.C., and coauthors, 1992: The Boardman Regional Flux Experiment. *Bull. Amer. Meteor. Soc.,* **73,** 1785-1795.

——, W.J. Shaw, and J.M. Hubbe, 1995: Boundary layer characteristics over areas of inhomogeneous surface fluxes. *J. Appl. Meteor.,* **34,** 559-571.

Dowell, D.C., F. Zhang, L.J. Wicker, C. Snyder, and N.A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.,* **132,** 1982-2005.

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP short-range ensemble forecasting (SREF) system: Inclusion of WRF members. Preprints, *WMO Expert Team Meeting on Ensemble Prediction Systems,* Exeter, UK, 6-10 Feb 2006.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077-3107.

——, 1993: A nonhydrostatic version of the Penn State-NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.,* **121,** 1493-1513.

——, D. Gill, K. Manning, W. Wang, and C. Bruyere, 2005: *PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and Users' Guide (MM5 Modeling System Version 3).*

Dyer, A.J., 1974: A review of flux-profile relationships. *Bound.-Layer Meteor.,* **7,**

363-372.

——, and B.B. Hicks, 1970: Flux-gradient relationships in the constant flux layer. *Quart. J. Roy. Meteor. Soc.,* **98,** 206-212.

Entekhabi, D., and P.S. Eagleson, 1989: Land surface hydrology parameterization for atmospheric general circulation models including subgrid scale spatial variability. *J. Climate,* **2,** 816-831.

Ek, M.B., K.E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J.D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108(D22),** 8851, doi:10.1029/2002JD003296.

Epstein, E.S., 1969: Stochastic dynamic prediction. *Tellus,* **6,** 739-759.

——, and E.J. Pitcher, 1972: Stochastic analysis of meteorological fields. *J. Atmos. Sci.,* **29,** 244-257.

Errico, R., and D. Baumhefner, 1987: Predictability experiments using a high-resolution limited-area model. *Mon. Wea. Rev.,* **115,** 488-504.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99(C5),** 10 143-10 162.

——, 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343-367.

——, and P.J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.,* **124,** 85-96.

——, and ——, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.,* **128,** 1852-1867.

Fleming, R.J., 1971: On stochastic dynamic prediction. Part I: The energetics of uncertainty and the question of closure. *Mon. Wea. Rev.,* **99,** 851-872.

Friedl, M.A., 1996: Relationships among remotely sensed data, surface energy balance, and area-averaged fluxes over partially vegetated land surfaces. *J. Appl. Meteor.,* **35,** 2091-2103.

Fujita, T., D.J. Stensrud, and D.C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.,* **135,** 1846-1868.

Gao, W., R.L. Coulter, B.M. Lesht, J. Qiu, and M.L. Wesely, 1998: Estimating clear-sky regional surface fluxes in the southern Great Plains Atmospheric Radiation Measurement site with ground measurements and satellite observations. *J. Appl. Meteor.,* **37,** 5-22.

Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.,* **128,** 997-1015.

Gleeson, T.A., 1970: Statistical-dynamical predictions. *J. Appl. Meteor.,* **9,** 333-344.

Godfrey, C.M., 2006: The influence of improved land surface and soil data on mesoscale model predictions. Ph.D. dissertation, University of Oklahoma, 128 pp.

Guyer, J.L., and D.R. Bright, 2008: Utility of short-range ensemble forecast (SREF) guidance for forecasting the development of severe convection. Preprints, *24th Conf. on Severe Local Storms,* Savannah, Georgia, Amer. Meteor. Soc., CDROM13A.1.

Hack, J.J., B.A. Boville, B.P. Briegleb, J.T. Kiehl, P.J. Rasch, and D.L. Williamson, 1993: Description of the NCAR Community Climate Model (CCM2). NCAR Tech. Note NCAR/TN-382-STR, 108 pp. [Available from National Center for Atmospheric Research Boulder, CO 80307.]

Hacker, J.P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.,* **133,** 3260-3275.

Hamill, T.M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905-2919.

——, ——, and R.E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.,* **128,** 1835-1851.

——, J.S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776-2790.

Henderson-Sellers, A., Z.-L. Yang, and R.E. Dickinson, 1993: The Project for Intercomparison of Land-surface Parameterization Schemes. *Bull. Amer. Meteor. Soc.,* **74,** 1335-1349.

Hess, G.D., B.B. Hicks, and T. Yamada, 1981: The impact of the Wangara experiment. *Bound.-Layer Meteor.,* **20,** 135-174.

Holt, T.R., D. Niyogi, F. Chen, K. Manning, M.A. LeMone, and A. Qureshi, 2006: Effect of land-atmosphere interactions on the IHOP 24-25 May 2002 convection case. *Mon. Wea. Rev.,* **134,** 113-133.

Holtslag, A.A.M., and A.P. van Ulden, 1983: A simple scheme for daytime estimates of the surface fluxes from routine weather data. *J. Clim. Appl. Meteor.,* **22,** 517-529.

——, E.I.F. de Bruijn, and H.-L. Pan, 1990: A high-resolution air mass transformation model for short-range weather forecasting. *Mon. Wea. Rev.,* **118,** 1561-1575.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322-2339.

Horst, T.W., and J.C. Weil, 1994: How far is far enough? The fetch requirements for micrometeorological measurement of surface fluxes. *J. Atmos. Oceanic Technol.,* **11,** 1018-1025.

Houtekamer, P.L., and H.L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796-811.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123-137.

——, ——, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.,* **133,** 604-620.

——, and J. Derome, 1994: Prediction experiments with two-member ensembles. *Mon. Wea. Rev.,* **122,** 2179-2191.

——, and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.,* **123,** 2181-2196.

——, and L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.,* **125,** 2416-2426.

——, ——, J. Derome, H. Ritchie, and H.L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225-1242.

Ide, K., P. Courtier, M. Ghil, and A.C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteor. Soc. Japan,* **75(1B),** 181-189.

Jones, A.S., I.C. Guch, and T.H. Vonder Haar, 1998a: Data assimilation of satellite-derived heating rates as proxy surface wetness data into a regional atmospheric

mesoscale model. Part I: Methodology. *Mon. Wea. Rev.,* **126,** 634-645.

——, ——, and ——, 1998b: Data assimilation of satellite-derived heating rates as proxy surface wetness data into a regional atmospheric mesoscale model. Part II: A case study. *Mon. Wea. Rev.,* **126,** 646-667.

Kain, J.S., and J.M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 24, Amer. Meteor. Soc., 165-170.

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.,* **82D,** 35-45.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 341 pp.

Katz, R.W., A.H. Murphy, and R.L. Winkler, 1982: Assessing the value of frost forecasts to orchardists: A dynamic decision-making approach. *J. Appl. Meteor.,* **21,** 518-531.

Keppenne, C.L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 1971-1981.

——, and M.M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal general circulation model. *Mon. Wea. Rev.,* **130,** 2951-2965.

Kong, F., M. Xue, K.W. Thomas, K.K. Droegemeier, Y. Wang, K Brewster, J. Gao, J. Kain, S.J. Weiss, D. Bright, M.C. Coniglio, and J. Du, 2008: Real-time storm-scale forecast experiment - Analysis of 2008 Spring experiment data. Preprints, *24th Conf. on Severe Local Storms,* Savannah, Georgia, Amer. Meteor. Soc., CDROM 12.3.

——, ——,——, Y. Wang, K. Brewster, J. Gao, K.K. Droegemeier, J. Kain, S.J. Weiss, D. Bright, M.C. Coniglio, and J. Du, 2009: A real-time storm-scale ensemble forecast system: 2009 Spring experiment. Preprints, *23rd Conf. on Wea. Anal. Forecasting/19th Conf. on Num. Wea. Pred.,* Omaha, Nebraska, Amer. Meteor. Soc., CDROM 16A.3.

Koster, R.D., and M.J. Suarez, 1992: A comparative analysis of two land surface heterogeneity representations. *J. Climate,* **5,** 1379-1390.

——, and ——, 1996: The influence of land surface moisture retention on precipitation statistics. *J. Climate,* **9,** 2551-2567.

Kurkowski, N.P., D.J. Stensrud, and M.E. Baldwin, 2003: Assessment of

implementing satellite-derived land cover data in the Eta model. *Wea. Forecasting,* **18,** 404-416.

Kustas, W.P., J.H. Prueger, K.S. Humes, and P.J. Starks, 1999: Estimation of surface heat fluxes at field scale using surface layer versus mixed-layer atmospheric variables with radiometric temperature observations. *J. Appl. Meteor.,* **38,** 224-238.

Lakhtakia, M.N., and T.T. Warner, 1994: A comparison of simple and complex treatments of surface hydrology and thermodynamics suitable for mesoscale atmospheric models. *Mon. Wea. Rev.,* **122,** 880-896.

Lanicci, J.M., T.N. Carlson, and T.T. Warner, 1987: Sensitivity of the Great Plains severe-storm environment to soil-moisture distribution. *Mon. Wea. Rev.,* **115,** 2660-2673.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409-418.

LeMone, M.A., and coauthors, 2000: Land-atmosphere interaction research, early results, and opportunities in the Walnut River watershed in southeast Kansas: CASES and ABLE. *Bull. Amer. Meteor. Soc.,* **81,** 757-779.

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177-1194.

——, 2003: The potential of the ensemble Kalman filter for NWP-a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183-3203.

Lorenz, E.N., 1960: Maximum simplification of the dynamic equations. *Tellus,* **12,** 243-254.

——, 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130-141.

Mahfouf, J.-F., 1991: Analysis of soil moisture from near-surface parameters: A feasibility study. *J. Appl. Meteor.,* **30,** 1534-1547.

——, E. Richard, and P. Mascart, 1987: The influence of soil and vegetation on the development of mesoscale circulations. *J. Clim. Appl. Meteor.,* **26,** 1483-1495.

Mahrt, L., 1987: Grid-averaged surface fluxes. *Mon. Wea. Rev.,* **115,** 1550-1560.

Manabe, S., 1969: Climate and the ocean circulation. Part I: The atmospheric circulation and the hydrology of the earth's surface. *Mon. Wea. Rev.,* **97,** 739-774.

Margulis, S.A., and D. Entekhabi, 2003: Variational assimilation of radiometric

surface temperature and reference-level micrometeorology into a model of the atmospheric boundary layer and land surface. *Mon. Wea. Rev.,* **131,** 1272-1288.

Marshall, C.H., K.C. Crawford, K.E. Mitchell, and D.J. Stensrud, 2003: The impact of the land surface physics in the operational NCEP Eta model on simulating the diurnal cycle: Evaluation and testing using Oklahoma Mesonet data. *Wea. Forecasting,* **18,** 748-768.

McCumber, M.C., and R.A. Pielke, 1981: Simulation of the effects of surface fluxes of heat and moisture in a mesoscale numerical model. Part I: Soil layer. *J. Geophys. Res.,* **86(C10),** 9929-9938.

McNider, R.T., A.J. Song, D.M. Casey, P.J. Wetzel, W.L. Crosson, and R.M. Rabin, 1994: Toward a dynamic-thermodynamic assimilation of satellite surface temperature in numerical atmospheric models. *Mon. Wea. Rev.,* **122,** 2784-2803.

McPherson, R.A., C. Fiebrich, K.C. Crawford, R.L. Elliott, J.R. Kilby, D.L. Grimsley, J.E. Martinez, J.B. Basara, B.G. Illston, D.A. Morris, K.A. Kloesel, S.J. Stadler, A.D. Melvin, A.J. Sutherland, and H. Shrivastava, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.,* **24,** 301-321.

Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037-1056.

Mitchell, H.L., and P.L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416-433.

Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73-119.

Monteith, J.L., 1965: Evaporation and environment. *Symp. Soc. Exp. Biol.,* **XIX,** 205-234.

Murphy, J.M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.,* **114,** 463-493.

Nappo, C.J., 1975: Parameterization of surface moisture and evaporation rate in a planetary boundary layer model. *J. Appl. Meteor.,* **14,** 289-296.

Nickerson, E.C., and V.E. Smiley, 1975: Surface layer and energy budget parameterizations for mesoscale models. *J. Appl. Meteor.,* **14,** 297-300.

Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.,* **117,** 536-549.

Oklahoma Climatological Survey, 2005: *Estimates of soil moisture from the Oklahoma Mesonet, version 3.0*, 20 pp.

Ookouchi, Y., M. Segal, R.C. Kessler, and R.A. Pielke, 1984: Evaluation of soil moisture effects on the generation and modification of mesoscale circulations. *Mon. Wea. Rev.,* **112,** 2281-2292.

Paegle, J., K.C. Mo, and J. Nogués-Paegle, 1996: Dependence of simulated precipitation on surface evaporation during the 1993 United States summer floods. *Mon. Wea. Rev.,*
124, 345-361.

Palmer, T.N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.,* **55,** 633-653.

Pan, H.-L., 1990: A simple parameterization scheme of evapotranspiration over land for the NMC medium-range forecast model. *Mon. Wea. Rev.,* **118,** 2500-2512.

——, and L. Mahrt, 1987: Interaction between soil hydrology and boundary layer development. *Bound.-Layer Meteor.,* **38,** 185-202.

Parrish, D.F., and J.C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747-1763.

Penman, H.L., 1948: Natural evaporation from open water, bare soil, and grass. *Proc. Roy. Soc. London,* **193,** 120-145.

Peters-Lidard, C.D., E. Blackburn, X. Liang, and E.F. Wood, 1998: The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *J. Atmos. Sci.,* **55,** 1209-1224.

Philip, J.R., 1957: Evaporation, and moisture and heat fields in the soil. *J. Meteor.,* **14,** 354-366.

Pielke, R.A., 1974: A three-dimensional numerical model of the sea breezes over South Florida. *Mon. Wea. Rev.,* **102,** 115-139.

——, G.A. Dalu, J.S. Snook, T.J. Lee, and T.G.F. Kittel, 1991: Nonlinear influence of mesoscale land use on weather and climate. *J. Climate,* **4,** 1053-1069.

Pinty, J.-P., P. Mascart, E. Richard, and R. Rosset, 1989: An investigation of mesoscale flows induced by vegetation inhomogeneities using an evapotranspiration model calibrated against HAPEX-MOBILHY data. *J. Appl. Meteor.,* **28,** 976-992.

Pitcher, E.J., 1977: Application of stochastic dynamic prediction to real data. *J. Atmos. Sci.,* **34,** 3-21.

Pleim, J.E., and A. Xiu, 1995: Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. *J. Appl. Meteor.,* **34,** 16-32.

——, and ——, 2003: Development of a land surface model. Part II: Data assimilation. *J. Appl. Meteor.,* **42,** 1811-1822.

Price, J.C., 1982: On the use of satellite data to infer surface fluxes at meteorological scales. *J. Appl. Meteor.,* **21,** 1111-1122.

Priestley, C.H.B., and R.J. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.,* **100,** 81-91.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.,* **124,** 1861-1887.

Rabin, R.M., S. Stadler, P.J. Wetzel, D.J. Stensrud, and M. Gregory, 1990: Observed effects of landscape variability on convective clouds. *Bull. Amer. Meteor. Soc.,* **71,** 272-280.

Reichle, R.H., D.B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103-114.

——, J.P. Walker, R.D. Koster, and P.R. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeor.,* **3,** 728-740.

Rodell, M., and coauthors, 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.,* **85,** 381-394.

Ruggiero, F.H., K.D. Sashegyi, R.V. Madala, and S. Raman, 1996: The use of surface observations in four-dimensional data assimilation using a mesoscale model. *Mon. Wea. Rev.,* **124,** 1018-1033.

Sato, N., P.J. Sellers, D.A. Randall, E.K. Schneider, J. Shukla, J.L. Kinter III, Y.-T. Hou, and E. Albertazzi, 1989: Effects of implementing the Simple Biosphere Model in a general circulation model. *J. Atmos. Sci.,* **46,** 2757-2782.

Schaake, J.C., V.I. Koren, Q.Y. Duan, K. Mitchell, and F. Chen, 1996: A simple water balance model (SWB) for estimating runoff at different spatial and temporal scales. *J. Geophys. Res.,* **101,** 7461-7475.

Schmugge, T., 1978: Remote sensing of surface soil moisture. *J. Appl. Meteor.,* **17,**

1549-1557.

Schroeder, J.L., W.S. Burgett, K.B. Haynie, I. Sonmez, G.D. Skwira, A.L. Doggett, and J.W. Lipe, 2005: The West Texas Mesonet: a technical overview. *J. Atmos. Oceanic Technol.,* **22,** 211-222.

Segal, M., R. Avissar, M.C. McCumber, and R.A. Pielke, 1988: Evaluation of vegetation effects on the generation and modification of mesoscale circulations. *J. Atmos. Sci.,* **45,** 2268-2292.

——, and R.W. Arritt, 1992: Nonclassical mesoscale circulations caused by surface sensible heat-flux gradients. *Bull. Amer. Meteor. Soc.,* **73,** 1593-1604.

——, W.E. Schreiber, G. Kallos, J.R. Garratt, A. Rodi, J. Weaver, and R.A. Pielke, 1989: The impact of crop areas in northeast Colorado on midsummer mesoscale thermal circulations. *Mon. Wea. Rev.,* **117,** 809-825.

Segele, Z.T., D.J. Stensrud, I.C. Ratcliffe, and G.M. Henebry, 2005: Influence of a hailstreak on boundary layer evolution. *Mon. Wea. Rev.,* **133,** 942-960.

Sellers, P.J., D.A. Randall, G.J. Collatz, J.A. Berry, C.B. Field, D.A. Dazlich, C. Zhang, G.D. Colello, and L. Bounoua, 1996: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. *J. Climate,* **9,** 676-705.

——, F.G. Hall, G. Asrar, D.E. Strebel, and R.E. Murphy, 1988: The First ISLSCP Field Experiment (FIFE). *Bull. Amer. Meteor. Soc.,* **69,** 22-27.

——, and J.L. Dorman, 1987: Testing the simple biosphere model (SiB) using point micrometeorological and biophysical data. *J. Clim. Appl. Meteor.,* **26,** 622-651.

——, Y. Mintz, Y.C. Sud, and A. Dalcher, 1986: A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.,* **43,** 505-531.

Smith, C.B., M.N. Lakhtakia, W.J. Capehart, and T.N. Carlson, 1994: Initialization of soil-water content in regional-scale atmospheric prediction models. *Bull. Amer. Meteor. Soc.,* **75,** 585-593.

Smith, E.A., W.L. Crosson, and B.D. Tanner, 1992: Estimation of surface heat and moisture fluxes over a prairie grassland. 1. In situ energy budget measurements incorporating a cooled mirror dew point hygrometer. *J. Geophys. Res.,* **97(D17),** 18.557-18.582.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **131,** 1663-1677.

Stensrud, D.J., H.E. Brooks, J. Du, M.S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433-446.

——, J.-W. Bao, and T.T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077-2107.

Stewart, T.R., R.W. Katz, and A.H. Murphy, 1984: Value of weather information: A descriptive study of the fruit-frost problem. *Bull. Amer. Meteor. Soc.,* **65,** 126-137.

Tippett, M.K., J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485-1490.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133,** 1789-1807.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297-3319.

Trier, S.B., F. Chen, and K.W. Manning, 2004: A study of convection initiation in a mesoscale model using high-resolution land surface initial conditions. *Mon. Wea. Rev.,* **132,** 2954-2976.

Tracton, M.S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: practical aspects. *Wea. Forecasting,* **8,** 379-398.

van den Hurk, B.J.J.M., W.G.M. Bastiaanssen, H. Pelgrum, and E. van Meijgaard, 1997: A new methodology for assimilation of soil moisture fields in weather prediction models using Meteosat and NOAA data. *J. Appl. Meteor.,* **36,** 1271-1283.

Walker, J.P., and P.R. Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res.,* **106(D11),** 761-774.

Wang, X., and C.H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60,** 1140-1158.

Weiss, S.J., J.S. Kain, D.R. Bright, J.J. Levit, G.W. Carbin, M.E. Pyle, Z.I. Janjic, B.S. Ferrier, J. Du, M.L. Weisman, and M. Xue, 2007: The NOAA Hazardous Weather Testbed: Collaborative testing of ensemble and convection-allowing WRF models and subsequent transfer to operations at the Storm Prediction Center.

Preprints, *22nd Conf. on Wea. Anal. Forecasting/18th Conf. on Num. Wea. Pred.,* Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 6B.4.

Wetzel, P.J., D. Atlas, and R.H. Woodward, 1984: Determining soil moisture from geosynchronous satellite infrared data: A feasibility study. *J. Clim. Appl. Meteor.,* **23,** 375-391.

——, and J.-T. Chang, 1988: Evapotranspiration from non-uniform surfaces: A first approach for short-term numerical weather prediction. *Mon. Wea. Rev.,* **116,** 600-621.

——, and R.H. Woodward, 1987: Soil moisture estimation using GOES-VISSR infrared data: A case study with a simple statistical method. *J. Clim. Appl. Meteor.,* **26,** 107-117.

Whitaker, J.S., and A.F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.,* **126,** 3292-3302.

——, and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913-1924.

Wilson, M.F., A. Henderson-Sellers, R.E. Dickinson, and P.J. Kennedy, 1987: Sensitivity of the biosphere-atmosphere transfer scheme (BATS) to the inclusion of variable soil characteristics. *J. Clim. Appl. Meteor.,* **26,** 341-362.

Xinmei, H., and T.J. Lyons, 1995: The simulation of surface heat fluxes in a land surface-atmosphere model. *J. Appl. Meteor.,* **34,** 1099-1111.

——, T.J. Lyons, R.C.G. Smith, J.M. Hacker, and P. Schwerdtfeger, 1993: Estimation of surface energy balance from radiant surface temperature and NOAA AVHRR sensor reflectances over agricultural and native vegetation. *J. Appl. Meteor.,* **32,** 1441-1449.

Xiu, A., and J.E. Pleim, 2001: Development of a land surface model. Part I: Application in a mesoscale meteorological model. *J. Appl. Meteor.,* **40,** 192-209.

Xu, Q., and B. Zhou, 2003: Retrieving soil water contents from soil temperature measurements by using linear regression. *Adv. Atmos. Sci.,* **20,** 849-858.

Xue, M., M. Tong, and K.K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. *J. Atmos. Oceanic Technol.,* **23,** 46-66.

Yan, H., and R.A. Anthes, 1988: The effect of variations in surface moisture on mesoscale circulations. *Mon. Wea. Rev.,* **116,** 192-208.

Zhang, D.-L., and R.A. Anthes, 1982: A high-resolution model of the planetary boundary layer-sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.,* **21,** 1594-1609.

——, and W.-Z. Zheng, 2004: Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *J. Appl. Meteor.,* **43,** 157-169.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132,** 1238-1253.

——, Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments. *Mon. Wea. Rev.,* **134,** 722-736.

Zhang, H., and C.S. Frederiksen, 2003: Local and nonlocal impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study. *J. Climate,* **16,** 2117-2137.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.,* **83,** 73-83.