J8.3 AN EVALUATION OF USING REAL-TIME, SATELLITE-DERIVED VEGETATION FRACTION IN THE ETA MODEL

David J. Stensrud and Nicole P. Kurkowski
NOAA/National Severe Storms Laboratory, Norman, OK
and
Michael E. Baldwin
Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK

1.0 INTRODUCTION[1]

Vegetation plays an important role in land-atmosphere interactions by helping to determine the partitioning of the surface sensible and latent heat flux. Ookouchi et al. (1984) show that solenoidal circulations are able to develop between patches of moist and dry soil. Similarly, areas of very dense vegetation next to bare soil surfaces under favorable environmental conditions promote sea breeze-type circulations (Segal et al. 1986). Observations indicate that harvesting winter wheat over Oklahoma alters the surface sensible heat flux, such that afternoon cumulus clouds develop over harvested fields before they form over adjacent areas with an active green canopy (Rabin et al. 1990). Markowski and Stensrud (1998) further show that the harvesting of winter wheat over Oklahoma and Kansas plays a major role on the spatial distribution of monthly mean diurnal cycles of conserved variables in the surface layer.

Schwartz and Karl (1990) find there are statistically and practically significant relationships between the timing of the onset of vegetation and surface daily maximum temperature. Their study demonstrates at least a 3.5°C reduction in surface daily maximum temperature at an agricultural inland area over any two-week period subsequent to first leaf compared to a two-week period prior to first leaf. Stations generally near major bodies of water show a smaller (1.5°C) reduction.

As a result of the important role that vegetation plays in land-surface and land-atmosphere interactions, it needs to be represented adequately in numerical weather prediction models. Thus, it is essential that the vegetation conditions be gathered in an accurate and timely manner. Remote sensing devices such as the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) represent the preferred method for gathering vegetation data.

There are several advantages to utilizing the 1-km AVHRR data over that of the 30-m land remote sensing satellite system (LANDSAT) or 10-m Systeme Probatoire d'Observation de la Terre (SPOT). First, the revisit periods of LANDSAT and SPOT are near two weeks. Since cloud-free images are not always possible when the satellite passes over a given location (Crawford et al. 2001), a month or longer may pass before the land surface can be observed clearly. Second, AVHRR data have nominal cost, whereas high-resolution data from LANDSAT and SPOT are costly and cover only limited regions of the globe episodically (Gutman and Ignatov 1995). Other notable advantages include the availability of spectral information for vegetation studies, global coverage, and the long-term continuous observational period.

It is possible to compute both the vegetation fraction and leaf area index (LAI) from the AVHRR through calculations based upon the Normalized Difference Vegetation Index (NDVI) (see Crawford et al. 2001). However, according to Gutman and Ignatov (1998), it is less difficult to compute the horizontal density, or vegetation fraction, than the LAI from the NDVI. Thus, for example, the LAI is assigned a constant value of 1 in the Eta model and only the

[1] Corresponding author address: David Stensrud, NOAA/NSSL, 1313 Halley Circle, Norman, OK 73069. Email: David.Stensrud@nssl.noaa.gov

vegetation fraction is allowed to vary in time and space. Other land surface models allow both parameters to vary.

Currently, the NCEP Eta model uses a 0.144 degree (approximately 14 km) resolution monthly database for vegetation fraction, based upon a five year climatology (Black et al. 1997) developed at the National Environmental Satellite Data and Information Service (NESDIS). The monthly vegetation fraction values apply to the 15th of every month and these data are interpolated for daily values. These estimations are based upon observations of NDVI, which are obtained from NOAA's polar-orbiting satellites that have the AVHRR.

Vegetation characteristics change from year-to-year, however, and are influenced by circumstances such as fires, irrigation, deforestation, desertification, crop harvesting, drought-affected vegetation, hailstorms, and the early onset of spring vegetation (Crawford et al. 2001). Using near real-time satellite-derived values of vegetation fraction and LAI improves the 2-m temperature simulations of the Penn State- National Center for Atmospheric Research Mesoscale Model version 5 (MM5) when using the Parameterization for Land-Atmosphere-Cloud Exchange (PLACE; Wetzel and Boone 1995) for the land surface (Crawford et al. 2001). Thus, it is hypothesized that an improved initialization of vegetation fraction in the Eta model can lead to smaller surface temperature forecast errors, that likely result in part from the use of climatological land use information (Mitchell et al. 2000). Comparisons of numerical forecasts of the Eta model are made, using both climatological and satellite-derived values of fractional vegetation coverage from the NOAA AVHRR data, in order to examine the potential importance of real-time vegetation fraction information to forecasts of 2-meter temperatures and dewpoint temperatures from 0 to 48 h. The cases selected span the growing season of 2001.

2.0 METHODS

a. Eta Model Description

The workstation version of the Eta model (Black 1994) is used in this study. This model uses a domain that is approximately 20% of the operational Eta domain, although it has the same horizontal grid spacing of 22 km and the same 50 vertical layers that were used operationally in the Eta model during most of 2001. The initial and boundary conditions are obtained directly from a subset of the operational Eta model fields. However, there are two major differences between the operational Eta model and the version run for this study. This workstation version uses the Kain-Fritsch convective parameterization scheme (Kain and Fritsch 1993) and uses fourth-order horizontal diffusion instead of second-order horizontal diffusion. Neither of these changes is expected to influence significantly the results of the present study regarding the importance of vegetation fraction on the model forecasts.

The land-surface model (LSM) used within the Eta model is a multi-layer soil-vegetation-snowpack model. It was originally developed at Oregon State University (Mahrt and Pan 1984; Pan and Mahrt 1987) and was then modified at NCEP for use in the Eta model (Chen et al 1996; Black et al. 1997). This LSM has one canopy layer and eight prognostic variables: soil moisture and temperature in the three soil layers, water stored on the canopy, and snow stored on the ground (Chen et.al.1996). Vegetation fraction influences the surface latent heat flux calculations directly by weighting the relative contributions of bare soil and canopy transpiration. Thus, a larger value of vegetation fraction indicates that the vegetated surface contributes a greater portion of the total latent heat flux than a smaller value of vegetation fraction. This also has a large influence on the model predictions of 2-m dewpoint temperature. In addition, since the vegetation fraction influences the latent heat flux, it also influences the sensible heat flux through the surface energy balance. Thus, vegetation fraction has a potentially large influence on both the 2-m temperature and dewpoint temperature, and thus affects the development of the planetary boundary layer in the model forecasts.

b. Calculation of vegetation fraction

Since 1989, United States Geological Survey's Earth Resources Observing Systems

(EROS) Data Center (EDC) has used the 1-km AVHRR daily observations to produce biweekly maximum NDVI composites over the conterminous United States (Gutman and Ignatov 1995). Since clouds are associated with low values of NDVI, biweekly maximum NDVI composites tend to have minimal cloud contamination. Biweekly composites of NDVI from USGS/EROS for 2001 are available on the web approximately 4 months after the date of interest. These data are downloaded and navigated using a geographic information system. The biweekly period that contains the forecast day of interest is selected for use in order to mimic the data that could be available if daily updates of this information were provided routinely.

The values for vegetation fraction (fveg) are derived from NDVI using the following equation (Chang and Wetzel 1991):

$$fveg = \begin{cases} 1.5(NDVI - 0.1), NDVI \le 0.547 \\ 3.2(NDVI) - 1.08, NDVI > 0.547 \end{cases}$$

where the values of vegetation fraction are constrained to lie between 0 and 1. Since the Eta model has a spatial resolution of 22 km, while the NDVI data resolution is 1 km, an objective analysis is necessary to determine a representative value for the vegetation fraction on the Eta grid. Data points for the 1-km vegetation fraction that lie within a 10 km radius of each Eta grid point are averaged and then applied to that Eta grid point. While more sophisticated analyses are possible, this approach is reasonable for an initial study into the importance of vegetation fraction in the Eta model. Locations outside of the region of the biweekly composite data use the Eta model climatology for the vegetation fraction, which can produce large gradients in vegetation fraction in southern Canada.

c. *Running the Eta model*

Two different runs of the Eta model are performed: one with the climatological value for vegetation fraction (Eta) and the other with the NDVI-derived vegetation fraction (VegEta). The model grid, physical process schemes, and initial and boundary conditions are identical in these two runs except for the vegetation fraction defined in the surface parameter input file. The model provides output every 3 h, beginning at the 1200 UTC initialization time and ending at 48 h.

A total of 12 case studies are examined for the months of April to August of 2001 (Table 1). Two forecasts are completed for each of these months, except for April, which has four forecasts. Each day is chosen because a majority of the U.S. for that day is free of deep convective clouds, thereby minimizing the influence of the convective parameterizations. This case selection maximizes the potential for the effects of vegetation to be seen and diagnosed in the model forecasts, as the influences of many of the other model physical process schemes are minimized.

Case	Days
04.18.01	06.17.01
04.19.01	06.18.01
04.26.01	07.05.01
04.27.01	07.07.01
05.13.01	08.04.01
05.16.01	08.07.01

Table 1. List of all days used in this study.

Once the forecasts and post-processing are complete, the model 2-m temperature and dewpoint temperature are bilinearly interpolated to National Weather Service (NWS) surface observing sites. Since the model domain covers the contiguous 48 states, there are over 1600 NWS surface stations that provide hourly reports and can be used to compare the performance of the two forecasts at each output time. The number of NWS stations actually used in the analyses is somewhat less, varying from approximately 1350 to 1600. While the hourly surface observations are examined for obvious problems, errors may exist. However, any errors in the observational data influence the results of both of the forecasts in the same manner and should not influence the subsequent comparisons. The bias, root mean-square-error (rmse), and Pearson (ordinary) correlation (see Wilks 1995 for documentation on all three of these measures) are computed both for both the operational forecasts (Eta) and the NDVI-derived vegetation fraction forecasts (VegEta), and the results are compared. We now turn to the forecast comparisons.

3.0 RESULTS

a. Description of vegetation fraction differences

Comparisons between the vegetation fraction from the Eta and VegEta model initial conditions (Figs. 1, 2 and 3) indicate that the VegEta vegetation fraction contains more horizontal structure than the values from the Eta. Thus, the real-time satellite data are capturing the temporal and horizontal variations in vegetation growth that likely are masked by the use of a 5-year monthly climatology. Early in the growing season the largest differences in vegetation fraction are over the southeast U.S. (Fig. 1c) with the values from the VegEta being larger than those from the Eta, but these differences decrease as the growing season progresses. By mid-June (Fig. 2c) the largest differences in vegetation fraction are over the plains states, stretching from Oklahoma to North Dakota and in southeastern Canada. The region of largest differences shifts farther northward into the northern plains during July (Fig. 2f) and August (Fig. 3c) and a more localized difference is also seen in western Mexico associated with vegetation green-up during the North American monsoon. Average absolute differences in vegetation fraction calculated across the entire model grid from the Eta and VegEta are slightly more than 0.10 for the 12 days examined.

Since the method used by Gutman and Ignatov (1998) to calculate vegetation fraction for the Eta model produces slightly larger values of vegetation fraction for a given NDVI value than the method used in this study, it is somewhat surprising that the VegEta often has larger vegetation fractions than the Eta. However, vegetation fractions can increase by 0.30 or more over a one-month period, suggesting that one explanation for the larger vegetation fractions in the VegEta is the improved temporal and spatial resolution of the data. Crop moisture estimations during 2001 indicate that much of the eastern U.S. had moist to abnormally moist soil conditions, and surface temperatures during April and May over the eastern U.S. exceeded normal values by 1 to 3°C, with the largest temperature anomalies in April centered over Illinois, suggesting that an early start to the growing season is likely and would explain many of these differences in vegetation fraction. This assessment reinforces the conclusion of Crawford et al. (2001) on the need for real-time vegetation information.

b. Comparisons between Eta and VegEta

Strong relationships exist between the spatial distribution of the vegetation fraction differences between Eta and VegEta and the corresponding differences in surface fluxes, 2-m temperature, and 2-m dewpoint temperature. For example, recall that for mid-May 2001 the VegEta contains broad regions with vegetation fractions larger than the Eta, with differences approaching 40% in the southeast and stretching northward along the east coast into Canada (see Fig. 1f). In addition, areas in the far northwest also contain larger values of vegetation fractions in VegEta. Differences in the instantaneous flux values between the two models show large, coherent regions in which both the VegEta latent and sensible heat fluxes are different from those of the Eta at 9 and 33 h (not shown), near the time of daytime maximum heating during forecast days 1 and 2. Many of these coherent areas of flux difference correspond well with areas of vegetation fraction difference between the two forecasts.

To explore how the vegetation fraction information influences the forecasts, both the 2-m temperature and dewpoint temperature errors at 9-h and 33-h are binned with respect to the value of the VegEta vegetation fraction. Results indicate that both the Eta and VegEta provide the worst 2-m temperature forecasts for bare soil conditions (Fig. 4). Both runs, however, have better 2-m temperature forecasts as the vegetation fraction increases, while both forecasts also have worsening dewpoint temperature forecasts as the vegetation fraction increases from 0 to 60%.

The Eta and VegEta mean bias curves are nearly identical for vegetation fractions below 60% and separate for vegetation fractions

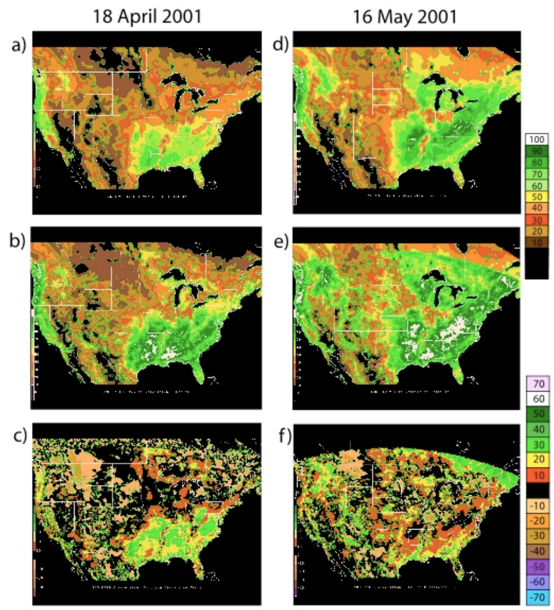

Fig. 1. Vegetation fractions from 18 April 2001 from (a) the Eta model climatological database, and (b) the NDVI-derived 1 km data set as used in the VegEta. The difference (VegEta-Eta) in the vegetation fractions is shown in (c). Similarly, the same fields from 16 May 2001 are shown in (d) through (f). Color bar on the upper right represents the vegetation fractions in % at intervals of 10%. The bottom right color bar is for the vegetation fraction difference fields.

above this value, with the VegEta biases smaller than those from the Eta. For vegetation fractions between 90 and 100%, the VegEta temperature forecast bias is smaller by 0.5°C compared to the bias in the Eta forecasts and, similarly, the VegEta dewpoint temperature bias is reduced by 1.0°C compared to bias in the Eta forecasts. As the growing season progresses, the influence of the real-time vegetation information decreases as illustrated by analyses from 18 April and 7 August (Fig. 5).

In April, the VegEta bias in 2-m temperature is 1°C smaller than those from the Eta, and for 2-m dewpoint temperature is 3°C smaller than those from the Eta, for vegetation fractions above 90%. However, by August these differences are only a fraction of a degree. This decrease in the forecast differences is expected; the average absolute differences in vegetation

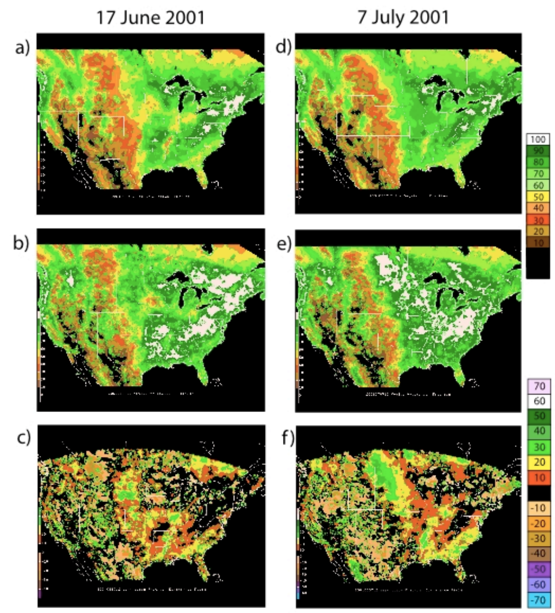

Fig. 2. As in Fig. 2, but for (a) to (c) 17 June 2001 and (d) through (f) 7 July 2001.

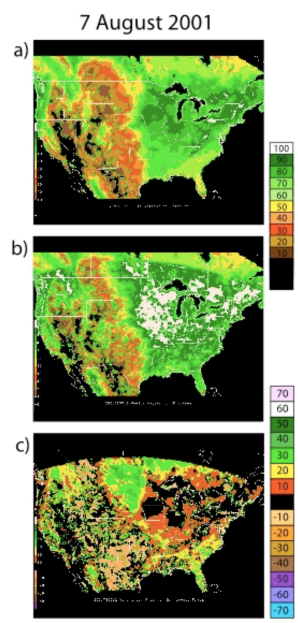

Fig. 3. As in Fig. 2, but for (a) to (c) 7 August 2001.

fraction between the VegEta and Eta also decrease during the growing season, and large regions showing only small differences in the vegetation fraction between climatology and the real-time data become apparent by July (Fig. 3f). However, for most vegetation fractions the

VegEta biases continue to be less than those from the Eta.

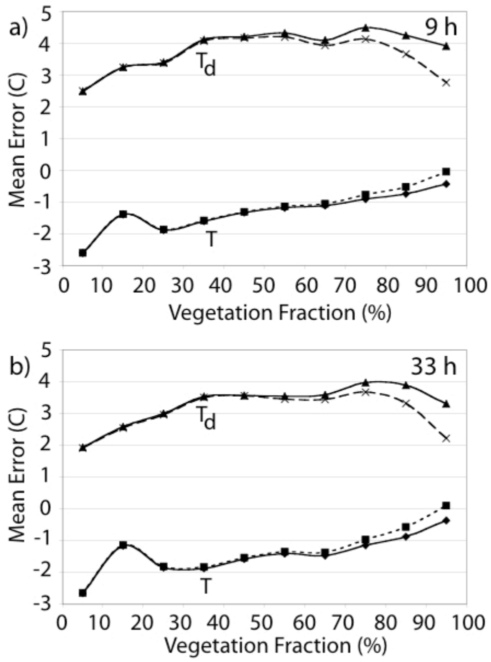

Fig. 4. Mean error, or bias (forecast – observation), vs. VegEta vegetation fraction (as derived from the NDVI data) for all 12 cases for (a) 9 h (2100 UTC) and (b) 33 h (2100 UTC) forecast times. Solid lines are for the Eta forecasts and dashed lines for the VegEta forecasts. T indicates the 2-m temperature curves, and T_d the 2-m dewpoint temperature curves. Data are averaged over 10% vegetation fraction bins based upon the VegEta vegetation fraction values.

The Wilcoxon signed-rank test, a nonparametric test for the significance of the difference between the distributions of two non-independent samples involving matched pairs (Wilks, 1995), is performed on the 9- and 33-h forecast 2-m temperatures and dewpoint temperatures for each case day. Results show that the differences in the temperatures between the VegEta and Eta forecasts are significant at the 99% level for the forecasts during April and May. However, the significance level decreases below 99% for the remaining half of the forecast days, with the exception of the 9-h forecasts for July 7th and August 7th and the 33-h forecasts for June 16th and July 5th. Differences in the dewpoint temperatures between the VegEta and Eta forecasts are significant at the 99% level for all the cases from April to July, but are not significant for the August days. This analysis supports the conclusion that the real-time vegetation information is most important early in the growing season and has decreased importance later during the summer when the vegetation growth largely is complete.

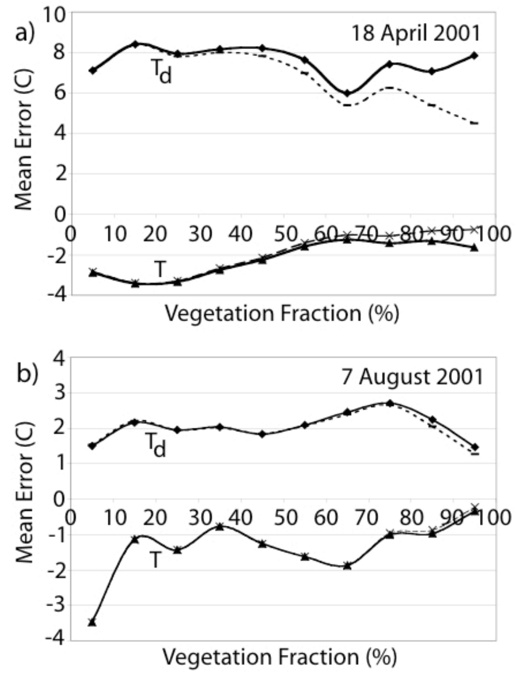

Fig. 5. As in Fig. 4, but from the 9 h forecast time only from (a) 18 April 2001 and (b) 7 August 2001. Note how the added vegetation information in the VegEta has a greater impact in April than in August.

4.0 DISCUSSION

The main goal of this study is to investigate the impact of using NDVI-derived vegetation fraction information in the Eta model. Since methods could be developed to provide these data in near real-time, it is important to evaluate what would happen if we could detect drought-affected vegetation, the early onset of spring green-up, regions of burned vegetation from forest fires, and crop harvesting, which otherwise are missed by the climatological vegetation data used in the operational Eta. Thus, values of the biweekly NDVI-derived vegetation fraction as determined using satellite data are inserted into the Eta model for 12 cases spanning most of the growing season during 2001 and forecasts using these data compared to those from both the operational version of the model and observations.

Results show that the use of the satellite-derived vegetation fraction improves the forecasts of both the 2-m temperature and dewpoint temperature at the 99% significance level for much of the growing season. This result clearly illustrates the value of satellite-derived vegetation information and the need for a system to get this type of information into the Eta model. While this study focuses solely on the influence of the implementation of near real-time NDVI-derived vegetation fraction in numerical models, the improved specification of other land-surface parameters that influence the forecasts also are worth investigation. Soil moisture specification, LAI, surface roughness, and surface albedo also are important in partitioning the surface energy budget and methods for improving their initialization, such as the Land Data Assimilation System (Mitchell et al. 2000), should be vigorously pursued. However, until these data are incorporated more completely into the operational models, the model flux predictions and low-level temperature forecasts may continue to be problematic.

References

Black, T. L., 1994: The new NMC mesoscale Eta model: Description and forecast experiments. *Wea. Forecasting*, **2**, 266-278.

_____, M.Baldwin, K. Brill, F. Chen, G. DiMego, Z. Janjic, G. Manikin, F. Mesinger, K. Mitchell, E. Rogers, and Q. Zhao, 1997: Changes to the Eta forecast systems. *NWS Technical Procedures Bulletin*, **441**, 31 pp. [Available from NCEP/EMC, 5200 Auth Road, Camp Springs, MD 20746.]

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.

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

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.

Gutman, G., and A. Ignatov, 1995: Global land monitoring from AVHRR: Potential and limitations. *Int. J. Remote Sens.*, **16**, 2301-2309.

______, and _____, 1998: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int. J. Remote Sens.*, **19**, 1533-1543.

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. 46, Amer. Meteor. Soc., 165-170.

Mahrt, L., and H. –L. Pan, 1984: A two-layer model of soil hydrology. *Bound.-Layer Meteor.*, **29**, 1-20.

Markowski, P. M., and D. J. Stensrud, 1998: Mean monthly diurnal cycles observed with PRE-STORM surface data. *J. Climate.*, **11**, 2995-3009.

Mitchell, K., and Coauthors, 2000: Recent GCIP-sponsored advancements in coupled land-surface modeling and data assimilation in the NCEP Eta mesoscale model. *Preprints,* 15th Conf. on. Hydrology, Long Beach, CA, Amer. Meteor. Soc., 180-183.

Ookuchi, Y. M., 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**, 2281-2291.

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

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.

Schwartz, M. D., and T. R. Karl, 1990: Spring phenology: Nature's experiment to detect the effect of "green-up" on surface maximum temperatures. *Mon. Wea. Rev.*, **118**, 883-890.

Segal, M., J. F. W. Purdom, J. L. Song, R. A. Pielke, and Y. Mahrer, 1986: The predictability of sea breeze generated thunderstorms. *Preprints, WMO International Workshop on Rain-Producing Systems in the Tropics and Extra-Tropics*, San Jose, Costa Rica, 101-107.

Wetzel, P. J., and A. Boone, 1995: A parameterization for land-atmosphere-cloud exchange (PLACE): Documentation and testing of a detailed process model of the partly cloudy boundary layer over heterogeneous land. *J. Climate*, **8**, 1810-1837.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences, Academic Press, New York, 467 pp.