9.1

APPLICATION OF ATMOSPHERIC AND LAND DATA ASSIMILATION SYSTEMS TO AN AGRICULTURAL DECISION SUPPORT SYSTEM

William Myers* and Fei Chen
National Center for Atmospheric Research, Boulder, CO

James Block
DTN Meteorlogix, Burnsville, MN

1. INTRODUCTION

Agriculture is a critical sector of the US economy. Both weather and soil conditions are important inputs to agricultural decision-making processes. However, both the weather and the soil models necessary to adequately predict the agricultural environment at field scales are currently lacking. For example, phenological pest models predict the evolution of an organism's life stages based on the temperature of its environment. These models often use only daily maximum and minimum air temperatures to estimate the continuum of conditions affecting the organism. These gross temperature bounds can be poor surrogates for temporally higher resolution air and soil temperature forecasts that are specific to a farm's microclimate.

The National Center for Atmospheric Research (NCAR) and DTN/Meteorlogix are working together on a NASA-funded project to improve soil condition predictions critical for agricultural applications. The approach is to optimize weather forecasts from several real-time weather prediction models using NCAR's Dynamic Integrated ForeCast (DICast®) system and to combine that with the High-Resolution Land Data Assimilation System (HRLDAS) and the Noah Land Surface Model (LSM) to generate high-resolution soil temperature and moisture forecasts. These forecasts will be used to drive agriculture-specific models, such as pest and crop models. The output of these models will be provided to over 60,000 agricultural users via the DTN/Meteorlogix Decision Support System (DSS), DTN Online.

An important goal of this project is to assess the impacts of NASA new-generation remote-sensing data on predicting soil temperature. Real-time NASA MODIS satellite data will be used to improve the HRLDAS initial land and vegetation conditions. Many of the current HRLDAS vegetation parameters are based on climatological data sets. MODIS leaf area index (LAI) and green vegetation index called the Fraction of Photosynthetically Active Radiation (FPAR) provide much more timely and accurate estimates of the current environmental state. They are also available at a much higher spatial resolution than other data sets. The incorporation of these additional data is expected to lead to a better description of vegetation development and hence better estimates of plant transpiration and soil state across the forecast domain. The difference between climatological and remotely sensed LAI is obvious in Figures 1 and 2.

2. PROJECT ORGANIZATION

The project is divided into two main areas. The first is a scientific research effort geared toward base improvements in the HRLDAS/Noah modeling system. This work includes evaluation of different thermal transfer approaches and the incorporation of NASA MODIS data sets. These retrospective studies have used NCEP weather analyses and observational data from 2005-2007. The weather and soil observations were collected by the Soil Climate Analysis Network (SCAN) and the Oklahoma Mesonet.

The second effort has been the development of an operational forecast system which takes advantage of the latest HRLDAS advances. This system generates soil temperature and moisture forecasts out to 48 hours into the future. It is run once per day at 0900Z so that its forecast conditions are available to the agricultural community early in the morning. DTN currently uses these forecasts internally as a resource to its agricultural advisory staff and eventually expects to publish the data online in an end-user friendly format.

The project domain covers the east and central US. In this area, dry land (non-irrigational) farming is practiced and human effects on the environment, such as soil moisture are lessened. This domain covers much of the DTN user base. Unfortunately, there are relatively few soil observational sites within the domain. This is the

* *Corresponding author address:* William Myers, NCAR, 3450 Mitchell Lane, Boulder, CO 80302; e-mail: myers@ucar.edu

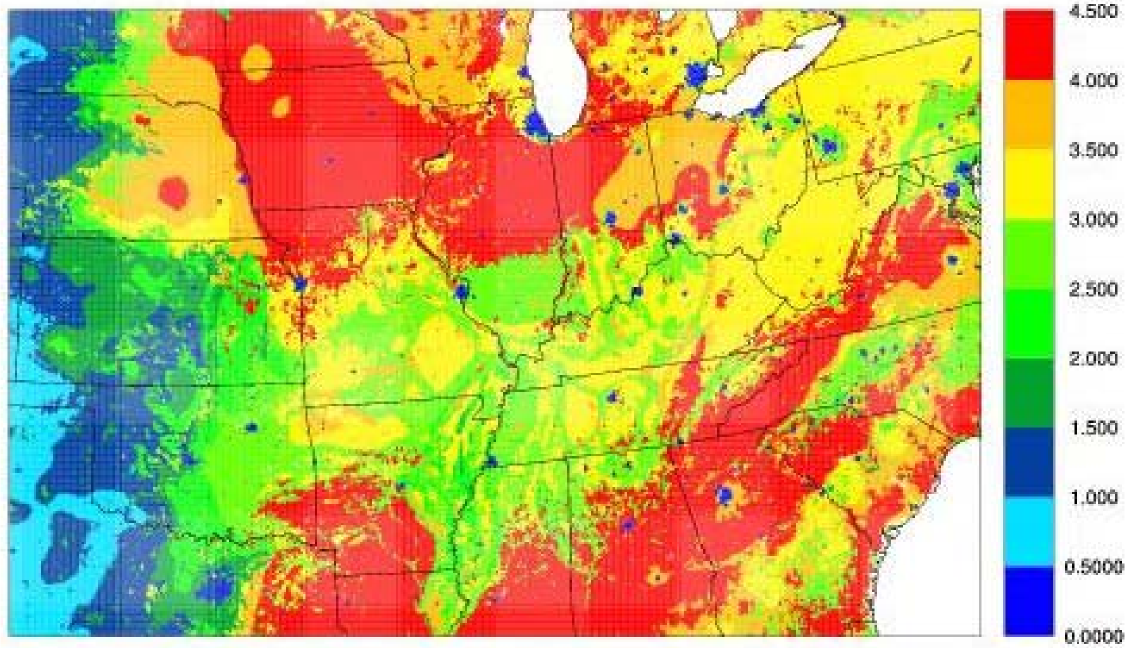

Figure 1: HRLDAS climatological Leaf Area Index (LAI) for July 1.

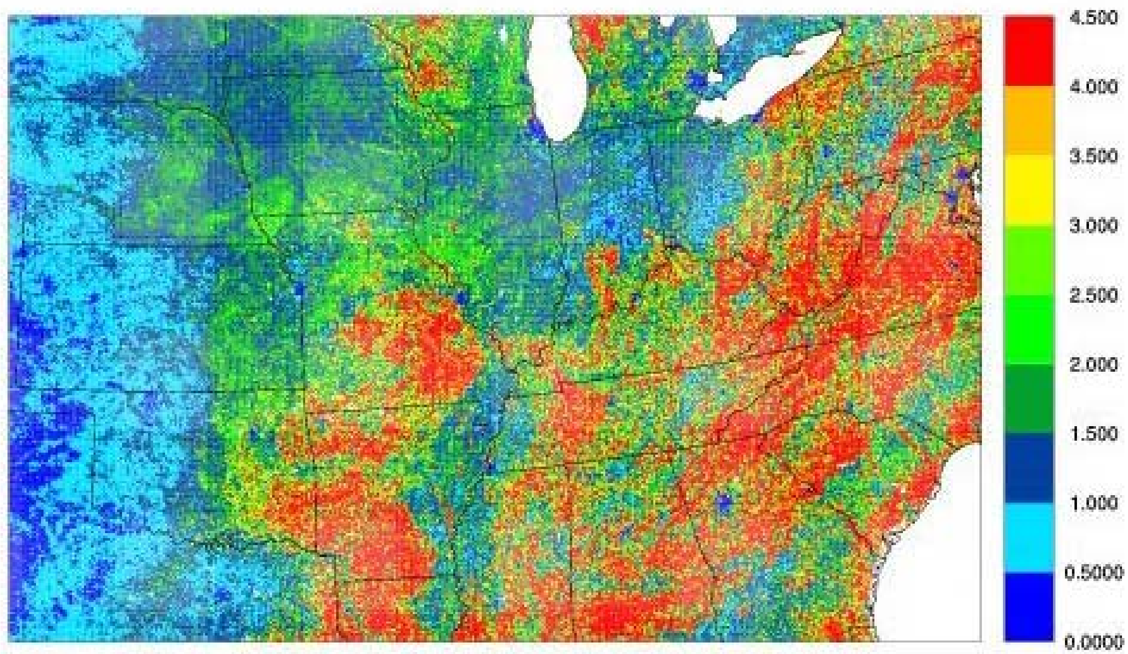

Figure 2: NASA-MODIS remotely sensed LAI valid around July 1, 2006.

main limiting factor in verification of the soil model output. The domain is shown in Figure 3 along with a 5 cm soil temperature forecast at 4.5 km spatial resolution.

The verification of the modeled soil temperatures and moistures requires a high quality observational data set with complete metadata. While the OK Mesonet provides the highest standard data set encountered thus far, for the HRLDAS/Noah development and verification project, we have focused on the SCAN data. The spatial distribution of the SCAN sites is necessary to effectively tune the HRLDAS/Noah system to run under a wide variety of soil types and vegetation conditions. To increase the number and distribution of observational sites, we continue to investigate other networks and will ingest their data as time permits. The SCAN sites in our domain of interest are shown in Figure 4. For verification purposes, the domain has been broken into 4 quadrants (NW, NE, SW, and SE). The results vary significantly between regions. Due to the project's agricultural focus, the verification has mainly focused on the growing season (Spring and Summer) at the 5 and 10 cm soil depths where early plant development is occurring.

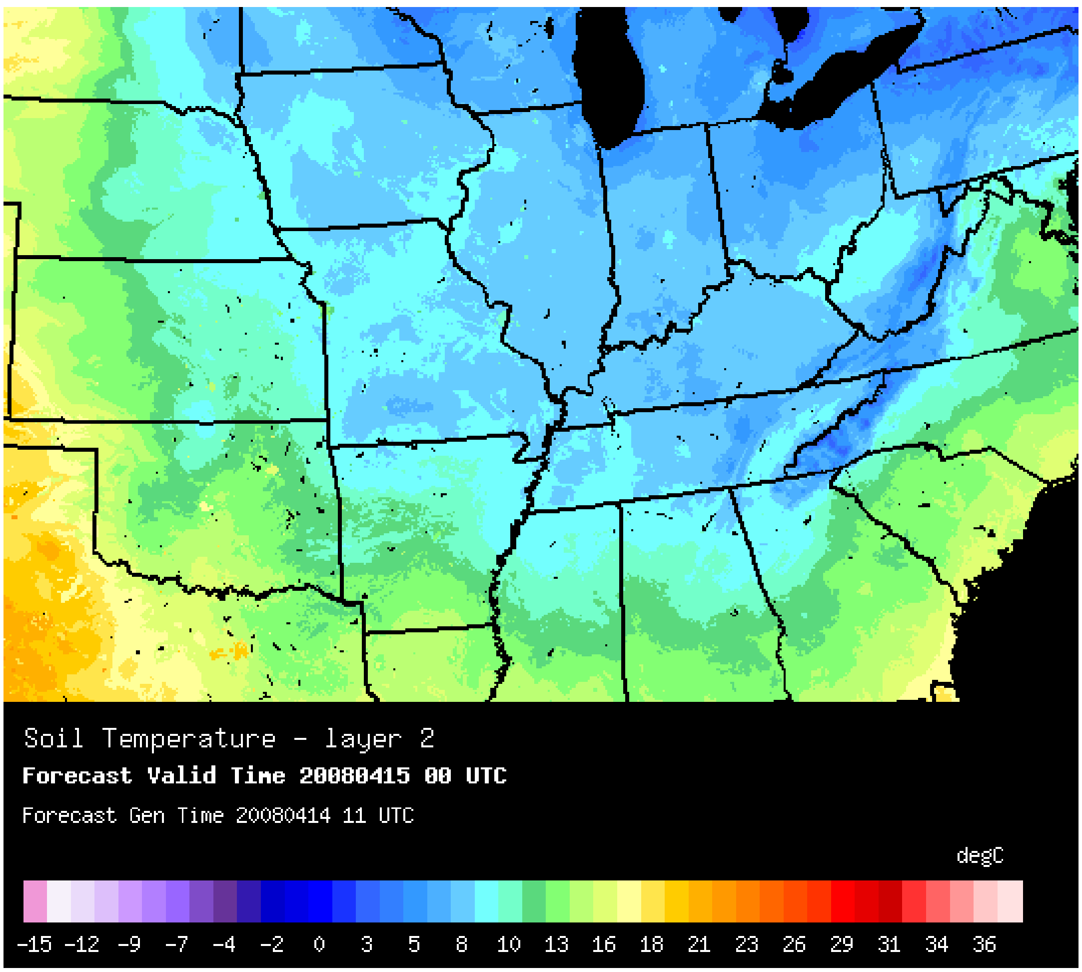

Figure 3: HRLDAS/Noah LSM domain. The grid consists of 543x350 points at roughly 4.5 km resolution.

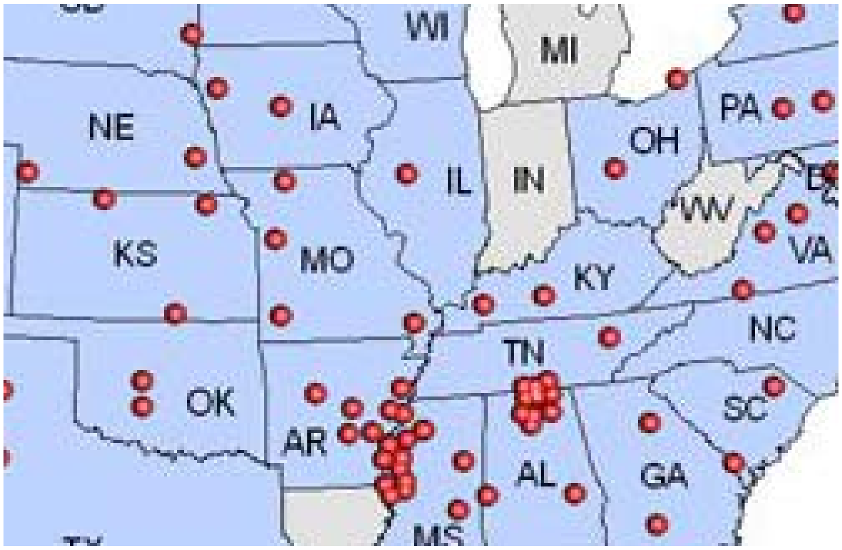

Figure 4: SCAN sites within the project domain. Note that not all these sites are active. Data quality problems have been uncovered at others, making them unusable for verification purposes.

3. RESULTS TO DATE

At the start of the project, in September 2006, HRLDAS was configured to run with 4 subsurface nodes that matched the NCEP North American Model (NAM) subsurface soil depths (5cm, 25cm, 75cm, and 150cm). Early verification results indicated that the thermal transfer near the surface was not being adequately modeled in some regions. After experimenting with a number of other thermal conductivity models and seeing little improvement, it was decided to add more subsurface nodes. A node near the surface (at 1cm depth) was added. The other nodes were adjusted to better match the standard SCAN and OK Mesonet observational depths (5cm, 20cm, 50cm, and 100cm). The deepest node, at 166cm, is not verifiable. The new near-surface node reduced the mean absolute errors (MAE) for the 5 cm node by roughly 6%, a nice incremental improvement. The 10 cm node saw a huge improvement in modeled soil temperature with a reduction of over 40% in the MAE. The error reductions were fairly uniform across all regions. These results can be seen in Tables 1 and 2.

Still, in several regions the 5 cm soil temperature diurnal cycle was much too large compared to the observations. The Noah model has a parameter called “czil”, which describes the turbulence near the surface and affects the surface heat transfer. This parameter was globally defined. That is, a single value was used across the domain. In the past, the HRLDAS/Noah system had been run over smaller domains. This global parameter usually had to be tuned for a project’s domain. In this project, it seemed that the results in some regions were very sensitive to the value of czil. By varying this parameter, large improvements could be seen in some regions for certain seasons while others were less significantly affected. Eventually a primitive algorithm was developed to set czil values on a point-by-point basis. The main inputs to the algorithm are location and vegetation type. The development of such an algorithm is difficult due to the limited number of observational sites and incomplete variety of soil and vegetation types at these locations. The improvement due to the addition of the czil algorithm was significant for several regions and seasons. It actually had a negative impact in the SE quadrant. However, in this region, the 6-layer results were already quite good. Several sites in Mississippi, along the river, seem to be worsened by the refined czil algorithm. This will merit further investigation in the coming months.

The addition of the NASA-MODIS data further improved the modeled soil temperatures overall. Again, the improvement varied regionally, with some regions showing slight increases in forecast error. One reason for this may be the inexact use of the MODIS data up to now. The 1 km resolution MODIS data were first extracted at the centers of the HRLDAS (4.5 km) grid, then applied to the observation sites. Thus the MODIS data applied at the observation sites could come from up to 3 km away. With this caveat, it is still encouraging to see incremental improvement in forecast quality from the MODIS data. At the 10 cm depth, the improvement in the SE region nearly overcomes the setback there due to the czil modification.

	4-Layer	6-Layer	CZIL	MODIS
NE	2.17	2.14	2.14	2.00
NW	2.53	2.44	2.17	2.21
SE	1.77	1.66	2.00	2.12
SW	2.28	2.02	1.80	1.65
Overall	2.188	2.065	2.028	1.995

Table 1: 5 cm Soil temperature MAEs for the planting and early growing season (April 1 – July 1, 2006) in degrees C.

	4-Layer	6-Layer	CZIL	MODIS
NE	4.83	2.35	2.24	1.75
NW	2.51	1.54	1.59	1.63
SE	3.5	2.03	2.37	2.12
SW	2.96	1.96	1.96	1.72
Overall	3.450	1.970	2.040	1.805

Table 2: 10 cm Soil temperature MAEs for the planting and early growing season (April 1 – July 1, 2006) in degrees C.

Overall, the improvement in modeled soil temperatures from the initial 4-layer version to the current 6-layer version with the czil algorithm and MODIS data, is fairly significant. The 5 cm soil temperatures are improved by roughly 9% while the 10 cm temperature errors are nearly halved. Not including the SE region, the 5 cm errors are decreased by 16%.

The operational forecast system has been running continuously since late February 2008 using the 6-layer version of HRLDAS/Noah without the czil and NASA data upgrades. Weather forecasts have been generated using NCEP's NAM numerical weather prediction model and also using NCAR's DICast® system. DICast® is a consensus fuzzy logic forecast system which emulates the human forecast process and outperforms all the forecasts it considers (including NWS MOS forecasts). Using these two weather forecasts, soil forecasts have been generated daily. Soil forecast errors from the month of March 2008 are provided below. Note that in the retrospective (scientific research) work, weather observations were used to drive the Noah LSM, while in the operational system, weather forecasts are used and thus the results would not be expected to match reality as well. Also, note that several northern sites have been removed from this early season evaluation. It is difficult to accurately determine whether a site is affected by snow cover. Snow cover significantly effects the near surface soil temperature and moisture calculations. These sites showed unusually large errors with indications that snow cover had not been correctly determined.

	5 cm	10 cm
NAM	2.06	1.52
DICast	1.96	1.55

Table 3: Soil temperature MAEs for March 2008 forecasts using two different weather forecast systems.

DICast's improvement in the weather forecast translates to only a small improvement in the 5 cm soil temperatures but slightly worsens the 10 cm errors. In all, these are encouraging results with soil temperature errors of 2°C or less for these 48 hour prediction systems. The soil moisture errors were around 7% for both forecast systems and depths.

4. A DATA MINING APPROACH TO SOIL TEMPERATURE AND MOISTURE PREDICTION

A data mining approach was developed to predict soil temperature/moisture at observing sites. This is an alternative to the HRLDAS physically-based model. At each site, an observational history of relevant driving data from 2005-2006) was generated. Each site's data were fed into a data mining package called "Cubist". Cubist performs a Classification and Regression Tree (CART) analysis on the observational history and generates "rules" to make a forecast based on new observational input data. These rules consist of classification of the data (e.g. air temperatures rising and downward solar radiation in a certain range) and a linear regression equation to be applied to each class. These linear equations approximate the complex non-linear processes

	0	1	Median	3	4
5 cm HRLDAS/Noah	1.13	1.55	1.96	2.52	6.18
5 cm Data Mining	0.53	0.66	1.03	1.31	2.67
10 cm Data Mining	0.40	0.48	0.84	1.12	2.13
20 cm Data Mining	0.29	0.36	0.59	0.82	2.05

Table 4: Soil temperature MAEs for HRLDAS/Noah and data mining for the 2007 growing season. Errors are provided for each quartile. The errors were calculated over a 60 hour forecast period. Note that the 5 cm machine learning errors are roughly half of those produced by the physical model.

and provide a prediction for the soil temperature and moistures one hour in the future.

A data mining based system was crudely put together to generate forecasts out to 60 hours. The forecasts had to be bootstrapped. That is, using the latest observations (hour 0), the first hour's estimates of soil temperature and soil moisture were generated. Using these one-hour forecasts and the observed weather, the second hour's soil forecasts were generated, and so on. This is a "perfect prognosis weather forecasting" approach. That is, we assume that the weather forecasts are perfect. Forecasts were generated for the 2007 growing season and compared to HRLDAS forecasts (without czil and MODIS improvements).

The surprising result was that these data mining generated forecasts showed errors that are roughly half of what the physically-based model produces. Table 4 shows the 60 hour forecast mean absolute errors for each quartile of the 34 sites used in the study. These sites were required to have at least 365x24 complete observations in 2005-2006 and produce forecasts for at least half the days in the 2007 growing season.

The median error for the 5 cm soil temperature prediction was about 1.0ºC for the data mining approach and about 2.0ºC for HRLDAS/Noah. The deeper soil temperatures showed less error as would be expected since soil temperature generally varies more slowly at the lower levels.

For soil moisture, the results were much better for the data mining approach than for HRLDAS/Noah. This is because the data mining forecast system starts with an observed set of soil moistures. HRLDAS does not have this feedback and drifts slightly from reality over time. If the soil moisture observations were directly inserted into HRLDAS, its forecasts would probably be similar to those generated by the data mining system.

While these results are encouraging, it should be made clear that the data mining system has its drawbacks. Mainly, it can only be used at locations where there is an observational history – and a longer history is better. DTN/Meteorlogix's initial interest is in grid forecasts and there is not a clear path to improve the grid forecasts using this technique. It may be possible to develop a set of "rules" for each soil and vegetation type combination. This set of rules could be applied at each grid point having that characterization. Using this approach, it might be possible to generate a grid populated with data mining forecasts. However, the weakness in this plan is that there are not enough soil observation sites to develop rules for all soil and vegetation pairs. Still, we think there is probably a way to use the data mining approach to improve the HRLDAS/Noah forecasts.

A main goal of this data mining effort is to independently evaluate the potential value of the NASA-MODIS data. Once extracted, it will be relatively easy to add the LAI and FPAR (as well as other MODIS data sets) to the training data used to generate the CART rules and equations. By regenerating the simulated soil state using these new rules, it will be possible to quantify what effect, if any, the MODIS data has on the soil temperature and moisture calculations.

5. CONCLUSIONS AND FUTURE WORK

Exploratory efforts in a number of areas have led to improvements in modeled soil temperature and moisture values. The addition of more nodes, the czil algorithm, and the NASA-MODIS LAI and FPAR data have led to significant reductions in errors at the depths most important to agricultural decision making. Further verification on other independent data sets will be required to validate the flavor of the results presented above.

These upgrades to HRLDAS are being incorporated into the operational forecast system as appropriate. The current operational forecast system (6-layer with no czil or MODIS upgrades) provides a soil forecast product that is more accurate than any other known product. It is also available at higher spatial and temporal resolution than other products. The value of this operational product to end-users is being evaluated by the largest US commercial agricultural weather provider with a base of more than 60,000 users who already receive their information via DTN Online, an agriculturally oriented DSS.

While the use of the NASA-MODIS LAI and FPAR data sets can still be refined, other MODIS data sets merit consideration. The most obvious next MODIS data set to investigate will be albedo. This would replace the spatially and temporally coarse climatological albedo data set currently in use. The problem is that MODIS provides albedo at several wavelengths while the Noah LSM expects one albedo value. The mapping from MODIS albedos to Noah albedo is not yet understood.

It is also expected that upgrades to the Noah LSM physics such as multi-layer snow and vegetation canopy models will be developed in the near future. This model enhancement work will be funded by other sponsors but its results will hopefully be incorporated into this project.

6. ACKNOWLEDGEMENTS

This research is sponsored by the NASA Earth Observing System, in response to NRA #NNH05ZDA001N, Research Opportunities in Space and Earth Sciences (ROSES-2005). The material within is based upon work supported by NASA under award No. NNA06CN03A titled "A Soil Temperature and Moisture Decision Support System for Agriculture". Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This project would like to acknowledge the data sets and help provided by the USDA Natural Resource Conservation Service (NRCS) who manage the SCAN data sets, the Oklahoma Mesonet, and various NASA labs who organize the MODIS data.

7. REFERENCES

7.1 *HRLDAS and Noah LSM*

Chen,F., K. Manning, M. LeMone, S. Trier, J. Alfieri, R. Roberts, M. Tewari, D. Niyogi, T. Horst, S. Oncley, J. Basara, and P. Blanken, 2005: Description and Evaluation of the Characteristics of the NCAR High-Resolution Land Data Assimilation System, Vol.46, pp. 694-713

Trier, S., F. Chen, K. Manning, M. LeMone, and C. Davis, 2007: Sensitivity of the PBL and Precipitation in 12-Day Simulations of Warm Season Convection using Different Land Surface Models and Soil Wetness Conditions. Submitted to *Monthly Weather Review*, June 2007, revised October 2007.

Trier, S., F. Chen, and K. Manning, 2004: A Study of convection initiation in a mesoscale model using high-resolution land surface initiation conditions. *Mon. Wea. Rev., in press.* Five Other Publications

7.2 *DICast (sometimes referred to as the Road Weather Forecast System)*

Mahoney, W.P., and W. L. Myers, 2003: Predicting Weather and Road Conditions: An Integrated Decision Support Tool for Winter Road Maintenance Operations. *Transportation Research Record: Journal of the Transportation Research Board*, No 1824, 2003, pages 98-105.

Myers, Bill, M. Petty and J. Cowie, 2002: An Automated Road Weather Prediction System for Road Maintenance Decision Support. 18th AMS IIPS Conf 3.5.

Pisano, Paul, A. Stern, W. Mahoney, W. Myers, D. Burkheimer, 2004: Winter Road Maintenance Decision Support System Project (MDSS). Transportation Research Board: 6th International Symposium of Snow and Ice Removal

7.3 *Relevant web sites*

www.rap.ucar.edu/projects/
nasa-ag/hrldas/display_hrldas_animation.html

www.rap.ucar.edu/projects/dicast

www.rulequest.com /cubist-info.html