JP 1.16 VERIFICATION CASE STUDIES WITHIN THE 12KM NORTH AMERICAN LAND DATA ASSIMILATION SYSTEM (NLDASE) PROJECT

Charles J. Alonge*, and Brian A. Cosgrove

SAIC / Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD

1. INTRODUCTION

For the past seven years, researchers at the NASA Goddard Space Flight Center (GSFC) have been developing an LDAS under the auspices of the collaborative North American Land Data Assimilation System (NLDAS) project (Mitchell et al. 2004). Including partners from NOAA NCEP, NOAA OHD, NOAA CPC, NOAA NESDIS, Princeton University, Rutgers University, the University of Washington, and the University of Maryland, this project has produced high-quality retrospective and real time land surface fields over a 1/8th degree CONUS domain (Figure 1).

Validation efforts focused on NLDAS land surface states and fluxes have illustrated the high quality nature of NLDAS output (Robock et al. 2003), and cleared the way for the use of LDAS conditions in the NWP initialization process. In response to this, NASA GSFC and NOAA NCEP initiated a follow-on to the NLDAS project, named the NLDAS Arakawa E-grid (NLDASE) project. The NLDASE offline modeling system operates on the same 12km grid (Figure 1) used by NCEP's Eta mesoscale model (Black 1994), the National Weather Service's main regional weather prediction model. Through the use of the same LSM (the Noah LSM, Ek et al. 2003), grid, and parameter data sets, interpolation issues are avoided and NLDASE land surface states can be directly inserted into the coupled Eta model.

Because soil moisture, temperature and snow are integrated land surface states, biases in land surface forcing, model physics and parameterization accumulate in the land surface stores of fully coupled land surface models often used to initialize numerical weather prediction models (NWP). These biased land surface states have detrimental impacts on the partitioning of the surface energy and water fluxes which can ultimately lead to inaccurate weather forecasts. Through the use of NASA's Land Information System (Kumar et al. 2004), high quality meteorological forcing, and data assimilation, the NLDASE project aims to remove the biases that can accumulate in coupled modeling systems. A complete overview of the NLDASE modeling system, forcing data, and assimilation methods can be found in Cosgrove and Alonge (2005).

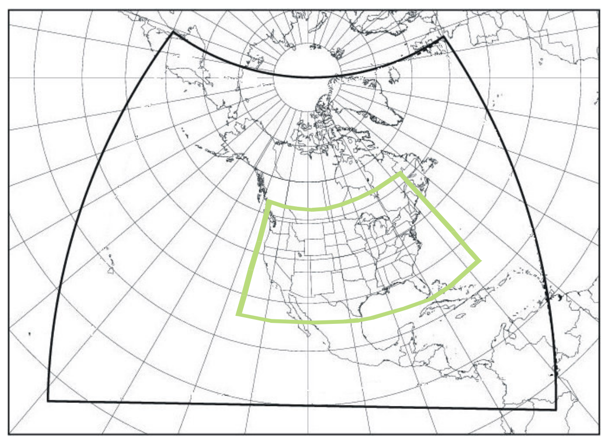


Figure 1. NLDASE / Eta (black) and NLDAS (green) domains.

2. NLDASE BENCHMARKING EFFORT

The NLDASE project recently underwent an effort to evaluate the impact of NLDASE land surface states on Eta model forecasts. In order to provide initial NLDASE land surface conditions to the Eta model, three retrospective offline Noah LSM simulations were executed from October 2000 through August 2003. The first of the three experiments (LIS1) utilized an optimal blend of NLDASE forcing data and did not utilize any data assimilation techniques. The second and third

*Corresponding Author: Charles J. Alonge
NASA Goddard Space Flight Center
Mail Code 614.3, Greenbelt, MD 20771
Email: Charles.Alonge@gsfc.nasa.gov

retrospective simulations (LIS2 and LIS3, respectively) are identical to LIS1 with the exception that each run includes the assimilation of daily MODIS snow cover values (Hall et al. 2002) according to the rule based approach of Rodell and Houser (2004). The LIS2 simulation applied a daily snow water equivalent (SWE) update amount of 5 mm while LIS3 used an amount of 10 mm. Output from the three offline NLDASE retrospective simulations mentioned above were then used to initialize the land surface states within a workstation version of the operational Eta model.

Computational constraints did not allow for a large number of Eta simulations, therefore, a 10 day benchmark experiment period was chosen to analyze the impact of NLDASE initial conditions on Eta model forecasts. In order to promote the inclusion of an active land surface, yet at the same time enable an examination of the impacts of MODIS snow assimilation, a convectively active spring season benchmarking period was deemed optimal. Previous studies have shown that the land surface can play a significant role in dictating convective initiation and intensity (Findell and Eltahir 2003; Clark and Arritt 1995). Additionally, small changes in the planetary boundary layer moisture of 1 g kg^{-1} can influence the triggering of convection (Crook 1996). A more active land surface and its associated increases in evaporation could account for such a difference in boundary layer moisture. It was therefore hypothesized that the NLDASE land surface states would have the largest impact in a convectively active period.

Spanning the period from May 1st to May 10th, 2003, the benchmark experiments were comprised of a total of 80 separate 84-hour Eta model simulations. Specifically, 84-hour Eta model simulations were executed each day at 00Z and 12Z over the 10 day time period. Four sets of Eta model runs were conducted in this fashion, each using different types of restart files that included: 1) NCEP operational restart files to establish a baseline control run, 2) NLDASE LIS1 restart files to test the basic effects of uncoupled NLDASE initial conditions, 3) NLDASE LIS2 restart files to test the effect of MODIS snow assimilation using a 5mm SWE value, and 4) NLDASE LIS3 restart files to test the effect of MODIS snow assimilation using a 10mm SWE value.

All 80 Eta model forecasts executed during the benchmark period were then compared with observations and with each other to gauge forecast improvement/degradation from the use of NLDASE uncoupled land surface states. NCEP's Forecast Verification System (FVS) was chosen as the centerpiece of the regional and national benchmarking effort, while a separate suite of site-specific benchmarking metrics was utilized to provide complementary local analyses.

3. RESULTS

Overall, initialization of Eta land surface states with NLDASE output had a mixed impact on forecasts (Cosgrove and Alonge, 2005). Although detailed results from the use of NCEP's Forecast Verification System on the benchmark Eta experiments are detailed in presentation J5.4 (Cosgrove and Alonge) a brief summary is given here:

- The use of NLDASE land surface states greatly improved the surface forecasts of temperature and relative humidity overall while the impacts on wind speed forecasts were mixed.
- The use of MODIS snow cover assimilation (LIS2, LIS3) generally improved surface forecasts.
- Upper air forecasts showed mixed results with 300 mb temperature being the only field which consistently improved overall.
- The root mean square error (RMSE) of upper air forecasts was more often reduced than was the bias.
- The impact on 24 hour precipitation forecasts was mixed and generally small.
- The NLDASE initialized forecasts showed small improvements in the equitable threat score and probability of detection statistics indicating improvements in precipitation placement and in the fraction of time the Eta model issued a non-zero

precipitation forecast given the occurrence of an observed precipitation event.

- The use of MODIS snow cover assimilation led to small improvements in precipitation bias, but had only mixed impacts on the equitable threat score, probability of detection and false alarm ratio statistics.

The aforementioned FVS benchmark metrics are frequently used by NCEP to evaluate Eta model performance, and provide copious amounts of useful information as to where weaknesses are present in numerical forecast guidance. However, these benchmarks are regional in nature and cannot depict the true impact that the forecast improvement/degradation may have on a single location. For example, the 2 m relative humidity bias may be 2% too high in a particular verification region in the LIS1 forecasts, and 4% too high in the control Eta forecasts. This is not a large difference in the regional bias, but at a specific location, the biases in relative humidity may be larger and may have significant ramifications on local forecast guidance.

A selection of surface station time series are presented to highlight the impacts that NLDASE initialization has on site-specific ETA forecasts of temperature, and relative humidity. Precipitation timing, placement, and intensity errors are also presented graphically to illuminate some of the differences that may not be depicted in the aforementioned traditional regional FVS skill scores. Finally, time series of surface downwelling/upwelling shortwave and longwave radiation forecasts and observations are presented to evaluate the impact NLDASE land surface states have on these fields. Hourly surface temperature, humidity, and mean sea level pressure station data were extracted from the same observational data sets used in creating the regional NCEP FVS statistics. Surface fields from the Eta forecasts were bilinearly interpolated to station locations which allowed for a direct comparison of all four forecasts to each another and to the surface station observations. Data from the Surface Radiation Budget Network (SURFRAD) were used to evaluate the surface radiation fluxes against Eta data bilinearly interpolated to the SURFRAD sites. For precipitation forecasts, NCEP Stage II hourly precipitation data were compared to hourly precipitation output from the Eta forecasts to illustrate improvements/degradations.

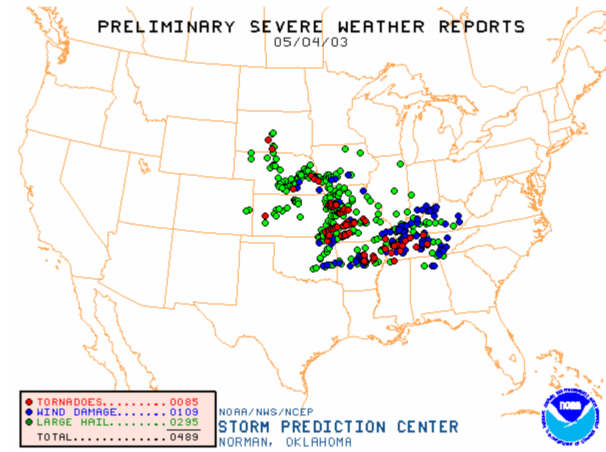


Figure 2. Tornado, Wind Damage, and Large Hail reports for May 4th, 2003 from the Storm Prediction Center (SPC). Source: NOAA / SPC

3.1. *Surface Temperature and Relative Humidity Forecast Evaluation*

The regional benchmarks of the retrospective Eta forecasts indicated that the surface temperature and relative humidity fields were most sensitive to the use of uncoupled NLDASE land surface states. Selected verification sites throughout this section will show forecast improvement in these variables from four retrospective Eta model forecasts (control, LIS1, LIS2, and LIS3) initialized at 12 Z on May 3rd, 2003 and run out to 84 hours (ending 00 Z on May 7th, 2003), a period which encompassed copious amounts of convective activity and severe weather (Figure 2).

A significant number of sites showed good agreement between all four forecasts, with some sites leaving little room for improvement with respect to the observations. However, the four retrospective Eta forecasts were not always in such good agreement with one another. Figure 3 shows a great deal of variation between the four forecasts at Great Trout Lake, a site in northwestern Ontario, Canada. This was consistent

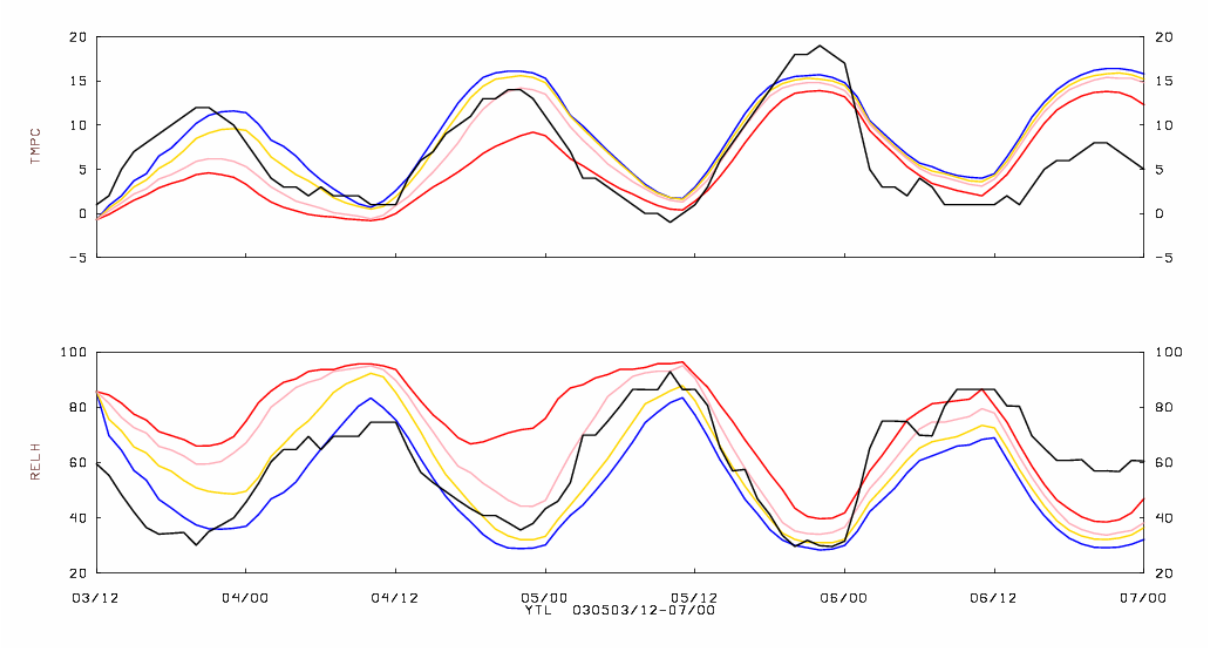


Figure 3. Surface temperature (TMPC), and relative humidity (RELH) forecasts compared to observations at Great Trout Lake, Ontario (YTL). Temperature is in degrees Celsius, and relative humidity in percent. Observations are plotted in black, the control ETA forecast in red, LIS1 forecast in blue, LIS2 forecast in yellow, and LIS3 forecast in pink

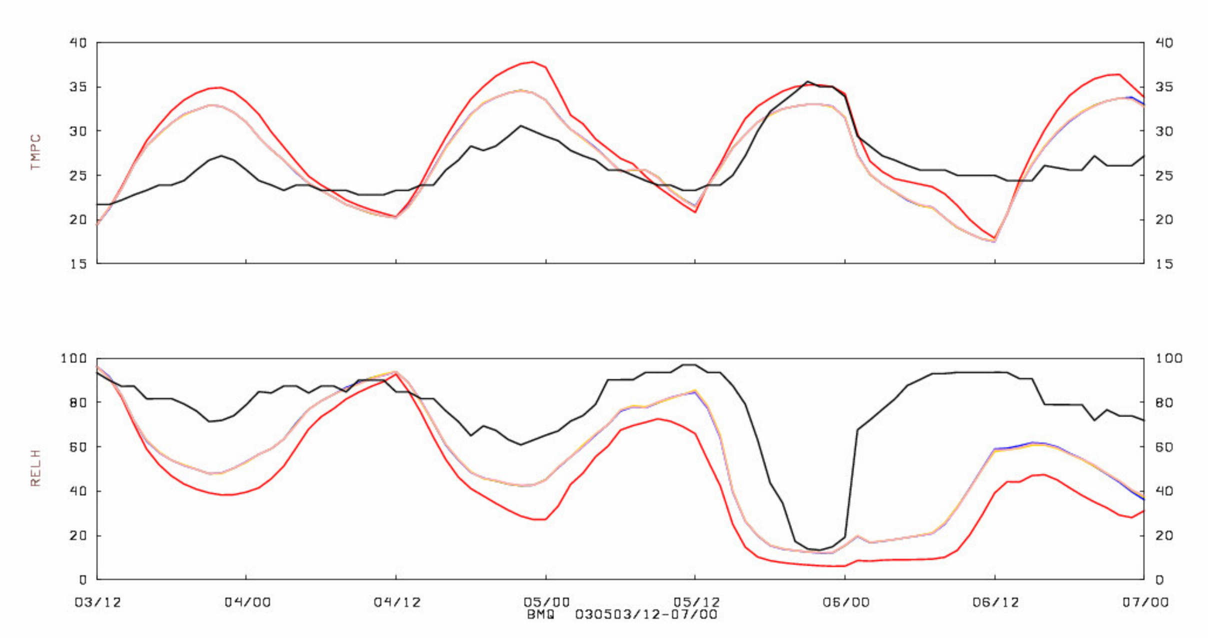
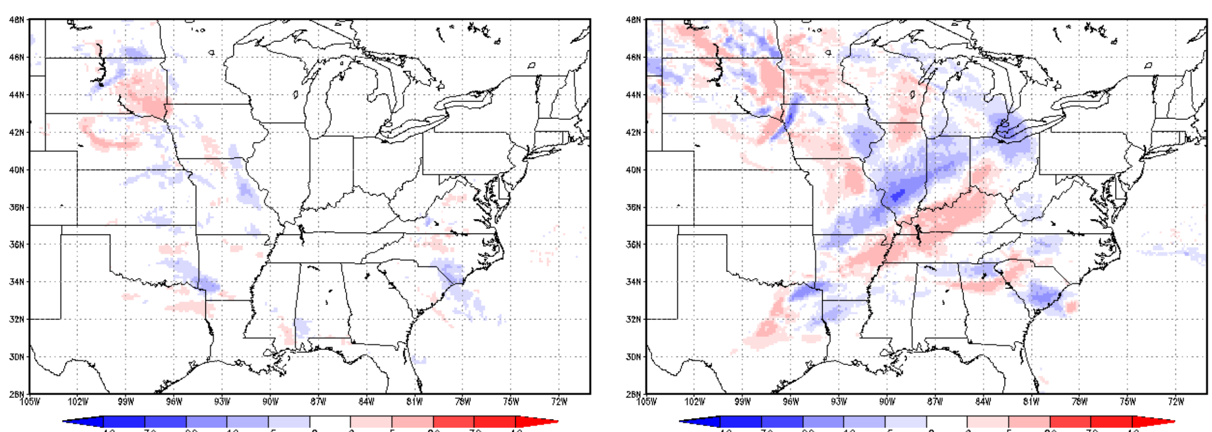


Figure 4. Surface temperature (TMPC), and relative humidity (RELH) forecasts compared to observations at Burnet, Texas (BMQ). Temperature is in degrees Celsius, and relative humidity in percent. Observations are plotted in black, the control ETA forecast in red, LIS1 forecast in blue, LIS2 forecast in yellow, and LIS3 forecast in pink.

with the regional verification statistics which showed that MODIS snow assimilation had the largest impacts on surface temperature and relative humidity throughout much of Canada. The LIS1, LIS2, and LIS3 forecasts performed better than the control forecast in terms of relative humidity and temperature for a large portion of the forecast period. Only in the last 24 hours of the forecast did the control run perform noticeably better with respect to temperature and relative humidity.

The Great Trout Lake, Ontario site exhibited a large amount of variability between all four forecasts. Again, this was generally not the case for most verification sites. Most verification sites exhibited the largest differences between the control forecast and the three NLDASE forecasts. Figure 4 evaluates the four Eta retrospective forecasts in the small town of Burnet, Texas, located approximately 70 km northwest of Austin, Texas. Small differences exist between the three NLDASE forecasts. The control run differs largely from the NLDASE runs, particularly with respect to 2 m relative humidity forecasts. Throughout most of the forecast, the NLDASE runs verified much better than the control run with respect to temperature. The same was true for the relative humidity forecast values. The observations indicated that a dryline passed through this site and then retrograded westward between 12 Z on May 5th and 00 Z on May 6th. The NLDASE runs more effectively capture the intensity of the initial dryline passage through this particular site and several others during this time period. However, all four forecasts failed to retrograde the dryline westward during this period.

As discussed previously, the NLDASE land surface states showed general improvement in the surface temperature and relative humidity forecasts in most of the verification regions. A majority of the sites examined in this study featured behavior similar to the verification presented in Figure 4, with only very small differences existing between the NLDASE initialized forecasts and the control forecast. When large differences did exist, such differences were most often between the NLDASE runs as a whole and the control forecast. Only in specific regions (e.g. Eastern and Western Canada, Figure 3) did the NLDASE forecasts differ greatly from one another. The regional verification statistics depicted in Cosgrove and Alonge (2005) are robust and account for a very large number verification sites. However, many of the verification sites exhibit little differences between all four forecasts, and these small magnitude differences can overwhelm the larger impacts at a smaller number of sites due to the area averaging done in the FVS system. As such, differences in bias and RMSE were often small for many of the FVS verification regions. These site specific verification efforts have shown that the surface temperature and humidity forecasts can be quite different at times. These large differences illustrate strong impacts on the forecasts that are sometimes



00-24 hr

24-48 hr

Figure 5. Precipitation differences for the sum of the 0-24 hour, and 24-48 hour precipitation forecasts. Forecast differences are between the control run and LIS1 (LIS1-control).

averaged out in the regional verification statistics.

3.2. *Individual Eta Precipitation Forecast*

The regional FVS precipitation verification provided a great deal of information on precipitation bias, probability of detection, false alarm ratio, and equitable threat scores. This verification system gauged model forecast performance in 24 hour blocks of accumulated precipitation, and missed many details associated with short lived precipitation events. While precipitation differences are large when evaluating the forecast in this manner (Figure 5), details of the timing and intensity of individual precipitation events are lost due to individual events being summed together into 24 hour bins of precipitation. This section will highlight an individual event within a single forecast where timing, placement, and intensity were important to the accuracy of the precipitation forecasts. The forecasts examined are the same as those presented in the previous section.

Figure 6 highlights the Stage II radar/gauge hourly precipitation product on 12 Z May 4th, and hourly precipitation forecasts valid at the same time for the control and LIS1 simulations. Also depicted is the difference between the two Eta forecasts (LIS1 – control). There exists a noticeable difference in precipitation between the two forecasts in central Missouri (4 mm). This is associated with a small shift in both timing and intensity between the two forecasts. The LIS1 forecast placed the maximum precipitation in Missouri further westward and also yielded higher precipitation amounts than the control simulation. Figure 7 is a time series of observed versus modeled rainfall at St. Louis, Missouri and shows that control run handled the overall timing of the precipitation better than the NLDASE runs. However the NLDASE runs better

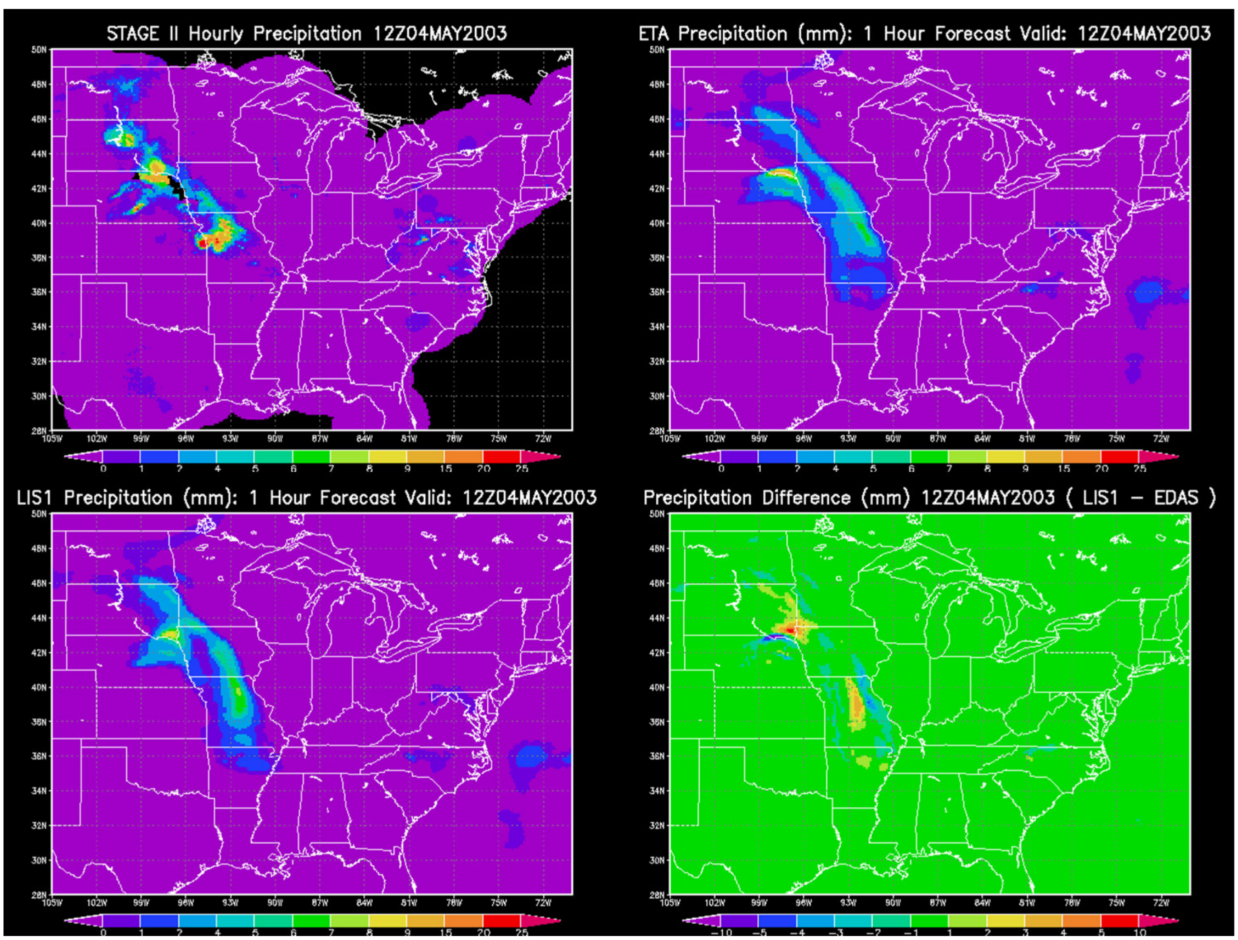


Figure 6. Stage II, control, and LIS1 precipitation amounts (mm) and differences between LIS1 and the control forecast (LIS1 – control) valid the hour ending 12 Z May 4th, 2003.

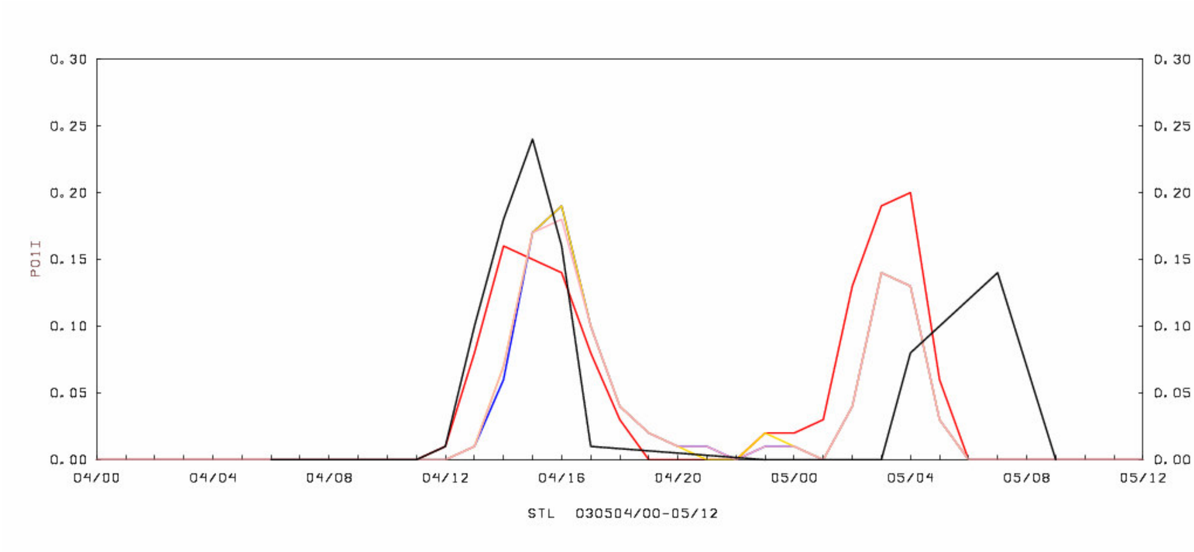


Figure 7. Precipitation time series for St. Louis, MO (inches). Observed precipitation is plotted in black, LIS1 forecast in blue, LIS2 forecast in yellow, LIS3 forecast in pink, and the control run in red.

predicted the magnitude of the maximum rainfall values at this particular location. As seen in the site-specific surface temperature and humidity forecasts, the NLDASE runs exhibited a tendency to be similar to one another in the precipitation forecasts (Figure 7). Overall, this event was well forecasted by both the control and NLDASE initialized forecasts, but the NLDASE runs more accurately maximum amount of the precipitation during this event.

In general, the NLDASE forecasts were very similar to one another and at times differed greatly from the control precipitation forecasts. This was particularly true later in the forecasts as the NLDASE simulations diverged further from the control forecast (Figure 5). The precipitation example presented here highlighted the many differences that can occur not only at smaller temporal scales, but within FVS verification regions as well. These significant differences in precipitation cannot be demonstrated within the regional forecast statistics presented earlier. The regional statistics group forecasts into 24 hour blocks and average them over several forecasts. FVS analysis results in very robust evaluations of long term average model skill, but the example shown above illustrates the large differences that can occur between individual forecasts.

The positive and negative aspects of the forecast illustrated above shows that individual forecast details must be considered when referencing different sources of forecast guidance. For example, the Eta forecasts initialized with NLDASE land surface states showed an overall decrease in precipitation forecast skill throughout the Midwest in the regional precipitation verification statistics; however, when the individual precipitation event that occurred in St. Louis on May 4th, 2003 (Figure 7) was analyzed as discussed above, the 24-36 hour NLDASE forecasts were to superior to the control run forecasts in terms of forecasting the heaviest precipitation amounts.

3.3. *Surface Radiation Forecast Evaluation*

The NCEP FVS was designed to validate surface and upper air meteorological fields and does not evaluate surface radiation fluxes. The same four sample forecasts used in the prior two sections were validated against downwelling/upwelling shortwave and longwave radiation data from the SURFRAD network (Figure 8) to assess the impact NLDASE land surface states have on Eta model forecasts of these fields. This is not intended to be an exhaustive evaluation of Eta model surface radiation fluxes, and serves the purpose of highlighting some of the impacts that uncoupled land surface state initialization can have on these forecast fields.

Large differences in surface radiation fluxes emerge from the use of NLDASE land surface states in the Eta model. Figure 8 shows the differences between the control run and the three NLDASE runs at forecast hour 30, valid at 18 Z on May 4th, 2003 over the continental United States. Large differences exist in many areas (> 300 Wm^{-2}), although differences between all four forecasts were small at most of the SURFRAD observation locations (circles) during this time.

Figure 9 shows a time series of downwelling/upwelling shortwave and longwave radiation fluxes at the Table Mountain, Colorado, SURFRAD site. Differences between the control forecast and NLDASE forecasts were generally very small. All four Eta forecasts displayed large errors throughout the forecast period in all of the surface radiation fields. Downwelling shortwave radiation was severely underestimated throughout a significant portion of the forecast. Conversely, downwelling longwave radiation was overestimated by all of the Eta model forecasts throughout a large majority of the forecast period. Errors in upwelling longwave and shortwave radiation were not as severe as the downwelling fluxes. Overall, the Eta model did a generally poor job at simulating surface radiation fluxes at this site during the forecast period.

Overall, the NLDASE initialized forecasts of longwave and shortwave radiation were very similar to the control forecasts throughout all of the verification sites. In general, all four forecasts performed poorly in the simulation of downwelling shortwave radiation, with errors at times exceeding 500 Wm^{-2}. However, the forecasts performed well under clear sky conditions (not shown). Large differences occurred between the control forecasts and the NLDASE forecasts across much of

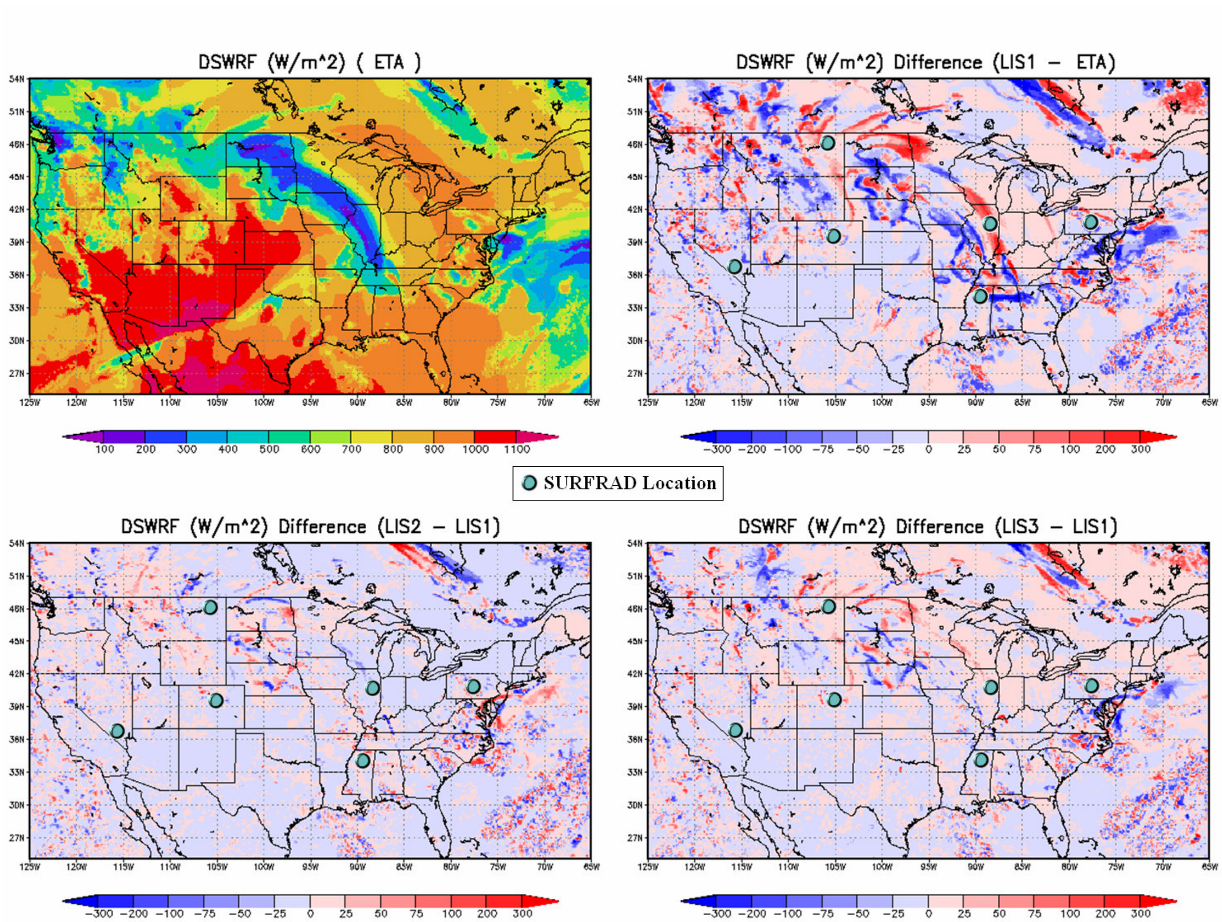


Figure 8. 30 hour forecast of downwelling shortwave radiation for the control forecast and differences between the control and LIS1 (LIS1-control), LIS2 and LIS1 (LIS2-LIS1), and also LIS3 and LIS2 (LIS3-LIS1) valid on 18 Z May 4th, 2003. Circles indicate SURFRAD sites.

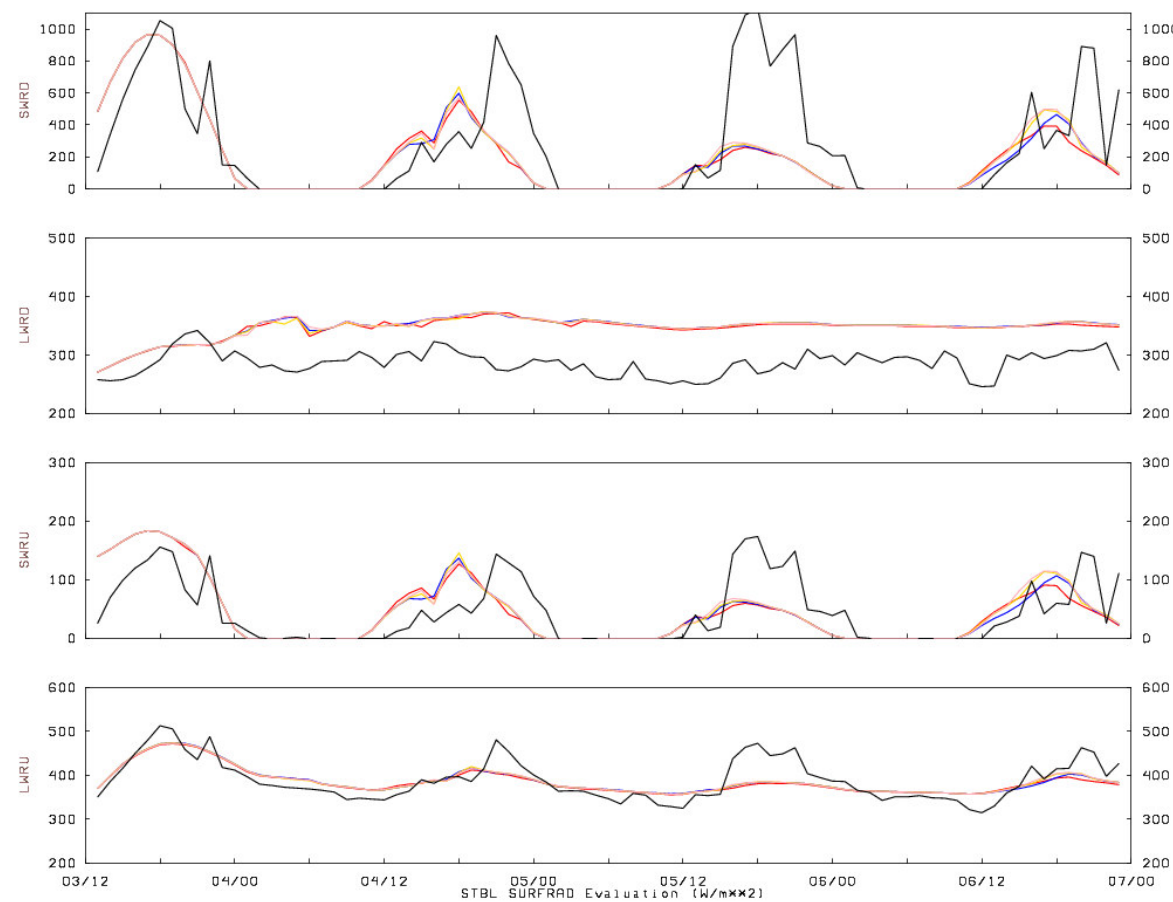


Figure 9. Verification time series of surface downwelling shortwave radiation (SWRD), downwelling longwave radiation (LWRD), upwelling shortwave radiation (SWRU), and upwelling longwave radiation (LWRU) fluxes (W/m^2) at the Table Mountain, Colorado (STBL) SURFRAD site. Observations are plotted in black, the control ETA forecast in red, LIS1 forecast in blue, LIS2 forecast in yellow, and LIS3 forecast in pink.

the domain (Figure 8), but the location of the SURFRAD validation sites failed to coincide with these differences. Therefore, more surface radiation sites need to be examined to fully gauge the impact of the NLDASE offline land surface states, particularly in regions where MODIS snow cover assimilation may have had a large impact.

4. SUMMARY AND CONCLUSIONS

The regional FVS benchmarking activities indicated the initialization of Eta land surface states with NLDASE output had a mixed impact on forecasts. Surface fields including 2m temperature and 2m humidity greatly benefited from the uncoupled initialization process, while the upper air and precipitation fields featured a mix of desirable and undesirable impacts.

The FVS benchmark metrics are frequently used by NCEP to evaluate Eta model performance, and provided useful information regarding the forecast improvement/degradation due to the inclusion of NLDASE land surface states. However, the broad nature of these metrics fails to represent the impact that the forecast improvement/degradation may have on a single location. The site specific verification activities presented herein highlight the impacts that NLDASE initialization has on site-specific ETA forecasts of temperature, relative humidity, precipitation, and radiation. The examples presented here show significant improvements in forecasts of 2 m temperature and relative humidity and precipitation.

However, there existed several sites where the inclusion of NLDASE land surface states degraded the Eta forecasts of these fields. Examples of forecast degradations will be presented on the poster which accompanies this preprint. General conclusions of the on going site specific verification activities are given here:

- In general, the largest differences in forecasts occurred between the NLDASE simulations as a whole and the control forecast.
- Mean sea level pressure forecasts (not depicted above) were only slightly impacted by NLDASE initialization.
- Inopportune location of SURFRAD observation sites failed to capture the large differences in control run and LIS1, LIS2, and LIS3 radiation forecasts present across domain.
- In general, downwelling shortwave radiation was poorly forecast in all simulations during cloudy conditions, while upwelling and downwelling longwave radiation fluxes were well forecast in all simulations.
- In some cases, the timing of dry-lines and fronts was impacted (both positively and negatively) by the use of NLDASE land surface states.
- Short term forecasts were impacted less than long lead (> 48hr) forecasts.
- In general, precipitation distribution was impacted less than precipitation timing and intensity.
- NLDASE-based forecasts featured improved precipitation magnitude or timing at some locations and degraded precipitation timing or magnitude at others.

5. ACKNOWLEDGEMENTS

This research was supported through the GAPP program through funds from the NASA Terrestrial Hydrology Program.

6. REFERENCES

Baldwin, M., and K. Mitchell, 1997, The NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research, Preprints, 13th AMS Conf. on Hydrology, Long Beach, CA, 54-55.

Black, T. D., 1994: NMC NOTES, The New NMC Mesoscale Eta Model: Description and Forecast Examples. *Wea. Forecasting*, **9**, 265-278.

Cosgrove, B.A., and C. J. Alonge 2005: Recent results from the 12 km North American land data assimilation system, Preprints 86th AMS Annual Meeting, Atlanta, GA.

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 NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**(D22), 8851, doi:10.1029/2002JD003296.

Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr, 2002: MODIS snow-cover products. *Remote Sens.Environ.*, **83**, 181–194.

Kumar, S. V., and Coauthors (2005), LIS - An Interoperable Framework for High Resolution Land Surface Modeling. Environmental Modeling and Software, in press.

Mitchell, K. E., D. Lohmann, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, D. P. Lettenmaier , C. H. Marshall, J. K. Entin, M. Pan, W. Shi, V. Koren, J. Meng, B. H. Ramsay, and A. A. Bailey, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, doi:10.1029/2003JD003823.

Robock, A, and Coauthors, 2003: Evaluation of the North American Land Data Assimilation System over the Southern Great Plains during the warm season. *J. Geophys. Res.*, **108** (D22), 8846, doi:10.1029/2002JD003245.

Rodell M., P. R. Houser, 2004: Updating a land surface model with MODIS-Derived snow cover. *J. Hydrometeor.*, **5,** 1064 -1075.