EUROPEAN COMMISSION

RESEARCH DG

MARIE CURIE MOBILITY ACTIONS

REINTEGRATION ACTIONS
PERIODIC SCIENTIFIC/MANAGEMENT REPORT

Project no. 517581
Project acronym XFLOOD
Project full name
Advancing Quantitative Precipitation Estimation and Short-to-Medium Range Forecasting on the Basis of Remotely Sensed Data Assimilation

International Reintegration Grants (IRG)

Deliverable No 3

Model improvements and the development of the data assimilation schemes

by

Anastasios Papadopoulos, Emmanouil Anagnostou,

Ioannis Vamvakas and Efthymios Serpetzoglou

Period Number: 2

Due date of deliverable: 15/04/2007

Period covered: from 15/04/2006 to 14/04/2007

Date of preparation: 21/05/2007

Date of submission: 21/05/2007

Start date of project: 15/04/2005

Duration: 24 months

Project coordinator name: Dr. Evangelos Papathanassiou
Project coordinator organisation name: Hellenic Centre for Marine Research

Organisation name of lead contractor for this deliverable: Hellenic Centre for Marine Research

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)
Dissemination in level PU (Public)

1. INTRODUCTION

Rapidly developing mesoscale convective systems (MCS) are usually responsible for the heaviest and most destructive rainfall and flood events in the Mediterranean region. Rainfall from an MCS is spatially heterogeneous such that convective cells (typically associated with lightning activity) would produce areas of intense rain rates imbedded within larger precipitation regions. This pattern can persist over several hours and cover a large area producing large accumulations of rain and often result in flash floods and other problems of surface water drainage. Such problems are exacerbated in urban areas where run-off occurs more rapidly and interference in human activity is greatest. To moderate these hazards, systems have been developed for issuing short-range (up to 48 hrs) quantitative precipitation forecasts (QPF) on the basis of numerical weather prediction (NWP) models and observational inputs (e.g., synoptic observations, surface and remotely sensed data). An important component of QPF is the simulation of deep moist convection processes and the consequential precipitation. These moist processes are represented by convective parameterization schemes. Significant effort has been devoted in developing such schemes over the past decades (e.g., Tripoli and Cotton 1982; Betts and Miller 1986; Kain and Fritsch 1990; Janjic 1994). At the resolutions typically used by regional NWP (grid increments ranges from 10 to 50 km), numerical models exhibit low skill at forecasting highly variable convective precipitation events. Two error sources have long been recognized: (1) the lack of sufficient data and coarse grid resolutions that limit our ability to correctly specify those features acting to trigger convection in model initialization (e.g., Kain and Fritsch 1992), and (2) weak assumptions used in developing the convective parameterization schemes (e.g., Mellor and Yamada 1982). Clearly, the high non-linearity in modeling atmospheric processes causes sensitivities to even small perturbation in the atmospheric initial state, the quality of which dictates the accuracy of QPF. An aspect that has shown convincing signs of improving numerical weather prediction is data assimilation. Data assimilation aims at obtaining the optimal initial state of the atmosphere in order to define a better starting point for model integration.

Collective evidence from past studies indicates that improved initialization of the local environment in NWP models (particularly this of the atmospheric moisture distribution) is an avenue for potentially improving QPF as it relates to simulated convective precipitation. Brenner (2004) have shown high correlation between atmospheric humidity profiles and convective rainfall patterns for summertime storms. Other studies, like the one by Gallus and Segal (2001), have shown that improved representation of mesoscale features in model initialization may benefit the forecasts. More specifically, this study showed that the greatest positive impact in model forecasts was achieved by adding water vapour in relatively dry regions of the model where radar echoes were present. In an earlier study, Rogers et al. (2000) was able to improve QPF in a warm season environment by forcing the model convective parameterization scheme (CPS) to trigger (or suppress) convection when and where it was observed according to radar reflectivity data. To this end, continuous lightning observations may provide wealth of information about the growth, location, lifecycle and ice microphysics of convection in MCS regimes. Charge separation leading to lightning is a physical process that takes place in regions of a thunderstorm associated with rigorous vertical forcing. The sound connection of lightning occurrence with the existence of strong updrafts (>10 m/s) and graupel growing from super-cooled cloud water makes the continuous measurement of this meteorological parameter potentially useful for improving the prediction of storm evolution and intensification. Alexander et al. (1999), assessing the impact of assimilating different rain rate retrievals in model forecasts, examined retrievals based on satellite only versus combining with continuous lightning data. Using these data they continuously assimilated latent heating profiles into numerical simulations of the 1993 Superstorm. They found that the

improvement in forecasts of precipitation patterns, sea level pressure fields, and geopotential height fields was more pronounced when information from all data sources was combined, concluding that lightning data had a greater positive impact than the other data sources. Chang et al. (2001) demonstrated further improvements in quantitative precipitation forecasting of extratropical cyclones on the basis of data assimilation using lightning-based rain estimates as gap-filling information. The challenge, however, remains as to how useful lightning information can be as a sole data source to directly nudge model fields such as local environmental properties.

Arguably, developing a way to modify model-generated variables (e.g., atmospheric specific humidity) on the basis of lightning intensity could result in more accurate parameterizations of the convective environment. Making use of the sound link between lightning occurrence and convection, a technique was recently developed by Papadopoulos et al. (2005) to amend the initialization data inadequacies and limitations in formulating sub-grid scale processes using only continuous lightning observations. The Papadopoulos study showed that nudging the model's humidity profiles to an empirical profile related to flash rates, leads to more realistic model soundings and consequential improvements in convective precipitation forecasts.

Another avenue for improving the convective precipitation forecasting skill is through a better estimation of the energy and moisture exchanges between land surface and the atmosphere. Soil moisture is known to play an important role in the partitioning of energy between sensible and latent heat fluxes as well as in driving the total moisture budget and temperature variations of the lower atmosphere (e.g., Clark and Arritt 1995; Betts 2004). Koster and Suarez (2001) found that soil moisture is an important source of forecast skill for the predictability of precipitation, and Zhang and Frederiksen (2003) have shown that the initial soil moisture conditions supplied to a NWP model affect both temperature and precipitation forecasts. Several other studies have demonstrated that soil moisture strongly influences the boundary layer dynamics. However, the role of soil temperature in the evolution of the lower atmosphere should not be underestimated. For example, the outgoing longwave radiation is a function of soil temperature and directly affects the surface radiation budget. Also, the ground heat flux depends on soil temperature and soil moisture conditions, as well as vegetation coverage, atmospheric conditions, and the thermophysical properties of the soil. Apparently, the characterization of the spatial and temporal variability of soil moisture and soil temperature fields is a critical factor in formulating and modifying the mesoscale atmospheric circulations. Even though efforts are under way to provide more extensive networks of soil moisture and soil temperature data from a variety of remote sensing and direct observational sources (Leese et al. 2001; Seuffert et al. 2004), routine observations of these quantities suitable for data assimilation are currently unavailable over large areas of the world. Thereby, Land Surface Models (LSMs) are considered as valuable tools for the definition of the land surface conditions that are required in the atmospheric NWP models.

The accuracy of the performance of the LSMs depends on the quality of (1) the initial conditions of the soil state (i.e., soil moisture, soil temperature and snow cover), (2) the atmospheric forcing (i.e., precipitation, radiation budget, air temperature, wind, humidity, and pressure), (3) information of the topography, soil properties and land cover, and (4) the internal model parameterization schemes. Commonly, the atmospheric forcing terms are provided by the NWP models. However, this coupling procedure is subject to substantial model biases and errors that may negatively impact the quality of their output due to the positive feedbacks caused by the non-linear land-atmosphere interactions. This has motivated the development of offline Land Data Assimilation Systems (LDAS), like the North American Land Data Assimilation System (Mitchell et al. 2004). This system requires

information such as precipitation, downward shortwave and longwave radiation, near surface air temperature, humidity, and wind speed. Precipitation and solar radiation are considered as the most important forcing terms due to their significant impact on the water and energy budgets. Water enters the land surface system through precipitation and leaves through evapotranspiration, which is mostly driven by incoming solar radiation. Other variables such as near surface humidity, near surface air temperature and wind speed also affect evapotranspiration, but their role is less important (Cosgrove et al. 2003). Therefore, the success of LDAS relies on the accuracy of the estimated precipitation and radiation forcing fields.

In this report, we present the potential of improving the convective precipitation forecasting skill of the POSEIDON weather forecasting system (Papadopoulos et al. 2002) through a lightning data assimilation technique and through an offline physically-based LDAS driven by surface forcing anchored by estimated rainfall fields.

2. LIGHTNING DATA ASSIMILATION

2.1 The Papadopoulos et al. assimilation technique

As discussed above the Papadopoulos et al. scheme makes direct use of lightning data to indicate areas of convection in a model's domain. Based on that information the model humidity profiles are nudged relative to the observed flash rates. The adjusted humidity profiles are then used to compute, based on the model's own convective parameterization scheme (MCPS), heating rate profiles that are more compatible with the local scale convective environment. The technique starts by projecting the flash rate data on the model's spatio-temporal grid structure. Convective areas are then delineated from pixels with flash rate exceeding a certain threshold (in Papadopoulos et al. that threshold is set to 2 and 10 flashes/15min over sea and land, respectively) and compared to model estimations. Four possible scenarios are considered. The first scenario concerns model grid cells where lightning observations indicate that there is no significant electrical activity. In this case the MCPS is allowed to carry on, as it would normally do without nudging. The other three scenarios are associated with grid cells where flash rate data are above the threshold denoting convection. These scenarios are then referred to as deep, shallow or no-convection, according to the model output. Subsequently, a nudging technique is applied into the MCPS following the decision-making procedure showed in Figure 1.

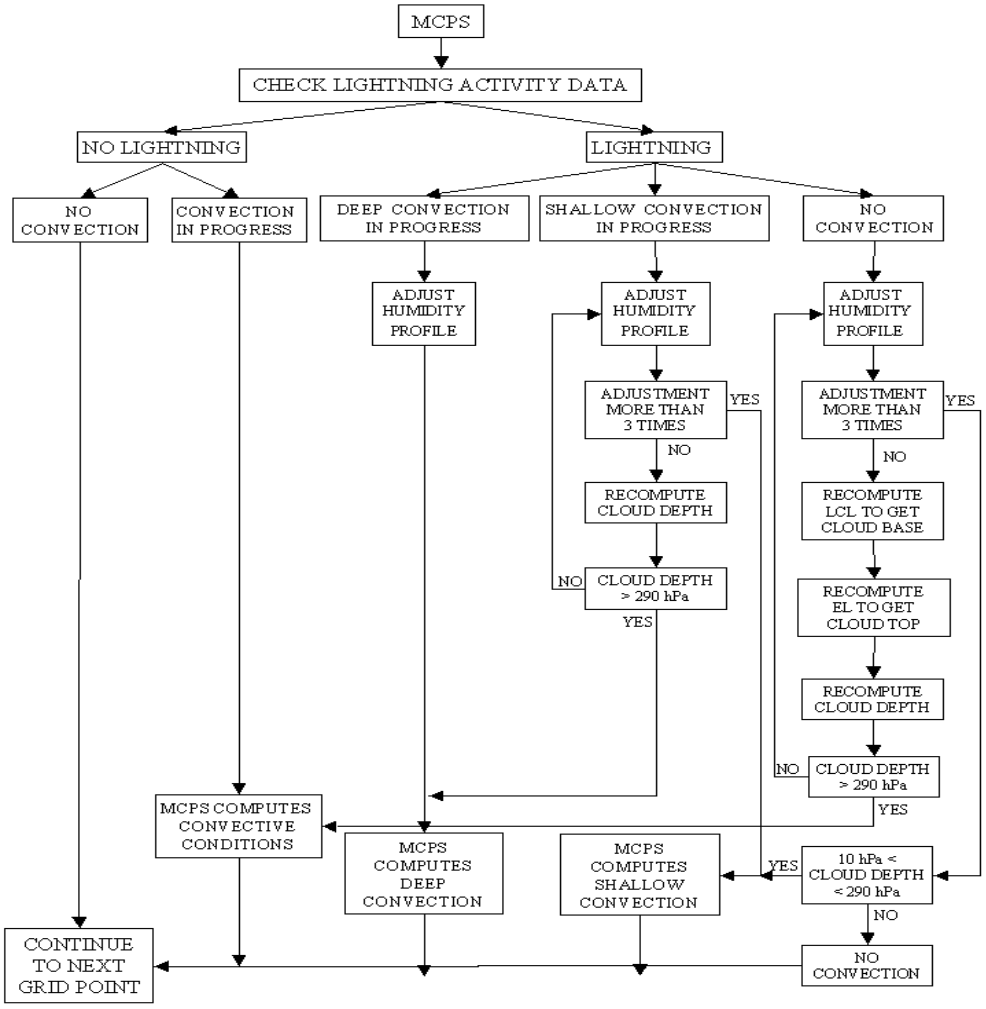

Figure 1: Flow chart of the assimilation technique (adapted from Papadopoulos et al. 2005).

A short description of how the nudging is applied is presented below. In the scenario where the model produces deep convection, the humidity profile within the deep convective cloud is scaled to an empirical profile as:

$$q_l^{ass} = q_l^{mod} + F \cdot (q_l^{emp} - q_l^{mod}), \qquad q_l^{emp} > q_l^{mod}, \qquad (1)$$

where q_l^{mod} is the model-simulated specific humidity at pressure level *l*, q_l^{emp} is the empirical specific humidity value at corresponding level, and q_l^{ass} is the specific humidity value after nudging. The adjustment is applied only when predicted values are below the empirical profile values. The nudging term *F* is a dimensionless number that in Papadopoulos et al. was taken equal to the observed flash rate from the 15-min interval containing the model's integration time. An important aspect of this approach is the evaluation of q_l^{emp}. In Papadopoulos et al. q_l^{emp} was determined in a simplistic way through trial and error having as criterion maximizing the assimilation impact on QPF. In the case of shallow or no convection predicted by the model's convective parameterization scheme an iterative procedure is

applied. First, the simulated humidity profiles are scaled to the same empirical humidity profile (q_l^{emp}) using the above stated equation. Subsequently, cloud depth is recomputed on the basis of q_l^{ass} profile; if the convective cloud exceeds 290 hPa in depth, then deep convection is assigned, otherwise moisture is added by nudging again the humidity profile on the basis of the above equation. The iterative procedure is set to repeat up to three times. If after three iterations a sufficiently deep cloud is not produced, the model's convective parameterization scheme is set to compute shallow or no convective effects at that model grid cell, based on the new model environment. Because of the potential artifacts associated with any kind of nudging, special care is given to prevent numerical noise and unrealistic results. A large nudging term would rapidly force the model environment towards the observed fields leading to an uncontrollable state. This may be due to the decreased ability of the model to control mass conservation as well as incompatibilities in the mesoscale features caused by artificial changes in the model profiles. To moderate such effects the following constrains are imposed: (1) the nudging term *F* cannot exceed an upper threshold value corresponding to 50 (25) flashes per 15-minute time intervals over 0.1x0.1km^2 grid area, respectively over land and sea; (2) the increase in specific humidity per time step (or iteration) is limited to a maximum value of 0.5 g kg^{-1}; and (3) the humidity value after nudging cannot exceed the model's saturation humidity value for that pressure level. These constraints are intended to minimize the possibility that precipitation enhancement occurs because of the artificial saturation of the water vapour content, and not due to the model environment favouring an impending condensation. After any adjustment the latent heating profiles are computed to ensure the closure of the model's convective parameterization scheme.

2.2 Improving the Papadopoulos et al. assimilation technique

The goal of this work was to further develop the Papadopoulos et al. lightning assimilation technique. Our goal was achieved through the following steps:

1. Implementation of the technique in the latest version of the POSEIDON weather forecasting system. The POSEIDON system is in operational use since 1999 providing daily 72-hour forecasts for the Mediterranean region. The results are available at http://www.poseidon.ncmr.gr. When lightning data are available through this system and suitable for operational use, we will have the unique opportunity for the potential development of the integrated system.
2. Investigation of the possible dependence of the nudging terms on the different sources of lightning data retrievals. For example, the nudging coefficient *F* may have some dependence on the lightning network's cloud-to-ground (CG) detection efficiency.
3. Investigation and improvement of the technique in order to be used in different climate regimes. For example, the empirical humidity profiles may vary according to the region where the model is applied. We have to consider that the empirical profiles were determined empirically (through trial and error analysis) on the basis of maximizing the assimilation impact on QPF accuracy.

One important aspect of this work was to incorporate the related numerical sub-modules of the data assimilation technique in a more robust scheme. Thus, the assimilation methodology has been designed and written in terms of computer code structure to make it suitable for easier implementation in future versions of the POSEIDON weather forecasting

system and generally in any mesoscale meteorological model. The code has already been implemented and tested in the latest version of the POSEIDON weather forecasting system. The storm cases that are presented in Papadopoulos et al. (2005) have been reproduced and gave similar results.

Figure 2 demonstrates an example of the performance of the technique for a case of warm rain activity in Europe.

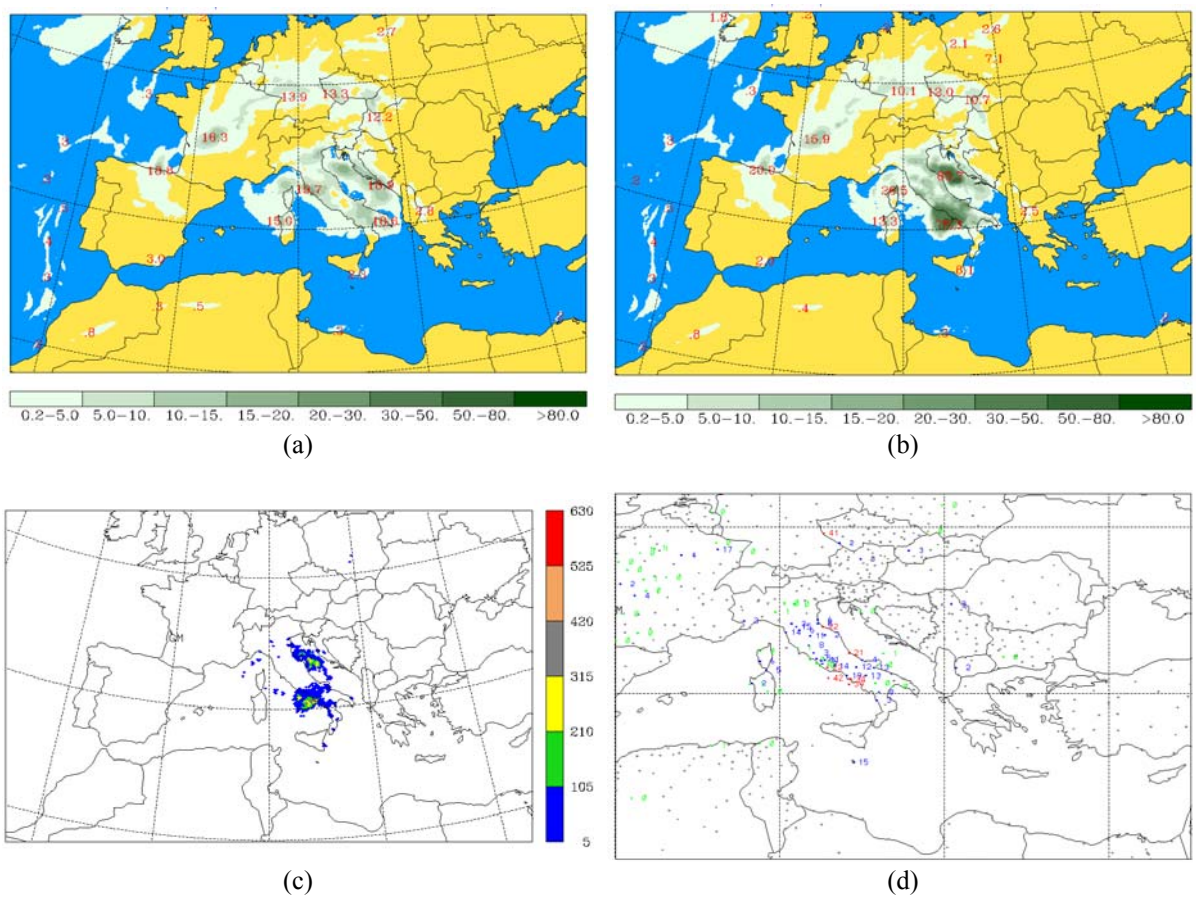

Figure 2. Rainfall fields (mm/6hr) from (a) the control run in which no lightning data were assimilated, (b) the simulation applying the Papadopoulos et al. (2005) assimilation technique, (c) the observed lightning activity and (d) ECMWF rain gauge data for the period from 06 to 12 UTC at 28 August 2002.

The simulations were performed for 48 hours with 0000 UTC initialization time. In the vertical, 32 levels were used stretching from the ground to the model top (15800 m). In the horizontal, a grid increment of 0.10-degree was applied using a time step of 36 seconds. For the initial meteorological conditions (geopotential height, wind components and specific humidity), the ECMWF reanalysis gridded data on a 0.5-degree horizontal grid increment at 11 standard pressure levels (1000, 925, 850, 700, 500, 400, 300, 250, 200, 150 and 100 hPa) were interpolated at the model grid points using optimal interpolation analysis. The boundary conditions were linearly interpolated at each model time step from the ECMWF data available every 6 hours. To define the initial state of the sea surface temperature (SST) field, as well as the soil temperature and moisture availability at the six model soil layers, the ECMWF 0.5x0.5-degree gridded data were used. In the current configuration the soil layers were defined at the depths of 5, 15, 28, 50, 100, and 255 cm. For each model grid point we

calculated topography, soil and vegetation types, slopes and azimuths of the sloping surfaces, utilizing high resolution datasets. Namely, the topographic dataset and the vegetation dataset both available from USGS at 30x30 arc sec resolution, and the UNEP/FAO dataset after being converted from soil type to soil textural ZOBLER classes at 2x2-min resolution (Papadopoulos et al. 1997).

In the assimilation technique the lightning data are CG lightning observations obtained by the ZEUS regional network covering Europe and Africa. The current locations of the ZEUS receivers are shown in Figure 3. These sites, selected for their low man-made electric noise interferences, include: Birmingham [UK], Roskilde [Denmark], Iasi [Romania], Larnaka [Cyprus], Evora [Portugal], Addis Ababa [Ethiopia], Dar e Salaam [Tanzania], Bethlehem [South Africa], Lagos [Nigeria], and Dakar [Senegal]. Each receiver samples the time series of the vertical electric field of sferic waveforms. The system retrieves the location of CG lightning activity occurring over a very large region (Europe, North Africa and part of the Atlantic and West Asia) based on the arrival time difference (ATD) method (Lee 1986a;b). An error analysis of ZEUS lightning location retrieval on the basis of simulation and through in-situ validation data from Spain's regional lighting network has shown that the lightning location error in Europe varies from 5 to 25 km with a mode at 15 km (Chronis and Anagnostou 2003). Within the periphery of the network's receivers the system's detection efficiency is expected to range between 70% and 90%.

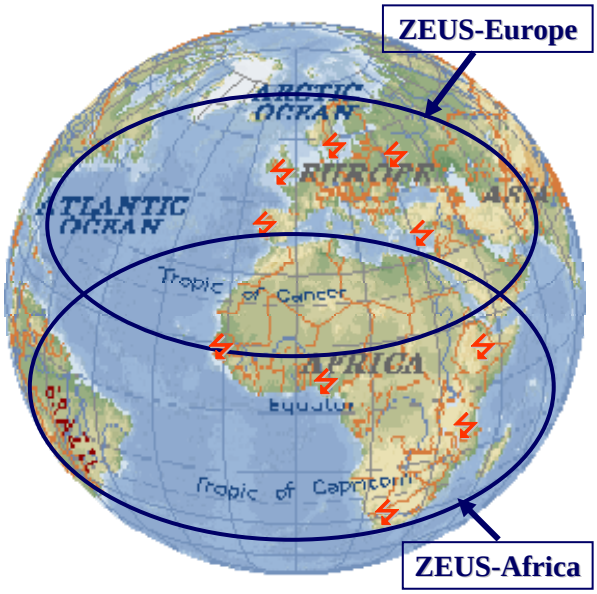

Figure 3. The current locations of the ZEUS network receivers.

Special software has been developed for projecting the flash rate data on the model spatio-temporal grid structure. Namely, CG lightning data as recorded by the ZEUS system are gridded at selected resolution and then projected from the mercator to the Eta coordinates. For the specific needs of this work we selected the 0.1x0.1-degree spatial and 15-minutes temporal resolutions. Applying a linear interpolation between sequential time intervals, flash rates are then interpolated to the corresponding time step of the model integration.

To meet our second objective we investigated the possible dependence of the technique on the source of the lightning data. For this purpose we performed numerical experiments for a storm case in the US utilizing data from the US National Lightning Detection Network

(NLDN). The POSEIDON weather forecasting system was set up to work in an area covering the entire continental US and surrounding waters. It is important to note that the assimilation technique has been applied using the same empirical humidity profiles as those used in the simulations over the Mediterranean region. In both the control (CTRL) and the lightning-nudged (CASE) model runs we used the ECMWF reanalysis fields to define the initial and boundary conditions. For the definition of the physiographic characteristics of the region we used the same datasets (utilizing the appropriate tiles) as for the operational POSEIDON weather forecasting system. The model output was compared against reference rainfall fields derived from a mosaic of hourly raingauge-calibrated WSR-88D rain estimates.

To verify the simulated precipitation at the various validation points, precipitation from the model grid was interpolated to each radar pixel using bilinear interpolation:

$$M = \frac{\sum_{k=1}^{4} w_k \cdot M_k}{\sum_{k=1}^{4} w_k} \qquad (2)$$

where M_k are the values at the four model grid points surrounding the observation, while the weighting factor w_k is taken as the reverse squared distance, thus making nearer points more influential.

The verification scores used in this work are derived using the contingency table approach (Wilks 1995). This is a two-dimensional matrix where each element counts the number of occurrences in which the observations and the model forecasts exceeded or failed to reach a certain threshold for a given forecast period. The table elements are defined as: A-model forecast and observation exceeded the threshold; B-model forecast exceeded the threshold but observation did not; C-model forecast did not reach the threshold but observation exceeded it; and D-model forecast and observation did not reach the threshold. Considering the above elements the forecast skill can be measured by evaluating the bias score (BS) and the equitable threat score (ETS). The bias score is defined by:

$$BS = \frac{A + B}{A + C} \qquad (3)$$

where BS defines the ratio of the number of occurrences that model forecasts exceed a specified threshold versus the respective number for observations. The ET score is defined as

$$ETS = \frac{A - E}{A + B + C - E} \qquad (4)$$

where E is defined as

$$E = \frac{F \cdot O}{N} = \frac{(A + B) \cdot (A + C)}{N} \qquad (5)$$

where N is the total number of observations being verified (N=A+B+C+D). The introduction of the E term (Messinger 1996) is an enhancement to the normal threat score (as defined in Wilks 1995), acting to reduce it by excluding the number of randomly forecast "hits". Computing the bias and the equitable threat scores, a measurement of the model accuracy on the frequency of occurrences at or above a certain precipitation threshold amount can be revealed. Consequentially, at given thresholds the bias score can represent a systematic overestimation (when BS>1) or underestimation (when BS<1), and the ET score can present not good forecasts (when ETS≈0) or the perfect forecasts (when ETS=1). In past studies where quantitative precipitation forecasts were evaluated (e.g., Messinger 1996; Colle et al. 1999; 2000), ETS values of about 0.1-0.2 for moderate amounts of rainfall (10-20 mm in 6 hours) have been considered adequate for operational forecasts.

Another widely used score for verifying precipitation forecasts is the Heidke Skill Score, HSS (Heidke 1926). It is computed based on the contingency table elements from the expression:

$$\mathrm{HSS} = \frac{2 \cdot (\mathrm{A} \cdot \mathrm{D} - \mathrm{B} \cdot \mathrm{C})}{(\mathrm{A} + \mathrm{C}) \cdot (\mathrm{C} + \mathrm{D}) + (\mathrm{A} + \mathrm{B}) \cdot (\mathrm{B} + \mathrm{D})} \qquad (6)$$

Since the aforementioned statistical measures do not use the magnitude of the precipitation errors, they are not strictly influenced by the variability of forecasting error. To measure the magnitude of the difference between model forecast and observed precipitation we calculate the root mean square error (RMSE) as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{\mathrm{NOBS}} (\mathrm{MP}_i - \mathrm{OP}_i)^2}{\mathrm{NOBS}}} \qquad (7)$$

where MP_i and OP_i are the model-estimated and the observed precipitation, respectively, and the NOBS is the total number of observations at a specific location reaching or exceeding a certain threshold amount. Combining these statistical criteria we attempt to provide a comprehensive evaluation of model performance. For example, a greater ETS will represent a significant model improvement only if it is accompanied by a BS with value close to one and a lowering RMSE. The skill scores are presented for 6-hourly periods. Using rain accumulations from longer periods (e.g., daily) would result in smoothing the local scale details occurring in high-resolution data.

Having this in mind, we are now able to comprehend the meaning of the maps presented in Figure 4. This figure shows (a) the rain accumulations by the WSR-88D estimates, and the rain accumulation forecasts from (b) the CTRL and (c) the CASE simulations, overlaid by contours of lightning intensity for a characteristic 6-hourly period of the storm. First of all, we note that the intense rainfall observed by radar is well correlated with regions of lightning intensity. Then, comparing the CTRL and CASE rainfall fields we observe that the technique enhances convection in regions with lightning while depresses precipitation in other regions. The precipitation enhancement (and depression) on the basis of lightning nudging improves the agreement with radar rainfall. This is also evident when comparing the radar, CASE and CTRL rain accumulation plots: CTRL misses completely the convection developed over New Mexico (the western area of the radar domain delineated by the solid black rectangle). Figure 4d summarizes the BS, ETS, HSS, and RMSE statistical

scores for the respective period, focusing on the afore-mentioned area. It is noted that the strong underestimation of rainfall in the CTRL run (all the threat scores are zero) turns into an acceptable estimate in the CASE run, increasing the BS to a value close to one and reducing significantly the RMSE.

These numerical experiments indicate the significance of lightning in triggering (and/or enhancing) convection in regional/mesoscale models. The most important conclusion, however, is that the nudging terms (more specifically the nudging coefficient *F*) do not depend on the lightning network's CG detection efficiency.

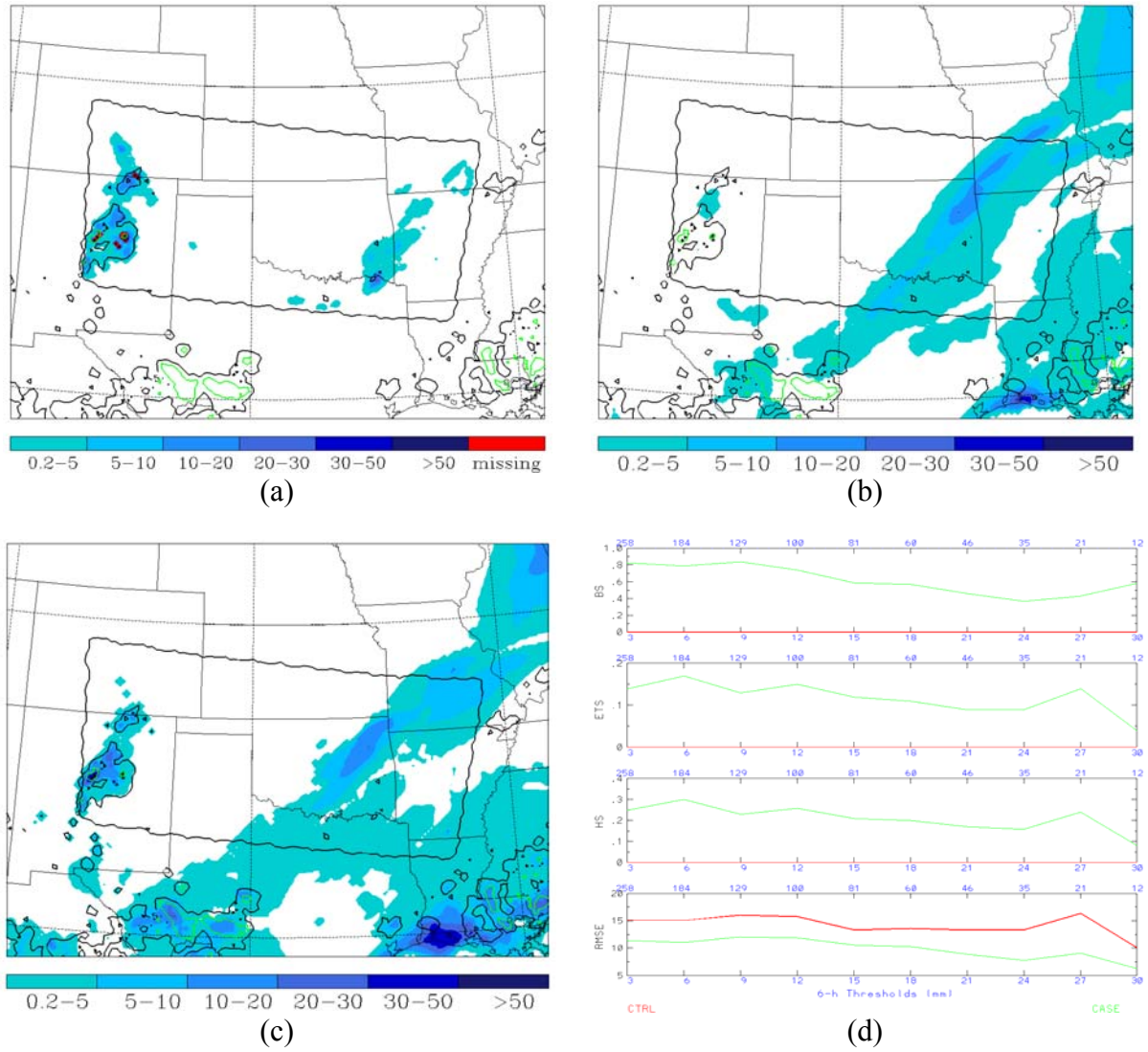

Figure 4. Rainfall fields (mm/6hr) from (a) WSR-88D rainfall estimates (b) the control run (CTRL) in which no lightning data were assimilated, (c) the simulation applying the Papadopoulos et al. (2005) assimilation technique (CASE). The radar domain is shown by solid line, while black and yellow contours delineate areas of low (5 flashes/6-hr/0.1deg) and high (50 flashes/6-hr/0.1deg) lightning intensity. Panel (d) shows the statistical values calculated from the CTRL and CASE experiments against radar rainfall fields. Numbers above each panel denote the total number of observations reaching the corresponding threshold value. The presented data correspond to a 6-hr period ending at 00 UTC of July 30, 2002.

To address the third objective we investigated the performance of the technique over tropical regions. For this purpose, we set up the POSEIDON weather forecasting system over West Africa, using the same datasets as in the previous experiment to define the initial and boundary conditions and the physiographic characteristics, for both the CTRL and CASE runs. The simulation was performed for a period of three days (13-15 July 2004). Model output was compared against reference gridded rainfall fields derived from Tropical Rainfall Measurement Mission (TRMM)-calibrated Special Sensor Microwave/Imager (SSM/I) rain rates (as have been produced in Task 2).

Since it is known that the characteristics of the tropical convective systems are not similar to those in middle latitudes we did not use the same empirical humidity profiles as in the previous experiments. Additionally, we compiled a number of atmospheric soundings released from stations in West Africa and we determined the humidity profile to which the model-generated humidity profiles should be scaled. In Figure 5 we present the humidity profiles we used for mid-latitudes and for tropical regimes.

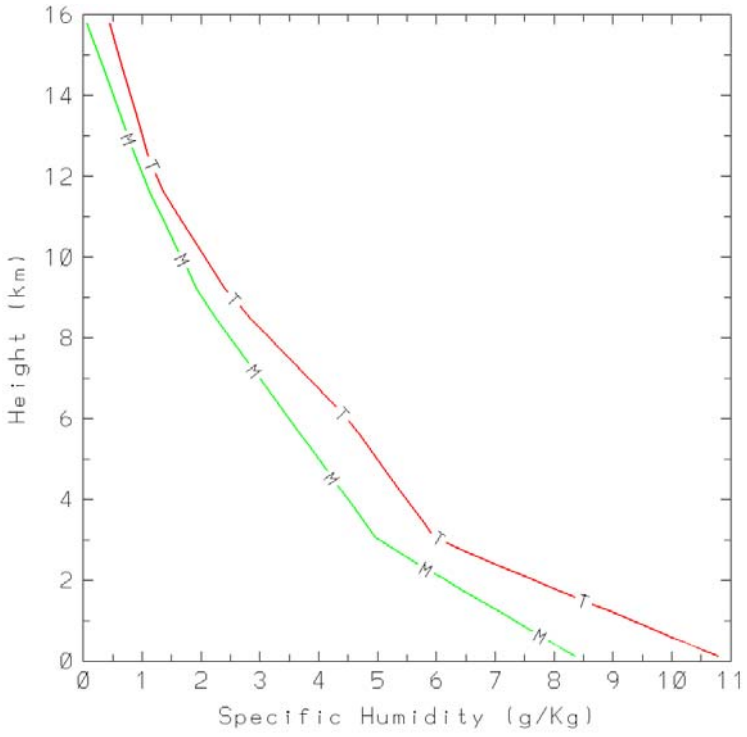

Figure 5. The humidity profiles used for mid-latitude (green line) and for tropical regions (red line).

In Figure 6a we note that the intense rainfall observed by SSM/I is well correlated with regions of ZEUS detected lightning activity. Significant improvement is demonstrated when comparing CTRL and CASE rainfall fields (Figure 6b and c) and the error statistics (Figure 6d).

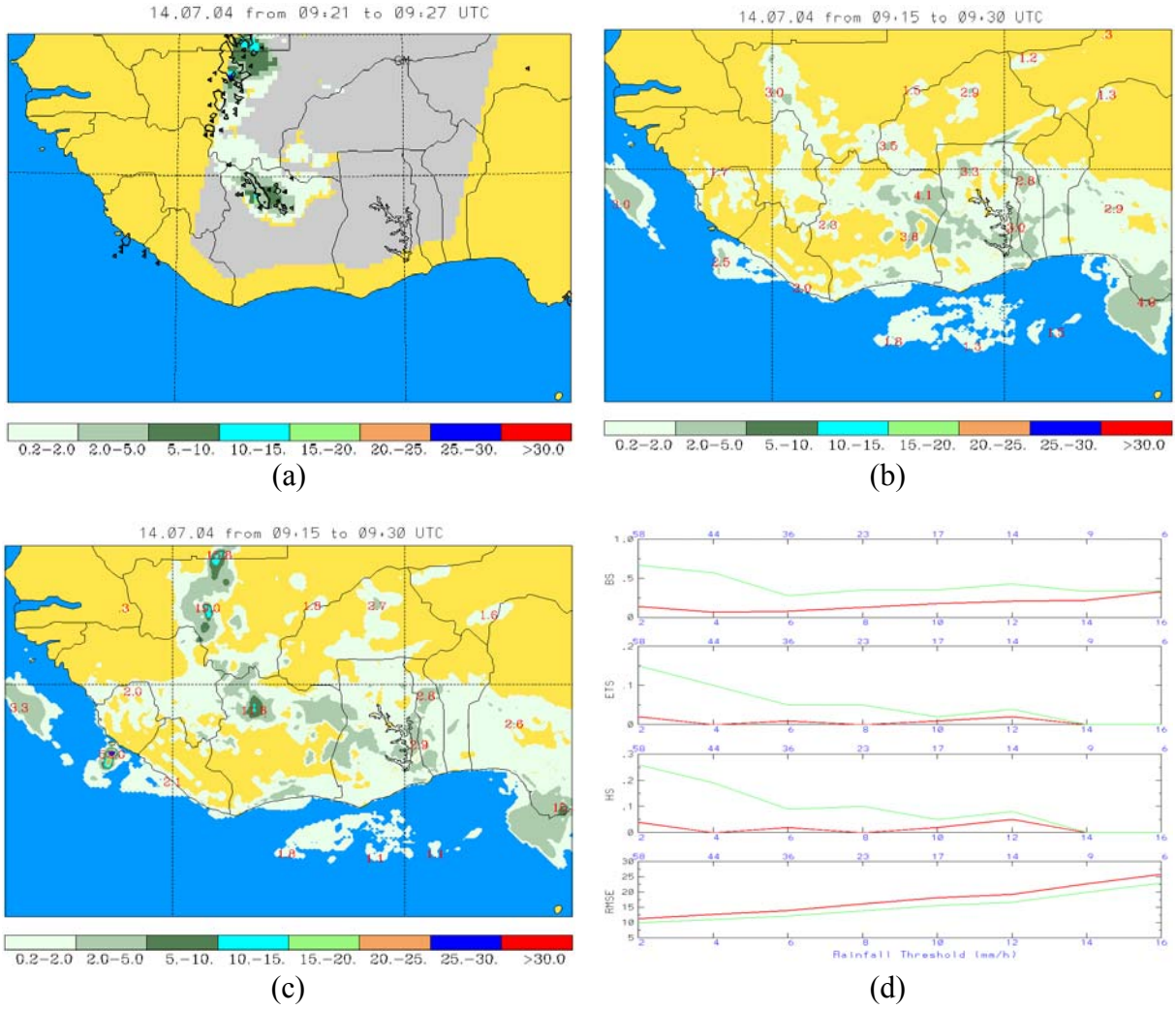

Figure 6. Rainfall fields from (a) TRMM PR-calibrated SSM/I rain rates (contours delineate areas of lightning intensity of 10 flashes/1-hr/0.1deg) (b) CTRL and (c) CASE experiment. The accumulation period for the rainfall fields is denoted on each panel. Panel (d) shows the statistical values calculated by comparing the CTRL and CASE experiments for the entire storm period against TRMM PR-calibrated instantaneous SSM/I rain rate fields. Numbers above each panel denote the total number of observations reaching the corresponding threshold value.

The performance of the technique was originally verified in the Papadopoulos et al. (2005) study on the basis of three major thunderstorms that occurred in the Mediterranean region in the warm seasons of 2002 and 2003. In this study we further developed and evaluated the technique on the basis of storms occurred in three different continents: Europe, Central-West Africa, and continental USA. The statistical evaluation denotes that the regional lightning assimilation can help improve the model's efficiency to simulate convective precipitation, and that this improvement is consistent across different continental regimes. Consequently, the definition of the empirical humidity profiles can have a global applicability.

3. LAND DATA ASSIMILATION SYSTEM

As discussed in the introduction, the accurate description of the surface energy budget plays an important role in the performance of the NWP models to simulate the atmospheric boundary dynamics and particularly the development of warm season convective systems. The computations of land processes are accomplished by coupling advanced LSMs into the atmospheric numerical models. Due to the non-linear character of these processes, the production of accurate simulations and forecasts for the soil state (i.e., soil temperature and soil moisture distributions) requires accurate specification of (1) the initial conditions, (2) the atmospheric forcing parameters (i.e., air temperature, precipitation, radiation budgets, winds) and (3) information of the topography, soil properties and land cover. Ongoing research investigates the potential benefits of assimilating ground-based and remote-sensed data into atmospheric and land surface models.

Following this trend, our goal is to investigate the potential of improving NWP models through data assimilation into land surface models. To accomplish this, we

1. investigate the sensitivity of the performance of the POSEIDON weather forecasting system to different types of soil state conditions.
2. implement and test the NCAR Community Land Model version 3.0 (CLM3.0).
3. examine the benefits of developing an offline physically-based LDAS driven by atmospheric forcing based primarily on remote-sensed rainfall fields.

3.1 Assimilation of surface data in the land model (OSU) of the POSEIDON weather forecasting system

As discussed before, soil moisture, soil temperature and precipitation are considered to be very important parameters in land-atmosphere interactions, having significant impact on surface water and energy budgets and intensely affecting the predictability of the atmospheric conditions. Having this in mind, we initially investigated the simulation performance of the POSEIDON weather forecasting system (especially the convective precipitation forecasting skill) through a first-level assimilation (nudging) of ground-based and remote-sensed data. As mentioned before, the POSEIDON system is based on the ETA model and land processes are parameterized using a revised version of the Oregon State University (OSU) model.

For the specific numerical experiments the POSEIDON system was integrated over the data-rich region of the continental USA covering a 5-day forecast period (from 22 July 2004 at 00 UTC to 27 July 2004 at 00 UTC) with a horizontal grid increment of 0.1x0.1 degree. For the initial and boundary meteorological conditions the ECMWF reanalysis gridded data on a 0.5x0.5 degree were used. Three experiments were designed. In the control run (named CTRL hereafter) the initial soil conditions were computed at the 6 soil layers (currently defined at the depths of 5, 15, 28, 50, 100 and 255 cm) from the ECMWF gridded data through a typical interpolation allowing the OSU to carry on as it normally does. In the second experiment (named ECMWF), the ECMWF re-analysis fields of soil temperature and moisture were introduced through a nudging scheme to dynamically update the ETA surface boundary conditions during model simulation. In the final experiment (named NEXRAD) observed rainfall fields (derived from a mosaic of hourly raingauge-calibrated radar (WSR-88D) rainfall estimates) were assimilated into the OSU land model as forcing term instead of the model-generated fields, while the initial soil conditions were defined in the same way as in the CTRL experiment.

The accumulated rainfall fields obtained from the three model experiments for a 6-hour period (ending at 18 UTC, 24 July 2004) are shown in Figure 7 (panels a, b, and c). The fourth map (Figure 7d) corresponds to the respective rainfall fields as observed by the NEXRAD radar network. The results from the numerical experiments were evaluated against the reference rainfall fields (radar) using the contingency table approach and evaluating the bias score (BS), the threat score (TS), the Heidke skill score (HSS) and the root mean square error (RMSE). The left panels of Figure 8 depict the calculated values of these statistical scores for the three numerical experiments, while in the right panels the relative differences (in %) of ECMWF and NEXRAD experiments versus CTRL score values are shown. A slight improvement is noted in all statistical scores in the case of ECMWF soil state conditions being assimilated into the model, while more significant improvement was observed in the case of rainfall data assimilation into the OSU model. The RMSE score assessment shows that the assimilation technique as performed in the frame of the NEXRAD experiment provides moderate (6%-8%) reduction in error variance for the entire 5-day period and up to 25% for the 6-h period.

Additionally, in Figure 9 we show the high correlation between soil moisture variability and rainfall patterns. Indeed, the regions where differences in soil moisture distributions between CTRL and NEXRAD experiments are more pronounced are well correlated with regions where intense rainfall was observed up to 18 hours before by the NEXRAD network.

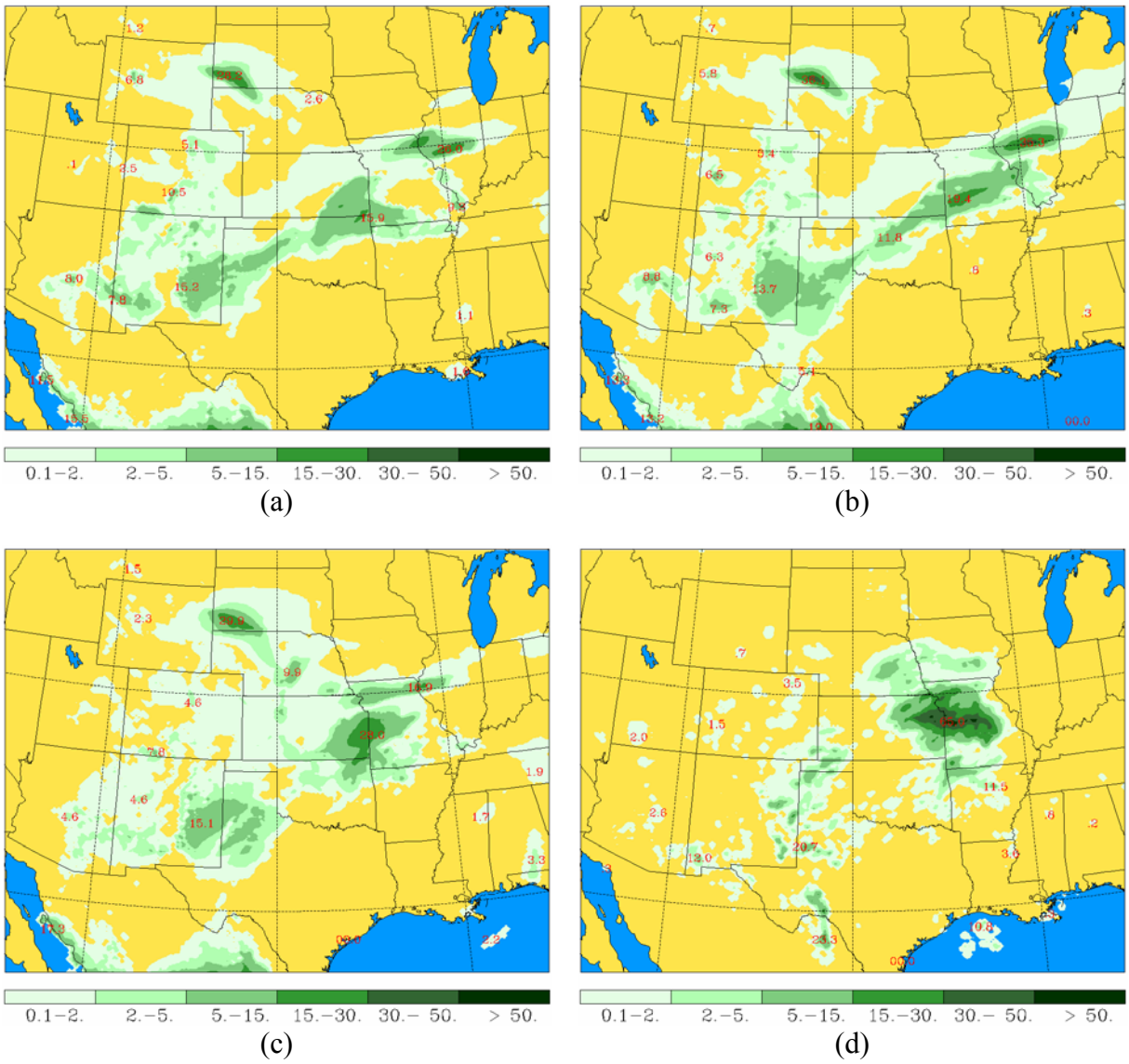

Figure 7. Rainfall fields as predicted by (a) CTRL (b) ECMWF and (c) NEXRAD experiments and (d) as observed by the NEXRAD radar network, for a 6-hour period ending at 18 UTC, 24 July 2004.

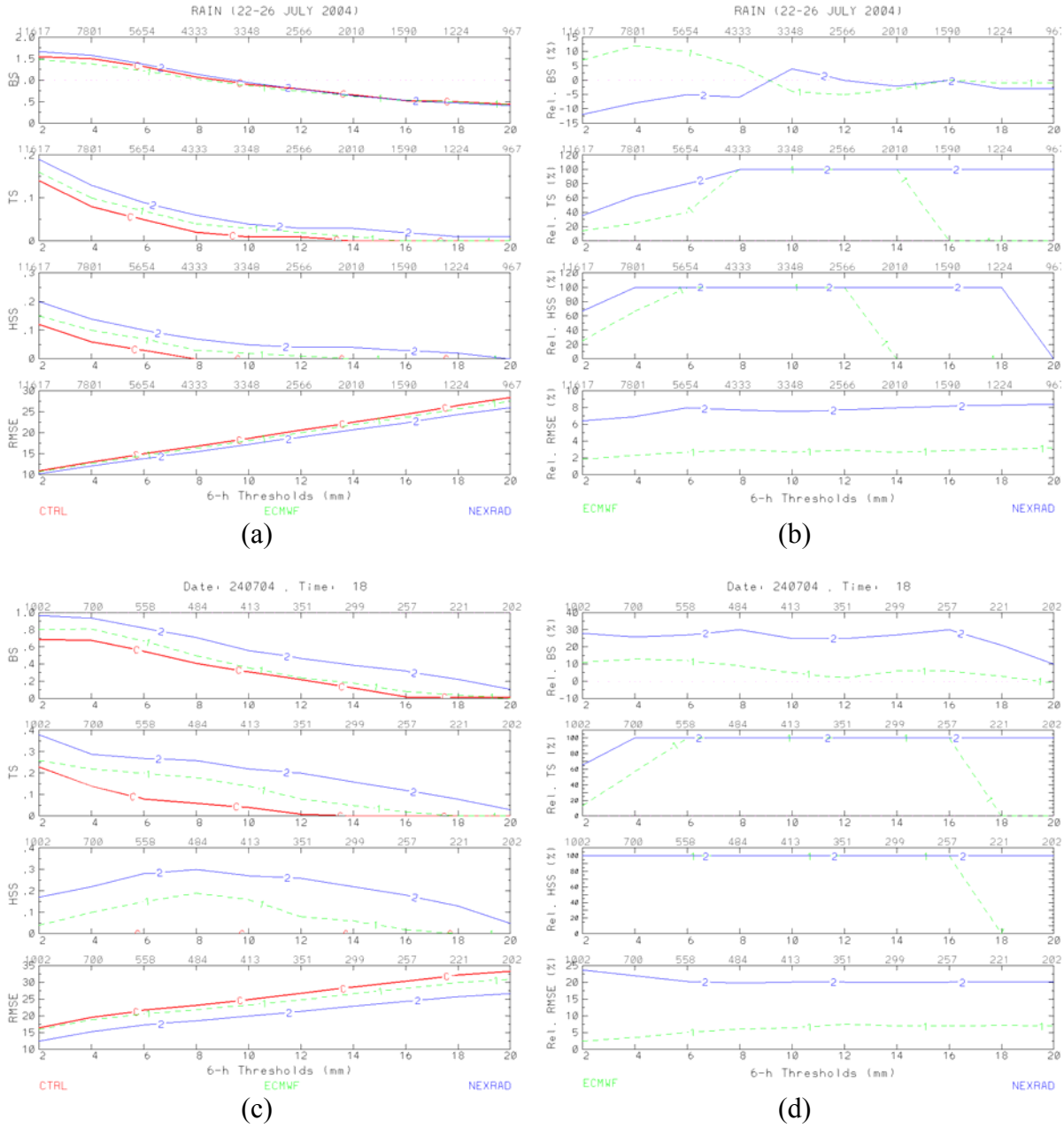

Figure 8. Statistical evaluation of the CTRL, ECMWF and NEXRAD experiments for the entire forecast period (upper panels a and b) and for a 6-h period ending at 18 UTC 24 July 2004 (lower panels c and d) against radar rainfall fields. Numbers above each panel denote the total number of observations reaching the corresponding threshold value.

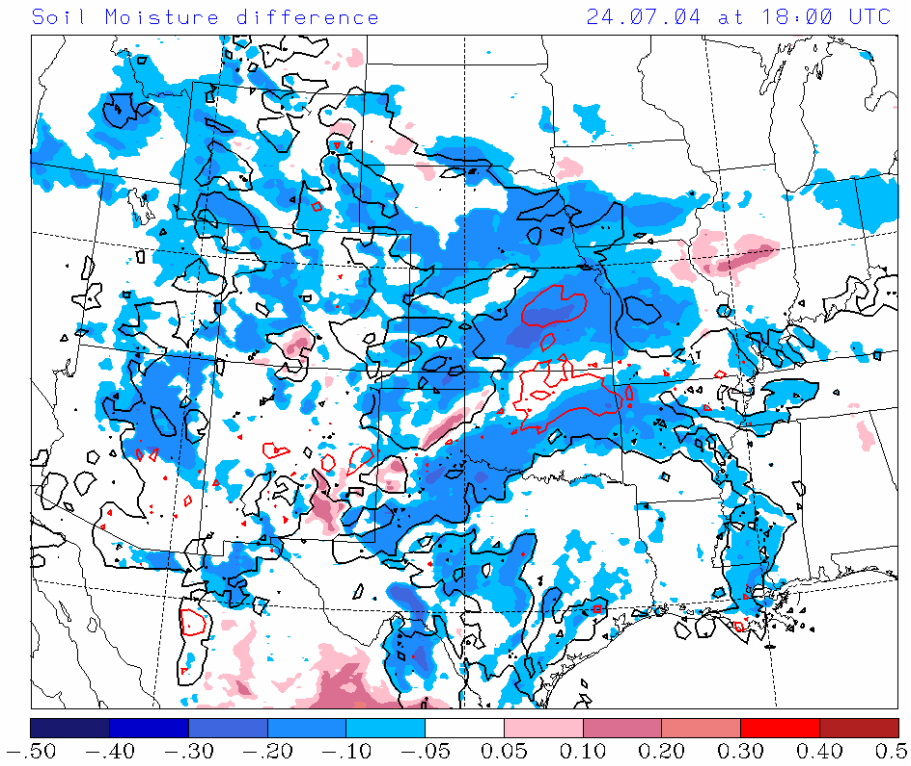

Figure 9. Differences in soil moisture fields between CTRL and NEXRAD experiments (colorbar) and rainfall fields as observed by NEXRAD radar network (contours). Black and red contours denote rainfall rates of 1 mm/18h and 20 mm/18h, respectively.

3.2 THE NCAR Community Land Model version 3.0 (CLM3.0)

Since the results discussed in 3.1 showed significant evidence of improving convective precipitation forecasting in the POSEIDON system through data assimilation into its internal land surface model (OSU), our next step was to implement and test a more elaborate, state-of-the-art land surface model. NCAR Community Land Model version 3.0 (CLM3.0) is a well-documented model (see http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLM_Tech_Note.pdf), which has been designed to integrate all land processes into a single modeling system. The model components consist of: biogeophysics (i.e., surface fluxes of energy, moisture, and momentum), hydrologic cycle, biogeochemistry and dynamic vegetation. CLM3.0 initially utilizes only the biogeophysics and hydrology components, which were perfectly suitable to our work and thus more appropriate for us to use.

Based on externally-provided atmospheric forcing data (e.g., precipitation, radiation, wind speed, air temperature, and humidity fields) CLM3.0 computes a number of prognostic surface variables that include runoff, soil moisture and temperature in various soil layers, water intercepted on the canopy, leaf temperature, latent and sensible heat fluxes. CLM3.0 has one vegetation layer like most land surface models, ten unevenly spaced vertical soil layers (the respective depths are defined at 0.7, 2.8, 6.2, 11.9, 21.2, 36.6, 62.0, 103.8, 172.8, and 286.4 cm) with variable hydraulic conductivity, and up to five snow layers depending on the

total snow depth. The land surface is represented by 5 primary sub-grid land cover types (glacier, lake, wetland, urban, vegetated) for each grid cell. The vegetated portion of a grid cell is further divided into patches of plant functional types (PFTs), each with its own leaf and stem area index and canopy height. Each sub-grid land cover type and PFT patch is a separate column for energy and water calculations at every time step. Most surface processes such as evaporation from the ground, transpiration from the plants' rooting zone, soil and snow water propagation, leaf temperature and fluxes, soil and snow temperature, and phase change are parameterized through physical equations. The parameterization of runoff-related processes is based on TOPMODEL concept (Beven and Kirkby, 1979). The required surface data used for the definition of the physiographic characteristics for each grid cell are currently available in the spatial resolution presented in Table 1.

Table 1

Surface data, base spatial resolution, and method of aggregation to the model's grid.

Surface Field	Resolution	Aggregation Method
Percent glacier	0.5°	Area average
Percent lake	1°	Area average
Percent wetland	1°	Area average
Percent sand and clay	5-minute	Soil mapping unit with greatest areal extent in grid cell
Soil colour	2.8° (T42)	Soil colour class with greatest areal extent in grid cell
PFTs (percent of vegetated land)	0.5°	Area average, choosing 4 most abundant PFTs
Monthly leaf and stem area index	0.5°	Area average
Canopy height (top, bottom)	0.5°	Area average

Although this resolution is suitable for global simulations, in our specific applications (high resolution simulations) it is essential to utilize much finer datasets. This was ascertained through a series of preliminary tests which demonstrated insufficient heterogeneity of land-surface characteristics at the required resolvable scale. Recently, the Moderate Resolution Imaging Spectroradiometer (MODIS) scientific team has developed new land surface datasets, maintaining the multiple PFT, canopy and herbaceous layer representation used in CLM3.0 (Lawrence and Chase 2007). The data, kindly provided to us by Dr. Lawrence, are listed in Table 2 along with their improved resolution and method of aggregation.

Figure 10 shows the near-surface clay and sand percentage distributions associated with the high-resolution land surface datasets in the two model domains used in our work. The first covers parts of Europe, North Africa, and West Asia while the second was over North America. The inhomogeneity of these distributions is a good indicator of the high resolution of this new dataset, and is definitely an important factor associated with the ability of both land and atmospheric models to describe more realistically mesoscale circulations.

Table 2

High resolution surface datasets used in the version of CLM3.0 for mesoscale simulations.

Surface Field	Resolution	Aggregation Method
Percent glacier	0.05°	Area average
Percent lake	0.05°	Area average
Percent wetland	0.05°	Area average
Percent sand and clay	0.05°	Soil mapping unit with greatest areal extent in grid cell
Soil colour	0.05°	Soil colour class with greatest areal extent in grid cell
PFTs (percent of vegetated land)	0.05°	Area average
Monthly leaf and stem area index	0.05°	Area average
Canopy height (top, bottom)	0.05°	Area average

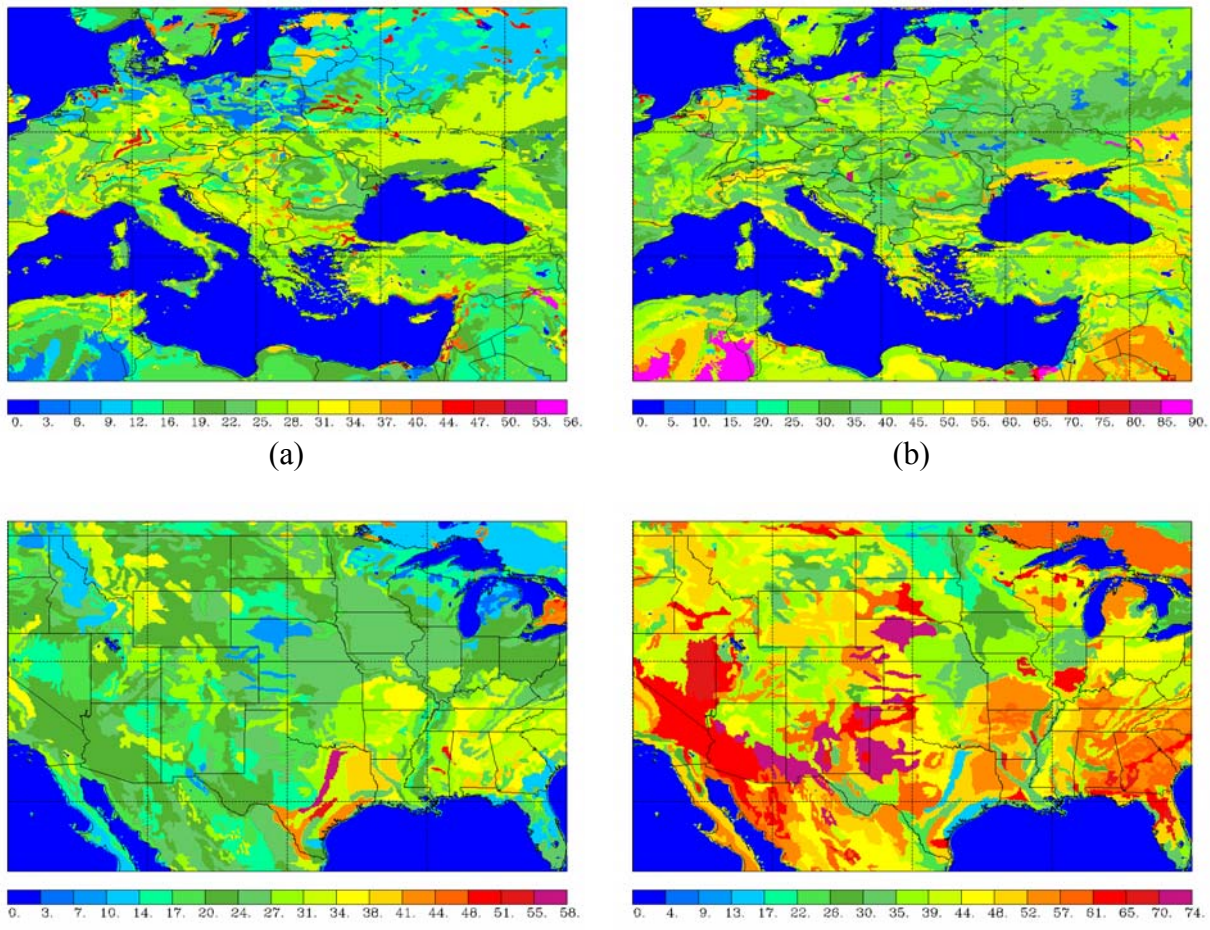

Figure 10. Near-surface clay and sand distributions (%) from the high resolution land surface dataset described in Table 2, over Europe/North Africa/West Asia (panels a and b, respectively) and North America (panels c and d, respectively).

Before using CLM3.0 as the core of the offline LDAS for the POSEIDON weather forecasting system, it was necessary to examine its general performance when applied for mesoscale simulations over a large domain. It is noted that CLM3.0 has been originally developed for global climate simulations, and that the new high resolution surface datasets have not been fully tested (regarding the complex interaction with the internal model physics). One important aspect of this investigation, which was facilitated through numerous simulations and analyses of the results, was the understanding of the model's spin-up process.

To take advantage of the data-rich regions of continental US (including remote-sensed and in-situ observations), we selected the same model domain as used in the experiments in section 3.1 (defined by 80^o W, 120^o W, 25^o N and 50^o N) using a grid increment of 0.1-degree. Two main sets of experiments were conducted regarding the supplied atmospheric forcing (precipitation, incident solar radiation, wind, surface pressure and air temperature):

1. ECMWF reanalysis data for all atmospheric forcing fields (named ECMWF hereafter).
2. ECMWF reanalysis data for all fields except precipitation which is provided by the NEXRAD radar network (named NEXRAD experiment).

As mentioned earlier, the ECMWF data were available at 0.5x0.5 degree and 6-hour time interval. Furthermore, we projected the hourly NEXRAD rainfall data (which are available at 4-km spatial resolution) on the CLM3.0 spatial grid structure. Since the NEXRAD data were gridded at a spatial resolution of 0.1 degree and a temporal resolution of 1 hour, we converted the ECMWF data (including rain data) into the same higher resolution using optimal interpolation methods. This enabled us to take advantage of the high resolution of the NEXRAD precipitation data without compromising the accuracy of the atmospheric forcing data assimilation, and simplified significantly the comparison procedure between the two different sets of experiments. Since the NEXRAD data were available for a 4-month period (May through August 2004), we performed all the simulations for this period. For both sets of numerical experiments described above, four long-term simulations were performed initialized with different soil moisture (θ) value for the entire spatial domain and for all CLM soil layers:

(a) $\theta = 0.1$ (completely dry run)

(b) $\theta = 0.2$ (dry run)

(c) $\theta = 0.3$ (middling run)

(d) $\theta = 0.4$ (wet run)

Likewise, soil temperatures were set to 290K everywhere.

For each simulation we computed the model domain-averaged soil moisture at 6-hour intervals for all ten CLM3.0 soil layers. Then, we interpolated these values to the six soil layers of the ETA model (defined at the depths of 5, 15, 28, 50, 100 and 255 cm, respectively), based on the following relationship:

$$\theta_i^E = \frac{\sum_{k=1}^{n} \theta_k^C \cdot d_k^C}{d_i^E} \qquad (8)$$

where d denotes depth, superscripts C and E denote CLM3.0 and ETA soil layers, respectively, and the sum is over all corresponding soil depths.

For example, the respective soil moisture for the first ETA soil layer is:

$$\theta_1^E = \frac{\theta_1^C \cdot 0.07 + \theta_2^C \cdot 2.8 + \theta_3^C \cdot 1.5}{5} \qquad (9)$$

The temporal evolution of the computed model domain-averaged soil moisture for the equivalent ETA soil layers as computed from the NEXRAD experiment are shown in Figure 11.

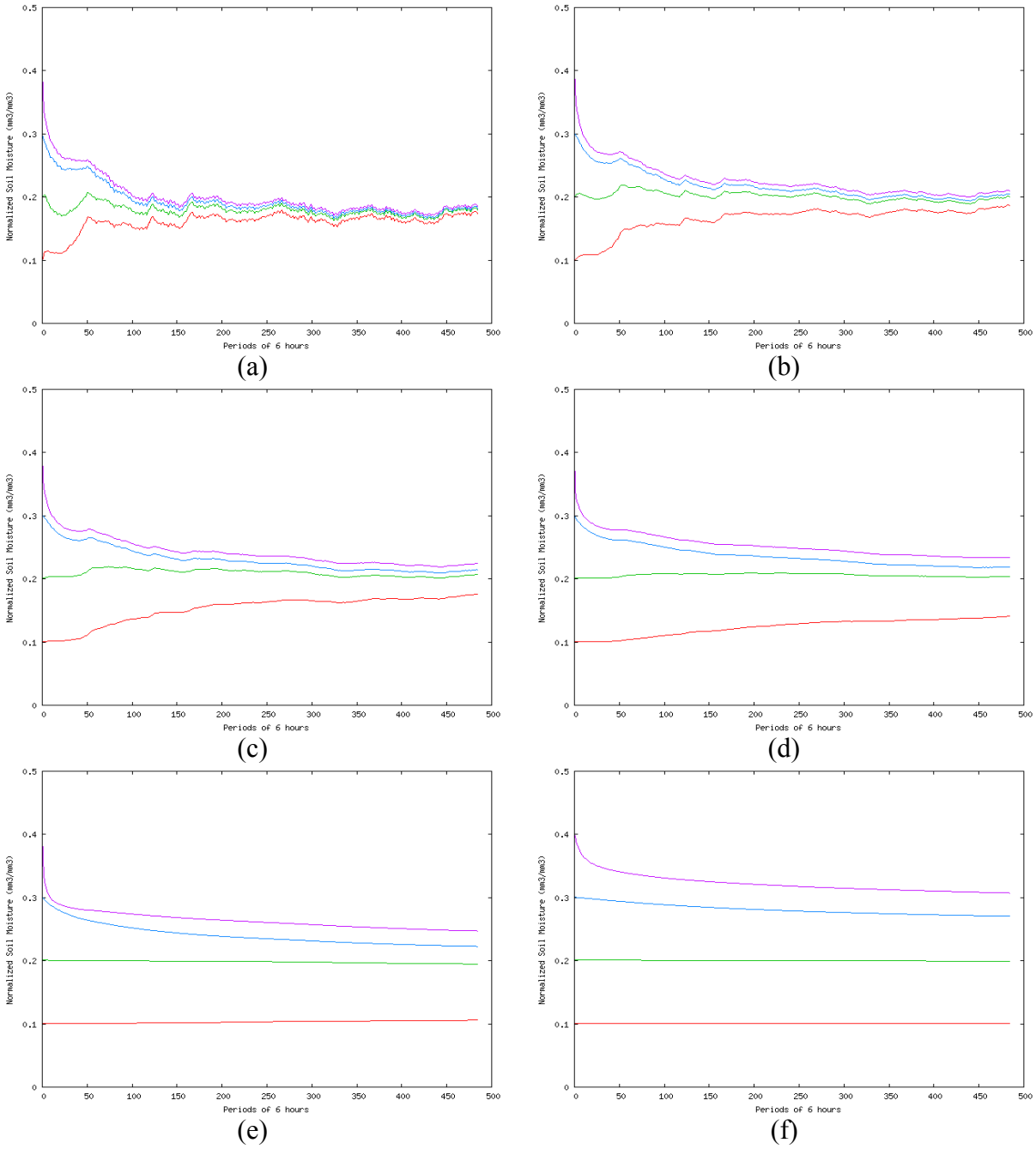

Figure 11. Time series plot of the CLM3.0 domain-averaged soil moisture values interpolated to the six ETA soil layers [(a) 5 cm, (b) 15 cm, (c) 28 cm, (d) 50 cm, (e) 100 cm, (f) 255 cm] for the entire 4-month period. The four different runs are denoted with red (completely dry), green (dry), blue (middling) and cyan (wet).

As illustrated in Figure 11, when the soil layer is closer to the surface, the convergence to a common solution is reached faster. With the exception of the completely dry run, soil moisture in the first two ETA soil layers reaches equilibrium sooner than in the next two soil layers. Figure 12 shows the percent difference between soil moisture fields simulated by the dry and middling run for the first four ETA soil layers at the end of the 4-month period. It can be seen that the soil moisture differences at the upper soil layers are small, varying from 0% to approximately 8-10%. This indicates that the completion of the spin-up process results in a physically realistic balance between external forcing (mainly precipitation) and simulated surface parameters (including soil moisture). Indeed, Figure 13 shows that the rainfall rate controls the near-surface soil moisture evolution, no matter what the initial soil state conditions are. Although there is no apparent equilibrium reached for the two deeper layers by the end of the simulation period, we have to take into account that the upper four soil layers contain the root zone, while the sixth layer acts as the reservoir for long-term moisture storage.

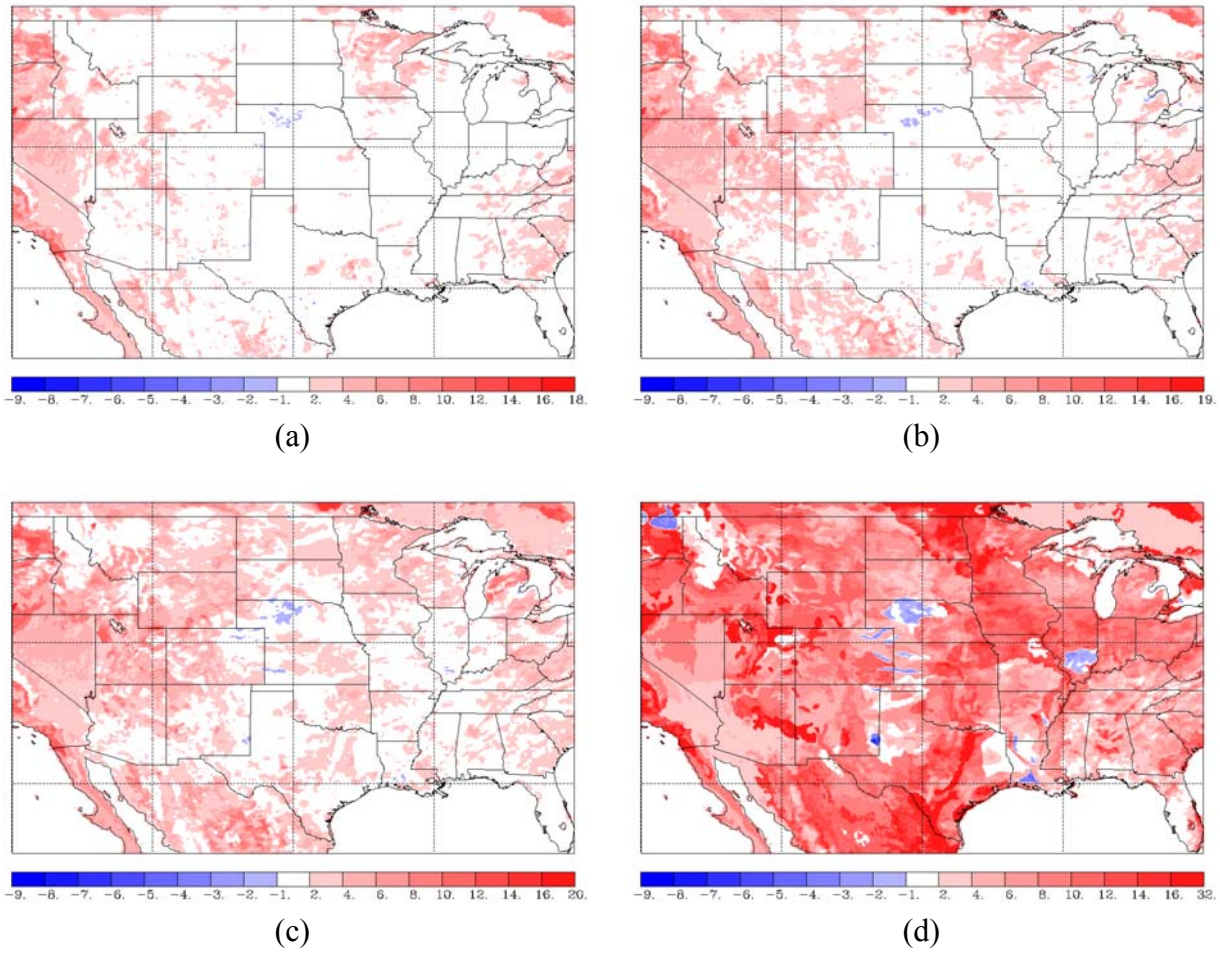

Figure 12. Maps of soil moisture difference (in %) between the dry (θ_{init}=0.2) and the middling (θ_{init}=0.3) initialization experiments for the first four ETA soil layers (as in Figure 11) at the end of the 4-month period.

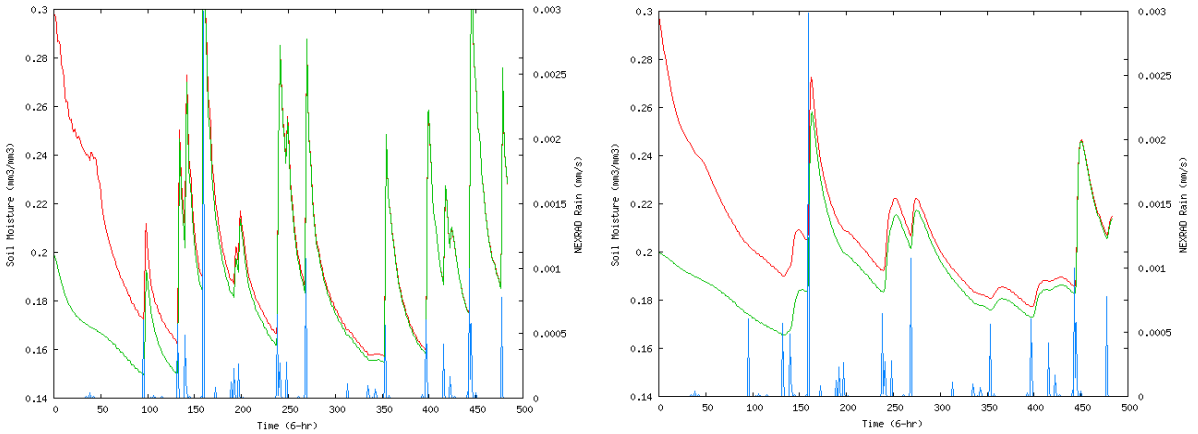

Figure 13. Time series plot of CLM3.0 soil moisture and respective NEXRAD rainfall rate for the entire 4-month period (in 6-hour intervals) for a vegetated grid cell at (a) first and (b) second soil layers.

3.3 Implementation of a coupling scheme between POSEIDON and CLM3.0

The implementation of a coupling procedure between the POSEIDON weather forecasting system and the CLM3.0 model was our third objective in the context of developing a Land Data Assimilation System. As in 3.1, the POSEIDON system was integrated over the data-rich region of the continental USA covering a 5-day forecast period (from 22 July 2004 at 00 UTC to 27 July 2004 at 00 UTC) with a horizontal grid increment of 0.1x0.1 degree. For the initial and boundary meteorological conditions the ECMWF reanalysis gridded data on a 0.5x0.5 degree were used. In our experiment (named CLM), the CLM3.0 soil temperature and moisture fields as simulated from the NEXRAD experiment were introduced through a nudging scheme to dynamically update the ETA surface boundary conditions during the model integration.

The results from the numerical experiments were evaluated against the reference rainfall fields (radar) using the contingency table approach and evaluating BS, TS, HSS and RMSE, as described in 3.1. The left panels of Figure 15 depict the calculated values of these statistical scores for the CLM experiment as compared to the CTRL and NEXRAD experiments performed in 3.1. The right panels show the relative differences (in %) of CLM and NEXRAD experiments versus CTRL score values. Significant improvement is noted in all statistical scores in the case of CLM soil state conditions being assimilated into the model when compared to the CTRL experiment. NEXRAD experiment retains higher statistical scores than CLM throughout the 5-day forecast period, although CLM seems to be reaching the high performance of the NEXRAD assimilation technique for greater rainfall accumulation values, which are usually associated with mesoscale convective systems. Thus, the LDAS implemented in our study with the use of CLM3.0 land surface model and POSEIDON weather forecasting system could be considered an effective system for improving precipitation forecasting, especially in the case of MCS.

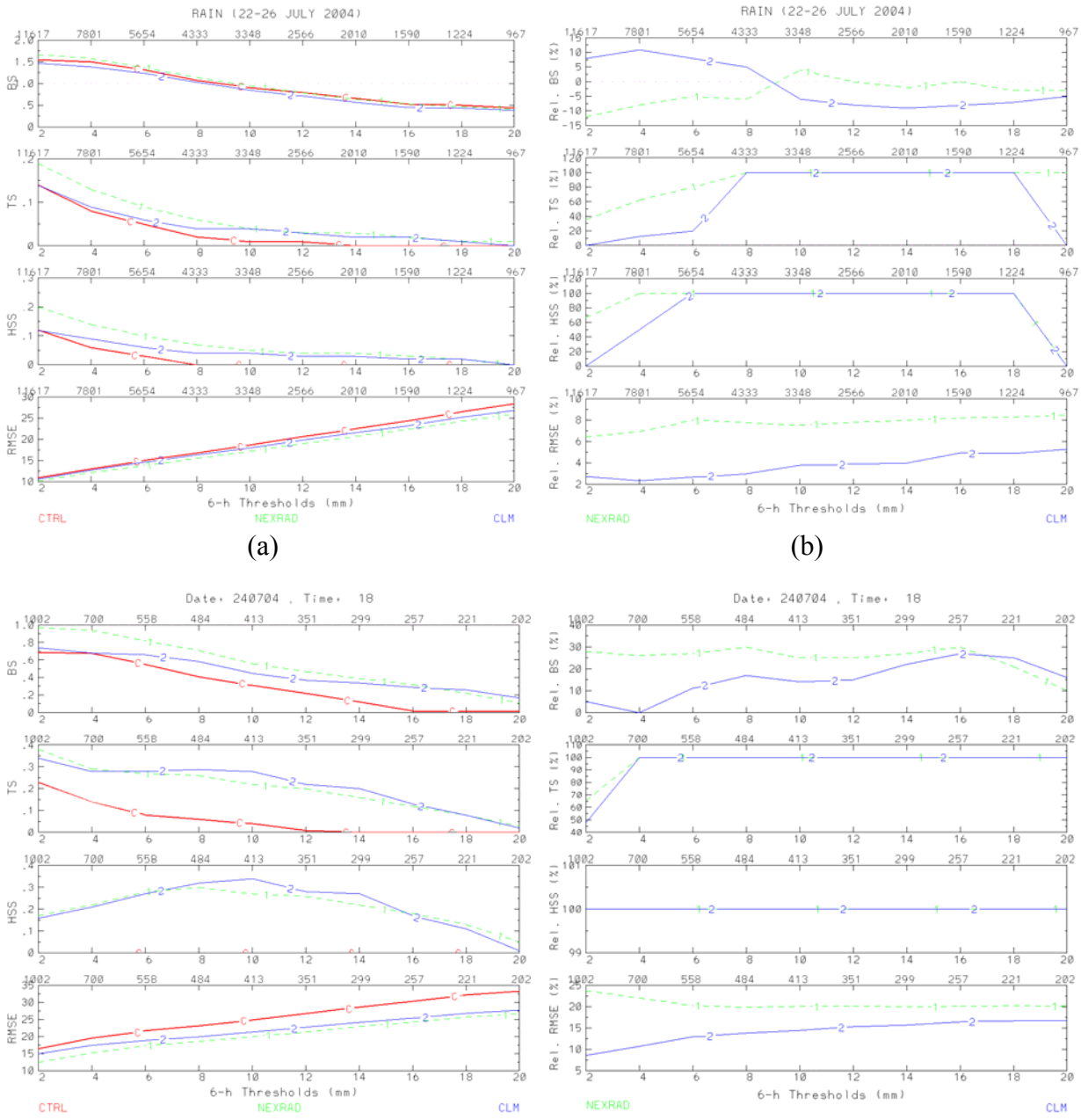

Figure 15. Statistical evaluation of the CTRL, NEXRAD and CLM experiments for the entire forecast period (upper panels a and b) and for a 6-h period ending at 18 UTC 24 July 2004 (lower panels c and d) against radar rainfall fields. Numbers above each panel denote the total number of observations reaching the corresponding threshold value.

References

Alexander, G. D., J. A. Weinman, V. M. Karyampudi, W. S. Olson, and A. C. L. Lee, 1999: The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 Superstorm. Mon. Wea. Rev., 127, 1433–1457.

Betts, A.K., 2004: Understanding hydrometeorology using global models. Bull. Amer. Meteor. Soc., 85, 1673-1688.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and Arctic Air mass data sets. Quart. J. Roy. Meteor. Soc., 112, 693-709.

Beven, K.J., and M. J. Kirkby, 1979. A physically based variable contributing area model of basin hydrology. Hydrol. Sci. Bull, 24, 43-69.

Brenner, I. S., 2004: The relationship between meteorological parameters and daily summer rainfall amount and coverage in west-central Florida. Wea. Forecasting, 19, 286-300.

Chang, D.-E., J. A. Weinman, C .A. Morales, and W. S. Olson, 2001: The effect of spaceborne microwave and ground-based continuous lightning measurements on forecasts of the 1998 Groundhog Day Storm. Mon. Wea. Rev., 129, 1809–1833.

Chronis T. G., E. N. Anagnostou, 2003: Error analysis for a long-range lightning monitoring network of ground based receivers in Europe. J. Geophys. Res., 108, 4779, doi:10.1029/2003JD003776.

Clark, C. A., and R. W. Arritt, 1995: Numerical simulations of the effect of soil moisture and vegetation cover on the development of deep convection. J. Appl. Meteorol., 34, 2029–2045.

Colle, B. A., C. F. Mass, and K. J. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest dusting the 1997-99 cool seasons. Wea. Forecasting, 15, 730-744.

Colle, B. A., K. J. Westrick, and C. F. Mass, 1999: Evaluation of MM5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. Wea. Forecasting, 14, 137-154.

Cosgrove, B. A., et al., 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. J. Geophys. Res., 108(D22), 8842, doi:10.1029/2002JD003118.

Gallus, W. A., Jr., and M. Segal, 2001: Impact of improved initialization of mesoscale features on convective system rainfall in 10-km Eta simulations. Wea. Forecasting, 16, 680-696.

Heidke, P., 1926: Berechnung des erfolges und der gute der windstarkevorhersagen im sturmwarnungsdienst. Georg. Ann., 8, 310-349.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. Mon. Wea. Rev., 122, 927-945.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization, J. Atmos. Sci., 47, 2784-2802.

Kain, J. S., and J. M. Fritsch, 1992: The role of the convective ‘‘trigger function’’ in numerical forecasts of mesoscale convective systems. Meteor. Atmos. Phys., 49, 93–106.

Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models, J. Hydrometeor., 6, 558– 570.

Lawrence, P. J., and T. N. Case, 2007: Representing a new MODIS Consistent Land Surface in the Community Land Model (CLM 3.0). J. Geophys. Res., 112, G01023, doi: 10.1029/2006JG000168.

Lee, A. C. L., 1986a: An experimental study of the remote location of lightning flashes using a VLF arrival time difference technique. Quart. J. Roy. Meteor. Soc., 112, 203-229.

Lee, A. C. L., 1986b: An operational system for remote location of lightning flashes using a VLF arrival time difference technique. J. Atmos. Oceanic Technol., 3, 630-642.

Leese, J., T. Jackson, A. Pitman, and P. Dirmeyer, 2001: GEWEX/BAHC international workshop on soil moisture monitoring, analysis, and prediction for hydrometeorological and hydroclimatological applications. Bull. Amer. Meteor. Soc., 82, 1423–1430.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys. Space Phys., 20, 851-875.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. Bull. Amer. Meteor. Soc., 77, 2637-2649.

Mitchell, K. E., et al., 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.

Papadopoulos, A., G. Kallos, S. Nickovic, D. Jovic, M. Dacic, and P. Katsafados, 1997: Sensitivity studies of the surface and radiation parameterization schemes of the SKIRON system. Proceedings of the International Symposium on Regional Weather Prediction on Parallel Computer Environments, 15-17 October 1997, Athens, Greece, 155-164.

Papadopoulos, A., G. Kallos, P. Katsafados, and S. Nickovic, 2002: The Poseidon weather forecasting system: An overview. GAOS, 8, 219-237.

Papadopoulos, A., Th. Chronis, and E. Anagnostou, 2005: Improving Convective Precipitation Forecasting Through Assimilation of Regional Lightning Measurements in a Mesoscale Model. Mon. Wea. Rev., 133, No. 7, 1961–1977.

Rogers, R. F., J. M. Fritsch, and W. C. Lambert, 2000: A simple technique for using radar data in the dynamic initialization of a mesoscale model. Mon. Wea. Rev., 128, 2560–2574.

Seuffert, G., H. Walker, P. Viterbo, M. Drusch, and J.-F. Mahfouf, 2004: The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis. J. Hydrometeor., 5, 516–531.

Tripoli, G. J., and W. R. Cotton, 1982: The Colorado State University three-dimensional cloud/mesoscale model-1982. Part I: General theoretical framework and sensitivity experiments. J. de Rech. Atmos., 16, 185-220.

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

Zhang, H., and C. S. Frederiksen, 2003: Local and nonlocal impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study, J. Clim., 16(13), 2117– 2137.