Improving Convective Precipitation Forecasting Through Assimilation of Regional Lightning Measurements in a Mesoscale Model

Anastasios Papadopoulos
Hellenic Centre for Marine Research
Athens, Greece

Themis G. Chronis and Emmanouil N. Anagnostou
Department of Civil and Environmental Engineering
University of Connecticut
Storrs, CT 06269, USA

Submitted to Monthly Weather Review

January 2004

Corresponding author: Prof. Emmanouil N. Anagnostou
Civil and Environmental Engineering, U-37
University of Connecticut, Storrs, CT 06269
(860) 486-6806 (voice), (860) 486-2298 (fax)
manos@engr.uconn.edu
ABSTRACT

A technique developed for assimilating regional lightning measurements into a meteorological model is presented in this paper. The goal is to improve the forecast of convective precipitation rates by dynamically modifying the model’s vertical air moisture distribution during the early stages of the numerical simulation through assimilating Cloud-to-Ground lightning information. Utilizing real-time location, timing and flash rate data retrieved from a long-range lightning detection network in Europe, a regional/mesoscale meteorological model is informed about the deep moist convection spatio-temporal development and intensity within the model domain. This information is then used to modify the moisture profiles in the model leading to more accurate parameterizations of the convective effects. An empirical relationship between lightning density (flash rate) and the vertical distribution of air moisture is established by investigating numerous experimental moisture adjustment profiles under different convective conditions. The analysis is based on evaluating the model performance for three flood inducing storm cases associated with deep convection in a warm-season environment over the Mediterranean region. Results show that assimilation of flash rate data can significantly improve the model’s convective precipitation forecasting ability in the assimilation period, while maintaining the subsequent forecasts at higher levels comparing to the control simulation. The approach is general enough to be applied to any mesoscale model, but with an expectable varying degree of success. Its advantage when applied in an operational setting is that real-time lightning data responding promptly to the occurrence of convection would continuously get assimilated to update the moist state of the atmosphere in the model.
1. **Introduction**

Convective events produce very intense rainfall that often leads to river floods and flash floods causing significant loss of life and property damage, soil erosion, and other socio-economic problems. In most cases, rapidly developing mesoscale convective systems (MCS) are responsible for the heaviest and most destructive rainfall and flood events in the Mediterranean region. To moderate these hazards, nowcasting systems have been developed using appropriate inputs (e.g. synoptic observations, surface and remotely sensed data) combined with forecasters’ own knowledge. The decision-making process is integrated with model simulations of deep moist convection and consequential precipitation, an important component of operational numerical weather forecasting. Several parameterization schemes have been suggested for applications in three-dimensional mesoscale numerical models in order to represent the moist processes that occur in convective systems (e.g., Kuo 1965; Fritsch and Chappell 1980; Tripoli and Cotton 1982; Betts and Miller 1986; Kain and Fritsch 1990; Janjic 1994). However, the present operational numerical models exhibit quite low skills at forecasting the development and evolution of convective precipitation. Two sources of error have long been recognized: (1) the lack of sufficient data to correctly specify mesoscale features that act to trigger convection in the model initialization (e. g., Kain and Fritsch 1992), and (2) the assumptions used in developing the physical parameterization schemes (e.g., Mellor and Yamada 1982). It is the general consensus that model simulations are sensitive to small uncertainties in the analysis of the atmospheric initial state, the quality of which dictates the accuracy of model forecasts (Lorentz 1963). The differential conservation equations used to resolve the highly nonlinear processes involving in the formulation of MCSs require temporal boundary values (representing the initial-value problem), therefore small errors in the initial conditions even using the “perfect” model might be enlarged later in the simulation resulting in inconsequential forecasts.
In the last decades, numerous of techniques have been developed to obtain the optimal initial state of the atmosphere in order to define a better starting point for model simulations. The increasing availability of different types of observations motivates research on more advanced data assimilation techniques. For example, assimilating rain rates into numerical mesoscale models was found that could result in an improved forecast of convective precipitation. Wang and Warner (1988) using radar-based rain rate estimates defined latent heating profiles. Assimilating these profiles into a numerical model they were able to demonstrate a considerable improvement of the very-short-range precipitation forecasts, compared to control simulation in which no assimilation was used. Manobianco et al. (1994) assimilated rain rates derived from combination of the Special Sensor Microwave/Imager (SSM/I) observations and Geostationary Operational Environmental Satellite Infrared (GOES-IR) imagery. They recomputed the model-generated latent heating profiles to reflect the satellite-derived rain rates resulting in improved forecasts of frontal positions and low-level vertical motion patterns of a rapidly intensifying Atlantic cyclone. Even though assimilating rain rate data have proved their useful applications, there are still issues due to the sparse coverage by ground systems (radar and rain gauges) and the very low overpass frequency of space-based microwave sensors. Alternative observations are in need to provide frequent and large regional information on convective processes.

Charge separation leading to lightning is a physical process that takes place in regions of a thunderstorm associated with rigorous forcing. The strong connection of lightning occurrence with the existence of strong updrafts (>10 m/s) and graupel growing from super-cooled cloud water in a convective storm makes the continuous measurement of this meteorological parameter potentially useful for improving the prediction of storm evolution and intensification. Indeed, as cloud particles grow around nuclei centers some of them become charged through collisions and interaction with ice polarized crystals and graupel
curried aloft by strong updrafts. The smaller particles tend to acquire positive charges, while the largest particles acquire more negative charges (MacGorman and Rust, 1998). These particles tend to separate under the influences of updrafts and gravity resulting a net positive charge at the upper portion of the cloud and negatively charged at lower portions of the cloud. Separation of charge produces enormous electrical resistance within the cloud and between the cloud and the earth’s surface that generates lightning. Empirical relationships between lightning intensity and convective parameters (updraft velocity, precipitation, ice water content) have been established on the basis of data from field campaigns (e.g., Goodman et al. 1998; Petersen and Rutledge 2001), and justified by numerical experiments as reported in Goodman (2003). Furthermore, using lightning observations the convective regime within the cloud shields can be identified, thus, improving its delineation on satellite IR imagery (Chronis et al., 2004). In a recent study, Morales and Anagnostou (2003) using long-range lightning observations have enhanced the rain rate estimation derived from multi-satellite infrared and microwave rainfall. Alexander et al. (1999) assessing the impact of assimilating different rain rate retrievals in improving model forecasts, they examined the distributions derived from SSM/I only, to combining SSM/I with IR, and SSM/I, IR and lightning. 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 of the sources is 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 measurements as gap-filling information. The challenge, however, is on how might lightning data be used in assimilation techniques to nudge directly
model fields such as updraft/downdraft velocities and environmental properties (temperature and humidity profiles).

In the present study, using lightning information as an indicator of the areas of rigorous convection in a thunderstorm, the model-generated variables (e.g., humidity profiles) are modified according to the lightning intensity resulting in a more accurate parameterization of the convective environment. In an earlier study, Rogers et al. (2000) using radar reflectivity data as a signature of timing and location of convection occurrence, forced a model convective parameterization scheme to trigger (or suppress) convection when and where it was observed. With their approach they produced an improved forecast of convective precipitation in a warm season environment. In our approach, taking advantage of the strong link between lightning occurrence and convection, a technique is developed to amend the initialization data inadequacies and limitations in formulating subgrid-scale processes on the basis of real-time lightning observations available within the model domain. Acquainting that the initial vertical moisture distribution is a crucial factor affecting the overall model performance, especially in convective conditions, nudging the model humidity profiles to empirical profiles related to flash rates, more realistic model soundings are computed resulting in improved forecasting of convective precipitation.

The purpose of this paper is to demonstrate the assimilation technique into a mesoscale model (SKIRON/Eta) applied for operational use (POSEIDON weather forecasting system) using cloud-to-ground (CG) lightning data recorded by a regional network of ground-based receivers (ZEUS) in Europe. In section 2 and 3 brief descriptions of the ZEUS and the POSEIDON weather forecasting systems are provided. Section 4 describes the assimilation technique. The technique performance assessed based on three warm season thunderstorms is presented in section 5 for a number of empirical adjustment profiles. Finally in section 6 we discuss our overall conclusions and future research directions.
2. **Study area and data**

The study region covers the main part of the European continent and the North Africa (Figure 1). The data used include lightning location and timing retrievals from a long-range lightning network and 6-hourly rainfall reports from the European Center for Medium-range Weather Forecasts (ECMWF) rain gauge network in Europe. Figure 1 shows the locations of ECMWF rain gauges and Zeus lightning network receiver stations overlaid the model’s mesh grid. Three storm cases were used in this study falling within a typical period for deep convective activity in Europe (Prodi et al., 2000). Sample statistics from each storm case are shown in Table 1, including the storm periods, the number of meteorological stations used, the number of the 6-hourly rain gauge reports, the number of occurrences in which the observations meet or exceed the threshold of 3, 12, 21, and 30 mm, the maximum number of flashes observed by ZEUS system within intervals of 15 minutes over area of 100km² and the mean flash rate at this site, and the total flashes observed within the storm period and the mean flash rate in the computational domain. It is noted, that the storm occurred in 28-29 Aug 2002, exhibiting too intense lightning activity over specific locations, while during the two other storms the intensity of the lightning was lower and its distribution more uniform.

Zeus network (described in http://sifnos.engr.uconn.edu) consists of six Very Low Frequency sferic receivers, located in Birmingham [UK], Roskilde [Denmark], Iasi [Romania], Larnaka [Cyprus], Mt. Etna [Italy], and Evora [Portugal], each sampling the time series of the vertical electric field of sferic waveforms. The system retrieves the location of cloud-to-ground (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, 1986). An error analysis of Zeus lightning location retrieval based on simulation and through 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, 2004). Within the periphery of network’s receivers the systems detection efficiency is expected to be in the range of 70% to 90%.

3. The meteorological mesoscale model

The Hellenic Centre for Marine Research (HCMR, former NCMR) has developed the POSEIDON system, an operational monitoring, forecasting, and information system for the marine environmental conditions of the Greek seas (described in Nittis et al., 2001 and http://www.poseidon.ncmr.gr). The monitoring network of the system consists of 11 oceanographic buoys and 9 wave buoys. To satisfy forecasting requirements a fully operational modeling system has been developed consisting of meteorological, wave, ocean hydrodynamic, surface pollutant dispersion and shallow water wave models.

The POSEIDON weather forecasting system is based on the SKIRON/Eta model, which is fully operational since September 1999, providing daily 72-hour weather forecasts. The SKIRON/Eta is an evolution of the 1997 version of the NCEP/Eta (National Centers for Environmental Prediction) model, developed at the University of Athens (Kallos et al. 1997; Nickovic et al. 2001; Papadopoulos et al. 2002). The Eta is a gridpoint, step-mountain vertical coordinate model. The convective effects in the Eta model are parameterized using the revised Betts-Miller-Janjic (BMJ) convective scheme (Betts, 1986; Betts and Miller, 1986; Janjic, 1994), while for the grid-scale precipitation a simplified explicit cloud water scheme (Zhao and Carr, 1997) is used. For the simulation of the surface processes the Eta uses the two-layer soil model developed at the Oregon State University (OSU), including surface hydrology with a vegetation canopy (e.g., Ek and Mahrt, 1991; Chen et al. 1996). The radiation package used in the model was developed at the Geophysical Fluid Dynamics Laboratory based on the work of Fels and Schwartzkopf (1975) and Schwarzkopf and Fels
(1991). More details on the model dynamics and physics packages can be found in previous studies (e.g., Messinger, 1973, 1977; Janjic 1984; Mesinger et al. 1988). The SKIRON/Eta shares the same physics packages as the Eta model with some modifications concerning the use of six layers in the soil model component incorporating fine data set of soil textural classes and land cover, and the introduction of the slopes and the azimuths of the sloping surfaces in the calculations of the incoming solar radiation on inclined surfaces (Papadopoulos et al. 1997).

On the basis of a long operational period (four years) as part of POSEIDON weather forecasting system, the SKIRON/Eta model has been evaluated in several high rainfall events. As shown in Papadopoulos et al. (2001) the model has been able to describe well the timing and spatial pattern of those events, but has a limitation in predicting the high rainfall rates (mainly in the extreme summertime storms). Past studies on Eta model have reached at similar conclusions reporting an overestimation of small rainfall occurrences and a decreasing skill for high rain rates (e.g., Mesinger 1996; Baldwin and Black 1998; McDonald and Horel 1998; Colle et al. 1999). It is a common conviction that modern operational numerical models are associated with significant error on quantitative precipitation forecasts, especially in predicting extreme events. Many processes and interactions in the atmosphere-ocean-land system can lead to precipitation, including large-scale motion of moist air, soil moisture availability and the consequential large-scale moisture transport in the atmosphere, sea surface temperature variability, and convection caused by heating moist air near the surface. Therefore, the lack of insufficient data to represent these processes in the model initialization may aggravate the model errors resulting in an inaccurate precipitation forecast. Since lightning activity is directly associated with deep convective conditions, assimilating flash rates into a meteorological mesoscale model is expected to define a better starting point for the model integration. A consistent assimilation during the early stage of the simulation
should also improve the model performance in subsequent simulations moderating effects from convective parameterization deficiencies.

For the simulations in this study, the SKIRON/Eta model was integrated over the domain covering the Mediterranean Sea, part of Europe and North Africa as illustrated in Figure 1. In the vertical, 32 levels were used stretching from the ground to the model top (15,800 m). In the horizontal, the grid increment of 0.10-degrees was applied using 36 seconds time step. 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 on 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 field (SST), and the soil temperature and moisture availability at the six soil model 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.

4. The Assimilation technique

The assimilation scheme developed here aims at improving the short-term convective precipitation forecasts making use of lightning activity information. In previous techniques, lightning data were blended with other remotely sensed data to derive heating profiles (indirectly from retrieved rain rates) that were assimilated in the thermodynamic equations. Unlike in those studies, lightning data (timing, location and flash rates) are used here to determine the areas of convection, and thereafter the model humidity profiles are nudged empirically as function of the observed flash rates. Making use of these adjusted humidity profiles the model can compute more realistic heating rate profiles based on its own
convective parameterization scheme. The assimilation scheme first collocates the model-estimated areas of convection with convective areas delineated from lightning data, and subsequently applies a nudging technique into the model prognostic equations. The two steps are described below.

\[ a. \text{Matching model convective areas with lightning observations} \]

The technique begins by projecting the flash rate data on the model’s spatio-temporal grid structure. Consequently, CG lightning data (location, timing and flash rates) as recorded by ZEUS system are gridded at 0.1 by 0.1 degrees spatial and 15 minutes temporal resolution and projected from the mercator to the Eta coordinates. Applying a linear interpolation between sequential time intervals, flash rates are then interpolated at the corresponding time step of the model integration. Price and Rind (1992) noted that marine convective clouds have a much lower lightning rate than equivalent-sized clouds over land, even though the precipitation rate may be similar. For this reason, threshold values of 10 and 2 CG flashes per 15 minutes have been selected to indicate possible presence of deep convection over land and sea, respectively. Therefore, grid points with lightning frequency values greater or equal to 0.4 and 0.08 flashes per model time step (36sec) delineate the convective areas over land and sea, respectively.

In the Eta model the BMJ cumulus parameterization scheme implements two approaches of convective conditions regarding to deep (precipitating) and shallow (non-precipitating) convection, handling differently the land and the sea grid points. At each model time step and for each grid point the cloud layer is computed between the cloud base (which is set at the model level just below the Lifting Condensation Level (LCL)) and the cloud top (which is set at the highest model level where the temperature of a freely rising parcel becomes again equal to the surrounding air, the Equilibrium Level (EL)). Grid points
in which cloud depths exceed the 290 hPa threshold are directly associated with deep
convective regimes. Similarly, grid points with cloud depth below the above threshold are
assumed to characterize shallow convective areas, while no convection is considered when
cloud depth is less than 10 hPa, or when no cloud is formed.

Crosschecking through all grid points we compare model predictions of the presence
of deep convection with convective areas delineated from lightning data (as described above).
Without any nudging of the model we noted that in the first hours (4-6 hours) of the
simulations the agreement with observations is high, followed by a moderate divergence in
subsequent hours. However, in terms of precipitation amount the agreement with rain gauge
rainfall measurements was low. Considering that the vertical air moisture distribution is not
introduced accurately in the model initialization a technique is developed to modify the model
humidity profiles when and where was deemed necessary according to the lightning
information.

The flowchart of Figure 2 presents this decision-making procedure. Five possible
scenarios are considered, which are defined by combinations of the above flash rate
thresholds and the corresponding model prediction for a grid cell. The first two scenarios
include areas where the observed flash rates indicate that there is no significant electrical
activity (therefore no deep convection). More specifically, in the first outcome the model also
predicts that there is no convection, so no action is taken and the model integration carries on
to the next grid cell. In the second scenario the model has predicted deep (or shallow)
convection in progress. Since the absence of electrical activity does not exclude the
possibility of convection, the BMJ scheme is allowed to carry on, as it normally would do
without any additional forcing. The last three scenarios are associated with grid cells where
the flash rates are above the defined thresholds denoting convection. These three scenarios
are then referred as deep, shallow or no-convection, which is according to the prediction of the BMJ scheme. A nudging is applied in those cases based on a technique described below.

\[ q_{\text{ass}}^l = q_{\text{mod}}^l + F \cdot (q_{\text{emp}}^l - q_{\text{mod}}^l), \quad q_{\text{mod}}^l > q_{\text{emp}}^l \]
\[ q_{\text{ass}}^l = q_{\text{mod}}^l, \quad q_{\text{mod}}^l \leq q_{\text{emp}}^l \]

where \( q_{\text{mod}}^l \) is the model specific humidity, \( q_{\text{emp}}^l \) is an empirical specific humidity profile value, \( q_{\text{ass}}^l \) is the adjusted specific humidity, and \( l \) is the index of the vertical model level at a grid cell. As shown in Equation (1), adjustment is applied only when the model prediction is below the empirical profile value. The scaling factor \( F \) is a dimensionless number that equals the total flash rate observed within the current 15min time interval.

In the case of shallow convection predicted by BMJ (the fourth scenario) an iterative procedure is applied. First, the simulated humidity profiles are scaled to the same empirical humidity profile using Equation (1). Subsequently, cloud depth is recomputed on the basis of
$q_{s}^{\text{ass}}$ 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 Equation (1). The iterative procedure is set to repeat up to three times. If after three iterations a shallow cloud is still computed then the procedure continues with the BMJ’s shallow convective parameterization.

When the BMJ scheme estimates no convection against the indication of the observed flash rates (the fifth outcome), then moisture is added to the entire water column on the basis of the above iterative procedure. If after three iterations a sufficient deep cloud is not produced, then according to the new model environment the BMJ scheme is driven to compute shallow or no parameterized convective effects at that model grid cell.

Because of the potential artificial contributions from any kind of nudging, special care was given to prevent numerical noise and unrealistic results. A nudging term too large would force the model environment rapidly to the observed leading in uncontrollable state. This might be caused by the decreased model’s ability to control mass conservation and by incompatibilities in the mesoscale features due to artificial changes in the model profiles. To moderate such effect we introduced the following constrains: (1) the scaling factor $F$ is limited to threshold values set to 100 flashes per 15 minutes over land and half of that over sea, yielding, for a model time step of 36sec, values of 4 (2) flashes for land (sea) grid points; (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 final nudging value cannot exceed the model’s saturation humidity value. With those constrains we expect to minimize the chance that precipitation enhancement would not occur because of an artificial saturation of the water vapor content, but only when the model environment favors an impending condensation. After any
adjustment the latent heating profiles are computed to ensure the closure of the BMJ parameterization scheme.

Examples of the nudging technique applied for a hypothetical empirical profile and sample model and model-saturation profiles are presented in Figures 3 and 4. The first example shows the humidity profile adjustment associated with two different nudging terms ($F=0.5$ and $F=2$), which correspond to ~12 and 50 flashes per 15 minutes, respectively. In the lower $F$ value (low flash rate scenario), the adjusted profile ranges between the initially simulated and the empirical humidity profile. In the larger $F$ value, which pertains to regimes with significant electrical activity, the model is forced to a profile that is beyond the empirical profile, but below the saturation profile. The upper threshold limiting the maximum moisture increase per time step is noted at points in Figure 3 indicated with letter A. In those locations Equation (1) had predicted amounts above the threshold and the constraining factor reduced those to the specified threshold. Points B and C indicate locations in the profile where the adjusted values get above the corresponding model-saturated values, and the constraining factor reduced the increase to the saturated value level. In Figure 4, we show cases where the model does not agree with observations regarding the occurrence of deep convection; consequently, the iterative procedure is applied. Profiles with the indications of 1, 2, and 3 denote three sequential nudging iterations resulting to the adjusted vertical distribution of moisture. In the example of Figure 4, letter A denotes locations in the profile where the adjusted values of second iteration exceeded the empirical so the third iteration would not modify any further those values as defined by Equation (1). We also note locations in the profile (indicated with letter B) where the adjusted values exceeded the model-saturated value; consequently, the adjustment was reduced to the model saturation level.

5. Application of the assimilation technique
The impact of the proposed assimilation technique on improving the convective precipitation forecasting is assessed based on the three aforementioned warm season thunderstorms in the Mediterranean region. We first evaluate the contribution of the assimilation technique to the model simulations. For this purpose we performed simulation experiments for the 28-29 August 2002 convective event that was associated with intense rainfall. During that period the atmosphere was very unstable, exhibiting significant electrical activity with CG flash rates reaching 317 flashes/15min within a model’s grid cell (100 km²). Simulations were performed with and without assimilating the lightning data. In the case of lightning assimilation, the technique was applied to the entire storm period using the first empirical humidity profile indicated in Figure 5. As indicated in Figure 6, after 12 hours of model integration (time enough for the model spin up) the main patterns of the 6-hourly accumulated precipitation fields from both simulations look alike. However, there are regions in the storm exhibiting significant differences in the magnitude of rain accumulations. Those are the regions associated with the most intense lightning activity, hence convection (see Figure 7, right panel). In particular, as indicated in Figure 7, which shows the rain accumulation differences (left panel) and the corresponding flash rate in the 6-hourly time interval, the technique increases the rainfall accumulation in areas with higher electrification. It is noted that the control run was successful to simulate convection in regions where lightning was observed. With the assimilation we increased the magnitude of rainfall in those convective regimes, while creating more convection in other regions at subsequent time steps (not shown here). The overall effect of assimilation in rain estimation is shown in Figure 8, which plots the relative frequency (spaced logarithmically) of rain accumulation differences between assimilated and control simulations. The mode is shown to be well above zero and the distribution is skewed to large positive values, while we observe fewer cases where rain rates are higher in the control simulation (negative differences). This is expected by the
implementation of the technique, which adds moisture in regions with lightning, while affecting neighbor regions through the model’s microphysical mechanism and moisture advection. The question we address next is as to whether this rainfall enhancement is quantitatively comparable with independent observations. Furthermore, we examine how sensitive the assimilation technique is to the lightning intensity, and how long after forcing the model can maintain the information provided by the assimilation. These aspects are discussed next.

A set of experiments has been performed to address the above issues: a control run in which no flash rate data are assimilated (hereafter named CTRL), a run where the flash rate data were continuously assimilated for the entire simulation (hereafter named CASE) period, and finally a run where flash rate data were assimilated for 6 hours (hereafter named ASE6). The CTRL experiments were designed as described in section 3, while in both of the assimilated experiments the only additional information is the use of matched flash rate data from ZEUS network (as described in section 4).

Since there is no experimental determination of the relationship between vertical humidity distribution and flash rate, several humidity profiles (not shown here) have been compiled by analyzing a large number of radiosonde measurements during thunderstorms. Each of the selected profiles was assimilated into the meteorological model. Evaluating the quantitative precipitation forecasts from CASE experiment relative to the control forecasts we identified the most suitable profiles to be used in the technique, i.e., the profiles shown in Figure 5. Subsequently, the ASE6 experiment was performed to investigate the efficiency of flash rate data assimilation in up to 12 hours of consequential simulations. In Figure 9 we present the ASE6 experiment. For each storm case, six samples of 6- and 12-hourly forecasts can be created by assimilating the lightning data from sequential 6-hourly time periods, while
keeping the past time steps of the simulation at CTRL mode. The experiments and the methodology for evaluating the results are discussed next.

\[ a. \text{Evaluating methodology} \]

As discussed earlier the model forecasts are verified against measured precipitation from more than 800 rain gauges of the ECMWF synoptic station network across Europe (Figure 1). Despite the known problems of the rain gauge errors (e.g., unrealistic measurements, interrupted transmission) and the issues associated with comparing point measurements with area-averaged estimates (Anagnostou et al. 1999), the measurements from this network are valuable considering their coverage and the continuous recording. The data sample statistics are presented in Table 1 where we show to have adequate observations at different rain accumulations for all three storms.

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

\[
M = \sum_{k=1}^{4} w_k \cdot M_k
\]

where \( M_k \) are the values at the four model grid points surrounding the observation, while the weighting factor \( w_k \) is the reverse squared distance giving to nearest points more influence.

The verification scores used in this study 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 gauge measurements 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 gauge measurement exceeded the threshold; B-model forecast exceeded the threshold but measurement not; C-model forecast did not reach the threshold but measurement exceeded it; and D-model forecast and measurement 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:

\[ BS = \frac{A + B}{A + C} \]  

(3)

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

\[ ETS = \frac{A - E}{A + B + C - E} \]  

(4)

where E is defined as,

\[ E = \frac{F \cdot O}{N} = \frac{(A + B) \cdot (A + C)}{N} \]  

(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), since it reduces 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 $ETS\approx 0$) 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/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:

$$HSS = \frac{2 \cdot (A \cdot D - B \cdot C)}{(A + C) \cdot (C + D) + (A + B) \cdot (B + D)}$$  \hspace{1cm} (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:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{NOBS} (MP_i - OP_i)^2}{NOBS}}$$  \hspace{1cm} (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.

\[ b. \textit{CASE Experiments} \]

We performed the CASE experiment to study the impact of the proposed assimilation technique on the model’s predictive accuracy and to understand its sensitivity on the intensity of lightning activity. The three thunderstorms selected for this study exhibit quite different convective conditions and associate lightning intensities (as showed in the sample statistics of session 2), which is representative of the above objective. The assimilation exercise is based on the three most efficient empirical humidity profiles shown in Figure 5. Figures 10, 11, and 12 summarize the BS, ETS, HSS, and RMSE statistical scores for the three storms. In the left panels we show the determined values of those scores for the CASE (three profiles) and CTRL experiments, while in the right panels we present the relative differences (in %) of CASE versus CTRL score values normalized by the corresponding CASE score value. We note a significant improvement in all the statistical scores when lightning data are assimilated into the model. It is also noted that the relative improvement increases as function of rain accumulation threshold, reaching to 100% improvement at rain accumulations exceeding 18-20 mm. This is attributed to the fact that lightning is a direct outcome of vigorous convection, which is associated with the most intense rain rates in the cloud. Consequently, as expected, assimilation of lightning data improves primarily the model predictions of the most intense rain rates in the cloud.

Specifically, the model bias (BS score) in CASE run for all three storms exhibits reduction with increasing the rain threshold. For example, in the CASE run of the 28-29 August 2002 storm BS starts at nearly 1.3 for the 3 mm threshold and converges to around one (unbiased prediction) at moderate to high rain accumulations, while the corresponding
CTRL run starts at around 1.5 for the lower threshold and gradually decreases reaching to zero at 21 mm threshold (i.e., failing to predict any rain above that threshold). Similar results can be obtained from the storm case of the 19-20 May 2002 (Figure 10). In the third case (Figure 12) the BS values are lower compared to the previous two cases, but even then CASE gives less biased (particularly at the higher thresholds) rain estimation than CTRL.

A noticeable improvement also occurs for the model ETS and HSS score values. The CASE scores are steadily greater than those computed from the CTRL runs, with the exception of the 28-29 July 2003 storm where these CASE and CTRL runs gave similar score values. This storm was associated with the least lightning frequency (see Table 1) indicating that the amount of lightning was not adequate to provide notable improvements to the model predictions. The RMSE score assessment, which gives a measure of the variability of prediction error, suggests that the assimilation provides more than 5% reduction in error variance. In the 28-29 August 2002 case, the RMSE reduced by up to 17%, when the first empirical humidity profile was used, with greater improvement occurring at the higher thresholds. Overall, the model was shown to exhibit better scores when the first empirical humidity profile was used in the assimilation technique, thus this profile was used in the ASE6 experiment described next.

c. ASE6 experiment

The ASE6 experiment was performed to investigate how long after the lightning forcing the model can maintain the information provided by the assimilation. Using the same methodology described above we evaluate the subsequent 6- and 12-hourly forecasts compared to the corresponding CTRL forecasts. The verifications scores for the three storm cases are showed in Figures 13, 14, and 15, respectively, with the left and right panels having the same significance as in Figures 10-12. Overall, there is a noticeable improvement for all
statistical scores, which is more pronounced in the first 6-hourly forecasts. Even though for specific thresholds the model BS, ETS and HSS obtained from the ASE6 experiments are not always better than CTRL, the CASE RMSE scores are constantly lower than CTRL in the order of 5%-7%. This suggests that the model can provide improved quantitative forecasting results even up to 12 hours after the lightning assimilation, with an expected better performance in the first 6 hours.

6. Conclusions

We have developed a technique for assimilating regional lightning data in a mesoscale meteorological model. In the study presented here we used data from the National Observatory of Athens long-range lightning network (Zeus) that covers primarily Europe, and the hydrostatic version of SKIRON/Eta mesoscale meteorological model that runs operationally in the POSEIDON system of the Hellenic National Marine Research Center. The assimilation technique uses the lightning location and flash rate information derived from Zeus to adjust the model predicted moisture profiles aiming at enhancing or creating deep convection where significant lightning is observed and is contrary to the model prediction. For this purpose an empirical relationship between lightning density (flash rate) and the vertical distribution of air moisture was established by investigating numerous moisture profiles in intense convective storms. The assimilation scheme is general enough to be implemented to any meteorological model and/or lightning network observations. Assessment was performed using as reference 6-hourly rain accumulation measurements from a network of ECMWF gauges in Europe. We explored the significance of lightning assimilation in the model prediction on the basis of three major thunderstorms that occurred in the Mediterranean region in the warm seasons of 2002 and 2003.
Three experiments were performed to evaluate the significance of lightning assimilation. In the first experiment the model was run at the control scenario (i.e., no lightning assimilation), while the other two were concerned with a continuous lightning assimilation (CASE scenario) and a six-hourly lightning assimilation (ASE6). In both cases, and most importantly in the CASE scenario, we showed that assimilation improves the error characteristics of the model predictions as compared to the 6-hourly ECMWF gauge measurements. The most important observation is that with assimilation the model is able to predict more accurately the higher rainfall accumulations. The overall performance of the model remained above the critical value of 0.1-0.2 for HSS for the whole dynamic range of rain accumulations. In terms of root mean square differences lightning assimilation introduces improvements of the order of 5-10%, while in the storm exhibiting the most lightning we showed RMS reduction reaching 17% in moderate to high rain accumulations.

A complementary extension of this work would be to assimilate latent heating profiles derived on the basis of rain rates retrieved from combined satellite Infrared and lightning data (e.g., Morales and Anagnostou 2002; Chronis et al. 2004). Alexander et al. (1999) and Chang et al. (2001) had demonstrated that assimilation of convective rainfall and adjusted latent heating diagnosed from lightning data led to improving the short-term forecast of major convective storms developed in the Gulf of Mexico. The physical basis for their approach relies on well-established links between convective updraft strength and the initiation of robust mixed phase precipitation processes, electric charge separation, and ultimately lightning production. Another aspect of future research is to investigate the viability of the land-surface alternative to coupled meteorological systems, which uses physically based land surface models driven by independent observation-based precipitation and radiation budget fields to provide improved land surface boundary conditions for the atmospheric model.
Acknowledgements: This study was supported by the *POSEIDON project* funded by the Financial Mechanism of the European Economic Area (EFTA) and the Greek Ministry of National Economy, and by the *NASA’s Global Water & Energy Cycle program*. The second author was supported by a *NASA Earth System Science Graduate Student Fellowship*. Zeus long-range lightning network was funded by the *Greek General Secretariat for Research and Development* through a ‘Center of Excellence’ grant to the National Observatory of Athens (NOA).
References


Ek, M., and L. Mahrt, 1991: A one-dimensional planetary boundary layer model with interactive soil layers and plant canopy. User Guide (version 1.0.4), Department of Atmospheric Sciences, Oregon State University, 54 pp. [Available from Department of Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331-2209 USA].


Papadopoulos, A., 2001: A numerical limited area atmospheric model with special abilities in handling the initial and boundary conditions atmospheric. Ph.D. dissertation, University of Athens, 280 pp. [Available (in Greek) from University of Athens, Department of Applied Physics, Athens, GR-15784.]


Table 1. Sample statistics of each storm case.

<table>
<thead>
<tr>
<th>Storm period</th>
<th># of stations</th>
<th>6-hourly reports (Acc≥0)</th>
<th>Acc&gt;3</th>
<th>Acc&gt;12</th>
<th>Acc&gt;21</th>
<th>Acc&gt;30</th>
<th>Max (mean) flashes/15min</th>
<th>Total # of flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 May 2002 - 21 May 2002</td>
<td>807</td>
<td>2889</td>
<td>141</td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>22 (0.3)</td>
<td>27,188</td>
</tr>
<tr>
<td>28 Aug 2002 – 30 Aug 2002</td>
<td>829</td>
<td>4740</td>
<td>280</td>
<td>84</td>
<td>33</td>
<td>14</td>
<td>317 (1.7)</td>
<td>277,271</td>
</tr>
<tr>
<td>28 July 2003 – 30 July 2003</td>
<td>815</td>
<td>7047</td>
<td>248</td>
<td>80</td>
<td>31</td>
<td>8</td>
<td>48 (0.4)</td>
<td>23,797</td>
</tr>
</tbody>
</table>
FIG. 1. Model domain (thin line) and lightning assimilation domain (rectangle) overlaid by the locations of ECMWF rain gauges (dots) and Zeus lightning receiver stations (solid circles).
FIG. 2. Flow chart of the lightning data assimilation technique.
FIG. 3. Hypothetical initially model-predicted (mod), saturated (sat) and adjusted (emp) profiles on the basis of two flash rate scenarios (F=0.5 and F=2).
FIG. 4. Same as in Fig. 1, but for a hypothetical iterative adjustment procedure.
FIG. 5. The three empirical humidity profiles used in the assimilation scheme.
FIG. 6. CTRL and CASE model predicted six-hourly rain accumulation fields (28 July 2002 at 12 UTC).
FIG. 7. Corresponding to Fig. 6 rain accumulation difference (CASE-CTRL) and flash density fields.
Fig 8. Relative frequency (%) of rain accumulation differences (CASE-CTRL) for the whole storm.
FIG. 9. Design of the ASE6 experiments.
FIG. 10. The values (left panels) and relative improvements in % (right panels) of bias scores, equitable threat scores, Heidke skill scores, and root mean square error for CASE experiment applied using the three empirical humidity profiles (19-21 May 2002 storm case). Numbers above each panel denote the total number of observations reaching the corresponding threshold value.
FIG. 11. Same as in figure 10 but for the 28-29 of August 2002 storm case.
FIG. 12. Same as in figure 10 but for the 28-29 of July 2003 storm case.
FIG. 13. Same as in Figure 10 but for the ASE6 experiment (19-21 of May 2002 storm case). The green and blue lines correspond to 6- and 12-hour forecasts.
FIG. 14. Same as in Figure 13 but for the 28-29 of August 2002 storm case.
FIG. 15. Same as in Figure 13 but for the 28-29 July 2003 storm case.