A Reanalysis System for the Generation of Mesoscale Climatographies

ANDREA N. HAHMANN,*+ DORITA ROSTKIER-EDELSTEIN,*,# THOMAS T. WARNER,*,@ FRANCOIS VANDENBERGHE,* YUBAO LIU,* RICHARD BABARSKY,& AND SCOTT P. SWERDLIN*

* National Center for Atmospheric Research, Boulder, Colorado**
+ Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Roskilde, Denmark
@ Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado
& National Ground Intelligence Center, Charlottesville, Virginia

(Manuscript received 22 July 2009, in final form 4 November 2009)

ABSTRACT

The use of a mesoscale model–based four-dimensional data assimilation (FDDA) system for generating mesoscale climatographies is demonstrated. This dynamical downscaling method utilizes the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5), wherein Newtonian relaxation terms in the prognostic equations continually nudge the model solution toward surface and upper-air observations. When applied to a mesoscale climatography, the system is called Climate-FDDA (CFDDA). Here, the CFDDA system is used for downscaling eastern Mediterranean climatographies for January and July. The downscaling method performance is verified by using independent observations of monthly rainfall, Quick Scatterometer (QuikSCAT) ocean-surface winds, gauge rainfall, and hourly winds from near-coastal towers. The focus is on the CFDDA system’s ability to represent the frequency distributions of atmospheric states in addition to time means. The verification of the monthly rainfall climatography shows that CFDDA captures most of the observed spatial and interannual variability, although the model tends to underestimate rainfall amounts over the sea. The frequency distributions of daily rainfall are also accurately diagnosed for various regions of the Levant, except that very light rainfall days and heavy precipitation amounts are overestimated over Lebanon. The verification of the CFDDA against QuikSCAT ocean winds illustrates an excellent general correspondence between observed and modeled winds, although the CFDDA speeds are slightly lower than those observed. Over land, CFDDA- and the ECMWF-derived wind climatographies when compared with mast observations show similar errors related to their inability to properly represent the local orography and coastline. However, the diurnal variability of the winds is better estimated by CFDDA because of its higher horizontal resolution.

1. Introduction

Atmospheric models have been used for decades for filling space and time gaps among observations with various motivations, for example, long-term reanalyses have been generated with global-model data assimilation systems. The resulting model-assimilated datasets (MADS) are used for analyzing trends, studying physical processes, and identifying erroneous observational data. On smaller scales, mesoscale models have been used in short simulations for “case study” analyses, where good model simulation skill at observation locations has been justification for believing the simulation in the space and time gaps between the observations. In both cases, the strength in using the models to fill the observation gaps is that the fields are dynamically consistent and they are defined on a regular grid. Additionally, the models respond to local forcing that adds information beyond what can be represented by the observations.

More recently, the availability of greater computing power has allowed the generation of long-term mesoscale reanalyses (Castro et al. 2007; Kanamitsu and Kanamaru 2007; Lo et al. 2008; among many others), where the resulting gridded datasets are used for various
applications. For example, mesoscale climatographies\textsuperscript{1} that are based on reanalyses can be used to define the statistical distribution of wind speeds for wind power assessment (Hagemann 2008). Boundary layer (BL) climatographies can be used with transport and diffusion (T&D) or air-quality models to define prevailing patterns in source–receptor relationships. An example of the latter application is the use of a multimonth mesoscale reanalysis of BL properties in Iraq to calculate the exposure of civilian and military personnel to potential releases of Sarin nerve gas during the 1991 Gulf War (Shi et al. 2004; Warner and Sheu 2000). In a similar but more-contemporary application, the U.S. National Ground Intelligence Center (NGIC) routinely performs 40-yr mesoscale reanalyses for various areas of the world to estimate the T&D of hazardous material released into the atmosphere for different seasons and various times-of-day. In addition, such BL climatographies can allow an assessment of the statistical risk to populations from the release of hazardous material into the atmosphere from chemical or nuclear power generation facilities. Last, the automated interpretation of long-period reconstructions of the atmosphere can be used to define physical processes in ways that are much more robust than what is obtainable from the use of a few case studies.

Other dynamical downscaling efforts have verified climatographies of near-surface model winds over land. Zagar et al. (2006) verified near-surface model winds obtained by the dynamical downscaling of the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr Reanalysis (ERA40) with a mesoscale model using a 10-km horizontal grid increment. Verification was performed against wind observations at 10-m height AGL for a period of 70 days. Kanamitsu and Kanamaru (2007) have compared near-surface winds from the 57-yr California reanalysis downscaling on a 10-km grid to the North American regional reanalysis that used a 32-km grid. Recently, dynamical downscaling has become a well-established method for wind power resource assessment purposes (e.g., Landberg et al. 2003), however, very few objective verifications have been reported in the literature. Jimenez et al. (2007) performed a verification of downscaled near-surface winds from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis Product by the Pennsylvania State University–NCAR Mesoscale Model (MM5; their procedure does not include the assimilation of observations as done in

\textsuperscript{1} The term climatology is used here to refer to a description of the climate based on an interpretation of a series of observations in contrast to the term climatology, which is the study of climate.
times that are months or years in the future. For this purpose, the CFDDA climatography is generated for a region of interest in the world and typical BL conditions are used to define likely directions and speeds of hazardous material transport for different seasons and times of day. In addition, by quantifying the variance in the wind fields over the historical period and inputting this data to the probabilistic T&D model one can represent the uncertainty in consequence assessment products that is attributable to historical weather variability.

### a. The basic CFDDA system configuration

The CFDDA system is currently based on the MM5, version 3.6 (Grell et al. 1995). The general configuration of MM5, as used in the CFDDA system, is summarized in Table 1—details about specific model characteristics can be found in the NCAR technical note on MM5 (Grell et al. 1995). In the simulations presented here, default values have been used for the various parameters in the parameterizations and no attempt has been made to tune them to better represent the climate of the region.

### b. The data assimilation technique

Data assimilation by Newtonian relaxation is accomplished by adding nonphysical nudging terms to the model predictive equations. These terms force the model solution at each grid point to observations, or analyses of observations, in proportion to the difference between the model solution and the observations or analysis. This approach is used because it is relatively efficient computationally; it is robust; it allows the model to ingest data continuously rather than intermittently; the full model dynamics are part of the assimilation system so that analyses contain all locally forced mesoscale features; and it does not unduly complicate the structure of the model code. The implementation of Newtonian relaxation in the CFDDA system forces the model solution toward observations (Stauffer and Seaman 1994) rather than toward a gridded analysis of the data. This approach was chosen because observations on the mesoscale are sometimes sparse and typically are not very uniformly distributed in space, making objective analysis difficult. With observation nudging, each observation is ingested into the model at its observed time and location, with proper space and time weights, and the model spreads the information in time and space according to the model dynamics. In addition, the observational values are quality controlled against the large-scale analysis value as described by Liu et al. (2004).

Studies using Newtonian relaxation include Stauffer et al. (1991), Stauffer and Seaman (1994), Seaman et al. (1995), and Fast (1995). A common finding in these studies is that analysis nudging works better than intermittent assimilation\(^2\) on synoptic scales. Stauffer and Seaman (1994) and Seaman et al. (1995) showed that nudging toward observations was more successful on the mesoscale than nudging toward analyses.

### c. Model configuration and setup

The model uses 36 computational levels, with approximately 12 levels within the lowest 1 km and with the model top at 50 hPa. Figure 1 displays the geographic location of the model grids and the surface elevation on each domain. The model has a horizontal grid increment of 45 km on the outer grid (D1) that covers most of the Mediterranean Sea and extends eastward to cover the Black and Caspian Seas. The first nested grid (D2), which has a grid increment of 15 km, covers the EM and Middle East regions. A third nested domain (D3), with a grid increment of 5 km, is located along the Levant (Lebanon, Syria, and Israel). One-way nesting is used to preserve the horizontal continuity of the results on each nest.

\(\text{Table 1. Summary of setup and parameterizations used in the CFDDA system.}\)

<table>
<thead>
<tr>
<th>Nonhydrostatic dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way interactive nesting procedure</td>
</tr>
<tr>
<td>Radiative upper-boundary condition that mitigates noise resulting from the reflection of vertically propagating waves</td>
</tr>
<tr>
<td>Coarse-domain time-dependent lateral-boundary conditions, relaxed toward large-scale reanalyses</td>
</tr>
<tr>
<td>Grell (1993) cumulus parameterization on 10-km grid increment, or coarser, grids</td>
</tr>
<tr>
<td>Reisner et al. (1998) mixed-phase microphysics parameterization that includes explicit prognostic equations for cloud water, rainwater, ice particles and snow processes</td>
</tr>
<tr>
<td>Modified medium-range forecast model (Hong and Pan 1996; Liu et al. 2006) boundary layer parameterization</td>
</tr>
<tr>
<td>Cloud effects on radiative transfer (Dudhia 1989) for shortwave and rapid radiative transfer model (Mlawer et al. 1997) for longwave</td>
</tr>
<tr>
<td>Noah land surface model (Chen and Dudhia 2001) with four soil layers</td>
</tr>
</tbody>
</table>

\(^2\) Intermittent data assimilation involves restarting the model at regular intervals, where the initial conditions are typically based on data within a certain time window being used to correct a first-guess field that is the forecast from the previous cycle. Restart intervals may typically be 1–6 h.
The CFDDA simulations cover 10 Januaries and 10 Julys for the period 1998–2007. Individual simulations are reinitialized every 8.5 days, using the large-scale reanalysis (2.5° × 2.5° grid) of the NCEP–U.S. Department of Energy reanalysis II (R2; Kanamitsu et al. 2002). These same analyses are used to calculate the lateral boundary conditions. Even though the rationale for the cold starts is that they prevent the accumulation of error in data-sparse areas, there has been no evidence of this problem. Each simulation overlaps the previous one by 12 h, to avoid using the time during which the model is spinning up mesoscale processes. The choice of 8.5-day integration segments is due to restrictions on the size of the lateral boundary condition file and the need to minimize the error growth in areas with sparse observations. Qian et al. (2003) and Lo et al. (2008) recommend this approach. An added advantage of segmenting the simulation is that it allows for simultaneous integrations for various time periods.

Initial land surface conditions are based on the National Aeronautics and Space Administration (NASA) Global Land Data Assimilation System (GLDAS; Rodell et al. 2004) dataset that is defined on a 1° × 1° grid. The GLDAS fields used here are substrate soil moisture and temperature at depths of 10 and 200 cm, ground skin temperature, and snow water equivalent. These fields are interpolated to all land grid points on each model domain and replace the standard values defined in the R2 analysis. Because the same land surface scheme is used in both GLDAS and MM5, no soil moisture conversion is required. Sea surface temperatures (SSTs) are interpolated to CFDDA ocean grid points on each model domain using the NCEP Version 2.0 global SST dataset (Reynolds et al. 2002), which is defined on a 1° × 1° grid and updated daily.

d. Data sources for assimilated observations

The type, spatial density, and frequency of observations used by the CFDDA system vary greatly in different parts of the world. The Advanced Data Processing datasets, available from the NCAR Computational and Information Systems Laboratory’s (CISL) Data Support Section, are employed. The data represent a global synoptic set of surface hourly and 6-hourly data reports, operationally collected by NCEP. The surface dataset includes mostly surface synoptic observations (SYNOP) and Meteorological Aerodrome Report (METAR) land reports but a few ship observations also exist. The upper-air data consist mainly of radiosonde soundings.

Figure 2 displays the geographic data distribution and the approximate frequency of data availability in January 2000 for the model’s outermost domain. The map representing surface data (Fig. 2a) shows a very inhomogeneous geographic distribution, with most of the observations having greater than 3-h time resolution located over western and central Europe—the lowest concentration of observations is found over northern Africa away from the coast. Both the Mediterranean and Red Seas show a few locations with oceanic data from ships, but their time resolution is poor (less than 1 observation per day on average). During January 2000 in this region, there were over 615,000 observations at
over 2000 stations. The geographic distribution of upper-air observations in Fig. 2b displays a similar pattern, with the most frequently reporting (2–4 times a day) stations over western Europe. Over the region covered by D3, only four radiosonde sites are available, with an averaged frequency of 1–2 observations per day.

3. Prevailing meteorological processes in the EM

The EM, located within the transition area between the subtropical high-pressure belt and the midlatitude westerlies, is very interesting and meteorologically complex. Superimposed on the large-scale processes are mesoscale effects related to the complex coastline and nearby mountains (see terrain elevation in Fig. 1b). The cold-season precipitation in this area is associated with Mediterranean cyclones, for which there are many climatological descriptions (e.g., Alpert et al. 1990; Trigo et al. 1999; Genovés et al. 2006). There are strong near-coastal gradients in observed precipitation, the positions of which have huge hydrologic consequences. These gradients are due to the preferred tracks followed by the cyclones, their intensity, and their interaction with the local topography (e.g., Saaroni et al. 2010). The northeastern and eastern coast of the Mediterranean basin is rich in cyclogenetic regions, resulting in considerable precipitation amounts. The opposite is true in the southeastern coast, leading to scarce precipitations and desert climate characteristics. Some adjacent lands to the northern and eastern Mediterranean coastal areas are rich in cyclonic activity as well, leading to significant precipitation amounts during the cold season—that is, the south coast of the Black sea and the Fertile Crescent extending throughout north Levant to the Persian Gulf. Thus, proper reproduction of the winter climate over this region compels correct simulation of the synoptic flow dominated by extratropical cyclones as well as the mesoscale flow dictated by the complex coast and terrain forcing.

Throughout the summer, the EM is dominated by a subtropical ridge extending from the north-African coast to the east and by a Persian trough extending from the monsoonal low through the Persian Gulf to the northeast Mediterranean and Turkey (Alpert et al. 1992). The synoptic-scale near-surface wind direction is generally westerly and is modulated by the sea breeze. This results in a westerly to northwesterly wind flow of nearly 7 m s$^{-1}$ during the daytime. In addition to the synoptic flow, the westerly daytime sea breeze is accelerated by anabatic winds forced by the inland mountainous regions (Alpert and Rabinovich-Hadar 2003; Doron and Neumann 1977; Berkovic and Feliks 2005). The easterly land breeze develops along the coast, because of the sea–land differential cooling after sunset. As the night progresses, katabatic flow originating on the western slopes of the mountain ridges (~30 km inland in the south of Israel and adjacent to the coast in the north) reaches the coast, increasing the wind intensity.

4. Verification of CFDDA for the eastern Mediterranean

This paper assesses the success of the downscaling system by comparing its climatographies to various datasets that have not been assimilated during the model integration. The verification focuses on the ability of the downscaling system to represent the frequency distribution of atmospheric states in addition to the simple time means.

Because precipitation data are not assimilated and many aspects of model physics typically need to operate properly for precipitation to be correctly simulated,
verifying the simulated precipitation fields is a good test of the CFDDA system’s ability to define unobserved or poorly observed fields. We verify the spatial distribution of monthly multiyear totals, the interannual variability of regional precipitation totals, and the frequency distribution of daily rainfall totals against gridded and gauge rainfall totals. Also evaluated are the CFDDA climatology of the 10-m winds over the Mediterranean Sea using NASA Quick Scatterometer (QuikSCAT) measurements (OBS). Coastal wind tower observations are also used for verification.

a. Large-scale circulation

Before focusing on the details of the validation of the CFDDA high-resolution simulations, it is important to assess the ability of the modeling system to correctly represent the large-scale features of the circulation over this particular region. To accomplish this, the 6-hourly real-time analyses from the ECMWF have been used to provide an estimate of the large-scale circulation. This ECMWF operational analysis is available at 1.125° × 1.125° resolution. Figure 3 displays the anomaly correlation between the ECMWF analysis and the corresponding CFDDA-D1 fields interpolated to the ECMWF grid. Correlations are plotted every 6 h for each January and July for 1998–2007 for the 700-hPa temperature and relative-humidity fields.

The large-scale thermal and moisture patterns simulated by the CFDDA modeling system closely correlate with those defined by the ECMWF analysis. The anomaly correlation values for 700-hPa temperature (Figs. 3a,b) average 0.985, with higher overall values in winter than in summer. The same seasonal trend is true for the anomaly correlations for 700-hPa relative humidity (RH; Figs. 3c,d), which has an average of 0.83. Higher in the troposphere, anomaly correlation values (not shown) are even larger. The anomaly correlations for each year of the simulation are displayed separately, and it can be seen that there is a tendency for higher correlations to exist for later years. This is probably a consequence of the increased number and quality of observations assimilated in both the ECMWF and CFDDA analysis and improvements in the ECMWF analysis cycle. For January 2006, a CFDDA simulation with no assimilation of point observations was conducted. The anomaly correlation values decreased to 0.85–0.9 for the 700-hPa temperature and to 0.55–0.7 for the RH. This result emphasizes the importance of the FDDA portion of
the system for constraining the simulation of the large scales.

b. Monthly precipitation climatography

The model-derived monthly rainfall totals are compared to three different gridded estimates of monthly rainfall. For gauge-only estimates, we employ the Global Precipitation Climatology Centre (GPCC; Beck et al. 2005) dataset, which provides monthly precipitation totals from 1901 to the present on a 0.5° × 0.5° global grid but only over land. Because a considerable fraction of the CFDDA domain is located over the sea, we also utilize the merged Tropical Rainfall Measuring Mission (TRMM) product (3B43; Huffman et al. 2007). These data combine precipitation estimates from multiple satellites (retrievals from measurements in the microwave and infrared regions of the spectrum) as well as gauge-based analyses on a 0.25° × 0.25° grid that extends from 50°N–50°S for the period from 1998 to the present. Finally, following the work of Ebert et al. (2007), we also use precipitation derived from the short-range forecasts (i.e., the accumulated precipitation during the forecast period between 12 and 36 h) from the ECMWF model, available daily on a 1.125° × 1.125° global grid. These daily estimates were aggregated to produce a monthly rainfall total and have an accuracy that is comparable to that of satellite estimates, especially in winter and at higher latitudes (Ebert et al. 2007). The focus here is mainly in the geographic distribution of rainfall; therefore, we have chosen not to interpolate the data from the various sources to a common grid. A more quantitative analysis will be done later.

Figures 4a–c show the average (1998–2007) January precipitation amount based on the three observational datasets described above. Large precipitation maxima over land are seen along the coastlines of the Levant,
Turkey, and along the west coast of the Balkan states. There is also precipitation over the Mediterranean Sea but in various degrees from TRMM (Fig. 4b) or ECMWF (Fig. 4c). A distinct line of maximum rainfall is also apparent along the Fertile Crescent. The precipitation gradient is very large perpendicular to the North African coastline, with values of about 50 mm along the Libyan coast, decreasing to less than 5 mm across the interior just 100 km inland. The gradient is somewhat less along the Levant and Turkey coasts, but it is still large. Overall, the precipitation estimated by CFDDA (Fig. 4d) represents most of these spatial patterns, except for the observed coastal maximum over Algeria and France (upper-left edge of the domain). There is less precipitation estimated by CFDDA over the central and southern Mediterranean than is estimated by TRMM (more tan color, Fig. 4b). However, verification of precipitation over the Mediterranean Sea is problematic because the TRMM dataset has a known negative bias to low-precipitation rates over the oceans (Huffman et al. 2007).

During July (Fig. 5), observations show areas of enhanced precipitation along the eastern Black Sea coasts of Turkey and Georgia. A secondary region is seen over Romania, extending northward and southwestward to the Balkan countries. The largest maximum is located over the Alps, along the northern Italian border. All observational estimates show no significant rainfall below about 35°N and considerably more rainfall over land than over oceanic regions. Overall, a very similar pattern is seen in the CFDDA rainfall climatography (Fig. 5d). However, the model rainfall has more small-scale variability than that observed by the blended TRMM product that is shown at a similar spatial resolution (Fig. 5b). For example, in the CFDDA-derived rainfall climatography, a series of northwest to southeast oriented areas of maximum precipitation are seen over Romania and southern and western Ukraine. Similar maxima, but more...
widespread in character, are seen in the GPCC and TRMM climatographies (Figs. 5a,b, respectively) and a similar pattern is seen in the ECMWF-forecasted rainfall (Fig. 5c).

c. Variability of daily rainfall

Daily rainfall totals simulated by the CFDDA system are verified against a database of daily precipitation totals from rain gauge stations for the geographic region covered by D3. The locations of the rainfall stations are shown in Fig. 1b. The data for a majority of these rain gauge observations were obtained from the Israeli Ministry of Agriculture. The dataset was augmented (in particular for the stations over Lebanon) with daily rainfall totals from the National Oceanic and Atmospheric Administration global surface summary of the day (GSOD). The analysis is conducted with only stations that have data for at least 70% of the days in the 10 Januarys covering the period from 1998 to 2007.

Rainfall in this region during the summer months is virtually nonexistent; thus, we have limited our analysis to the winter season.

During winter, a marked north–south gradient in rainfall exists in D3, with the larger totals in the northern Levant and lesser amounts toward the south. This gradient is mainly a consequence of the preferred path of extratropical cyclones (Alpert et al. 1990; Trigo et al. 1999; Schädler and Sasse 2006). The rainfall stations have been classified into three groups based on this north–south contrast. A summary of the rainfall statistics for each group of stations is presented in Table 2, with the lowest observed monthly rainfall totals in southern Israel (stations 1–6) and the largest observed values in northern Israel (stations 7–12). The CFDDA-D3 and the ECMWF operational 12–36-h forecast daily rainfall totals have been interpolated to the station locations for those days with observed values—only nonzero precipitation days in both models and observations are used in the statistics and subsequent frequency distributions. The simulated CFDDA monthly rainfall totals compare well with those observed over the two regions in Israel, but they overestimate the total over Lebanon. The ECMWF forecasts underestimate rainfall totals in all three regions. The mean daily total rainfall for rainy days is underestimated by CFDDA for Israel but values are better than those from ECMWF in all three regions. Table 2 also presents the percentage of “drizzle” days with precipitation of less than 0.1 mm but not including days with zero precipitation. This varies from 13% in southern Israel to almost zero over Lebanon. This value is reasonably well simulated for Israel by both CFDDA and ECMWF, but both models severely overestimated this percentage for Lebanon. Similar behavior is seen in other weather and climate models (Chen et al. 1996; Knievel et al. 2004).

The frequency distribution of observed daily rainfall totals was compared for each group of stations to the distributions from the CFDDA simulations and ECMWF forecasts. For southern Israel (Fig. 6a) the observed distribution shows that most of the rainfall events result in daily totals of 4–8 mm; for the northern Israel stations (Fig. 6b) and the stations in Lebanon (Fig. 6c) the distributions are shifted toward 8–16 mm of daily rainfall. The northern Israel stations show a few events (~10) with daily rainfall totals above 64 mm, with fewer such events in southern Israel and Lebanon. The CFDDA distribution of daily rainfall totals in categories above 1–2 mm is similar to the observed distribution; however, CFDDA consistently overestimates the number of light rainfall events (0.1–1 mm) in all regions, perhaps by a similar mechanism to that responsible for the overprediction of drizzle days over Lebanon. The consistency between CFDDA and the observed distribution is particularly good for the northern Israel stations (Fig. 6b), with errors of less than 15 events in all classes above 2 mm. In Lebanon, the CFDDA system overestimates the values in the higher-intensity classes (>16 mm). This may be due to a lack of representativeness of the observed rainfall totals in stations located in complex terrain (Chen et al. 2002). Furthermore, rainfall totals for the Lebanon stations were derived from the GSOD archive, which has only minimal quality control of the measurements. Similar error statistics have been found in other MM5 studies at similar resolutions over complex terrain (Colle and Mass 2000; Mass et al. 2002).
The rainfall from the ECMWF forecasts consistently has excessively high frequencies in the low-rainfall classes and too low frequencies for the higher precipitation amounts for all station groups. Daily amounts above 32 mm, which are not rare in the observed or CFDDA distributions, are almost nonexistent in the ECMWF daily rainfall totals. Therefore, although the monthly rainfall values forecast by ECMWF are reasonably close to those observed, the rainfall occurs too often and with a lower-than-observed intensity.

d. Interannual variability of rainfall totals

The gridded January rainfall data from observations and models are compared in terms of their interannual variability for two geographic regions: the Levant (31°–36°N, 36°–37°E) and southwest Turkey (36°–38°N, 25°–32°E), which are delineated by black boxes in Fig. 4a. These areas were chosen because of the large prevailing rainfall amounts in January (Figs. 4b,c). The area-averaged rainfall for these two regions during all Januaries for the period 1998–2007 is displayed in Fig. 7. Because the CPC rainfall estimate is available for land only, and TRMM data are available for land and water, for CFDDA we display both the values averaged over all grid points and over land-only grid points. In most years, these two values are very similar over coastal Turkey but the averages over the Levant using D2 rainfall data differ significantly. Over the Levant, the CFDDA-D2 and ECMWF rainfall totals are consistently larger than the TRMM estimates for all years. Over southwest Turkey, the TRMM estimates are often larger than the CFDDA and ECMWF values. The CFDDA-D2 precipitation values are uniformly larger than those for D1, for both regions and all years. Considerable interannual variability is seen in the observed precipitation in both regions, and the variation from one year to the next in both regions is well represented in the CFDDA and ECMWF rainfall.

e. Near-surface winds over the ocean

To verify the CFDDA-simulated low-level winds over the ocean, we use data from the NASA QuikSCAT satellite. The SeaWinds instrument on QuikSCAT is an active microwave radar that measures the backscatter from surface ocean waves, and winds can be obtained in
all conditions except for moderate to heavy rain (Hoffman and Leidner 2005). A function is used to relate the measured backscatter to the 10 m MSL neutral stability equivalent winds. The rms differences between quality-controlled ship winds and QuikSCAT winds are typically about 1 m s$^{-1}$ in speed and 15° in direction (Bourassa et al. 2003). Very high- and very low-speed winds have problematic retrievals. In particular, low speeds result in very poor directional skill, while winds above 25 m s$^{-1}$ are often underestimated (Hoffman and Leidner 2005). Also, because of the previous assumption of neutral stability, winds of less than 6 m s$^{-1}$ are underestimated by 20%–30% under unstable conditions, which are often found over the Mediterranean during winter (Zecchetto and De Biasio 2007).

In the validation presented here, we use the QuikSCAT level-3 daily, gridded ocean wind vectors. These data are provided on an approximate 0.5$°$ × 0.5$°$ global grid with separate maps for the ascending (0600 LT equator crossing) and descending (1800 LT equator crossing) passes. The data are available for the period 2000–present. The CFDDA 10-m winds for each domain are linearly interpolated to each individual QuikSCAT wind vector location, and matched to the closest valid time of the observation. Figures 8 and 9 compare the 10-m wind speed composites from QuikSCAT and CFDDA-D2 for January and July 2000–07, respectively. During January,
the observed wind climatology is dominated by the northwesterly Etesian winds (Ziv et al. 2004) in the Levantine Basin and the mainly northerly flow in the Black Sea area. During winter, the winds in the entire region are characterized by low steadiness (a measure of the constancy of the wind direction) resulting from the frequent passage of fronts (Zecchetto and De Biasio 2007). The observed mean 10-m wind speed is 6–10 m s\(^{-1}\) across the Mediterranean Sea (Fig. 8a), with speeds decreasing toward the southeast. Over the Black Sea, wind speeds range from up to 10 m s\(^{-1}\) in the west to 4 m s\(^{-1}\) in the east. The CFDDA 10-m winds (Fig. 8b) display a similar geographic distribution but the speed is underestimated by 1–2 m s\(^{-1}\). In summer (Fig. 9), when conditions over the Mediterranean Sea are dominated by steady local winds, the CFDDA system is able to capture very well the wind distribution observed by QuikSCAT. The area of wind funneling through the strait separating the islands of Crete and Rhodes (Zecchetto and De Biasio 2007) is clearly seen in the CFDDA-derived analysis. Mean absolute differences between QuikSCAT and the model are mostly below 1 m s\(^{-1}\) over more than 90% of the ocean grid points in D2.

To further examine the model-simulated winds as they relate to those observed by QuikSCAT, we show wind roses of model-derived 10-m winds and observed winds for two 5° × 5° regions in the EM. Solid lines in Fig. 8c show the two regions. In Fig. 10, the wind roses for January 2000–07 (combining both QuikSCAT ascent and descent times) for the area 32°–37°N and 25°–30°E (left box in Fig. 8c) are shown for QuikSCAT and CFDDA-D2. Observed winds are mostly from the west and northwest, with the most frequent high speeds (averaging 9 m s\(^{-1}\)) being from the northwest. The roses generated from the CFDDA-D2 winds show very similar directional frequency distributions but the model underestimates the frequency of the strongest wind class. Almost identical results are seen in the CFDDA-D1 winds (not shown). In Fig. 11, the wind roses generated with data from July 2000–2007 for a region closer to the EM coast (32°–37°N, 30°–35°E, right box in Fig. 8c) are shown. This time, separate wind roses are shown for the morning (ascent) and evening (descent) overpasses to isolate diurnal effects. During this time of the year, westerlies dominate with stronger winds more likely in the evening than in the morning. The CFDDA-derived winds reproduce this diurnal variation quite well, especially at higher resolution (Figs. 11c,f). As in winter, the model underestimates the speeds. However, the differences are more evident in the morning than in the evening. Because very few observations are available over the sea (Fig. 2a), only a limited impact of data assimilation is expected in these regions.

f. Boundary layer winds over land

Here, the CFDDA wind speed distributions for January and July are compared to observations on two towers. Data from these towers were not assimilated in the model integration. In addition, we verify the winds from the 6-hourly real-time ECMWF analyses at a resolution of 1.125°. The aim of this comparison is to assess the capability of CFDDA to reproduce wind speed statistics that are relevant to wind power generation applications.

The observations used in the verification consist of winds measured at the two locations depicted in Fig. 1b. Hourly observations at 65 m AGL were available for January and July 1999–2007 at station A, and half-hourly observations at 60 m AGL were available for January and July 1998–2007 at station B. Both stations are situated at a distance of about 8 km inland from the coast,
in areas characterized by a mixed canopy consisting of urban buildings, shrubs, and trees, with a roughness length of approximately 0.3 m. Station A is located in an area of rather flat terrain—the tower base is 62 m MSL and the coastline is relatively straight. Station B is situated in a region of complex terrain and complex coastline configuration and the tower base is at 60 m MSL. The latter area has been described in Rostkier-Edelstein and Givati (2003). Although neither of these observations is located in open rural areas suitable for wind power production, we opted to verify CFDDA against them because they provide a high-quality multiyear dataset at heights relevant to wind power applications.

The CFDDA-derived winds from D3 were interpolated to the location of the observations. The CFDDA terrain heights at the location of the stations are 36 m and 51 m MSL, for stations A and B, respectively. The lowest three CFDDA levels are located at approximately 15, 57, and 113 m AGL at both locations, and winds were vertically interpolated with respect to height using logarithmic weights, between the latter two levels, to the height of the masts. Therefore, the height AGL of the observed and interpolated wind is the same despite the elevation difference of the model terrain and the real terrain. Land grid squares (mainly urban type, with assigned roughness length of 1 m) surround the location of station A in D3. In contrast, station B is surrounded on three sides by land grid squares, except for the northwest quadrant, where over 40% of all modeled winds originate, which is an ocean grid square (with an assigned background roughness of 0.001 m).

Similarly, ECMWF analysis winds archived on the 1.125° grid were interpolated to the stations locations. Because of changes in model resolution, the terrain heights of the ECMWF analyses at the observation locations changed during the study period. The ECMWF terrain height values were appreciably higher than the actual ones (159–214 m MSL and 171–216 m MSL for stations A and B, respectively). The ECMWF wind values used for comparison with the observations are based on log–height interpolation from the pressure levels of 1000 and 925 hPa. The model that generated the operational ECMWF analysis system used here underwent changes in horizontal and vertical resolution, model physics, and data assimilation during the period 1998–2007. Therefore, no temporal consistency is expected, but the wind fields provide the “best” baseline analysis at the time.

Figure 12 displays the wind speed distributions (for all 6 hourly data) based on CFDDA-D3, ECMWF, and observations for both locations during January and July. Both models reproduce the observed differences in the wind speed distributions between January and July. The results for January (Figs. 12a,b) are characterized by
a wide distribution of wind speeds (relative to those for July) with a high-speed tail showing a few observed events that exceed 10 m s$^{-1}$. These events are associated with strong pressure gradients within extratropical cyclones. The periods between the cyclonic activity, which are mainly dominated by Red Sea troughs and winter highs, lead to weaker winds. The observed distributions for this month show no significant differences between the two sites, in spite of the differences in the orography and coastline characteristics. In winter, in the absence of strong solar heating, no pronounced land–sea temperature gradients develop and local thermal circulations are likely weak. Near-surface winds are thus dominated by the synoptic flow. Both models are able to reproduce the broad observed distributions. However, there is some model-dependent disagreement between the modeled and observed distributions, and some differences exist between the model solutions for the two sites. The ECMWF analysis tends to overestimate the frequency for the low-speed part of the January distributions at both sites and underestimates the frequency at higher speeds. The PDFs for CFDDA show underestimation of the speeds in the middle of the distribution and overestimation for winds above 7 m s$^{-1}$ at both sites. The CFDDA terrain and coastline (not shown, but partly described previously) is not able to correctly represent the ridge and bay in the vicinity of station B, as well as the correct distance to the coast. As a result, the site is located much closer to the coast in the model than in reality and is thus affected by lower roughness upstream during periods of strong winds.

The observed wind distributions for July (Figs. 12c,d) are narrower than for January, with speeds that do not exceed 11 m s$^{-1}$. This is a result of the summer weather

![Wind roses generated from July 2000–07 winds derived from (a) QuikSCAT ascent retrievals (OBS), (b) CFDDA-D1 (45 km) ascent times, (c) CFDDA-D2 (15 km) ascent times, (d) QuikSCAT descent retrievals (OBS), (e) CFDDA-D1 (45 km) descent times, and (f) CFDDA-D2 (15 km) descent times. The winds are restricted to the area between 32° and 37°N and between 30° and 35°E (right box on Fig. 8c). The wind rose radius represents the wind frequency (%) for each sector and speed range in the grayscale bar.](image)
regime that is characterized by weak westerly to northwesterly large-scale flow (Section 3), upon which sea–land breeze and terrain-induced circulations are superimposed. The observed distributions show some differences between the two nearby sites, presumably because of the local orography and coastline. For example, while station A (Fig. 12c) shows a maximum frequency for speeds of about 3 m s\(^{-1}\), station B (Fig. 12d) shows a bimodal behavior with maximum frequencies for speeds of 1–2 m s\(^{-1}\) and 4–6 m s\(^{-1}\). For summer, the ECMWF product overestimates some of the low- and medium-speed categories and underestimates higher speeds. The CFDDA wind distribution for station A is too flat, greatly underestimating the speeds in the 3–5 m s\(^{-1}\) range. For station B, the modeled distribution is insufficiently flat. The model does produce a bimodal distribution but the minimum is at too high a speed.

To better understand the role of the land–sea breeze in generating the observed wind distributions, and to assess the ability of the models to reproduce it, Fig. 13 displays the observed and model histograms for wind speeds at station A for 0000, 0600, and 1200 UTC for July. These times are not necessarily ideal for showing the phases of the land–sea breeze but rather are dictated by the availability of the ECMWF data. Figure 13a (0300 LT) shows the wind speed distributions at a time dominated by the land breeze, which is normally weaker than the sea breeze. The observed wind distribution shows a significant frequency of calm winds (<1 m s\(^{-1}\)) and speeds never exceed 6 m s\(^{-1}\). The ECMWF-derived winds fail to reproduce the calm speeds and most of the events fall in the 2–4 m s\(^{-1}\) category. In contrast, CFDDA-derived winds better capture the distribution. The coarse resolution of the ECWMF analyses most likely masks the land breeze
and terrain-induced flow at this time. In the absence of these, no opposing flow weakens the large-scale westerly to northwesterly flow. At 0900 LT (Fig. 13b), the sea breeze begins to develop, with higher observed wind speeds. The ECMWF winds remain rather similar to those at 0300 LT. In contrast, CFDDA winds correctly shift toward higher speeds but the peak is not sufficiently large. At 1500 LT (Fig. 13c), all three distributions show a narrow profile, with speeds in the range 3–9 m s$^{-1}$. The winds derived from the ECMWF analysis still under estimate the observed frequency peak and the higher speeds.

5. Summary and discussion

We present a method for generating high-resolution regional climatographies through the use of a mesoscale model driven by R2 and the assimilation of surface and upper-air observations by Newtonian relaxation. The results of 10 yr of simulations for January and July are verified in terms of the horizontal and seasonal distribution of rainfall and the near-surface wind field, both over the ocean and at two land sites.

The verification of monthly rainfall shows that CFDDA is able to correctly capture most of the geographic features and interannual variability of the precipitation field in this region. Over the oceans, CFDDA appears to underrepresent the monthly rainfall; however, the error in the CFDDA estimates as compared to those from TRMM is within the uncertainty of these satellite estimates over the oceans. Daily rainfall totals from the CFDDA system are verified against rain gauge observations, and the results show remarkable agreement both in terms of the mean amount but also in terms of their temporal distribution for three distinct regions. As seen in many other models, the CFDDA system appears to overestimate the amount of precipitation over higher elevations. It is unclear what the origins of this problem in the model simulation are, or whether the observations underestimate the true amount of rainfall of this region by undercatchment errors (Colle and Mass 2000) or preferential placement of the verifying gauges in lower elevation valley locations. Nevertheless, the high quality of the depiction of the rainfall distribution in most places by the CFDDA system demonstrates its possible usefulness in coupling to hydrological models.

Fig. 13. Frequency distributions of 6-hourly wind speeds from observation and simulated by the CFDDA-D3 (5 km) and ECMWF forecast model from station A during July 1999–2007 for (a) 0000, (b) 0600, and (c) 1200 UTC.
Wind roses constructed with wind vectors from CFDDA show considerable accuracy when compared with winds derived from QuikSCAT observations over the EM Sea. Although the directional distribution of winds is well represented, wind speeds simulated by the CFDDA system over the oceans are consistently slightly lower than those observed by QuikSCAT. The surface-layer formulation within the CFDDA system uses Charnock’s formula as written by Delsol et al. (1971) to compute the surface roughness over the ocean as a function of the wind speed. However, the constant used in MM5 is twice as large as that based on oceanic data (Wu 1969). Colle et al. (2003) reported similar biases in their verification of wind speeds over the Gulf of Mexico and the eastern Atlantic.

The comparison between CFDDA, ECMWF, and observed wind speed distributions at two mast locations is accomplished through the construction of wind speed histograms for observed and simulated data. In this comparison, representativeness error is not accounted for. Clearly, both the ECMWF and CFDDA analyses fail to truthfully represent the terrain (mainly at station B), the coastline (mainly at station B), and the roughness characteristics of the areas surrounding the masts observations. Wind measurements may be exposed to local influences limiting their spatial representativeness to a radius of only a few hundreds of meters. Some estimates of representativeness error for near-surface winds computed with a model grid increment of about 1 km lead to a value of about 1 m s$^{-1}$ for areas of complex terrain (e.g., Rife et al. 2004; Steinacker et al. 2000). Landberg et al. (2003) discussed the nature of these local effects and described methods to overcome them for wind resource estimation. These methods rely on dynamical downscaling using a mesoscale model and further fine tuning of the wind climatographies with a microscale model (Frank and Landberg 1997).

The verification results, despite the discussed caveats, confirm the potential of using CFDDA reanalyses as a tool for wind power resource estimation. Assimilated surface and upper-air observations are sparse in the vicinity of the masts, thus, showing the capability of the mesoscale reanalysis to reproduce near-surface wind statistics in poorly observed areas. To the best of our knowledge, this is one of the few published verifications of dynamical downscaled winds over land at turbine height, in the context of wind resource assessment. It demonstrates the benefit gained by high-resolution mesoscale reanalysis as compared with coarse-resolution global analyses, in particular when the circulation is dictated by mesoscale mechanisms. Some of the limitations encountered may be overcome by optimization of the CFDDA model parameters for the specific domain, for example, terrain, coastline, surface canopy, and specific parameters in the physical process schemes. Others limitations are imposed by the model horizontal grid size, that is, the representativeness errors due to local effects. These can be overcome by increasing the horizontal resolution further or by postprocessing the CFDDA reanalysis with statistical corrections or microscale modeling.

Many factors play a role in the quality of the mesoscale reanalysis system: the quality of the mesoscale model and its physical parameterizations, the quality of lateral and lower boundary conditions, the quality and spatial density of observations used in the data assimilation, the increased horizontal and temporal resolution, etc. It is beyond the scope of the paper to examine the individual contribution of each of these, but we believe all are necessary components of a credible system.

In the present study, no attempt has been made to optimize the model parameterizations for the weather conditions and climate of the EM. The choice of these is based on our experience from real-time mesogamma-scale weather forecasting in the United States (Liu et al. 2008). That type of optimization process, which is beyond the scope of this work, would very likely improve the accuracy of the climatography derived from the CFDDA system.

Acknowledgments. Andrew Negri of NASA provided assistance relative to the use of TRMM data. Mark Perel from the Israeli Ministry of Agriculture is thanked for providing the data from the rain-gauge stations. Dr. Ilan Setter from the Israeli Meteorological Service is thanked for providing the masts wind data. Gregory Roux and Ming Ge are thanked for their help with data processing and Jason Knievel is thanked for his help editing the manuscript. The suggestions of three anonymous reviewers greatly improved the quality of the manuscript. We thank the Army Testing and Evaluation Command (ATEC) and the NGIC for providing the financial support for this study. Dorita Rostkier-Edelstein acknowledges the NCAR Research Applications Laboratory visitor funds. The CFDDA system simulations were performed using NCAR/CISL supercomputers.

REFERENCES


970 JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY VOLUME 49


