A Long-Term WRF Meteorological Archive for Dispersion Simulations: Application to Controlled Tracer Experiments

FONG NGAN
NOAA/Air Resources Laboratory, and Cooperative Institute for Climate and Satellites, University of Maryland, College Park, College Park, Maryland

ARIEL F. STEIN
NOAA/Air Resources Laboratory, College Park, Maryland

(Manuscript received 18 October 2016, in final form 26 May 2017)

ABSTRACT

A long-term archive of meteorological data using the Weather Research and Forecasting (WRF) Model was created to provide data that are compatible with the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) dispersion model and to serve as initial and boundary conditions for simulations at a finer resolution. On the basis of these WRF data, generated with a variety of planetary boundary layer (PBL) schemes and nudging options, the HYSPLIT model was run to simulate four controlled tracer experiments—the Cross Appalachian Tracer Experiment (CAPTEX), the Across North America Tracer Experiment (ANATEX), a 1980 release in Oklahoma City, Oklahoma (OKC80), and the Metropolitan Tracer Experiment (METREX)—covering different time periods with diverse durations, including a summer day, several days in autumn, three months during winter, and one full year, respectively. The evaluation of the WRF results utilizing conventional observations showed a similar statistical performance for the different PBL schemes. Given the limited information the meteorological evaluation alone can provide, the authors used the dispersion evaluation with measurements from multiple tracer experiments to gain further insight into the most appropriate WRF configuration to generate reasonable data for dispersion applications. The dispersion simulations that were based on WRF data generated equal or slightly better statistical performance than did those driven by the North American Regional Reanalysis (NARR) dataset. The statistical comparison showed a mixed impact for the dispersion results driven by the nonnudged and nudged WRF data. The main advantage of the WRF data is the availability of hourly meteorological data from 1980 to the present and the inclusion of additional variables that are relevant to atmospheric dispersion and are not available from NARR. This WRF dataset will be accessible online, providing additional capabilities for using different meteorological inputs and a variety of options to compute the HYSPLIT mixing parameters.

1. Introduction

NOAA’s Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model is one of the most commonly used tools to simulate the transport and dispersion of pollutants for a variety of atmospheric applications. An objective performance evaluation against independent measurement datasets, such as tracer experiments, is fundamental to assess the model reliability to simulate transport and dispersion features under different meteorological conditions. For this reason, NOAA’s Air Resources Laboratory (ARL) developed a Data Archive of Tracer Experiments and Meteorology (DATEM; online at http://www.arl.noaa.gov/DATEM.php) that consists of standardized software and uniformly formatted data that include emissions, concentration measurements of chemical tracers, and meteorological inputs corresponding to multiple controlled tracer experiments (Rolph et al. 2017). DATEM provides a platform for HYSPLIT’s verification and development. In addition, the dataset allows the atmospheric-transport-modeling community to conduct various verification and sensitivity studies and to compare model results with each other on a common basis. HYSPLIT-compatible meteorological data for DATEM include the Global Reanalysis (Kalnay et al. 1996) and the North American Regional Reanalysis (NARR; Mesinger et al. 2006). These meteorological
datasets have not only coarse spatial resolution but also a low temporal frequency—2.5° and 32-km horizontal grid spacing available every 6 and 3 h for the Global Reanalysis and the NARR, respectively. Furthermore, some meteorological variables, such as the momentum flux, are not available in the NARR dataset and need to be rediagnosed in HYSPLIT on the basis of other state variables. Inconsistencies between the predicted and the diagnosed variables may be translated into dispersion-modeling errors.

The meteorological data produced from the Advanced Research dynamic core of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) have been frequently used by the research community to drive HYSPLIT simulations (e.g., Hegarty et al. 2013; Hernández-Ceballos et al. 2014; Klich and Fuelberg 2014; Simsek et al. 2014; Ngan et al. 2015a). In this work, we run the WRF meteorological model to create a long-term archive for driving dispersion applications. We intend to produce WRF data with 27-km horizontal grid spacing and hourly temporal frequency from 1980 to 2016 and eventually to extend the simulation to the present day. This new dataset, named North American Reanalysis Data for Dispersion Applications (NARDDA), has been developed for two purposes: to provide meteorological data compatible with the HYSPLIT dispersion model and to serve as initial and boundary conditions for WRF simulations at a finer resolution for dynamic downscaling. NARDDA consists of hourly output from meteorological simulations that includes all required variables for the dispersion calculations in ARL HYSPLIT format. In addition, since the inline version of HYSPLIT coupled with WRF has shown to be especially advantageous for episodes that require a high spatial and temporal resolution (Ngan et al. 2015b; see online at http://www.arl.noaa.gov/WRF_inline.php), the NARDDA can be the platform to initialize simulations at finer horizontal and vertical grid spacing. The inline HYSPLIT coupled with WRF takes advantage of the higher temporal frequency of the meteorological variables, avoids temporal and vertical interpolation of the data, and uses WRF’s vertical coordinate, resulting in a more consistent depiction of the state of the atmosphere being available for the dispersion computation.

As we consider creating a WRF dataset for dispersion applications that focus on near-surface pollutant releases, our main emphasis is on an accurate prediction of the planetary boundary layer (PBL) and its characteristics. Prior studies have shown that no PBL parameterization in the WRF Model can be singled out for better predicting the PBL wind and depth (Reen et al. 2014; Peltier et al. 2010; Angevine et al. 2014). Among many model variables, PBL wind and depth are two of the most fundamental meteorological inputs that determine the transport and mixing of tracer plumes in the atmosphere. Many studies conducted WRF sensitivity simulations with different PBL parameterizations and evaluated their performance on the basis of comparisons with meteorological variables such as wind, temperature, and precipitation (Reen et al. 2014; Peltier et al. 2010; Angevine et al. 2014; Kleczek and Holtslag 2014; Shin and Hong 2011). Meteorological evaluation using conventional observations only provides a minimum level of performance assessment and does not necessarily reveal the quality of the dispersion simulation driven by the corresponding meteorological inputs. In general, mixing and stability parameters, which are essential for the dispersion calculation, are not evaluated because they are usually not available. Therefore, a statistical evaluation of readily available meteorological variables, such as wind and temperature, alone is not expected to be sufficient to decide which meteorological dataset will produce better dispersion results. In this work, we included meteorological evaluations for wind and temperature but also focused on a dispersion evaluation for multiple tracer experiments to assess the most appropriate model configuration for dispersion applications.

A technique that can be applied for improving the quality of the predicted fields generated by the WRF simulation is four-dimensional data assimilation (Deng et al. 2009), also called nudging. Hegarty et al. (2013) demonstrated the benefit of using nudged meteorological data in tracer dispersion calculations, and other studies in air-quality modeling also showed the positive impacts of using nudging in WRF simulations for predicting ozone and fine particular matter (Rogers et al. 2013; Gilliam et al. 2015; Li et al. 2016). In addition to analysis nudging, which is commonly recommended for meteorological simulations that drive air-quality modeling, we also included wind and temperature observations to reduce the growth of model errors through observational nudging. We examined the effect that the observational-nudged meteorological inputs produced on the dispersion simulations by comparing the modeled tracer concentrations driven by nonnudged and nudged data.

The objective of this work is to generate a WRF configuration that is tailored for dispersion applications on the basis of statistical evaluation of different controlled tracer experiments, and the end product is a long-term archive of WRF data that is available for HYSPLIT modeling. Toward that end, the WRF runs were set with different PBL schemes and nudging options to simulate the meteorological conditions corresponding to four tracer experiments: the Cross Appalachian Tracer Experiment (CAPTEX), the Across North America Tracer Experiment (ANATEX), a 1980...
release in Oklahoma City, Oklahoma (OCK80), and the Metropolitan Tracer Experiment (METREX). HYSPLIT was set to use the hourly WRF data and to run forward in time to simulate the four controlled tracer experiments. The meteorological results were compared with wind and temperature observations at the surface and upper levels, and the dispersion results were evaluated against tracer concentration measurements. The WRF dataset created in this study will be available, alongside NARR data, to provide additional options for HYSPLIT users to perform dispersion modeling with different meteorological data and a variety of HYSPLIT setups for mixing calculations.

Section 2 presents a brief overview of all of the controlled tracer experiments that were used in this study. Model configurations for WRF and HYSPLIT, as well as simulation designs for different sensitivity cases, are presented in section 3. Section 4 describes the observations used in this study and the statistical evaluations for the meteorological and the dispersion results. The results and evaluations for the tracer experiments are shown in section 5. Section 6 presents the conclusions.

2. Experimental data description

Four tracer experiments that were conducted in the 1980s were simulated with HYSPLIT. A brief overview of each experiment is presented below. The release locations and sample network for each experiment are shown in Fig. 1. Other tracer experiments in DATEM that were performed in the 1970s, for which different boundary conditions were needed or that took place outside the North American domain, are not included in this study.

a. CAPTEX

CAPTEX (Ferber et al. 1986) used perfluoromonomethyl-cyclohexane (PMCH) as a tracer to understand long-range transport and diffusion of pollutants. This experiment lasted from mid-September to the end of October of 1983. Six 3-h releases took place from Dayton, Ohio (DAY; marked as C* in Fig. 1), during the afternoon (releases 1–4) and from Sudbury, Ontario, Canada (SUD; marked as C in Fig. 1), during the nighttime (releases 5 and 7). The tracer concentrations were measured at 84 ground-level stations distributed 300–800 km from the source at 3- and 6-h average intervals.
b. ANATEX

ANATEX (Draxler and Hefter 1989) used two inert tracers: perfluoro-trimethyl-cyclohexane (PTCH) released from Glasgow, Montana (GGW; marked as A* in Fig. 1), and perfluoro-dimethyl-cyclohexane (PDCH) emitted from Saint Cloud, Minnesota (STC; marked as A in Fig. 1). The releases started on 5 January and ended on 29 March 1987. The release times alternated between afternoon and nighttime so that there were 33 releases at each location over 2.5-day intervals. The sampling network consisted primarily of rawinsonde stations with a total of 77 sites over the eastern half of the United States and southern Canada. The concentration samples were taken in 24-h averages collected at 1400 UTC daily.

c. OKC80

On 1900 UTC 8 July 1980, a 3-h-duration single release of PMCH and PDCH tracers took place from Oklahoma City (OKC80; Ferber et al. 1981). Only the PMCH episode is included in this study since the two tracers shared the exact same pathway. The release location is marked as “O” in Fig. 1. The sample network included two arcs over the downwind area from the release location that measured tracer concentrations from 8 to 11 July. The first arc consisted of 10 sites located at a distance of 100 km (north of Oklahoma City) collecting 10 consecutive 45-min samples starting 2 h after the release. The second arc included 35 sites located at a distance of 600 km (in Missouri and Nebraska) taking 3-h average samples beginning 13 h after the tracer release.

d. METREX

During METREX (Draxler 1987), inert tracers (PMCH and PDCH) were released at two locations, Rockville, Maryland (RMD; marked as M* in Fig. 1), and Mount Vernon, Virginia (MVA; marked as M in Fig. 1), about 20 km outside Washington, D.C., at 36-h intervals alternating between nighttime (starting at 0300 UTC) and daytime (starting at 1500 UTC). METREX started at the end of December 1983 and ran through one whole year. The sampling network collected data at one urban and two suburban sites with 8-h average intervals. Even though METREX covered a relatively smaller spatial scale, it includes a longer period of time than other experiments, providing a varying range of weather conditions that could, in principle, produce a wide variety of plume scenarios.

3. Model configurations

a. WRF

WRF, version 3.5.1, was configured using a domain with 27-km grid spacing that covers the contiguous United States (Fig. 1). A total of 33 vertical layers were used, with the highest resolution near the surface and 100 hPa for model top. The thickness of the lowest layer was around 16 m, and 20 layers were included below 850 hPa (~1.5 km). The initial conditions (IC) and lateral boundary conditions (LBC) for the WRF simulation originated from the NARR, which is available every 3 h at a 32-km spatial resolution. The model was initialized every day at 0600 UTC, and the first 18 h of spinup time in the 42-h simulation were discarded. Thus, the daily simulation started at nighttime when the PBL was stable. A conversion program was applied to the WRF hourly output to obtain the required meteorological data for the HYSPLIT dispersion modeling (Stein et al. 2015a). The WRF simulations are initially available for the four tracer experiments, but eventually the end product (NARDDA) will cover a period from 1980 to the present.

The physics options used in the WRF runs included the single-moment 3-class scheme for microphysics (Hong et al. 2004), the Rapid Radiative Transfer Model for longwave radiation (Mlawer et al. 1997), the Dudhia scheme for shortwave radiation (Dudhia 1989), the Grell–Freitas ensemble for the subgrid cloud scheme (Grell and Freitas 2014), and the unified Noah land surface model (Chen and Dudhia 2001). Table 1 lists all of the PBL schemes and their associated surface schemes used in the simulations. By examining the meteorological results in conjunction with their corresponding dispersion simulations, we intend to understand the sensitivity of the WRF predictions to different PBL parameterizations and the subsequent impact on the dispersion calculations.

Grid nudging was applied to all simulations. In addition, we conducted a separate set of WRF simulations with observational nudging to examine how the use of assimilated meteorological information affects the dispersion results. Surface and sounding data were used for improving the reanalysis data for IC/LBC through the objective analysis package (OBSGRID) available in the WRF Model system. This approach is expected to generate better IC/LBC for WRF simulations. OBSGRID also generates surface grid–nudging files and observational-nudging files. Thus, the nudged configuration ingested the improved IC/LBC files and ran with surface and observational nudging in addition to the grid nudging. The meteorological observations are available every 6 h (NOAA/NWS/NCEP 1980a,b), and variables used in this study include temperature, wind speed, and wind direction.

b. HYSPLIT

We performed HYSPLIT simulations driven by different meteorological data (i.e., WRF and NARR) for the four controlled tracer experiments introduced in section 2. The temporal frequency of the WRF data is 1 h,
and the NARR data are available every 3 h. The dispersion results were evaluated against tracer concentration measurements taken during these experiments. The tracer release height occurred at the surface for all four experiments. A concentration grid with 25-km horizontal resolution was set for the simulations with one vertical layer extending from 0 to 100 m above ground level except for METREX, which used 0.5 km in the horizontal plane and 0–50 m for the vertical level in the concentration grid. The time average for tracer concentration output varies among the different experiments according to the frequency of the measurements. Table 2 shows additional details about the dispersion-model setup, including the number of Lagrangian particles, release date, emission rates, and emission frequency.

4. Statistical evaluations

a. Meteorological

Statistical metrics computed for WRF results against surface and upper-level wind and temperature measurements will help us to identify the runs that more accurately reproduce the prevailing weather conditions and will point to possible errors that can be propagated into the dispersion modeling. With limited observations in time and space, however, this kind of statistical metrics can only reveal error information at certain levels. Meteorological variables relevant to the mixing of pollutants are usually not available in conventional observations. The conventional data available during the studied experiments are wind and temperature at surface stations and from soundings with 3- and 12-h intervals, respectively. Indeed, determining how well the model predicts the mixing in the atmosphere and which variables are the main drivers for the dispersion of pollutants cannot be fully assessed by merely comparing the modeled and measured winds and temperatures. Therefore, we intend to use the evaluations of the HYSPLIT results against measurement data obtained from multiple controlled tracer experiments to infer the adequacy of the meteorological data for driving dispersion calculations.

The statistical evaluation of the meteorological results is presented through a Taylor diagram (Taylor 2001).
that indicates the model performance in terms of correlation $R$, centered root-mean-square (RMS) difference, and standard deviation. Taylor diagrams provide a statistical summary for the intercomparison of multiple simulations by showing the agreement between predicted and measured fields (Katragkou et al. 2015). In this study, the centered RMS and standard deviation normalized by the observed standard deviation are shown on the diagram. Modeled patterns that agree well with observations lie within the 0.25 circle and imply relatively high $R$ and low centered RMS differences between predicted and observed values (e.g., Fig. 3, described in more detail below). Model results that lie on the dashed line (marked as “reference”) have a good normalized standard deviation. On the right side of the line the model data have a larger variation than the observations, whereas on the left side the model data have a smaller variability than the observations. The surface- and upper-level observations were obtained from the Research Data Archive at the National Center for Atmospheric Research (NOAA/NWS/NCEP 1980a,b). Surface data and soundings were available every 3 and 12 h, respectively. About 690 surface stations and 97 soundings were in the WRF domain (Fig. 2).

b. Dispersion

For the evaluation of the dispersion results, we adopted the cumulative score (rank) introduced in Draxler (2006):

$$\text{rank} = R^2 + (1 - |FB/2|) + (\text{FMS}/100) + (1 - \text{KSP}/100).$$

(1)

It consists of four equally weighted and normalized statistical parameters—correlation coefficient $R$, fractional bias FB, figure of merit in space FMS, and Kolmogorov–Smirnov parameter KSP. The rank ranges from 0 (worst) to 4 (best). We assessed the performance of the dispersion modeling results through the four components representing different aspects of the model errors, whereas the single parameter “rank” expresses the overall model performance. The rank has been the statistical measure used for evaluating results from different dispersion models in various research papers (e.g., Hegarty et al. 2013; Eslinger et al. 2016;
In this work, we use the rank so that the statistics presented here are consistent with those shown elsewhere.

For the intercomparison among different dispersion simulations, we also calculated the normalized root-mean-square error (NRMSE), which is the root-mean-square error of the tracer concentrations divided by the mean of the measured and predicted values:

\[
\text{NRMSE} = \frac{\text{RMSE}}{\sqrt{\frac{M}{P}}} = \sqrt{\frac{\sum (M - P)^2}{nM P}},
\]

where \( M \) represents the measured values and \( P \) is the predicted values. In addition to the FB in the rank, the NRMSE provides further insight into the accuracy of the dispersion simulations driven by different meteorological data. The tracer concentration measurements are, as of the time of writing, open to the public online (http://www.arl.noaa.gov/DATEM.php), and the number of measurements paired with model results for the statistical calculation is listed in \( \text{Table 2} \).

5. Results and discussion

a. Evaluation for CAPTEX

The meteorological evaluation of the WRF results in compared with surface wind speed and temperature observations for the period covering the six episodes for CAPTEX is summarized in the normalized Taylor diagram in Fig. 3a. Most PBL schemes generated similar statistical scores. Among all configurations, the “TEMF” parameterization (see \( \text{Table 1} \) for all configuration explanations) presented the worst performance in terms of normalized standard deviation (NSD) and the lowest correlation coefficient for both surface wind speed and temperature (Fig. 3a), with values even worse than those for NARR. The wind speed corresponding to the “BouLac” configuration showed a better NSD score but a relatively low correlation coefficient that resulted in an RMS difference similar to that of the NARR case. The other seven configurations clustered together on the diagram, indicating similar performance for simulating the wind fields. Furthermore, separating the \( u \) and \( v \) components of the wind in the Taylor diagram (not shown) does not provide additional information to distinguish the better-performing WRF configuration. All cases except the TEMF and “ACM2” configurations showed good surface temperature predictions.

Wind direction and PBL height are important parameters for dispersion modeling since errors in those parameters may lead to a misplacement of the plume and an increase or decrease in the tracer concentration. The potential temperature vertical profile (not shown) provides information about how the various PBL schemes predict PBL heights. In general, the largest differences in potential temperature among the different simulations were found near the surface, but similar values were obtained above 200 m. Most of the WRF runs agreed well on PBL height except two configurations: the “MYJ” and the BouLac for CAPTEX 1 and 7, respectively. For the comparison with sounding data, we interpolated model data to the height of the observations and calculated the statistics with data below 1000 m (\(~800 \) data points) and all levels (\(~3000 \) data points). On the Taylor diagram corresponding to upper-level wind speed and temperature (not shown), all WRF runs showed more spread when only using data below 1000 m than when all levels were considered in the statistical calculation. No major outliers, except the wind speed in BouLac, could be distinguished. This is an indication that the upper-level patterns modeled by the different meteorological simulations were very similar. This result is expected since the predicted fields are less sensitive to the PBL parameterization as the free troposphere is approached.

Since the meteorological evaluation exposed some limitations in terms of distinguishing the best WRF configuration for performing the multiyear simulation, we rely on the evaluation of dispersion results against tracer measurements to provide further insight into this assessment. HYSPLIT simulations for the six CAPTEX episodes were driven by NARR data as well as the nine WRF meteorological runs. \( \text{Table 3} \) shows the statistical rank for all runs. The best statistical score, in terms of the rank, was found to be one of the simulations driven by WRF data although the HYSPLIT result using the NARR data was not necessarily the worst when compared with the rest of the cases that used WRF meteorological inputs that are based on different PBL schemes. Note that in \( \text{Stein et al. (2015b)} \), the statistical significance between dispersion simulations was estimated by the uncertainty associated with the rank value. They found that the significant difference between two ranks should be larger than 0.07–0.11 for the six CAPTEX episodes. When we used all data points from the six episodes for computing the statistics of tracer concentrations (2281 data points), the best three ranks were those driven by ACM2, “UW,” and “GBM” meteorological data. Consequently, these three configurations, together with “YSU,” MYJ and BouLac, were used for the next tracer experiment. The dispersion simulations driven by these six WRF meteorological datasets showed higher statistical scores, with ranks between 2.52 and 2.66, than the result obtained using the NARR data. The lowest three scores for the rank corresponded to the “QNSE” (2.39), “MYNN2” (2.44), and

Leadbetter et al. (2015).
TEMF (2.33) parameterizations, and they were worse than or similar to the rank of the NARR run (2.43). Thus, they were excluded from the rest of this study. Even though the evaluations of surface wind speed and temperature for QNSE and MYNN2 showed statistical scores that were similar to those of other WRF simulations, they did not necessarily produce good dispersion results.

TABLE 3. Rank [the cumulative statistical score; Eq. (1)] corresponding to HYSPLIT model results for six CAPTEX tracer releases.

<table>
<thead>
<tr>
<th>Release</th>
<th>NARR</th>
<th>MYJ</th>
<th>YSU</th>
<th>ACM2</th>
<th>QNSE</th>
<th>MYNN2</th>
<th>TEMF</th>
<th>BouLac</th>
<th>UW</th>
<th>GBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>2.12</td>
<td>2.82</td>
<td>2.81</td>
<td>2.72</td>
<td>2.78</td>
<td>2.88</td>
<td>2.77</td>
<td>2.73</td>
<td>3.04</td>
<td>3.03</td>
</tr>
<tr>
<td>R2</td>
<td>2.79</td>
<td>2.96</td>
<td>2.95</td>
<td>2.97</td>
<td>2.77</td>
<td>2.74</td>
<td>2.29</td>
<td>3.07</td>
<td>2.97</td>
<td>2.96</td>
</tr>
<tr>
<td>R3</td>
<td>1.92</td>
<td>1.95</td>
<td>1.95</td>
<td>1.94</td>
<td>2.02</td>
<td>1.94</td>
<td>1.88</td>
<td>1.87</td>
<td>1.97</td>
<td>1.97</td>
</tr>
<tr>
<td>R4</td>
<td>2.18</td>
<td>2.08</td>
<td>2.10</td>
<td>2.02</td>
<td>1.95</td>
<td>2.12</td>
<td>2.06</td>
<td>1.95</td>
<td>2.18</td>
<td>2.12</td>
</tr>
<tr>
<td>R5</td>
<td>2.57</td>
<td>2.67</td>
<td>2.56</td>
<td>2.49</td>
<td>2.56</td>
<td>2.60</td>
<td>2.50</td>
<td>2.60</td>
<td>2.49</td>
<td>2.50</td>
</tr>
<tr>
<td>R7</td>
<td>2.24</td>
<td>2.15</td>
<td>2.29</td>
<td>2.24</td>
<td>2.17</td>
<td>2.30</td>
<td>2.19</td>
<td>2.26</td>
<td>2.26</td>
<td>2.22</td>
</tr>
<tr>
<td>Alla</td>
<td>2.43</td>
<td>2.52</td>
<td>2.52</td>
<td>2.66</td>
<td>2.39</td>
<td>2.44</td>
<td>2.33</td>
<td>2.58</td>
<td>2.61</td>
<td>2.62</td>
</tr>
</tbody>
</table>

a All data points from the six releases (2281 points) are used.
The HYSPLIT concentration plots (Fig. 4) are the spatial patterns of model plumes driven by NARR and WRF with UW and ACM2 for the CAPTEX 1 representing an afternoon release from DAY and CAPTEX 7 showing an example of a nighttime release from SUB.

We selected WRF results using the UW and ACM2 configurations for Fig. 4 because they represent two classes of PBL schemes [turbulent kinetic energy (TKE) group and nonlocal K-profile group]. In general, all of the runs that are based on WRF meteorological inputs
(not all shown) had a similar spatial distribution of average tracer concentrations, but noticeable differences could be found when comparing the WRF- and NARR-driven runs. As shown in Fig. 4, for CAPTEX 7, the tracer release that took place at Sudbury in Canada behind a cold front moved southward into the United States. A high pressure system associated with a cold front caused the tracer to stagnate in western New York and central Pennsylvania (Ferber et al. 1986). The HYSPLIT runs, either driven by NARR or WRF data, predicted the location of the stagnant area in southern Pennsylvania, which was farther south than that shown by the measurements. In addition, the NARR-driven plume was wider and extended more to the west than the WRF-driven plume. For CAPTEX 1, the NARR-based plume went farther south than the WRF-based one. At the center of the plume and near the release location, the NARR-based HYSPLIT run simulated lower concentrations than those of the WRF-based HYSPLIT runs. In particular, the increase in the rank for the UW case was associated with a higher correlation coefficient and a lower fractional bias relative to the NARR case. The simulation using the UW parameterization was able to produce higher concentration values (orange color in Fig. 4) in the plume center that significantly improved the underprediction shown by the NARR-based results. The fractional biases corresponding to the HYSPLIT runs driven by NARR, UW, and AMC2 were −0.45, 0.13, and 0.54, respectively. The simulation based on the AMC2 PBL parameterization also predicted high tracer concentration on the south side of Lake Erie but over-predicted the concentration in the downwind area.

b. Evaluation for ANATEX

The six WRF configurations selected from the CAPTEX evaluation produced similar statistical scores in terms of surface wind speed and temperature for ANATEX. The normalized Taylor diagram (Fig. 3b) shows that all surface temperatures corresponding to the six WRF runs and the NARR data were clustered even closer for the ANATEX period than for the CAPTEX period. Also, for the surface wind speed evaluation and similar to the run for CAPTEX, the BouLac configuration and the NARR showed the worst results. Time series of daily mean absolute error of the $u$ and $v$ components of wind are presented in Fig. 5. In general, the NARR data presented the largest wind error relative to the WRF-based data throughout the ANATEX period. Among the six WRF runs, the modeled wind using the BouLac PBL scheme often generated a higher bias for both the $u$ and $v$ components of the wind than the rest of the simulations did.

Figure 6 shows a statistical summary of the four components of the rank corresponding to the HYSPLIT runs. The dispersion simulations using WRF data outperformed the one based on NARR data for the PTCH release from GGW. The BouLac parameterization produced the lowest rank, whereas other WRF-based simulations showed similar statistical scores. A representative case for the ANATEX-GGW release.
illustrated that the NARR-based HYSPLIT run did well simulating the main plume located in North Dakota, South Dakota, and Nebraska but failed to produce the plume stretching from the Great Lakes to Michigan and Ohio (Fig. 7; top panel). Note that the HYSPLIT runs driven by WRF inputs were able to capture the later plume even though this plume traveled far away from the release location and features low concentrations. On the other hand, for the tracer experiment released from STC, none of simulations that were based on WRF data generated results as good as the NARR in terms of the rank, mainly because of an increase in the fractional bias. Hegarty et al. (2013) obtained a similar result and pointed out that during the ANATEX-STC case the plumes moved west to east, only intersecting a few sampling locations. The spatial plots for the representative case (Fig. 7; middle and bottom panels) show that the HYSPLIT runs based on the UW and ACM2 PBL parameterizations overpredicted tracer concentrations over the sampling sites in the Great Lakes area. Notice, however, that, among all WRF cases, the UW simulation showed the lowest NRMSE for the tracer concentration, even lower than the simulation using NARR data for both tracer releases—0.571 (GGW) and 0.683 (STC) for the UW case and 0.744 (GGW) and 0.775 (STC) for NARR. These NRMSE scores indicate that the WRF meteorological conditions that used the UW PBL scheme generated tracer concentrations that presented smaller differences with the measurements than did the NARR-driven tracer concentrations.

c. Evaluation for OKC80

The synoptic pattern during this experiment was dominated by a heat-wave event in the central United States featuring clear skies, very dry conditions, and maximum temperatures exceeding 38°C in the study area. Southwesterly wind, associated with a persistent high pressure system centered over the southeastern United States (Ferber et al. 1981), was predominant in the boundary layer at the release site. Similar to CAPTEX and ANATEX, the meteorological evaluation shows that all six WRF results for the OKC80 experiment were very close to each other in the normalized Taylor diagram (Fig. 3c). The mean absolute errors (MAE) corresponding to the surface winds produced by the six different PBL schemes were in the range of 1.62–1.74. Among all, the UW meteorological configuration shows the smallest MAE for both the u and v components of the wind.

The dispersion model output evaluation was summarized in Fig. 8. All HYSPLIT runs using the WRF meteorological input generated better dispersion results than did the simulation driven by the NARR data. The ACM2 parameterization showed the best rank among all with an improvement in the rank of 0.44 units relative to the NARR-based dispersion simulation. Since the measurements were taken in two arcs over different periods of time, the model and observed concentrations were averaged for the two periods and put together in a composite plot (Fig. 9) for the spatial comparison. Given the spatial resolution used in this study, we do not expect the model to resolve the gradient shown in the 100-km arc measurements. Both runs (NARR and ACM2) shifted the plume toward the northeast while the center of the measured plume went northward. The concentration simulated using WRF-ACM2 meteorological inputs presented higher values at the center of the plume than did the case driven by NARR. As the plume traveled farther away to the 600-km arc, the veering of the model plume became larger. Moreover, the modeled...
FIG. 7. Tracer concentration plots for the (left) ANATEX-GGW episode for the 24-h average period starting 1400 UTC 13 Jan 1987 and (right) for the ANATEX-STC episode for the 24-h average period starting 1400 UTC 8 Jan 1987. HYSPLIT simulations used (top) NARR, (middle) UW, and (bottom) ACM2. The color background is model concentrations, and color-coded circles are measured concentrations (pg m$^{-3}$).
plume generated with NARR and AMC2 crossed northeastern Kansas to northwestern Missouri instead of southeastern Nebraska as the measurements indicated. The composite plots also show that track error for the ACM2 plume was smaller than that for NARR. The ACM2 plume moved slightly farther north and extended into southeastern Nebraska, close to the southeastern edge of the observed plume.

d. Evaluation for METREX

A WRF simulation using the UW configuration was conducted for METREX. Figure 3d presents the normalized Taylor diagram corresponding to surface temperature and wind speed. Over the yearlong duration of the simulation, the WRF-UW showed slightly better results than the NARR in terms of the surface wind prediction, whereas the NARR performed slightly better than the WRF when simulating the amplitude of surface temperatures. HYSPLIT simulations driven by NARR and WRF meteorological data were conducted for the two METREX release locations shown in Fig. 1. The dispersion simulations from the northern location (M* in Fig. 1, referred to as METREX-RMD) showed better statistical scores when using WRF than
with NARR data. For the southern location (M’ in Fig. 1, referred to as METREX-MVA), however, the NARR-driven simulations were slightly better than the WRF-driven simulation (Fig. 11, described below). Note that METREX is a local-to-urban-scale tracer experiment with a source–receptor distance of about 15–20 km. Indeed, Draxler (2006) concluded that using higher spatiotemporal resolution [e.g., the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) with 4-km grid spacing] was more appropriate for METREX to improve the dispersion calculations for shorter-duration samples than were simulations using coarse-grid meteorological data such as the NCEP–NCAR reanalysis with a 2.5° grid spacing. Consequently, a meteorological input with a 27-km grid spacing may not be able to accurately capture the transport and mixing of the tracer plume, which was driven by local meteorological features. The NARDDA dataset, however, can be used as a platform from which finer-scale meteorological simulations can be configured for dispersion modeling with adequate spatial and temporal scales for METREX.

e. Dispersion results using nudged data

The statistical evaluation of the tracer experiments showed that the HYSPLIT simulations that used WRF meteorological data with the UW and ACM2 PBL schemes generated better dispersion results than did other runs using different PBL schemes. Taking advantage of the available surface- and upper-level observations, we ran OBSGRID to improve IC/LBC and turned on observational nudging during the simulation. For this sensitivity test, we chose to use the UW configuration, which includes grid nudging.

Figure 10 shows the normalized Taylor diagram summarizing the improvement in the model performance between configurations with and without OBSGRID/nudging. It is evident that the RMS difference between the simulated and observed fields (for both surface wind speed and temperature) was reduced for the “nudged” simulations for time periods corresponding to all four tracer experiments (shown with different symbols in the figure). For this evaluation, we did not use any extra observations that were not included in the nudging, and, as expected, the statistical summary showed that the model wind and temperature fields were brought closer to the measurements by the OBSGRID and observational nudging procedures. The predicted temperatures showed larger variations in the base simulation while the nudged case converged closer to the measurements. For the surface wind speed, we can also see an improvement for the simulations using OBSGRID/nudging even though, for all cases, the wind variations still feature smaller amplitudes when compared to with the observations. Despite the better statistical performance of the WRF-nudged meteorological inputs, the dispersion outputs that use the nudged WRF inputs do not necessarily show better statistical performance than the HYSPLIT runs driven by the meteorological dataset without nudging.

Figure 11 shows the rank corresponding to the dispersion results using NARR, nonnudged (base), and nudged WRF data for all four tracer experiments. For OKC80 and METREX-MVA, the nudged WRF meteorological inputs yielded better dispersion results than the base WRF data did, but it degraded the results for CAPTEX and METREX-RMD. On the other hand, no significant difference was found in the rank statistics for the two ANATEX episodes. Considering each individual release in CAPTEX (not shown), 2 and 5 show degradation in their rank by 0.4 and 0.2, respectively. The other four CAPTEX releases show no noticeable changes in the rank values. Among all tracer experiments in this study, the ANATEX-STC and METREX-MVA showed a degradation in statistical scores when the WRF (base) was used in comparison with the results driven by NARR data. Note that the nudged meteorological input was able to produce an improvement in the dispersion simulation for METREX-MVA showing an increase in the rank of 0.21 and eventually resulting in slightly better results than the simulation driven by
NARR data. This rank comparison indicated that nudged WRF data produced dispersion runs that are slightly better for short-range/urban-scale tracer experiments than HYSPLIT runs that were based on WRF simulations without observational nudging.

We also examined the NRMSE for tracer concentrations produced by different meteorological inputs in Fig. 12. In general, nonnudged and nudged WRF data produced similar errors in the tracer simulations and most of the simulations driven by WRF meteorological conditions performed better than or similar to those driven by the NARR. The only exception was CAPTEX 7, in which the NARR run showed less error than the two WRF simulations. For the simulation with observational nudging, the model was nudged to improve the agreement of wind at the observation locations. These improvements only produce indirect and minimum impacts on variables that are important to the dispersion calculation. Furthermore, observations were limited in time and space (Fig. 2), and, as discussed in the previous section, only a few grid points around an observation were adjusted by the nudging process. Thus, even though a better statistical evaluation was obtained for the nudged meteorological inputs, it did not necessarily result in a better dispersion simulation. We also conducted another set of WRF runs using the ACM2 PBL scheme with OBSGRID/nudging for METREX. The dispersion result was similar or slightly worse than the one driven by the UW meteorological input.
6. Conclusions

WRF meteorological model simulations were conducted to create a long-term data archive in HYSPLIT-compatible format and available in a 1-h interval for driving dispersion applications. In addition, this dataset can provide IC/LBC for downscaling WRF simulations to a finer spatial and temporal resolution to drive HYSPLIT offline or inline using the coupled version of WRF-HYSPLIT. A domain with a 27-km horizontal grid spacing was configured with 33 vertical layers, and simulations were initialized with data from the NARR. We conducted WRF simulations with different PBL schemes and nudging options and used the meteorological data to model dispersion simulations with HYSPLIT for four controlled tracer experiments: CAPTEX, ANATEX, OKC80, and METREX. In addition to the meteorological evaluation using surface and upper-level observations of wind and temperature, we compared the dispersion results with tracer-concentration measurements to assess the most adequate meteorological data for atmospheric transport and dispersion applications. The ultimate goal of this study was to create a WRF dataset that provides an additional meteorological option for HYSPLIT uses that includes a variety of meteorological parameters that are particularly tailored to dispersion-modeling applications. Both meteorological datasets, NARR and WRF, have comparable spatial resolutions, but the WRF data provide hourly temporal frequency for the dispersion calculation. Furthermore, the friction velocity, one of the essential parameters for computing dispersion, is available in the WRF output. When using the NARR data, the HYSPLIT model needs to rediagnose the frictional velocity from other meteorological variables. Indeed, rediagnosing parameters and temporally interpolating data can contribute to errors in the dispersion calculation.

The four controlled tracer experiments used in the HYSPLIT simulations covered different geographical locations in the United States. These experiments were conducted for different time periods including a summer day (OKC80), several days in autumn (CAPTEX), three months during winter (ANATEX), and one full year (METREX). The evaluation for WRF simulations that were based on different PBL parameterizations was presented using Taylor diagrams for wind and temperature. Most of the results clustered together, except for the WRF runs with the TEMF and BouLac parameterizations and the NARR, indicating a similar statistical performance for most WRF configurations. Because of the spatiotemporal limitation of the conventional meteorological observations and unavailability of measured mixing parameters essential for the dispersion calculation, the statistical analysis of WRF results alone was insufficient to provide a clear picture on which WRF configuration would produce a better dispersion result. Thus, we relied on the evaluation of the transport and dispersion for multiple tracer experiments to provide further evidence for choosing a WRF configuration that improves the HYSPLIT simulations.

For the CAPTEX releases, the top three statistical scores corresponded to HYSPLIT results using WRF data with ACM2, UW, and GBM PBL schemes. The lowest three rank scores were associated with the dispersion runs driven by the QNSE, MYNN2, and TEMF parameterizations. These ranks were lower than or similar to those obtained with the NARR meteorological input. For the other three tracer experiments, the HYSPLIT results that were based on the WRF meteorological conditions show better or similar statistical ranks than the runs driven by the NARR data in most cases, except for the ANATEX-STC episode and METREX-MVA case. We conducted a WRF simulation with objective analysis to improve IC/LBC and observational nudging to minimize the error growth during the simulation by using available surface and upper-level observations. Our results show a mixed impact for the dispersion results driven by the base and nudged WRF data when compared with the NARR-driven simulations. For OKC80 and METREX-MVA, the nudged WRF meteorological input yielded better dispersion results than the base WRF data, but it degraded the results for CAPTEX and METREX-RMD. Even though the observational nudging can noticeably reduce the error in the predicted fields that were nudged, only indirect or minimum impacts can be noted on the mixing variables that are important for dispersion modeling. Thus, even though a better statistical performance was obtained for the nudged meteorological input, it did not necessarily result in a better statistical score for the corresponding dispersion simulation.

In general, the NARDDA dataset generates equal or slightly better dispersion results than the NARR data. The main advantage of using the NARDDA dataset is that it provides additional variables relevant to atmospheric dispersion that are not available from NARR, such as friction velocity, TKE, and time-averaged wind fields. With these new WRF meteorological data, HYSPLIT can calculate the dispersion using different options for mixing estimations available in the model. Conversely, the NARR data allow only one set of parameterizations for running HYSPLIT. Furthermore, for applications requiring precipitation data (e.g., involving wet-deposition processes), the WRF dataset provides hourly precipitation whereas NARR only includes
3-hourly data. The NARDDA covering 1980–2016 and beyond will be added to the DATEM, expanding the capabilities for using different meteorological inputs and providing a variety of options to compute the HYSPLIT mixing parameters. The NARDDA can provide users the capability to generate dispersion ensembles with variations in meteorological inputs and diverse configurations of the dispersion simulations (Stein et al. 2015b).

Given the spatial limitation of the WRF 27-km dataset, further investigation is needed for cases that require a higher spatial and temporal resolution, such as for tracer experiments covering an urban scale like METREX. For such cases, higher variability in meteorological parameters is expected when using different PBL schemes. Toward this end, WRF can be nested down from the NARDDA for finescale simulations. Although measurements of PBL turbulence parameters are fundamental to study dispersion modeling, they are very scarce and very limited in their spatial and temporal coverage. Indeed, no turbulence measurements were available in the four experiments presented in this study. Nevertheless, as more tracer experiments have been made publically available in recent years, (e.g., the Sagebrush tracer experiment; Finn et al. 2015), future research will be focused in analyzing turbulent mixing using tracer measurements and collocated turbulence data.

Acknowledgments. The authors express gratitude to the anonymous reviewers from NOAA/ARL and the Journal of Applied Meteorology and Climatology for valuable inputs and fine polishing of this article.

REFERENCES


