Validation of the Atmospheric Dispersion Model NAME against Long-Range Tracer Release Experiments

VIBHA SELVARATNAM,[a] DAVID J. THOMSON,[a] AND HELEN N. WEBSTER[a]

[a] *Met Office, Exeter, Devon, United Kingdom*

(Manuscript received 9 February 2023, in final form 7 July 2023, accepted 10 July 2023)

ABSTRACT: The Met Office's atmospheric dispersion model Numerical Atmospheric-Dispersion Modeling Environment (NAME) is validated against controlled tracer release experiments, considering the impact of the driving meteorological data and choices in the parameterization of unresolved motions. The Cross-Appalachian Tracer Experiment (CAPTEX) and Across North America Tracer Experiment (ANATEX) were long-range dispersion experiments in which inert tracers were released and the air concentrations measured across North America in the 1980s. NAME simulations of the experiments have been driven by both reanalysis meteorological data from European Centre for Medium-Range Weather Forecasts (ECMWF) and data from the Advanced Research version of the Weather Research and Forecasting (WRF) Model. NAME predictions of air concentrations are assessed against the experimental measurements, using a ranking method composed of four statistical parameters. Differences in the performance of NAME according to this ranking method are compared when driven by different meteorological sources. The effect of changing parameter values in NAME for the unresolved mesoscale motions parameterization is also considered, in particular, whether the parameter values giving the best performance rank are consistent with values typically used. The performance ranks are compared with analyses in the literature for other particle dispersion models, namely, Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT), Stochastic Time-Inverted Lagrangian Transport (STILT), and Flexible Particle (FLEXPART). It is found that NAME performance is comparable to the other dispersion models considered, with the different models responding similarly to differences in driving meteorological data.

KEYWORDS: Dispersion; Mesoscale processes; Global transport modeling; Model comparison; Model evaluation/performance

1. Introduction

The Met Office's atmospheric dispersion model Numerical Atmospheric-Dispersion Modeling Environment (NAME; Jones et al. 2007) is a Lagrangian model that calculates the transport and dispersion of pollutants by simulating the trajectories of a large number of model particles; each model particle represents a certain mass of the released material. Like other widely used Lagrangian models, the model is generally driven by meteorological data from a numerical weather prediction model and uses a stochastic parameterization for the effect of unresolved motions on the trajectories.

Validation of dispersion models against data from dispersion events is an important part of model development. Ad hoc events are valuable case studies (Draxler et al. 2015; Grant et al. 2012) giving the opportunity to consider the suitability of the model in real-world applications, which, given the use many are put to in managing hazards, is of critical use to responders and governments. But the use of controlled-release experiments is generally advantageous as more accurate release information is available, along with systematic measurements taken throughout the course of the experiments. This makes such experiments ideal for assessing model performance. Long-range dispersion experiments are expensive and, as a result, only a small number of such experiments are available. NAME has previously been evaluated with other dispersion experiments including the European Tracer Experiment (ETEX; Ryall and Maryon 1998) and Kincaid (Webster and Thomson 2002) datasets.

In this paper, NAME is validated against two long-range, controlled tracer release experiments conducted in North America. The two experiments used are the Cross-Appalachian Tracer Experiment (CAPTEX; Ferber et al. 1986) and the Across North America Tracer Experiment (ANATEX; Draxler and Heffter 1989). These experiments have been used for various dispersion model evaluations (Ngan and Stein 2017; Loughner et al. 2021; Hegarty et al. 2013). Several other Lagrangian dispersion models have been evaluated in Hegarty et al. (2013), namely, Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT; Draxler and Hess 1997), Stochastic Time-Inverted Lagrangian Transport (STILT; Lin et al. 2003), and Flexible Particle (FLEXPART; Brioude et al. 2013, 2012). This gives the opportunity for comparing the performance of NAME with that of other Lagrangian models that have already been evaluated against this dataset. To this end, we follow the approach of Hegarty et al. (2013), with the performance of NAME being evaluated using a ranking system consisting of four statistical parameters (Mosca et al. 1998; Stohl et al. 1998; Draxler 2006). We also investigate differences in performance when NAME is driven by alternative meteorological datasets. These meteorological datasets include most of those considered by Hegarty et al. (2013) (viz., various configurations of WRF; Skamarock et al. 2008), which enables us to directly compare NAME with other dispersion models, as well as with ECMWF's

Corresponding author: Vibha Selvaratnam, vibha.selvaratnam@metoffice.gov.uk

DOI: 10.1175/JAMC-D-23-0021.1

ERA-Interim dataset (Berrisford et al. 2011). In this, our aim is to add to the work of Hegarty et al. (2013) assessing whether skill is dominated by differences in the driving meteorological data or dispersion model.

As well as parameterizing the effects of turbulence, NAME represents the effects of atmospheric motions that are larger than the three-dimensional turbulence but too small to be resolved by the NWP model (Webster et al. 2018). We refer to these as the "unresolved mesoscale motions." At short distances from the source, the effect of the parameterization of these motions is to widen the modeled plume, while at long ranges, where the dispersion is likely to be dominated by the resolved scales of motion, the extra dispersion is expected to smooth out the smaller-scale features in the concentration field. This is consistent with the idea that the effect of unresolved motions is represented in the model as an average over an ensemble of such motions (and so produces smoothed concentration fields), while the effect of the resolved motions is represented deterministically. This also raises the possibility of adjusting the parameterization in order to account for how much small-scale structure is resolved by the NWP model. Conceptually, this is like altering the subgrid diffusion in large-eddy simulations, which has the effect of altering the filter scale of the simulation (Mason and Callen 1986). If the small-scale features are not well predicted, it is quite likely that increasing the diffusion—in effect, reducing the resolution of the model—will improve the performance measures. We investigate this by examining the sensitivity of the performance to the unresolved mesoscale motions parameterization. We consider how model performance rankings are affected by changing the parameterization values and compare this with the optimal parameter values estimated by Webster et al. (2018) using spectra of observed and NWP winds.

2. Experimental data

CAPTEX and ANATEX were controlled tracer release experiments conducted in the North American region in the 1980s (Draxler and Heffter 1989; Ferber et al. 1986; Draxler et al. 2001). The tracers used were effectively inert and nondepositing.

CAPTEX consisted of seven ground-level releases (referred to as CAPTEX-1 through CAPTEX-7) from 18 September to 29 October 1983. CAPTEX-6 was a short release of 30 min and each of the others was a 3-h release. The first four (CAPTEX-1–4) and CAPTEX-6 were releases from Dayton, Ohio, and the other two (CAPTEX-5 and CAPTEX-7) from Sudbury, Ontario, Canada. The first four releases were separated by periods ranging from 7 to 12 days and the latter two by 3 days, but with strong winds, and the tracer clouds were quickly advected out of the region. As such, the tracer clouds from each release were well separated, so each release can be treated as a separate experiment. A sampling network of 84 sites, 300–800 km from the source, collected ground-level samples of the tracer, perflouromonomethylcyclohexane (PMCH). From 19 September to 30 October, 3- and 6-h averages were retrieved for 48–60 h after each release. Here, we compare NAME predictions with data from the 3-h releases and omit the data from the short CAPTEX-6 release, in line with the work of Hegarty et al. (2013).

ANATEX consisted of 66 releases, each with a duration of 3 h, from 5 January to 26 March 1987. One-half were releases of perflourotrimethylcyclohexane (PTCH) from Glasgow, Montana (GGW), and the other one-half were releases of perflourodimethylcyclohexane (PDCH) from Saint Cloud, Minnesota (STC). Material was released at a height of 2 m above a one-story building. The releases from Saint Cloud included releases of PMCH, but these were coincident with PDCH and, following Hegarty et al. (2013), are not included in our comparison. The releases were at 2.5-day intervals, alternating between afternoon and nighttime, and occurred from both sites at each release time. The sampling network had 75 sites over the eastern United States and southeastern Canada, extending to about 3000 km from the source locations. From 5 January through 29 March, 24-h averaged air samples were collected at ground level.

Following Hegarty et al. (2013), only the first 10 releases (spanning 5–16 January) from ANATEX are included. This period of winter conditions contrasts with the summerlike conditions during CAPTEX, as well as having a similar number of measured-predicted data pairs for analysis as the CAPTEX experiments. As each site used a different tracer, releases from each were treated as separate experiments (called ANATEX-GGW and ANATEX-STC). However, we treat all the releases from one site as a single experiment. This is because, with the interval between releases being generally shorter than for CAPTEX and the range of the dispersion being longer, the different releases did not always stay well separated.

3. Dispersion models and meteorological data

As well as evaluating the performance of NAME against CAPTEX and ANATEX, we compare results with those obtained by Hegarty et al. (2013) for HYSPLIT, STILT, and FLEXPART. In these four Lagrangian particle-dispersion models, model particles are released from a source location and are advected by the mean wind obtained from input meteorological data, with random components added by the model to represent the effects of small-scale atmospheric motions unresolved by the input meteorological data. Unless otherwise stated, each CAPTEX release was represented by 50 000 model particles and the ANATEX releases by 25 000 particles (Hegarty et al. 2013). Increasing the number of particles seems to have little to no effect on the analyses and the fewer particles released for the ANATEX releases are compensated for by the longer averaging time for calculating the air concentrations. In the study by Hegarty et al. (2013), the models were driven by North American Regional Reanalysis (NARR) meteorological data and by four different sets of data from the Advanced Research, version 3.2.1, of the Weather Research and Forecasting (WRF) Model. We run NAME with the same four WRF datasets generated by Hegarty et al. (2013) for comparison with other models, as well as with data from the ECMWF ERA-Interim reanalysis. We note that since the analysis by Hegarty et al. (2013) was performed, there will have been improvements to the dispersion models. However, to compare results with the same meteorological

datasets, we consider the performance rankings obtained by Hegarty et al. (2013), and limited analyses with previous NAME versions suggest that the performance rankings of NAME would not have varied significantly over recent years. We now discuss the model configurations and meteorological data sources in more detail.

NAME, version 7.2 (2017), was used, with output calculated on a 0.25° × 0.25° grid in the latitude–longitude coordinate system and concentrations obtained over the lowest 100 m above ground level (AGL). Particle trajectories are calculated using the meteorological data's native horizontal coordinate system and height above ground, with the height converted to the native vertical coordinate system of the meteorological data for the purpose of interpolating the meteorological data. The model time step was set to 1 min, and a diffusive turbulence scheme with no velocity memory was used. Dispersion due to both turbulence and unresolved mesoscale motions was represented, while convection above the boundary layer was not.

The HYSPLIT runs by Hegarty et al. (2013) had the same output resolution and number of particles as chosen for the NAME simulations. Trajectory calculations used the same horizontal coordinate system as the meteorological data. HYSPLIT uses an internal terrain-following, or σ, vertical coordinate system, and meteorological fields are linearly interpolated to a grid defined in this coordinate system (Draxler and Hess 1997; Stein et al. 2015). The lowest vertical grid level was approximately 10 m above ground level, and the resolution decreases with height. The time step used in the HYSPLIT runs is 1 min (Hegarty et al. 2013).

STILT is built upon HYSPLIT and, hence, has many features in common, such as the mean advection scheme and the calculation grid. However, STILT simulates turbulence differently. The configuration used is the same as HYSPLIT (Hegarty et al. 2013). Although STILT is primarily used in backward mode, here we only discuss the performance of all the models in forward mode.

The version of FLEXPART used by Hegarty et al. (2013) is one modified to use meteorological data from the WRF model. Output concentrations were given on a 25 km × 25 km horizontal grid with the same projection as the meteorological data, which is similar to the 0.25° grid used by the other models. Concentrations were again output over the lowest 100 m AGL. FLEXPART uses the native horizontal coordinate system of the WRF data and, as for HYSPLIT and STILT, the vertical levels of the meteorological data are interpolated to an internal grid defined in a terrain-following coordinate system. The time step was calculated dynamically with a maximum of 90 s, and each release was represented by 100 000 particles in both CAPTEX and ANATEX (Hegarty et al. 2013).

ECMWF's ERA-Interim is a global atmospheric reanalysis starting from 1979, produced with a 2006 version of the Integrated Forecasting System (IFS; Berrisford et al. 2011). The horizontal spatial resolution is approximately 79 km with 60 vertical levels, and output is every 3 h.

The meteorological fields from the Advanced Research version (3.2.1) of the WRF Model (Skamarock et al. 2008) used in this study were obtained by Hegarty et al. (2013) using initial and boundary conditions from NARR (Mesinger et al. 2006). The configuration of WRF uses a conformal Lambert horizontal grid and a terrain-following, hydrostatic-pressure vertical coordinate system of 43 levels, with the lowest layer approximately 33 m thick. The model was configured with two nested horizontal resolutions of 30 and 10 km with one-way boundary conditions between the two. The dispersion models use the 10-km WRF data where available and the 30-km data elsewhere. The NAME model runs for the ANATEX experiments only had the 10-km WRF data available. This is still comparable to the other dispersion model runs, as the dispersion predominantly remains within the 10-km data domain for the duration of the experiment modeled here and all the measuring stations are well within that domain. Hourly output from the WRF model with wind nudging toward NARR in the boundary layer both turned on (denoted WRF pbl1) and off (WRF pbl0) are used to run the dispersion models. The WRF output contains instantaneous and (hourly) time-averaged wind fields, and both of these are used separately to drive different dispersion model runs. NARR is an extension of the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) global reanalysis, which has been run over the North American region. The reanalysis data span 1979 to 2014 and are on a conformal Lambert grid with a spatial resolution of approximately 32 km and 45 vertical levels, with output every 3 h.

4. Statistical measures

Assessing the accuracy of a dispersion model is difficult due to having both temporal and spatial discrepancies to consider in the dispersion forecast. Statistical parameters have different sensitivities to these variations. We use the ranking system used by Hegarty et al. (2013), which combines four statistical parameters to obtain an overall rank. The software used for the calculations is detailed in Draxler et al. (2001) and provided by the NOAA Air Resources Laboratory.

The four parameters are Pearson's correlation coefficient squared (R^2), the fractional bias (FB), the figure of merit in space (FMS), and the Kolmogorov–Smirnov parameter (KSP). The correlation coefficient R ranges from -1 to 1, where 1 is a perfect (positive) correlation between measured and predicted concentrations. R^2 gives the fraction of the variance in the observations that is explained by a linear regression on the model values. The FB is the difference between the mean observed and mean predicted concentrations expressed as a fraction of the average of the mean observed and mean predicted concentrations. A positive value indicates an overprediction by the model and a negative value indicates an underprediction. The FMS is the area, at a fixed time, with both measured and predicted concentrations above some threshold expressed as a percentage of the area with either measured or predicted concentration above the threshold. Here, following Stohl et al. (1998) and Hegarty et al. (2013), we replace the area simply with the number of sampling locations. The concentration threshold is taken to be zero, following Hegarty et al. (2013). It has been considered that a zero threshold could cause problems, as changing the number of particles in a Lagrangian particle dispersion model could, due to noise around the edges of the plume, change the number of

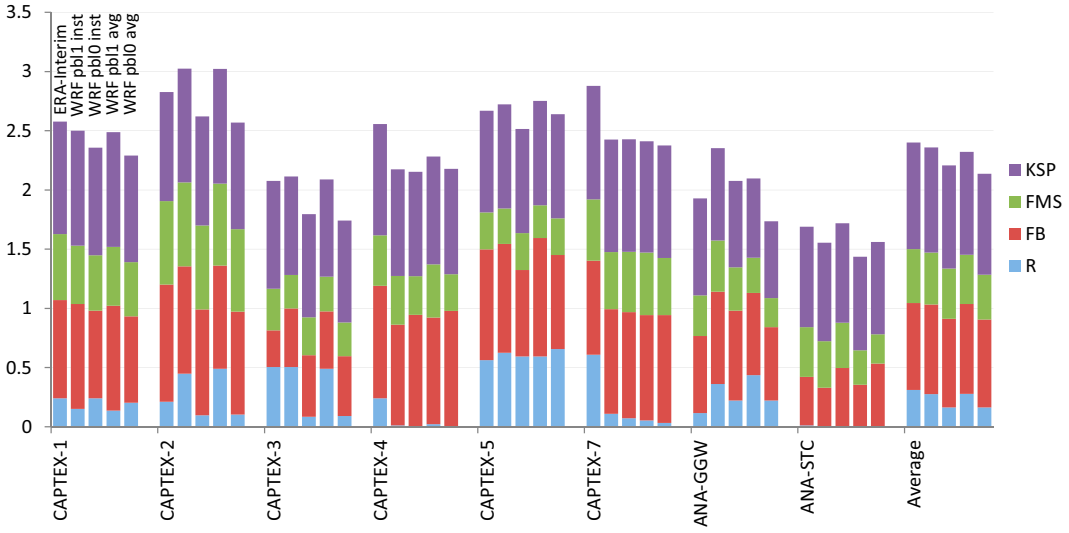

FIG. 1. Rankings of NAME driven by meteorological fields from various sources, showing the contribution of each of the statistical parameters.

nonzero concentrations output. However, with the sampling stations at fairly large distances from each other and averaging periods of 3 h or longer, we expect a change in the level of noise in the model run not to have a significant effect on the FMS, with the exception of any instances where the edges of the plume remain close to a sampling station for a prolonged period. A limited number of NAME runs with more particles supported this, so, for more comparable results, we use the same number of particles for the releases as the HYSPLIT runs by Hegarty et al. (2013). Similarly, there is a question of when measured concentrations would have been reported to be zero. For both experiments, a concentration slightly above estimated background level was subtracted from the measurements, and measurements were rounded to account for instrument accuracy (Ferber et al. 1986; Draxler and Heffter 1989). Any negative concentrations are then set to be zero. Although a high FMS indicates a good prediction, a low value does not necessarily imply a bad prediction, as the plume could have the correct shape but be slightly offset in location. This is particularly pronounced with narrow plumes. The KSP is the maximum absolute difference between the cumulative distributions of the observations and the measurements (expressed as percentages), so a smaller value implies a better prediction. FB and KSP depend only on the set of modeled values and the set of measured values considered separately, while R^2 and FMS depend on the paired values.

These four parameters are equally weighted so that each contributes a value between zero and one to the final rank. The final rank ranges from 0 to 4, with a higher rank implying a better prediction (Draxler 2006). The formula used is

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

5. Results

This section looks at the results from the statistical analyses of the different atmospheric dispersion model simulations. We consider how different driving meteorological data affect the performance of NAME, differences between various models being driven by the same meteorological data, and how changes in parameterizing the unresolved mesoscale eddies affect the performance of NAME.

a. Driving meteorological data

Here, we consider how NAME performs when driven by different meteorological fields. The performance of NAME will be compared with that of other models below, but we note here that it is similar to the other models. We look at runs of NAME driven by ECMWF ERA-Interim data and WRF with WRF pbl1 or WRF pbl0. We also use the WRF instantaneous wind fields and the time-averaged wind fields in separate runs. From Fig. 1, we can see that, generally, NAME seems to perform slightly better driven by the lower-resolution ERA-Interim meteorological data than with the higher-resolution WRF meteorological data, especially when WRF has the wind nudging toward NARR switched off. This improved ranking in the dispersion forecast driven by lower-resolution meteorological data is in contrast to the findings of Hegarty et al. (2013), where the dispersion models driven by higher-resolution meteorological data generally performed better. The improved ranking when NAME is driven by ERA-Interim data could be due to the NWP model performing better in some of these particular cases as opposed to a preference for lower-resolution data. However, it is also possible that, due to the experiments being long range, the smaller scales of motion are adding structure to the plume, but not accurately enough to improve the ranking. The effect of the small-scale structure on the overall ranking is discussed further in section 5c. The difference in performance between having the wind nudging switched on and off is very pronounced in all the experiments apart from ANATEX-STC, which is the only experiment where NAME performs better with the wind nudging switched off. We note

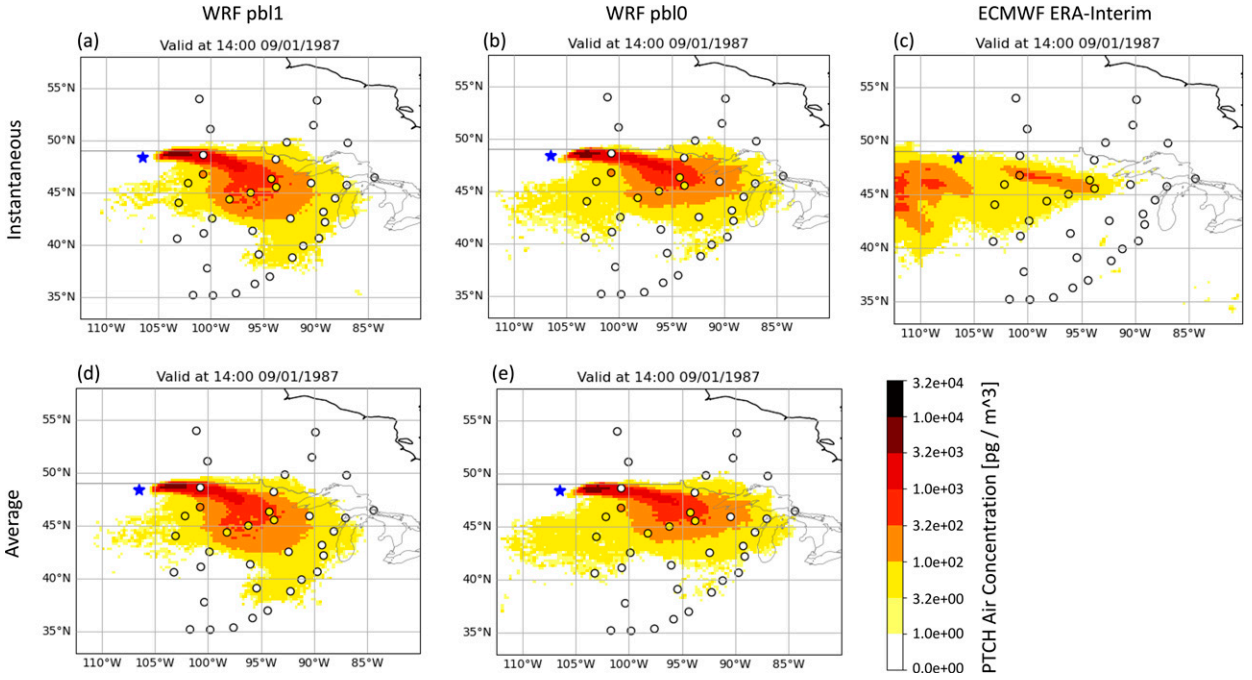

FIG. 2. ANATEX-GGW simulated average PTCH ground level concentrations (pg m^{-3}) for the 24-h period ending 1400 UTC 9 Jan 1987 resulting from tracer releases at Glasgow (blue star) from 1700 to 2000 UTC 5 Jan 1987 and from 0500 to 0800 UTC 8 Jan 1987. Simulations were run with NAME driven by WRF with wind nudging (a),(d) turned on and (b),(d) turned off and by (c) ERA-Interim, using (top) instantaneous meteorological fields and (bottom) time-averaged wind fields. Observed concentrations are shown in the circles, using the same color scale. The "e+0X" in the color scale here and in subsequent figures indicates multiplication by 10^X.

that this is the experiment in which NAME generally performs the worst. As this is the case when NAME is driven both by WRF and by ERA-Interim, this could be due to wind conditions, which are difficult to predict. In this experiment, we also note that the contribution to the rank of the correlation coefficient is almost zero, and this is consistent with the simulations by other models in Hegarty et al. (2013). We can also see that throughout the experiments, there is generally very little performance difference between using the instantaneous and time-averaged meteorological fields with WRF, and this is reflected in the average performance ranks.

Figure 2 shows NAME-predicted PTCH concentrations for the ANATEX-GGW experiment, when the model is driven by various meteorological fields. These plots are for the 24-h period ending at 1400 UTC 9 January 1987. We can see the similarities between all of the WRF-driven runs, although there are some subtle differences that can be seen by eye. With this particular example, the concentration plot for the run driven by ECMWF ERA-Interim reanalysis data differs significantly from all the runs driven by WRF. We see a cloud of tracer that appears much farther to the west than in the WRF runs, and this cloud is from the release on 5 January that occurred a few days earlier. Examining concentration plots from the day of that release, it is clear that the plume is advected in very different directions depending on which meteorological data are used to drive the model. These indicate that the ECMWF ERA-Interim data had easterly winds, whereas that for the WRF data had some easterly and westerly winds. This initial difference in where the tracer has been transported then creates legacy differences in the output at later times. Looking at the rank of the models, however, the ERA-Interim run is not the outlier despite this clear visual difference. This is most likely due to the length of experiment being long enough that this difference for a short period is not as significant as the smaller differences that are present throughout the experiment duration. It is, however, an exceptional example and, in the other experiments, the concentration plots when run using ECMWF ERA-Interim reanalysis data look similar to those using the various forms of WRF data (see Fig. 3). When we examine the contributions of the different statistical parameters to the final rank for the ANATEX-GGW experiment, we do see that the greatest difference between the ERA-Interim–driven run and the WRF-driven runs is the smaller correlation coefficient with ERA-Interim. In fact, throughout all of the experiments, the contribution of the correlation coefficient generally varies more when considering different driving meteorological data than that of the other statistical parameters, with the contribution to the overall ranking often being very small.

b. *Dispersion models*

Figure 4 shows how all the dispersion models perform being driven by both versions of the instantaneous WRF meteorological data, with the ranks for HYSPLIT, STILT, and FLEXPART from Hegarty et al. (2013). In all cases except ANATEX-STC, NAME and FLEXPART perform better with WRF pbl1 data in which the wind nudging toward

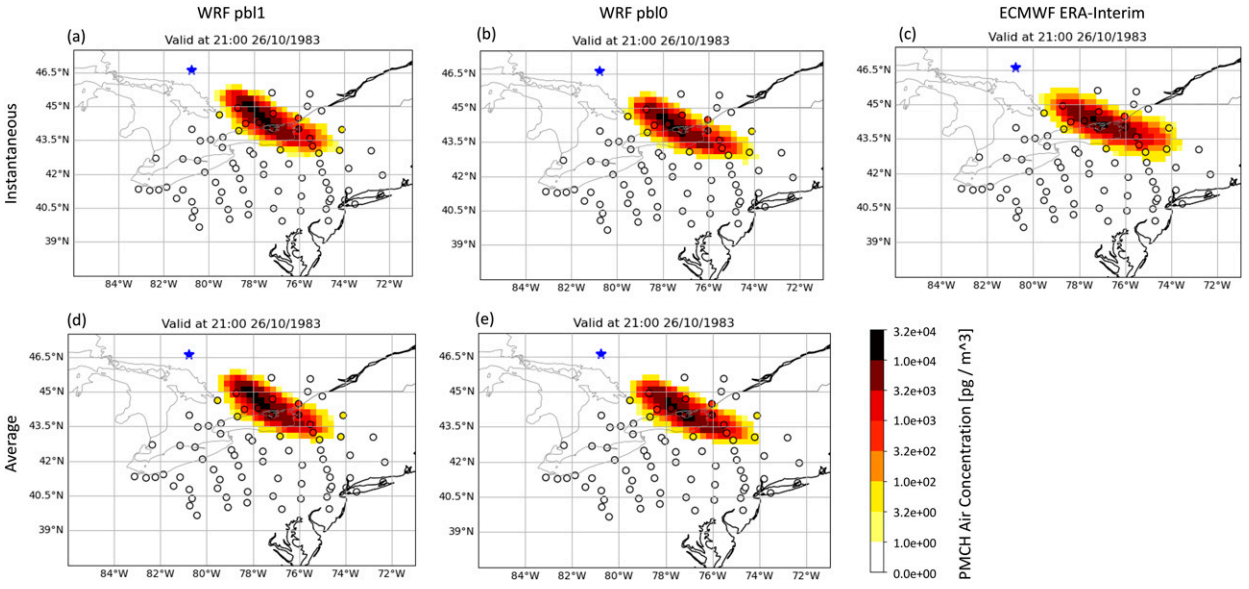

FIG. 3. CAPTEX-5 simulated average PMCH ground level concentrations (pg m^{-3}) for the 6-h period ending 2100 UTC 26 Oct 1983. The CAPTEX-5 release was from Sudbury (blue star) from 0345 to 0645 UTC 26 Oct 1983. Simulations were run with NAME driven by various meteorological sources as in Fig. 2, with observed concentrations shown in the circles.

NARR is turned on when compared with that when there is no wind nudging. HYSPLIT and STILT, however, do not seem to have a preference for one set of WRF meteorological data over the other and, on average, there is little difference between the two. Because STILT is built upon the HYSPLIT model, it is unsurprising that they react to a change in driving meteorological data in a similar way and, for each experiment, they both show a similar difference in performance using one set of meteorological data over the other.

On average, the performances of the various models are very similar, although there are some differences for individual experiments. It can be seen that FLEXPART driven by WRF pbl0 meteorological data generally has a slightly lower ranking than the other three models, apart from the CAPTEX-7 experiment in which it performs noticeably better than all the other models. It is also clear that the ANATEX-STC experiment seems to have been the most difficult to predict as this is the experiment where all the models had the weakest performance. Although this could be due to the WRF data having errors, we see in Fig. 4 that NAME runs driven by ERA-Interim show the same low ranking for this experiment, so it is likely that there were some difficult conditions to predict over this time and location.

We see from the previous section that the performance of NAME when driven with instantaneous meteorological fields and time-averaged fields is similar on average. Although there is variation between the experiments, the runs driven with time-averaged fields show the same preference of wind nudging being turned on and the average ranks are very similar.

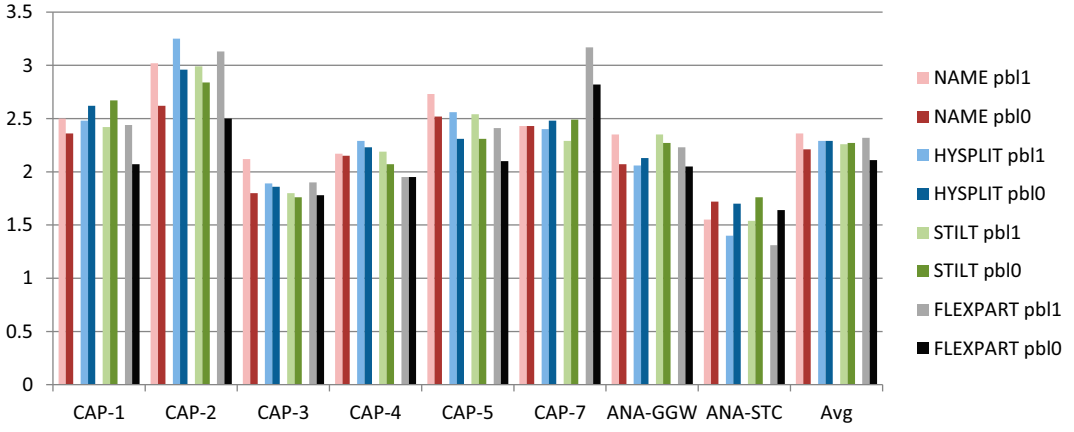

FIG. 4. NAME rankings with HYSPLIT, STILT, and FLEXPART rankings from Hegarty et al. (2013) run with WRF instantaneous wind fields.

TABLE 1. The unresolved mesoscale motions parameter values used within the NAME simulations.

	σ (m s^{-1})	τ (s)	K (m^2 s^{-1})
M1	0.44	4484	868
M2	0.70	8000	3920
M3	0.94	11 862	10 481
M4	1.18	15 500	21 582
M5	1.42	19 300	38 917
M6	1.66	23 100	63 654

From the work done by Hegarty et al. (2013), we see that other models show the same behavior on average.

c. *Unresolved mesoscale motions*

The NAME model has two parameters that determine the amount of diffusion the model adds to represent the effects of unresolved mesoscale motions (Webster et al. 2018). These parameters are the standard deviation σ and Lagrangian correlation time scale τ of horizontal random fluctuations, which are added to the particle velocity components. As we are considering long-range dispersion, it is actually the horizontal eddy diffusivity K that affects how the particles will be dispersed by the model, where $K = \sigma^2\tau$. Increasing K when running the model leads to a greater spread of the tracer and smoothing of small-scale features. There is a range of values estimated to be appropriate when using meteorological fields on specific temporal and spatial scales, as outlined in Webster et al. (2018). In this section, we consider the effect of using different diffusivity values on the performance of the model according to the statistical parameters used in this validation.

We use six sets of values for the unresolved mesoscale motions parameters labeled as runs M1–M6, where M1 runs use the smallest diffusivity values and M6 runs use the largest values. The unresolved mesoscale motions values used for M1 and M3 in this section are the smallest and largest values that would be suggested when using meteorological data with these temporal and spatial scales, based on the analysis of Webster et al. (2018). The M2 values consist of the average values of these ranges for both σ and τ and are the same values used throughout the comparisons above. We also consider three sets of larger diffusivity values (M4–M6), outside of the range recommended in Webster et al. (2018), and the case of turning the unresolved mesoscale motions scheme off (M0). The values used are shown in Table 1. We use these unresolved mesoscale motions parameters with NAME being driven by WRF meteorological data with time-averaged wind fields, with both wind nudging turned on and off.

We can see in Fig. 5 that, generally, the performance of the model increases as we turn the unresolved mesoscale motions scheme on and then as we increase the parameter values within the range recommended in Webster et al. (2018) (i.e., an increase in performance from M0 to M3). This corresponds to an increased spread of the tracer. This is not always the case; for example, the simulations of CAPTEX-5 run with wind nudging toward NARR switched off actually perform slightly better with smaller values. From Fig. 3, we see that the CAPTEX-5 predicted plumes match well with the observations. Therefore, an increase in the spread of the predicted plume would not have the benefit of smoothing errors. However, on average, the performance of the model improves as the unresolved mesoscale motions values are increased within this M0–M3 range, both when wind nudging toward NARR is switched on and off. We also note that the improved performance with the values increasing up to M3 generally occurs for all four of the statistical parameters that form the final rank, but, most notably, the correlation coefficient improves. The same conclusions are drawn from the results driven by instantaneous wind fields from WRF. Using values for the unresolved mesoscale motions parameters larger than recommended in Webster et al. (2018) was also considered and we see, for the M4 runs, this increase has resulted in an improved performance for some experiments, while it had a detrimental effect on others. For example, CAPTEX-7 shows an improvement in performance as the spread due to mesoscale motions is increased,

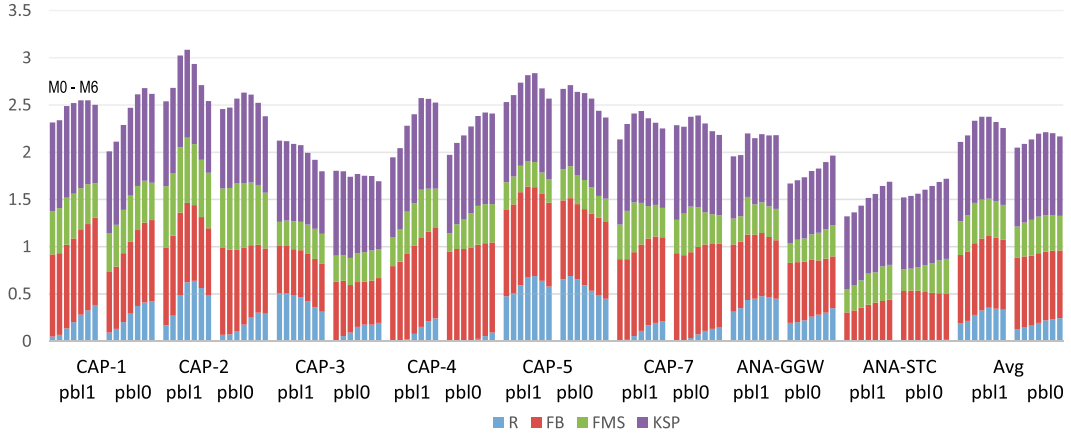

FIG. 5. Rankings of NAME run with WRF time-averaged wind fields using different unresolved mesoscale motions parameters (see Table 1), with wind nudging toward NARR switched on (pbl1) and off (pbl0). The case of turning the unresolved mesoscale motions scheme off is designated as M0.

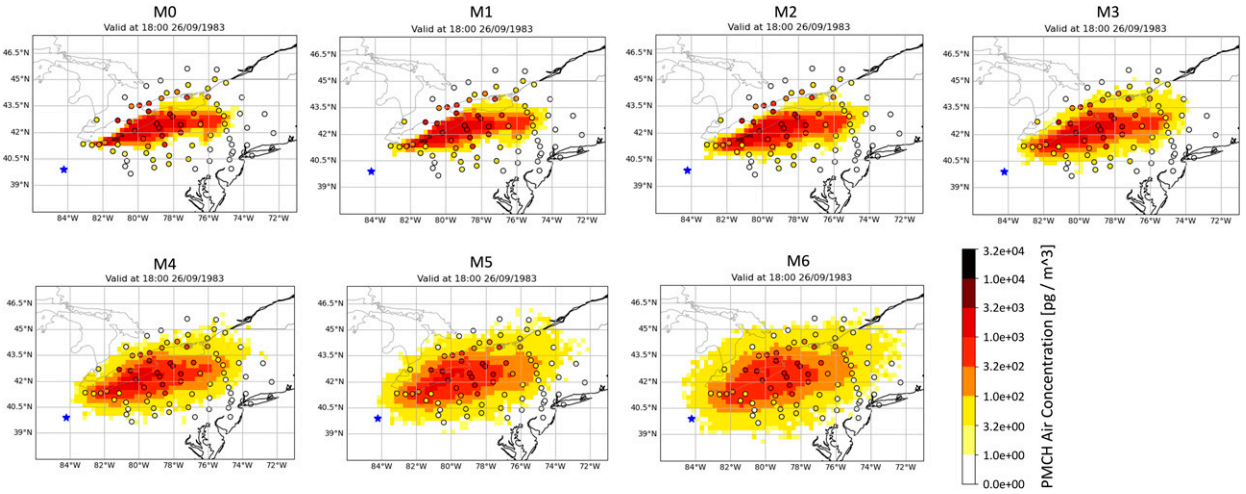

FIG. 6. CAPTEX-2 simulated average PMCH ground level concentrations (pg m^{-3}) for the 6-h period ending 1800 UTC 26 Sep 1983 following a tracer release from Dayton (blue star) from 1700 to 2000 UTC 25 Sep 1983. Simulations run with NAME driven by WRF pbl1 time-averaged fields using increasing unresolved mesoscale motions parameters denoted M1–M6, and with the parameterization off (M0). Observed concentrations are shown in the circles.

but only as far as the recommended values, after which the performance of the model suffers. However, in CAPTEX-4, this improvement continues a little beyond the recommended values. The largest values considered (M5 and M6) show that, on average, the performance of the model decreases, and this decrease is the case for every CAPTEX experiment for the largest M6 values and most CAPTEX experiments for the M5 values. For the ANATEX experiments, however, the performance continues to improve even to these largest values. This is probably because of the longer range of the ANATEX experiments and the poor performance generally for ANATEX-STC. This means that the additional spread compensates for errors in the predicted location of the plume and the smoothing out of the smaller features of the plume benefits the performance ranking of the model.

Figure 6 shows air concentrations from NAME simulations when the model is driven by WRF pbl1 meteorological data, using different unresolved mesoscale motions parameters. The example plots shown here are for the CAPTEX-2 experiment for the 6-h period ending at 1800 UTC 26 September 1983. The increase in the spread of the tracer is clearly seen as K increases in the unresolved mesoscale motions parameterization. We note from Fig. 5 that CAPTEX-2 is one of the experiments with the most clear difference as these parameters are changed, but similar visual differences between the concentration plots are seen for all the experiments, even where there is little difference in the rankings (e.g., see CAPTEX-3 without nudging or, for M0–M3, with nudging). In this case, Fig. 6 shows that the predicted plume with smaller unresolved mesoscale motions added appears to be centered slightly too far south relative to the observations. As the spread of the predicted plume increases, this positional error has less of an impact on the ranking as the predicted plume includes more of the observations. However, as the spread increases farther, this results in the highest observed concentrations not being reached in the predicted plume, as well as concentrations being predicted where zero concentrations were observed.

6. Conclusions

The Lagrangian particle dispersion model NAME was run to simulate controlled tracer release data from the CAPTEX and ANATEX experiments. The performance of the model was evaluated when driven by different meteorological fields (including ECMWF ERA-Interim reanalysis data and various configurations of WRF) as well as comparing the performance with other dispersion models as evaluated by Hegarty et al. (2013). The impact of changing the parameters within the NAME unresolved mesoscale motions parameterization on the performance of the model was also considered. The assessment used a system of ranking consisting of four statistical parameters to be consistent with the validation of the other models (Hegarty et al. 2013).

It was shown that the different dispersion models responded in a similar way to differences in driving meteorological data, in particular a tendency to perform better with WRF winds nudged toward NARR. NAME simulations were also conducted using ERA-Interim meteorological data, which generally performed better than the WRF-driven runs despite the lower resolution of the meteorological data. There was no distinctive difference in performance for a particular model, and differences in model performance over the various experiments were generally consistent (e.g., all the models achieved lower ranks for the ANATEX-STC experiment). It was also seen that the increase in the spread of the plume due to the effects of unresolved mesoscale motions improves the performance ranking of NAME. This improvement was consistent across the experiments when setting the parameters to values within the recommended range in Webster et al. (2018), but increasing the values further did not always improve the performance further and sometimes had a

detrimental effect. However, the primary usage of these models is often for the response to accidental releases of harmful material, where both the location and peak concentrations are of interest. Increasing the spread of the predicted plume may improve the performance ranking, but decreasing the predicted peak concentrations and changing the affected area could result in the population-weighted health effects being significantly different, with alternative action being taken during a dispersion event. As such, caution should be taken, as improving the ranking of the model performance may not always lead to more helpful dispersion predictions during emergency events.

Quantitatively evaluating dispersion model performance is difficult due to both the spatial and temporal differences needing to be taken into account, although controlled tracer release experiments such as CAPTEX and ANATEX, where the source parameters are well known, give the opportunity for a more thorough evaluation. This study shows that a range of dispersion models perform similarly for the various tracer release experiments considered here, and that decisions made when running the dispersion model, such as the source of driving meteorological data and parameterization choices for unresolved motions, play a significant role in the performance of the model.

Acknowledgments. We are grateful to Jennifer Hegarty for providing the WRF data. The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the Data Archive of Tracer Experiments and Meteorology (DATEM; https://www.arl.noaa.gov/research/dispersion/datem/) used in this publication. An earlier version of this work was presented at the *17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes* (Selvaratnam et al. 2016).

Data availability statement. The NAME model output used for the analysis in this publication is available via Zenodo (https://doi.org/10.5281/zenodo.7071763; Selvaratnam et al. 2022). The NAME model is available for use under license from the Met Office.

REFERENCES

Berrisford, B., and Coauthors, 2011: The ERA-Interim archive. ERA Rep. Series 1 version 2.0, 27 pp., https://www.ecmwf.int/sites/default/files/elibrary/2011/8174-era-interim-archive-version-20.pdf.

Brioude, J., W. M. Angevine, S. McKeen, and E.-Y. Hsie, 2012: Numerical uncertainty at mesoscale in a Lagrangian model in complex terrain. *Geosci. Model Dev.*, **5**, 1127–1136, https://doi.org/10.5194/gmd-5-1127-2012.

——, and Coauthors, 2013: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1. *Geosci. Model Dev.*, **6**, 1889–1904, https://doi.org/10.5194/gmd-6-1889-2013.

Draxler, R. R., 2006: The use of global and mesoscale meteorological model data to predict the transport and diffusion of tracer plumes over Washington, D.C. *Wea. Forecasting*, **21**, 383–394, https://doi.org/10.1175/WAF926.1.

——, and J. L. Heffter, 1989: Across North America Tracer Experiment (ANATEX) volume I: Description, ground-level sampling at primary sites and meteorology. NOAA Tech. Memo. ERL ARL-167, NOAA Air Resources Laboratory, 83 pp., https://repository.library.noaa.gov/view/noaa/9512/noaa_9512_DS1.pdf.

——, and G. D. Hess, 1997: Description of the HYSPLIT_4 modelling system. NOAA Tech. Memo. ERL ARL-224, NOAA Air Resources Laboratory, 27 pp., https://repository.library.noaa.gov/view/noaa/31133/noaa_31133_DS1.pdf.

——, J. L. Heffter, and G. D. Rolph, 2001: Data archive of tracer experiments and meteorology. NOAA Tech. Rep., NOAA Air Resources Laboratory, 25 pp., https://www.arl.noaa.gov/wp_arl/wp-content/uploads/documents/datem/datem.pdf.

——, and Coauthors, 2015: World Meteorological Organization's model simulations of the radionuclide dispersion and deposition from the Fukushima Daiichi nuclear power plant accident. *J. Environ. Radioact.*, **139**, 172–184, https://doi.org/10.1016/j.jenvrad.2013.09.014.

Ferber, G. J., J. L. Heffter, R. R. Draxler, R. J. Lagomarsino, F. L. Thomas, and R. N. Dietz, 1986: Cross-Appalachian Tracer Experiment (CAPTEX '83) final report. NOAA Tech. Memo. ERL ARL-142, NOAA Air Resources Laboratory, 60 pp., https://www.osti.gov/biblio/5695021.

Grant, A. L. M., H. F. Dacre, D. J. Thomson, and F. Marenco, 2012: Horizontal and vertical structure of the Eyjafjallajökull ash cloud over the UK: A comparison of airborne lidar observations and simulations. *Atmos. Chem. Phys.*, **12**, 10 145–10 159, https://doi.org/10.5194/acp-12-10145-2012.

Hegarty, J., and Coauthors, 2013: Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases. *J. Appl. Meteor. Climatol.*, **52**, 2623–2637, https://doi.org/10.1175/JAMC-D-13-0125.1.

Jones, A. R., D. J. Thomson, M. Hort, and B. Devenish, 2007: The U.K. Met Office's next-generation atmospheric dispersion model, NAME III. *Air Pollution Modeling and Its Application XVII*, C. Borrego and A.-L. Norman, Eds., Springer, 580–589, https://doi.org/10.1007/978-0-387-68854-1_62.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger, 2003: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. *J. Geophys. Res.*, **108**, 4493, https://doi.org/10.1029/2002JD003161.

Loughner, C. P., B. Fasoli, A. F. Stein, and J. C. Lin, 2021: Incorporating features from the Stochastic time-inverted Lagrangian transport (STILT) model into the hybrid single-particle Lagrangian integrated trajectory (HYSPLIT) model: A unified dispersion model for time-forward and time-reversed applications. *J. Appl. Meteor. Climatol.*, **60**, 799–810, https://doi.org/10.1175/JAMC-D-20-0158.1.

Mason, P. J., and N. S. Callen, 1986: On the magnitude of the subgrid-scale eddy coefficient in large-eddy simulations of turbulent channel flow. *J. Fluid Mech.*, **162**, 439–462, https://doi.org/10.1017/S0022112086002112.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Mosca, S., G. Graziani, W. Kulg, R. Bellasio, and R. Bianconi, 1998: A statistical methodology for the evaluation of long-range dispersion models: An application to the ETEX exercise.

Atmos. Environ., **32**, 4307–4324, https://doi.org/10.1016/S1352-2310(98)00179-4.

Ngan, F., and A. F. Stein, 2017: A long-term WRF meteorological archive for dispersion simulations: Application to controlled tracer experiments. *J. Appl. Meteor. Climatol.*, **56**, 2203–2220, https://doi.org/10.1175/JAMC-D-16-0345.1.

Ryall, D. B., and R. H. Maryon, 1998: Validation of the UK Met. Office's name model against the ETEX dataset. *Atmos. Environ.*, **32**, 4265–4276, https://doi.org/10.1016/S1352-2310(98)00177-0.

Selvaratnam, V., H. N. Webster, and D. J. Thomson, 2016: Validation of the atmospheric dispersion model NAME against long-range tracer release experiments. *Proc. 17th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Budapest, Hungary, HARMO, 56–60, https://www.harmo.org/Conferences/Proceedings/_Budapest/lishedSections/PPT/H17-062_oral.pdf.

——, D. J. Thomson, and H. N. Webster, 2022: Validation of the atmospheric dispersion model NAME against long-range tracer release experiments. Zenodo, accessed 8 September 2022, https://doi.org/10.5281/zenodo.7071763.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, 2015: NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Amer. Meteor. Soc.*, **96**, 2059–2077, https://doi.org/10.1175/BAMS-D-14-00110.1.

Stohl, A., M. Hittenberger, and G. Wotawa, 1998: Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data. *Atmos. Environ.*, **32**, 4245–4264, https://doi.org/10.1016/S1352-2310(98)00184-8.

Webster, H. N., and D. J. Thomson, 2002: Validation of a Lagrangian model plume rise scheme using the Kincaid data set. *Atmos. Environ.*, **36**, 5031–5042, https://doi.org/10.1016/S1352-2310(02)00559-9.

——, T. Whitehead, and D. J. Thomson, 2018: Parameterizing unresolved mesoscale motions in atmospheric dispersion models. *J. Appl. Meteor. Climatol.*, **57**, 645–657, https://doi.org/10.1175/JAMC-D-17-0075.1.