17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

VALIDATION OF THE ATMOSPHERIC DISPERSION MODEL NAME AGAINST LONG-RANGE TRACER RELEASE EXPERIMENTS

Vibha Selvaratnam, David Thomson and Helen Webster[1]

[1]Met Office, UK

Abstract: The Met Office's atmospheric dispersion model NAME has been validated against the long-range controlled tracer-release experiments CAPTEX and ANATEX. The model is driven by different sources of meteorology obtained from WRF and from ERA-Interim by ECMWF. The performance of NAME is assessed and compared with the validation of other Lagrangian particle dispersion models against the same experiments.

Key words: *model validation, model comparison, dispersion experiments*

INTRODUCTION

Validating dispersion models against controlled experiments gives the advantage of having more information about the releases than with ad-hoc events. This more accurate information about the releases along with the systematic measurements taken throughout the course of the experiments makes this data ideal for assessing how well the model is performing. It also gives the opportunity to consider how different sources of meteorological data used to drive the models might affect their performance.

The Met Office's atmospheric dispersion model NAME (Numerical Atmospheric-dispersion Modelling Environment) (Jones et al., 2007) is validated against controlled tracer release experiments conducted in North America. Here we will consider two long-range experiments that have also been used to validate other dispersion models (Hegarty et al., 2013). This will give the opportunity for both evaluating against the observations and comparing with the performance of other models.

EXPERIMENTAL DATA

The Cross-Appalachian Tracer Experiment (CAPTEX) and the Across North America Tracer Experiment (ANATEX) were controlled tracer-release experiments conducted in the North American region in the 1980's (Draxler and Heffter, 1989, Ferber et al., 1986).

CAPTEX consisted of seven releases (referred to as CAPTEX-1 through to CAPTEX-7) from 18 September to 29 October 1983. CAPTEX-6 was a short release of 30 minutes from Dayton, Ohio which has been omitted for this comparison to be consistent with the work done by Hegarty et al. (2013). Each of the others was a 3 hour release of perflouromonomethylcyclohexane (PMCH), the first four of which were from Dayton, Ohio and the last two (CAPTEX-5 and CAPTEX-7) from Sudbury, Ontario, Canada. Each release was separated by a few days so each release is treated as a separate experiment. A sampling network of 84 sites 300-800km from the source collected samples of the tracers at ground level. 3 and 6 hour averages were retrieved for 48-60 hours after each release over the period from 19 September to 30 October.

ANATEX consisted of 66 3 hour releases from 5 January to 26 March 1987. Half were releases of perflourotrimethylcyclohexane (PTCH) from Glasgow, Montana (GGW) and the other half were releases of perflourodimethylcyclohexane (PDCH) from St. Cloud, Minnesota (STC). The releases from St. Cloud included releases of PMCH but these were not included in the comparison by Hegarty et al. (2013) as they were coincident with the PDCH releases. As such, they are not included in this validation of NAME to be consistent with the validation of the other models. As each site released a different tracer, releases from each were treated as separate experiments (called ANATEX-GGW and ANATEX-STC). The releases were at 2.5 day intervals to alternate between afternoon and nighttime. The sampling network consisted of 75 sites over the eastern United States and southeastern Canada reaching up to about 3000km from the sources. Air samples averaged over 24 hours were collected at ground level from 5 January through to 29 March.

For this comparison, only the first 10 releases (spanning 5-16 January) from ANATEX are included to make it comparable in length to CAPTEX and to use the same period used in previous validations of particle dispersion models (Hegarty et al., 2013). This period at the start of the experiment also gives the contrast of winter conditions compared to the summer like conditions during CAPTEX. Each CAPTEX release was represented by 50,000 particles and the ANATEX releases by 25,000 particles which is consistent with the number of particles released by other model runs. Increasing the number of particles seems to have little to no effect on the analyses and the fewer particles released for the ANATEX releases is compensated for by the longer averaging time for calculating the air concentrations.

DISPERSION MODELS

All the models discussed here are Lagrangian particle dispersion models. Particles are released from a source location and are advected by the mean winds obtained from input meteorological data and a random component added by the model to represent turbulence.

Numerical Atmospheric-dispersion Modelling Environment (NAME)

The NAME output is given on a 0.25 x 0.25 degree grid in the lat-long coordinate system and concentrations are calculated over the lowest 100m agl. The meteorological data was used on its native horizontal grid with the vertical grid being interpolated and the model time-step was set to 15 minutes. Dispersion due to both turbulence and unresolved mesoscale motions were represented while convection above the boundary layer was not.

Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT)

The dispersion calculations in HYSPLIT conducted by Hegarty et al. (2013) were made on the same horizontal grid as the meteorological data. HYSPLIT uses an internal terrain following vertical coordinate so the meteorological fields are linearly interpolated to this grid. The lowest vertical grid level was at approximately 10m above ground level and the resolution decreases with height. The resolution of the output grid and the number of particles representing each release are the same as in the NAME runs and the time-step was 1 minute (Hegarty et al., 2013).

Stochastic Time-Inverted Lagrangian Transport (STILT)

STILT is built upon HYSPLIT and so has many of the same features such as the mean advection scheme and the calculation grid. Although STILT is primarily used in backward mode, for this comparison only the performance of the model run forward in time has been considered. The configuration used is the same as HYSPLIT (for example output concentration grid and number of particles released) but STILT simulates turbulence differently (Hegarty et al., 2013).

Flexible Particle (FLEXPART)

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

METEOROLOGICAL DATA

The models are run using reanalysed meteorological data from different sources and in some cases, different configurations of the same numerical weather prediction model. The meteorological data used for the two sets of comparisons are detailed below.

The ERA-Interim meteorological fields are from ECMWF. ERA-Interim is a global atmospheric reanalysis starting from 1979 produced with a 2006 version of the IFS (Integrated Forecast System). The spatial resolution is approximately 80km with 60 vertical levels and output is given every 3 hours.

The North American Regional Reanalysis (NARR) is an extension of the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) Global Reanalysis which has been run over the North American region. The reanalysis was run from 1979 to 2014 and is on a Lambert Conformal grid of approximately 32km and 45 vertical levels with output every 3 hours. This data is used to drive the WRF data used.

The Advanced Research version of the WRF (Weather Research and Forecasting) model (ARW) is used and the initial and boundary conditions are obtained from NARR. The configuration of ARW used is on a Lambert Conformal horizontal grid and uses a terrain-following hydrostatic-pressure vertical coordinate system of 43 levels, with the lowest approximately 33m thick. The model was configured with two nested horizontal resolutions of 30km and 10km with one way boundary conditions between the two and 3 hourly output. Output from the ARW model with wind nudging towards NARR in the boundary layer both turned on and off are used to run the dispersion models.

STATISTICAL MEASURES

Assessing the accuracy of a dispersion model is difficult due to having both temporal and spatial variations. Each statistical parameter has different sensitivities to these variations. We use the same system of ranking 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 NOAA Air Resources Laboratory.

The correlation coefficient (R) ranges from -1 to 1 where 1 is a perfect (positive) correlation between measured and predicted concentrations. The normalised sum of R^2 ranging from 0 to 1 is the value that contributes to the final rank. The fractional bias (FB) is a fraction of the average between paired predicted and measured values and ranges from -2 to 2. The values are paired in both space and time. A positive value indicates an overprediction by the model and a negative value an underprediction. The figure of merit in space (FMS) is a percentage overlap between measured and predicted areas at a fixed time. A fixed significant concentration level is set (although here a value of 0 g m^{-3} is used) and the percentage is the proportion of sites that agree to be either above or below this level. Because it is evaluated at a fixed time, 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 slightly shifted in space or time. This bias is particularly pronounced with narrow plumes. The Kolmogorov-Smirnov parameter (KSP) is the maximum absolute difference between two cumulative distributions (expressed as percentages), so a smaller value implies a better prediction (Mosca et al. 1998).

These four parameters are equally weighted so that each can contribute a maximum value of one to the final rank, which ranges from 0 to 4 and a higher rank implies a better prediction. The formula used is

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

RESULTS

Table 1 and Figure 1 show how well NAME performs when being driven by different meteorological data. The three sources of driving fields are from WRF, both with wind nudging towards NARR switched on (V1) and with it switched off (V2) and from ERA-Interim. Figure 1 shows how each of the statistical parameters contributes to the final rank given for each simulation as well as averages for each of the sources of meteorological data. We see that NAME driven by WRF generally performs better when the wind nudging towards NARR is switched on compared to when it is switched off. We also see that NAME seems to perform best when run with the ERA-Interim meteorology despite the lower resolution of the data.

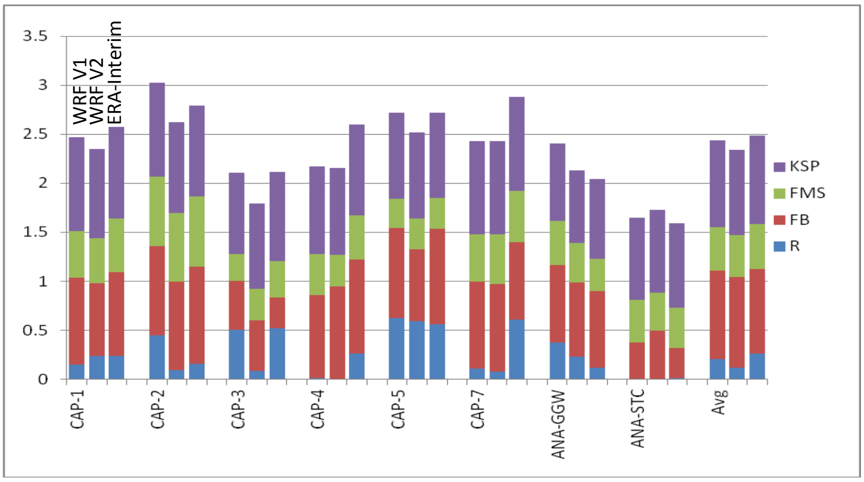

Figure 1. Bar plots of the normalised statistical parameters contributing to the rank as calculated in Eq. (1) for NAME driven by meteorological fields from WRF V1 (column 1), WRF V2 (column 2) and ERA-Interim (column 3).

The two cases where the model performs noticeably better when driven by ERA-Interim rather than with the WRF are for CAPTEX-4 and CAPTEX-7. In both of these experiments we see that the difference in the contribution from the correlation coefficient appears to make the most difference in the overall rank as it is almost zero for both the WRF driven simulations. The other statistical parameters seem to be largely the same for all three runs. However, the correlation coefficient does not always give an indication of the final rank, for example for ANATEX-STC where the correlation coefficient is almost zero for all of the runs but the final rank does differ. This reiterates the fact that a single statistical parameter may not give an accurate representation of how well the model is performing overall due to different biases but in combining the four, this should give a more robust indication.

Table 1. Rank results from the NAME model evaluation. Simulations are driven by WRF time-averaged fields with grid nudging of winds towards NARR either turned on (V1) or off (V2), and ERA-Interim fields.

Experiment	WRF V1	WRF V2	ERA-Interim
CAPTEX-1	2.47	2.35	2.58
CAPTEX-2	3.02	2.63	2.79
CAPTEX-3	2.11	1.79	2.12
CAPTEX-4	2.17	2.15	2.60
CAPTEX-5	2.73	2.51	2.72
CAPTEX-7	2.43	2.43	2.89
ANATEX-GGW	2.40	2.12	2.04
ANATEX-STC	1.65	1.73	1.59
Avg	2.37	2.21	2.42

Figure 2 shows how all the dispersion models perform being driven by both versions of the 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 V1 data where the wind nudging towards NARR is turned on compared to that when there is no wind nudging. HYSPLIT and STILT however do not seem to have a preference for one set of WRF meteorology over the other and on average there is little difference between the two. As STILT is built upon the HYSPLIT model, it is unsurprising that they react to a change in driving meteorology in a similar way and, for each experiment, they both show a similar difference in performance using one set of meteorology over the other.

On average, the performances of each of the models are relatively similar. It can be seen that FLEXPART driven by WRF V2 meteorological data generally has a slightly lower ranking than the other three models apart from for the CAPTEX-7 experiment where it noticeably performs 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 that all the models had the weakest performance for. Although this could be due to the WRF data having errors, we see in Figure 1 that NAME runs driven by ERA-Interim shows the same low ranking for this experiment so it is likely that there were some difficult conditions to predict over this time and location.

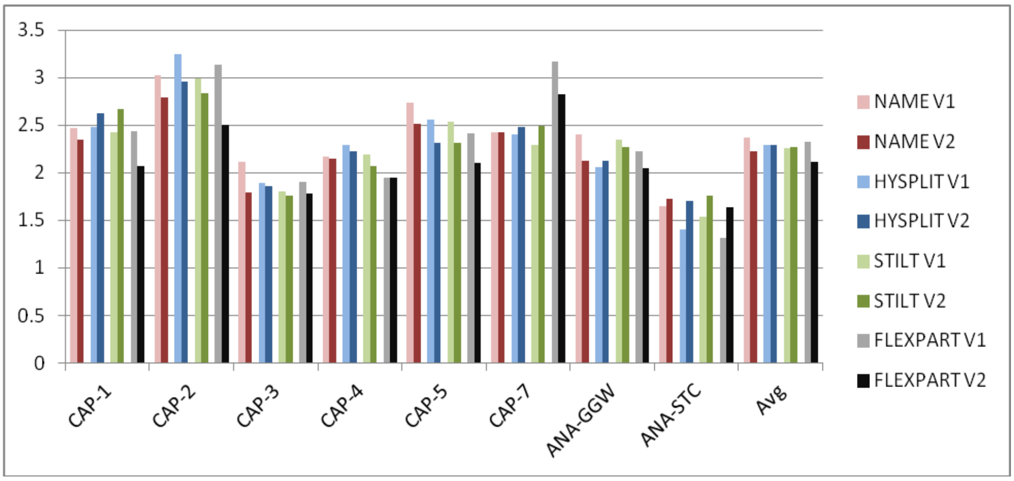

Figure 2. Bar plot of the rank for NAME, HYSPLIT, STILT and FLEXPART driven by meteorological fields from WRF with wind nudging on (V1) and WRF without wind nudging (V2).

CONCLUSIONS

The Lagrangian particle dispersion model NAME was run to simulate controlled tracer-release experiments and the performance evaluated when driven by different meteorological fields as well as comparing the performance with other dispersion models. 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 found that the different models responded in a similar way to differences in driving meteorology, in particular a tendency to perform better with WRF winds nudged towards NARR. NAME simulations were also conducted using ERA-Interim meteorology which generally performed better than the WRF driven runs despite the lower resolution. There was no distinctive difference in performance for a particular model and differences in model performance over the different experiments were generally consistent, for example all the models achieved lower ranks for the ANATEX-STC experiment.

REFERENCES

Draxler, R.R. and J.L. Heffer, 1989: Across North America Tracer Experiment (ANATEX) Volume I: Description, ground-level sampling at primary sites and meteorology. *NOAA Tech. Memo.* ERL ARL-**167**, 83 pp.

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 '3) final report. *NOAA Tech. Memo.* ERL ARL-**142**, 60 pp.

Hegarty, J., R.R. Draxler, A.F. Stein, J. Brioude, M. Mountain, J. Eluszkiewicz, T. Nehrkorn, F. Ngan and A. Andrews, 2013: Evaluation of Lagrangian Particle Dispersion Models with Measurements from Controlled Tracer Releases. *J. Appl. Meteor. Climatol.*, **52**, 2623-2637.

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.

Draxler, R.R., J. L. Heffter, G. D. Rolph, 2001: Data Archive of Tracer Experiments and Meteorology. [Available online at http:// http://www.arl.noaa.gov/documents/datem/document/datem.pdf].

Jones A.R., Thomson D.J., Hort M. and Devenish B., 2007: The U.K. Met Office's next-generation atmospheric dispersion model, NAME III. Borrego C. and Norman A.-L. (Eds) Air Pollution Modeling and its Application XVII (Proceedings of the 27th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application), Springer, 580-589.