A new method for evaluating regional air quality forecasts

H. F. Dacre

Department of Meteorology, University of Reading, Reading, RG6 6BB, UK.

Abstract

A Kriging interpolation method is combined with an object-based evaluation measure to assess the ability of the UK Met Office's dispersion and weather prediction models to predict the evolution of a plume of tracer as it was transported across Europe. The object-based evaluation method, SAL, considers aspects of the Structure, Amplitude and Location of the pollutant field. The SAL method is able to quantify errors in the predicted size and shape of the pollutant plume, through the structure component, the over- or under-prediction of the pollutant concentrations, through the amplitude component, and the position of the pollutant plume, through the location component. The quantitative results of the SAL evaluation are similar for both models and close to a subjective visual inspection of the predictions. A negative structure component for both models, throughout the entire 60 hour plume dispersion simulation, indicates that the modelled plumes are too small and/or too peaked compared to the observed plume at all times. The amplitude component for both models is strongly positive at the start of the simulation, indicating that surface concentrations are over-predicted by both models for the first 24 hours, but modelled concentrations are within a factor of 2 of the observations at later times. Finally, for both models, the location component is small for the first 48 hours after the start of the tracer release, indicating that the modelled plumes are situated close to the observed plume early on in the simulation, but this plume location error grows at later times. The SAL methodology has also been used to identify differences in the transport of pollution in the dispersion and weather prediction models. The convection scheme in the weather prediction model is found to transport more pollution vertically out of the boundary layer into the free troposphere than the dispersion model convection scheme resulting in lower pollutant concentrations near the surface and hence a better forecast for this case study.

Key words: pollution transport, Unified Model, NAME dispersion model, feature-based evaluation, ETEX, SAL

1. Introduction

Air quality models can be used for a variety of applications including the prediction of high pollution episodes, to determine the suitability of new pollution source sites and to inform decisions on air pollution strategy and regulation. Thus, it is important that they are evaluated against observations regularly to determine their predictive capability over the range of spatial and temporal scales for which they are applied. However, this evaluation is not straightforward, especially for Eulerian air quality models which predict the mean concentration of a given distribution of pollutant concentrations within a given area. Observations, however, represent a single member of this overall distribution at a specific location within this area. As the spread of concentrations around the mean is large, due to natural variability, it is not possible to predict the concentration observed at a given time and location downwind of a source using air quality models (Chatwin (1982), Weil et al. (1992)). Despite this representivity issue, traditional air quality verification methods are largely based on the gridpoint comparison of forecast and observed concentrations. Gridpoint-based verification methods, such as normalised-mean-square-error, can be problematic for pollutant fields with small-scale complex structure as forecasts are often unfairly penalised due to small positional errors (known as the 'double penalty' problem). This problem becomes worse as the resolution of the forecasts increases. In addition, many gridpoint-based verification methods fail to take into account the spatial correlation existing within pollution fields as they compare observations and forecast quantities at each location independently. Thus there is a growing need to develop new air quality evaluation methods that can take into account the spatial correlation and complex structures found in high resolution pollution fields.

The verification of precipitation forecasts suffers from the same problems as air quality verification due to the fact that both atmospheric dispersion and convective precipitation are stochastic phenomenon. In the field of quantitative precipitation forecasting there has been much work done recently to create new verification techniques which can be performed for forecasts of higher spatial and temporal resolution. These new techniques fall into three categories; neighbourhood based, scale decomposition and feature based (Casati et al. (2008)). Neighbourhood based techniques consider neighbouring observations in space and time in the forecast-observation evaluation thus they relax the requirements for perfect time-space matching (e.g. Roberts and Lean (2007)). However, these neighbourhood techniques rely on a dense network of observations such as radar observations which are rarely available for regional air quality evaluation. Scale decomposition techniques decompose the forecast and observed fields into the sum of spatial components on different scales. Verification is then performed on each scale component separately. These methods are useful for assessing the ability of forecasts to reproduce the spatial structure of the observed field and thus are able to determine the potential improvements due to high resolution forecasts. Finally, feature based methods identify features, such as intense rain events, in the predicted and observed fields and then assess different attributes

associated with each individual pair of forecast-observed features. Wernli et al. (2008) extended this method to avoid the problem of matching forecast and observed features by using a technique which assessed the structure, amplitude and location error for features without any matching requirements.

The main aim of this work is to show how spatial verification methods, that have been developed to evaluate NWP-gridded precipitation forecasts, can be used to evaluate air quality forecasts. This is achieved by using the object-based evaluation method of Wernli et al. (2008) to quantitatively evaluate the ability of the UK Met Office's NWP and dispersion models to predict regional pollution concentrations. The modelled predictions are compared against observations from the European Tracer Experiment (ETEX) that have been Kriged onto a regular grid. The results are used to highlight differences in the representation of pollution transport in the two models.

The ETEX field campaign is described in section 2. The models used in this paper are described briefly in section 3 and the SAL diagnostics are described in section 4. In section 5 an analysis of the observed tracer field during the first ETEX release is given. The SAL diagnostics are evaluated for the UM and NAME tracer experiments in section 6 and the role of convection is evaluated in section 7. Finally, in section 8 the main conclusions are given. A description of the Kriging method of interpolation is given in the appendix.

2. The European tracer experiment, ETEX

Two long-range dispersion experiments were carried out as part of ETEX during October and November 1994. During the releases, a non-toxic, non-depositing, non water-soluble, inert tracer (perfluoromethylcyclopentane) was released from a site near Monterfil in north-west France ($12°00'30''$W, $40°03'30''$N). 168 stations, all part of the synoptic network of national meteorological services, in 17 countries were equipped with air samplers and performed 3 hourly sampling (see Dacre (2010) figure 1). A complete description of the ETEX experiment was published by Van Dop et al. (1998) and Gryning et al. (1998).

Since the release rate was well known and deposition and chemistry processes did not occur the experiment provides a good test of atmospheric transport in models from a point source at continental scale. 24 institutions took part in real-time forecasting of plume evolution, with 28 long-range chemistry transport models, using meteorological data from various sources. The first ETEX experiment (ETEX 1) has been discussed in many papers in which the modelling results have been compared to observations (Ryall and Maryon, 1998; Stohl et al., 1998; Nasstrom and Pace, 1998; D'Amours, 1998). It was concluded in a review paper by Mosca et al. (1998), that almost all the models showed a satisfactory agreement with the measured values for ETEX 1. As such, this experiment provides a good case to test the ability of the UK Met Office's numerical weather prediction model (NWP), UM, to simulate the long-range transport of pollution from a point source release. Comparison of UM results with output from the UK Met Office's dispersion model, NAME, which took

part in the original ETEX intercomparison will be made along with comparison with the observations.

The ETEX 1 tracer release took place between 16 UTC on 23 October and 03:50 UTC on 24 October 1994. Figure 1(a) shows the UM mean sea level pressure field at 00 UTC on 24th October 1994 overlaid with frontal analysis. A mature low pressure system, was located north of the UK and approached Europe from the west. The cold front passed over the release site prior to the start of the tracer release. Low level winds were south-westerly behind the front. Figure 1(b) shows an infrared image from the AVHRR satellite at 07:26 UTC on 23 October 1994. The main polar-front cloud band lies south of the low pressure centre parallel to and ahead of the surface cold front and wraps cyclonically around the low pressure centre. Behind the cold front convective cloud can be seen in the location of the release site.

3. Model descriptions

Simulations of the ETEX 1 release have been performed using two models: The UK Met Office Lagrangian dispersion model, NAME, and the UK Met Office Eulerian NWP model, UM.

3.1. NAME

NAME III is the currently operational dispersion model used to perform air quality forecasts at the UK Met Office (Jones et al. (2007)). NAME (Numerical Atmospheric-dispersion Modelling Environment) is a Lagrangian particle trajectory model designed to predict the dispersion and deposition of gases and particulates in the atmosphere. Emissions of gases or particles are modelled by releasing a large quantity of particles into the model atmosphere, with each particle representing a mass of released pollutant. The particles are carried along passively by the 3D wind with turbulent mixing represented by random walk techniques using empirical turbulence profiles. In our simulations all meteorological data was obtained from the UK Met Office NWP model.

For the ETEX 1 simulation the tracer release is represented in NAME by the release of 2 million particles, released at a rate of 7.98gs^{-1} at a height of 20m. (This height was chosen to be consistent with the UM simulations, the actual release height is 8m). NAME was driven using $0.442°$ resolution meteorological data from the UM input into NAME every hour. NAME pollutant concentrations are computed by summing the mass of particles in a $1.0° \times 1.0°$ latitude/longitude area.

3.2. UM

The ETEX 1 tracer release has also been simulated using the UM, version 6.1. This model solves the non-hydrostatic primitive equations using a semi-implicit, semi-Lagrangian numerical scheme (Cullen, 1993). The model includes a comprehensive set of parameterisations, including boundary layer turbulent mixing (Lock et al., 2000), mixed phase microphysics (Wilson and Ballard, 1999)

and convection (Gregory and Rowntree, 1990). There is no explicit diffusion in the model. A limited area domain with horizontal gridlength of 0.442° was used over Europe extending from 37.5°N to 62.47°N and 9.5°W to 22.62°E. The model has 38 levels in the vertical on a stretched grid ranging from the surface to 5hPa. This corresponds to approximately 100m layer spacing in the boundary layer and 500m layer spacing in the mid-troposphere.

For the ETEX 1 simulation the tracer release is represented in the UM by a constant emission of tracer at the first model level, 20m, in a single gridbox. Tracers in our simulation are treated as passive substances, they are subject to advection, convection and turbulent transports but are neither deposited nor chemically transforming. A small amount of transport may occur due to numerical diffusion. This methodology has also been used by Donnel et al. (2001); Gray (2003); Dacre et al. (2007); Dacre (2010).

4. SAL diagnostics

In this section a non-numerical explanation of the SAL verification diagnostics are described. A detailed description of the mathematical formulation of the SAL diagnostics can be found in Wernli et al. (2008). The computation of the location and structure components require first the identification of individual objects within the considered domain, separately for the observed and predicted tracer fields. In this paper 'objects' are defined as regions in which tracer concentrations exceed a specified threshold value. A threshold value of 1/15th of the maximum tracer concentration value is used to identify coherent objects. This threshold is used for both the observed and modelled tracer fields. The choice of this threshold is not based on objective criteria but sensitivity tests have shown that the results are not sensitive to the choice of threshold value. SAL diagnostics calculated using a fixed threshold, set to the measuring instruments detection limit of 0.01ng m^{-3} also produce very similar results.

4.1. The structure component, S

The structure component of SAL compares the volume of the normalised tracer plume. It measures the size and shape of the objects. Values of S are within ±2 and 0 denotes a perfect forecast in terms of structure. S is positive if the model predicts widespread tracer in a situation of small tracer objects and negative if the model predicts a small peaked object in a situation of a large flat tracer object. The possibility to identify these kinds of errors is one of the key characteristics of SAL.

4.2. The amplitude component, A

The amplitude component of SAL measures the normalised difference of the domain averaged tracer values. It provides a measure of the quantitative accuracy of the total tracer in a specified region. Values of A are within ±2 and 0 denotes a perfect forecast in terms of amplitude. The value of $A = +0.67$ indicates that the model over-estimates the domain-averaged tracer by a factor of 3 and a value of $A = -0.67$ indicates an under-estimation by a factor of 2.

4.3. The location component, L

The location component of SAL consists of two parts: $L = L_1 + L_2$, and describes the accuracy with which tracer is distributed within the domain. L_1 measures the distance between the centers of mass of the predicted and measured tracer fields normalised by the largest distance between two boundary points in the domain (2816 km). The values of L_1 are in the range 0 to 1. In case of $L_1 = 0$, the centres of mass of the predicted and observed tracer fields are identical. Note that many different tracer fields can have the same centre of mass and therefore $L_1 = 0$ does not necessarily indicate a perfect forecast. For example, a predicted tracer field with two objects on opposite sides in the domain can have the same centre of mass as an observed tracer field with one object located in-between the two predicted events. The second part, L_2, considers the difference in the distribution of objects relative to the centre of mass between the modelled and observed tracer fields. If both the modelled and observed tracer fields contain only one object, then $L_2 = 0$.

5. Observed tracer transport

In this section the observed tracer concentrations from the ETEX 1 experiment are described. Because an existing network was used there were some limitations in the spatial resolution of the sampling. For example, the resolution of the sampling network close to the release site was too coarse to properly resolve the near source dispersion (Mosca et al. (1998)). The observations have been interpolated onto a $1.0^\circ \times 1.0^\circ$ latitude/longitude grid using a geostatistical Kriging technique. Kriging is an interpolation method for estimating values at locations which have not been sampled using a weighted average of neighbouring samples to estimate the unknown value at a given location. Details of the Kriging method are given in the appendix.

Figures 2(a)-(e) show the evolution of the Kriged tracer concentrations 12, 24, 36, 48 and 60 hours after the start of the tracer release respectively. The tracer is advected east-northeast across Europe by the low-level winds behind the cold front. During the initial stages of the tracer release (first 24 hours), the plume axis is orientated southwest-northeast and tracer concentrations up to 1.7ng m^{-3} are observed. During the second 24 hours of the tracer experiment the plume is more widespread and its axis becomes orientated northwest-southeast. Finally, 60 hours after the start of the tracer release, the Kriged tracer plume has been deformed into an elongated strip that extends from the west of Norway to southeast Europe.

6. Model results

In this section a subjective visual comparison will be made between the Kriged observations and model predictions from the UM and NAME. This will be followed by an objective quantitative comparison using the SAL diagnostics.

Figures 3(a)-(e) show the tracer concentrations predicted by the UM at a height of 20m, 12, 24, 36, 48 and 60 hours after the start of the tracer release respectively. Figures 4(a)-(e) show the tracer concentrations predicted by NAME at a height of 20m, 12, 24, 36, 48 and 60 hours after the start of the tracer release respectively. Comparison with the Kriged observations (figures 2(a)-(e)) shows that 12 hours after the start of the tracer release both the UM and NAME over-predict tracer concentrations. The extent of the plume is also more widespread in both model predictions compared to the observations. This over-prediction could be a result of the coarse observation network close to the source that is unable to capture the peak concentrations in the plume. 24 hours after the start of the tracer release, both the magnitude of the over-prediction and the over-estimation of the plume extent has reduced in both models. The modelled and observed tracer plumes compare well at this time. 36 hours after the start of the tracer release the magnitude of tracer concentrations is similar to the observed tracer concentrations but the modelled plumes are not as widespread as the observed tracer plume. The observed tracer plume is orientated in a west-east direction whilst the tracer plume in both the UM and NAME simulations is orientated in a southwest-northeast direction. 48 hours after the start of the tracer release this orientation error is even more pronounced. The tracer plume in both models has started to spread along the north-west/south-east axis but the plumes do not extend as far as the observed plume over east-central and south-east Europe. Finally, 60 hours after the start of the tracer release, the UM captures the transport of the observed tracer plume to the North Sea and the tracer transport south-eastwards but NAME does not. The plume in both models is also too widespread compared to the observations. It is hypothesised that this is a result of a failure of the meteorology to capture the anti-cyclonic transport of the tracer around the high pressure system situated over Europe. The plume orientation errors due to uncertainty in the meteorology are not investigated in this paper.

Figures 5(a)-(c) show the time evolution of the structure, amplitude and location components for both the UM and NAME simulations. The SAL components are calculated every 3 hours for the 3-hourly averaged tracer concentrations from 12 hours after the start of the tracer release onwards. Before this time, the plume is narrow and is not well sampled by the coarse observation network. This results in SAL diagnostics that are not representative of the model performance during the first 12 hours of the experiment.

6.1. Structure component

Figure 5(a) shows the time evolution of the structure component for both the UM and NAME simulations. 12-15 hours after the start of the tracer release, both the UM and NAME have negative structure component. This is a result of the models predicting a plume that is more peaked than the observed plume, due to higher maximum concentrations. Between 24 and 33 hours after the start of the tracer release both models have an structure component that is close to zero. This occurs because both the amplitude of the predicted tracer concentrations closely matches the Kriged observations, and because the size

of the predicted plume is similar to the Kriged observations. From 36 hours after the start of the tracer release onwards, the structure component again becomes negative. This is because the Kriged plume is more widespread than the modelled tracer plumes. The UM performs better than NAME over this period as the UM tracer plume is slightly more extensive than the NAME predicted plume. The peak positive structure component that occurs 39 hours after the start of the tracer release is because the Kriging technique produces a split plume (figure 6). The split plume occurs as a result of the coarse resolution of the observing network at longitudes $> 15°$E. A denser network of observations would probably result in a single plume being identified. In the modelled tracer fields tracer concentration data is available at higher resolution than the observations and hence only one object is identified. The Kriged observation tracer field contains two objects, one large and one small, compared to the modelled tracer fields which only contain one object. Thus the average size of the Kriged observation tracer objects is smaller than the modelled tracer object resulting in a positive structure component. From 48 hours onwards the structure component increases and eventually becomes positive 57 hours after the start of the tracer release indicating that the modelled plumes are too large and flat at the later stages of the simulation. This agrees with a visual inspection of the plumes (figures 3(e) and figure 4(e)). Overall the evolution of the UM and NAME structure components are very similar. However, the NAME structure component is significantly different to the UM structure component between 18 and 21 hours after the start of the tracer release. A visual inspection of the tracer concentration at 20m indicates that this is because the UM predicted tracer plume contains higher tracer concentrations (peak concentration of 5.1ng m^{-3} compared to a peak concentration of 3.8ng m^{-3} in NAME) within a similar sized plume resulting in a plume that is more peaked than the NAME predicted tracer plume (not shown). It is hypothesised that this difference is a result of the different convection schemes in the NWP and dispersion models. This is investigated further in section 7.

6.2. Amplitude component

Figure 5(b) shows the SAL amplitude component. It is a measure of the over- or under-estimation of tracer concentrations compared to the Kriged observations. 12 hours after the start of the tracer release both the UM and NAME simulations strongly over-predict tracer concentrations by a similar amount, amplitude component > 1.5. As the time since the start of the release increases the severity of the over-prediction reduces for both models. For both simulations tracer concentrations are within a factor of 2 of the Kriged observations from 30 hours onwards which is considered to be a good forecast. However, from 24 hours onwards the UM has a lower amplitude component than NAME. This difference occurs because NAME has higher concentrations at 20m than the UM. The difference in tracer transport responsible for this are discussed in section 7.

6.3. Location component

Figure 5(c) shows the location component of the SAL diagnostics. Between 12 and 36 hours after the start of the tracer release it is a measure of the difference in the location of the Kriged observations and the model predicted tracer field centre of mass as there is only one object identified. The L_2 part of the location component can only be non-zero if there is more than one object identified in either the observed or simulated fields. We can convert the L_1 component into a physical distance by multiplying by the maximum distance between gridpoints in the entire domain (2816km). Thus between 12 and 36 hours after the start of the release both the UM and NAME simulations have a centre of mass that is within 200km of the Kriged observed centre of mass. This is considered to be a good forecast since the resolution of the interpolated model output is approximately 100km. The peak in the location component that occurs 39 hours after the start of the release is because two objects are identified in the Kriged observations and only one in both the UM and NAME simulations, as discussed above. This results in a non-zero L_2 component and hence an increased location component at this time. From 42 hours after the start of the release onwards, there is a gradual increase in the location component for both the UM and NAME simulations. This is due to the fact that the orientation of the modelled and observed tracer plume differs from the observations and the plume is more widespread than the observations and is likely to be a result of errors in the meteorology.

7. Transport by convection

It was shown in section 6.2 that systematic differences in the amplitude component between the UM and NAME simulations occur from 21 hours after the start of the tracer release onwards. An important process for transporting tracer is deep convection. It has been shown that tracer can be rapidly transported out of the boundary layer into the free troposphere by convection (Dacre et al. (2007), Chagnon et al. (2007)). The representation of deep convection is treated differently in the UM and NAME models. This can result in different vertical distributions of tracer and hence differences in the SAL diagnostics. Details of the convection scheme used in NAME are given in Maryon et al. (1999). Details of the convection scheme used in the UM are given in Gregory and Rowntree (1990). It is hypothesised that differences in the transport of tracer by convection in these models is the cause of the amplitude component differences.

Figure 7 shows the amplitude component for simulations in which tracer is prevented from being transported by the convection schemes in both the UM and NAME. Preventing tracer from being transported by the UM has a large impact on the amplitude component from 21 hours after the start of the tracer release onwards. A more positive amplitude component results as concentrations are higher at 20m when tracer is not removed by convection. Preventing tracer from being transported by convection in NAME has a negligible effect on the amplitude component, infact the time-series of the amplitude component for

NAME simulations with and without convection are superimposed in figure 7. The effect of including or excluding convective transport also has a small effect on both the structure and location components for both the UM and NAME simulations (not shown).

Figure 8(a) shows convective rain rate 20 hours after the start of the tracer release. The dashed ellipse highlights the region of convection diagnosed in the UM that occurs over the region of tracer in figure 3. In NAME, convective mixing is triggered only where convective cloud is present with a depth greater than 100mb and a base below 800mb. Figure 8(b) shows the pressure at the convective cloud base 20 hours after the start of the tracer release. In the highlighted region the convective cloud base is below 900mb. Figure 8(c) shows the pressure at the convective cloud top. In the highlighted region the convective cloud top is above 600mb. Thus the convective scheme in NAME is active in this region.

Figure 9(a) shows the 3-hourly averaged UM tracer concentration > 0.001ng m^{-3} at heights of 20m and 3920m, 21 hours after the start of the tracer release. The contour at 3920m covers a large area indicating that the convection scheme has transported large amounts of tracer out of the boundary layer into the free troposphere. Figure 9(c) shows a vertical cross-section of 3-hourly averaged tracer concentration taken along the line shown in figure 9(a). It can be seen that whilst the highest tracer concentrations are still confined to the boundary layer (< 1km), a significant amount of tracer has been transported up to 6km in the atmosphere by the convection scheme. Figure 9(b) shows the 3-hourly averaged NAME tracer concentrations > 0.001ng m^{-3} at heights of 20m and 4000m, 21 hours after the start of the tracer release. In NAME the extent of the contour at 4000m is much smaller than in the UM. Figure 9(d) shows a vertical cross-section of 3-hourly averaged tracer concentration taken along the line shown in figure 9(b) overlaid with the diagnosed boundary layer height. As for the UM, tracer has been transported out of the boundary layer up to 5km. However, the amount of tracer transported vertically by the convection scheme in NAME is much less than that in the UM by a factor of 100. This results in NAME simulating higher peak concentrations than the UM at low-levels and explains why the amplitude component is higher for NAME than the UM and also why preventing tracer from being transported by convection scheme in NAME has little effect on the amplitude component.

8. Conclusions

In this paper an object-based evaluation method, SAL, has been combined with a Kriging interpolation method to quantitatively evaluate the ability of the UK Met Office's numerical weather prediction and dispersion models to predict the evolution of a plume of tracer as it was transported across Europe. The SAL method is able to quantify errors in the predicted size and shape of the tracer plume, through the structure component, the over- or under-prediction of the tracer plume, through the amplitude component and the position of the tracer plume, through the location component. The objectively determined

results of the SAL evaluation are similar to a subjective visual inspection of the predictions which is an attractive attribute of this method.

Although the UM and NAME predictions show a similar performance for plume structure and location, differences are identified in the amplitude component. By evaluating the transport of tracer by deep convection in both models, it has been shown that the differences in the amplitude component occur at times when convective transport is diagnosed. The UM convection scheme transports more tracer vertically out of the boundary layer into the free troposphere than the NAME convection scheme. This results in lower tracer concentrations within the boundary layer in the UM prediction than in NAME, which in turn leads to a lower amplitude-component for the UM compared to NAME and hence a better forecast. Thus, the SAL methodology can be used to identify differences in the transport of tracer between models.

In this paper a case study in which the emission rates are well known and chemical transport does not occur was simulated in order to diagnose differences in the representation of transport in the UM and NAME models. Although the evaluation in this paper has focused on a case study, in principle the SAL method could be used to evaluate the performance of models over a longer time period and hence to identify systematic errors on daily and hourly timescales and to determine predictability limits. In the future the SAL methodology could be used to compare forecasts from different resolution simulations in the same model or ensemble predictions with varying meteorology or emissions. The SAL diagnostics were used to compare forecasts for simulations performed at 50km, 12km and 4km resolution for the ETEX 1 case study. However, due to the coarse resolution of the observations it was not possible to identify differences in the tracer transport in simulations at different resolutions. It is anticipated that high resolution air quality datasets will be needed in the near future in order to evaluate high-resolution air quality forecasts.

9. Acknowledgments

I would like to thank the UK Met Office for use of the UM and NAME and Stefano Galmarini for providing the ETEX tracer measurements. I am grateful to Heini Wernli for providing the SAL diagnostics, to Helen Greatrex for help with the Kriging method and to Geovanni Leoncini for helpful comments on the paper.

A. Kriging the observed data

The Kriging software package used in this paper has been designed at the University of Reading to Krige rainfall datasets from rain gauges (Greatrex (2010)). It has been customised to read input from the ETEX dataset. The software follows 7 steps which are as follows:

- Step 1 formats the ETEX dataset so it can be read by the Kriging software.

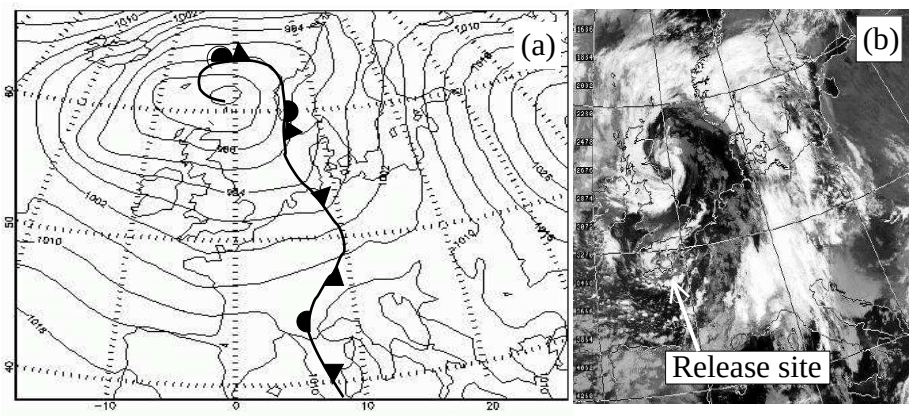

Figure 1: (a) UM mean sea level pressure at 00UTC on 24 October 1994, fronts overlaid. (b) AVHRR infrared at 07:26 UTC on 23 October 1994 courtesy of NASA Goddard Space Flight Centre.

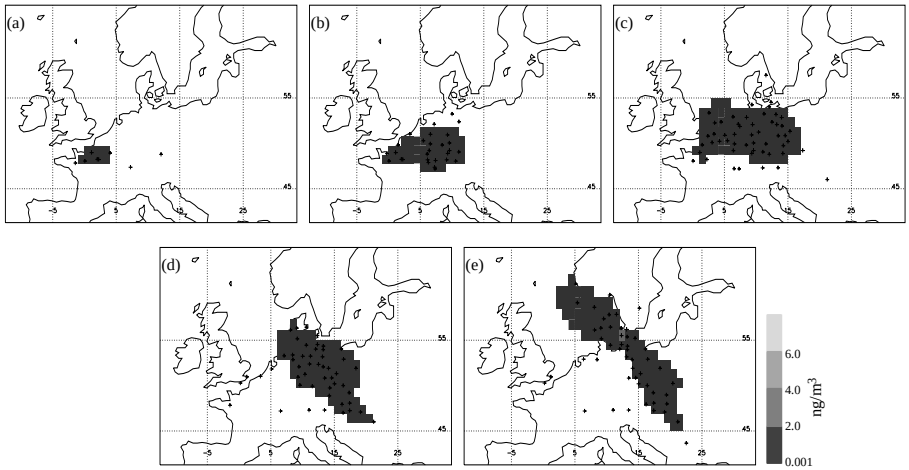

Figure 2: Kriged non-zero tracer concentrations 12, 24, 36, 48 and 60 hours after the start of the tracer release. The non-zero observations are superimposed as crosses.

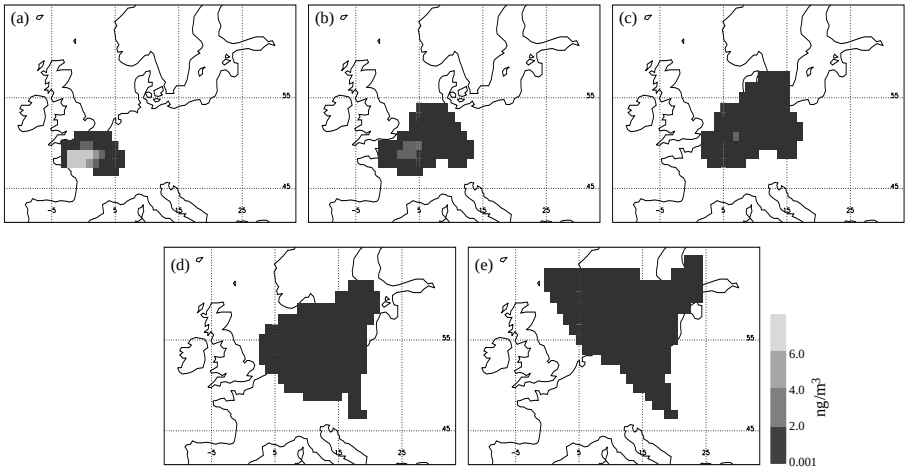

Figure 3: UM tracer concentration interpolated onto $1^\circ \times 1^\circ$lat/lon grid 12, 24, 36, 48 and 60 hours after the start of the tracer release.

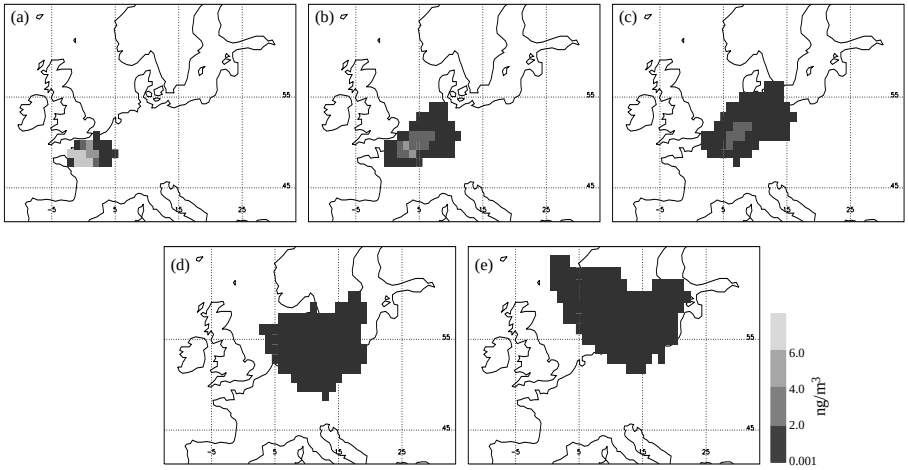

Figure 4: NAME tracer concentration interpolated onto $1^\circ \times 1^\circ$lat/lon grid 12, 24, 36, 48 and 60 hours after the start of the tracer release.

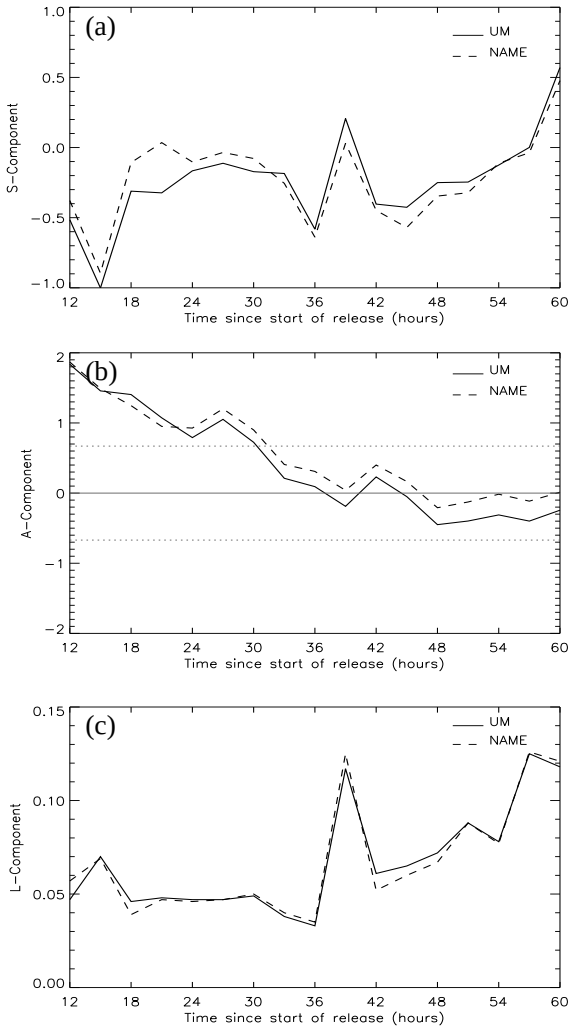

Figure 5: Time series of (a) structure component, (b) amplitude component and (c) location component for UM (solid) and NAME (dashed) simulations. The dotted lines in (b) are equivalent to an over/under-estimation by a factor of 2.

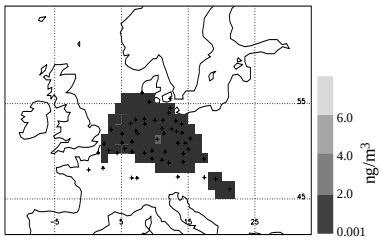

Figure 6: Kriged non-zero tracer concentrations 39 hours after the start of the tracer release. The non-zero observations are superimposed as crosses.

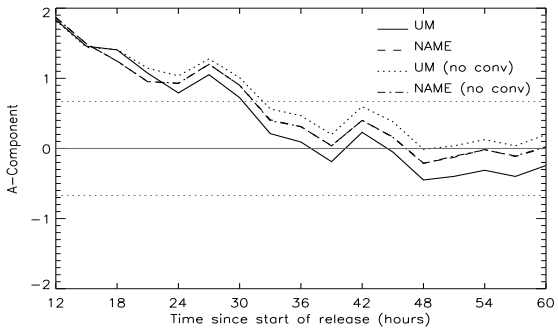

Figure 7: Time series of amplitude component for UM (solid), UM with no convection (dotted), NAME (dashed) and NAME with no convection (dash-dot) simulations. The dotted lines are equivalent to an over/under-estimation by a factor of 2.

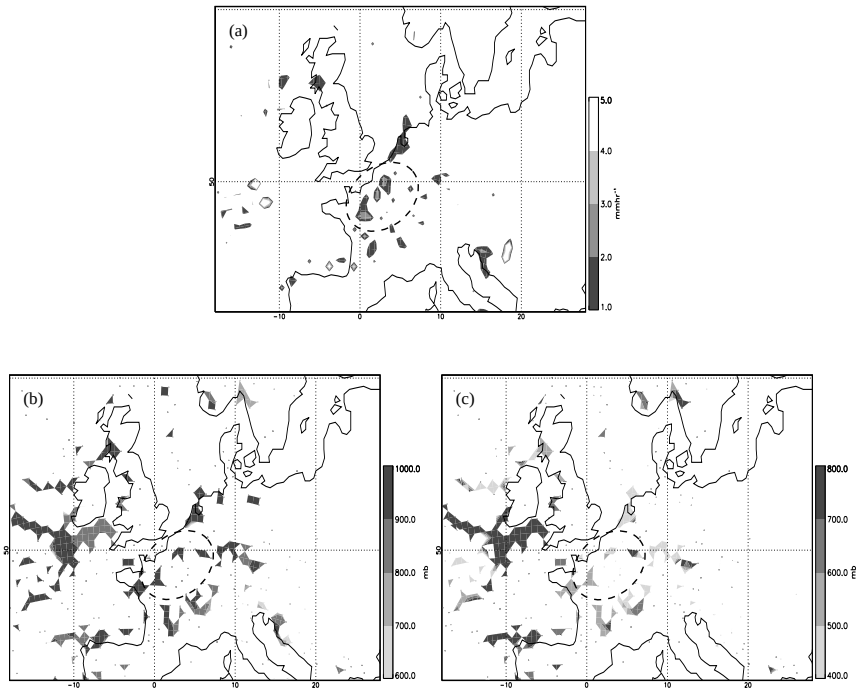

Figure 8: UM diagnosed output 20 hours after the start of the tracer release (a) convective rain rate, (b) pressure at cloud base and (c) pressure at cloud top.

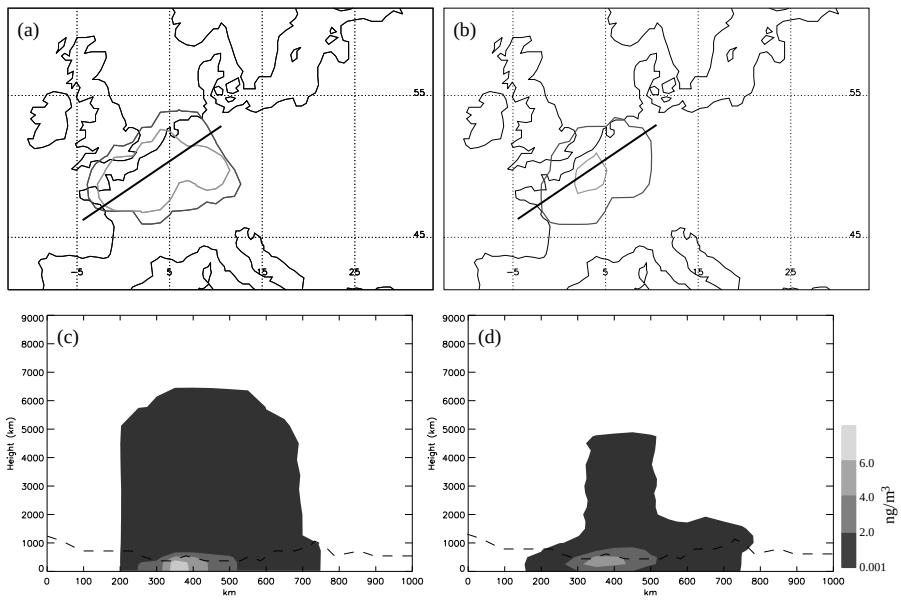

Figure 9: 3-hourly averaged tracer concentration 21 hours after the start of the tracer release for (a) 0.001ng m^{-3} contour for UM at heights of 20m (black) and 3920m (gray), and (b) 0.001ng m^{-3} contour for NAME at heights of 20m (black) and 4000m (gray). (c) UM vertical cross-section along the line shown in (a), (d) NAME vertical cross-section along the line shown in (b). Boundary layer height is overlaid (dashed)

- Step 2 performs the data analysis on the raw data with all zero observations removed. There are 939 non-zero data points in total with tracer concentrations ranging from 0.01ng m^{-3} to 12.57ng m^{-3}, and with a mean of 0.31ng m^{-3}.

- Step 3 creates a climatological variogram (figure 10(a)). A climatological variogram is used to overcome the problem of lack of data for a particular 3-hour time period. Climatological variograms allow information from the entire time period (90 hours) to be combined by normalising each event with respect to its variance (Lebel et al. (1987)). After the Kriging process has been completed, the tracer estimates are denormalised by multiplying the results by the variance for the time concerned. The bin size and maximum distance over which to model the variogram were chosen from the climatological variogram. A bin size of 15km was chosen as the variogram wasn't too noisy but the detail was captured. A range of 600km was chosen as the sill was flat at this range but there were still sufficient stations with a maximum distance of this range (> 250).

- Step 4 models the climatological variogram. A spherical model was chosen to model the climatological variogram (figure 10(b)), it has a range of 129.58km, a sill of 1.05 and a zero nugget. These parameters are used to perform the Kriging in step 5. They do not change significantly if a sub-sample of the observational data (between 12 and 60 hours only) is used indicating that no bias is introduced by normalising the data in step 3. Steps 3 and 4 were also performed for the indicator dataset. This is a binary dataset that contains 1's for observed tracer and 0's for no observed tracer.

- Step 5 performs Kriging over the entire grid for both the non-zero dataset and the indicator dataset. It also re-multiplies the final values by the time-step variance recorded in step 3.

- Step 6 performs double Kriging. One problem with Kriging data is that zero values can be smoothed out. Double Kriging involves, for each time step, finding the proportion of observation sites that recorded non-zero tracer. Then a threshold value from the indicator Kriged data is determined, which gives the same proportion of non-zero gridcells. A tracer/no tracer map is created using this threshold. Finally, the tracer amounts for the non-zero gridcells are filled in using the non-zero Kriged dataset.

- Step 7 creates maps of both the Kriging tracer estimates and the Kriging errors.

At present, it is not possible to estimate the errors associated with the double Kriged concentrations shown in figure 2. However, as almost all of the non-zero gridpoints in figure 2 are within the variogram range (129.58 km) of a non-zero observation the error on the Kriged estimate is likely to be small. A denser network of observations would help to reduce this error.

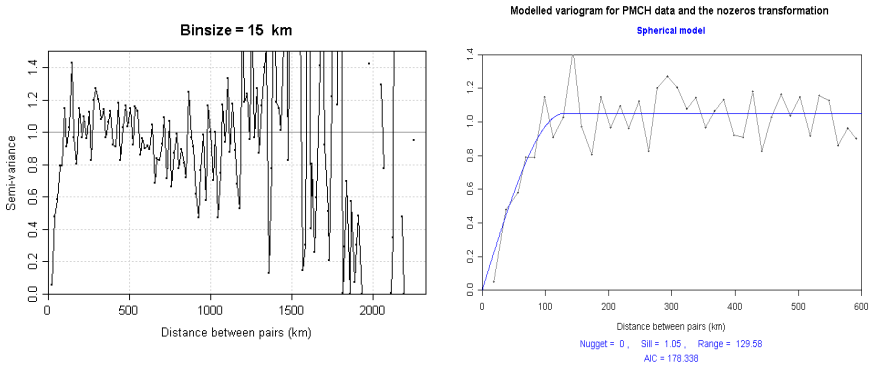

Figure 10: (a) Non-zero data climatological variogram, (b) Non-zero data modelled variogram.

References

Chatwin, P. C., 1982. The use of statistics in describing and predicting the effects of dispersing gas clouds, J. Hazard. Mater. 6, 213-230.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G. and Mason, S., 2008. Forecast verification: Current status and future directions, Met. Appl. 15, 3-18.

Cullen, M., 1993. The unified forecast/climate model, Meteorol. Mag., 122, 81-94.

Chagnon, J. M., and Gray, S. L., 2007. Stratosphere-troposphere transport in a numerical simulation of midlatitude convection, J. Geophys. Res. 112, D06314, doi:10.1029/2006JD007265

Dacre, H. F., 2010. Evaluating the ability of a numerical weather prediction model to forecast tracer concentrations during ETEX 2, Atmos. Env. 44, 294-303.

Dacre, H. F., Gray, S. L. and Belcher, S. E., 2007. A case study of boundary layer ventilation by convection and coastal processes, J. Geophys. Res. 112, D17106, doi:10.1029/2006JD007984.

D'Amours, R., 1998. Modeling the ETEX plume dispersion with the Canadian emergency response model, Atmos. Env. 32, 4335-4341.

Donnell, E. A., Fish, D. J. and Dicks, E. M., 2001. Mechanisms for pollutant transport between the boundary layer and the free troposphere, J. Geophys. Res. 106, 7847-7856.

Gray, S. L., 2003. A case study of stratospheric to troposphere transport: The role of convective transport and the sensitivity to model resolution, J. Geophys. Res. 108, 45990, doi:10.1029/2002JD003317.

Greatrex, H., 2010. The application of seasonal rainfall forecasts and satellite rainfall observations to crop yield forecasting for Africa. PhD thesis, University of Reading.

Gregory, D. and Rowntree, P. R., 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure, Mon. Weather Rev. 118, 1483-1506.

Gryning, S. -E., Batchvarova, E., Schneiter, D., Bessemoulin, P. and Berger, H., 1998. Meteorological conditions at the release site during the two tracer experiments, Atmos. Env. 32, 4123-4137.

Jones, A., Thomson, D., Hort, M. and Devenish, B., 2007. The UK Met Office's next-generation atmospheric dispersion model, NAME III. Air pollution modeling and its application XVII. Edited by C. Borrego and A-L. Norman.

Lebel, T., Bastin, G., Obled, C., and Creutin, J. D., 1987. On the accuracy of areal rainfall estimation: A case study, Water Resour. Res. 23, 2123-2134, doi:10.1029/WR023i011p02123.

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M. and Smith, R. N. B., 2000. A new boundary layer mixing scheme. Part 1. Scheme description and single-column model tests, Mon. Weather Rev. 128, 187-199.

Maryon, R. H., Ryall, D. B. and Malcolm, A. L., 1999. The NAME 4 dispersion model: science documentation. Met Office turbulence and diffusion note no.262.

Nasstrom, J. S. and Pace, J. C., 1998. Evaluation of the effect of meteorological data resolution on Lagrangian particle dispersion simulations using the ETEX experiment, Atmos. Env. 32, 4187-4194.

Mosca, S., Graziani, G., Klug, W., Bellasio, R. and Bianconi, R., 1998. A statistical methodology for the evaluation of long-range dispersion models: An application to the ETEX exercise, Atmos. Env. 32, 4307-4324.

Roberts, N. M. and Lean, H. W., 2007. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Mon. Wea. Rev. 136, 78-97.

Ryall, D. B. and Maryon, R. H., 1998. Validation of the UK Met. Office's NAME model against the ETEX dataset, Atmos. Env. 32, 4265-4276 .

Stohl, A., Hittenberger, M. and Wotawa, G., 1998. Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, Atmos. Env. 32, 4245-4264.

Wilson, D. R. and Ballard, S. P., 1999. A microphsically based precipitation scheme for the UK Meteorological Office unified model, Q. J. R. Meteorol. Soc. 125, 1607-1636.

Van Dop, H., Addis, R., Fraser, G., Girardi, F., Graziani, G., Inoue, Y., Kelly, N., Klug, W., Kulmala, A., Nodop, K. and Pretel, J., 1998. ETEX: A European tracer experiment; observations, dispersion modelling and emergency response, Atmos. Env. 32, 4089-4094.

Weil, J. C., Sykes, R. I., Venkatram, A., 1992. Evaluating air quality models: Review and outlook, J. Appl. Meteor. 31, 1121-1145.

Wernli, H., Paulat, M., Hagen, M., Frei, C., 2008. SAL - a novel quality measure for the verification of quantitative precipitation forecasts, Mon. Wea. Rev. 136, 4470-4487.