# *Prediction of visibility and aerosol within the operational Met Office unified model. I: Model formulation and variational assimilation*

QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
*Q. J. R. Meteorol. Soc.* **134**: 1801–1816 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.318

RMetS
Royal Meteorological Society

# Prediction of visibility and aerosol within the operational Met Office Unified Model. I: Model formulation and variational assimilation

P. A. Clark, S. A. Harcourt, B. Macpherson*, C. T. Mathison, S. Cusack and M. Naylor
*Met Office, Exeter, UK*

**ABSTRACT:** The formulation and performance of the Met Office visibility analysis and prediction system are described. The visibility diagnostic within the limited-area Unified Model is a function of humidity and a prognostic aerosol content. The aerosol model includes advection, industrial and general urban sources, plus boundary-layer mixing and removal by rain. The assimilation is a 3-dimensional variational scheme in which the visibility observation operator is a very nonlinear function of humidity, aerosol and temperature. A quality control scheme for visibility data is included. Visibility observations can give rise to humidity increments of significant magnitude compared with the direct impact of humidity observations. We present the results of sensitivity studies which show the contribution of different components of the system to improved skill in visibility forecasts. Visibility assimilation is most important within the first 6–12 hours of the forecast and for visibilities below 1 km, while modelling of aerosol sources and advection is important for slightly higher visibilities (1–5 km) and is still significant at longer forecast times. © Crown Copyright 2008. Reproduced with the permission of HMSO. Published by John Wiley & Sons, Ltd.

KEY WORDS fog; humidity; 3D-Var

*Received 10 December 2007; Revised 6 August 2008; Accepted 8 August 2008*

## 1. Introduction

Fog is hazardous to road, rail, marine and air traffic, with numerous documented fog events leading to deaths, damage to property and delays. According to Rosenfeld (1996), the number of deaths in the USA in fog-related traffic accidents in the period 1982–1991 was more than twice that due to the combination of flash floods, hurricanes, lightning and tornadoes. During persistent widespread fog in December 2006, the British Airports Authority was forced to cancel nearly half the flights at Heathrow airport over the space of three days because air traffic controllers doubled the distance between aircraft lining up to land and take off. Improved prediction of low-visibility events is valuable in safety and economic terms, for example in reducing fuel and diversion costs to airlines (Leigh, 1995).

Techniques for fog prediction can be categorised as manual, statistical, nowcasting or numerical modelling (1-dimensional or 3-dimensional). Manual techniques typically involve interpretation of the latest atmospheric profile from a radiosonde or (more likely) a numerical weather prediction (NWP) model, together with surface observations for the forecasting location, followed by application of rules for estimating fog point and rate of cooling. Statistical techniques may involve regression or neural networks (Pasini *et al.*, 2001) and rely on a training period of data which connect forecast profiles, local observations and observed visibility. This period may need to be several years to a decade for dense fog because of its relative rarity. Nowcasting techniques like that of the Nimrod system (Golding, 1998; Wright and Thomas, 1998) use satellite imagery and surface observations, with a very-short-range forecast produced by persistence or by extrapolating a feature of the initial distribution.

Fog prediction presents a formidable challenge to NWP models due to its often small spatial and temporal scales. As an illustration, the fog event analysed by Pagowski *et al.* (2004), which led to an 82-car pile-up in Ontario, covered only a few kilometres of road and reduced visibility to just a few metres for a period of only minutes. Numerical models require fine vertical resolution near the ground to resolve the evolution of the near-surface layers of the atmosphere in stable conditions. They also need fine horizontal resolution to represent the local orography, soil, water and vegetation properties which influence fog formation. One-dimensional models like that of Clark and Hopwood (2001a) do have high vertical resolution but need larger-scale input from a 3-dimensional model to capture the role of horizontal advection. Alternatively, a model like PAFOG (Bott and Trautmann, 2002) is initialised with a radiosonde and suited only to cases with negligible advection.

Operational three-dimensional NWP models cannot yet match the very fine resolution required to represent the

*Correspondence to: B. Macpherson, Met Office, FitzRoy Road, Exeter EX1 3PB, UK.
E-mail: bruce.macpherson@metoffice.gov.uk

most local effects, and so perform best when low visibility is widespread. Apart from the Met Office Unified Model (MetUM; Davies *et al.*, 2005) discussed in this paper, a few others give direct output of visibility. The HIRLAM model as run at the Danish Meteorological Institute (DMI) incorporates a diagnostic calculation of visibility at 2 m height (Petersen and Nielsen, 2000). It is based on a pseudo cloud water estimated from model information about solar zenith angle, cloud cover and wind velocity, temperature and specific humidity (both at screen level and at the lowest model level). Rain and snow intensity are also used. The parametrization has been tuned by a statistical analysis of observations from 29 Danish stations over a 2-year period. In the Rapid Update Cycle (RUC) model run in the USA, visibility is calculated from prognostic hydrometeor, cloud and relative humidity fields, as described by Smirnova *et al.* (2000), with further evaluation of performance in Smith *et al.* (2002) and mention of a visibility assimilation technique in Benjamin *et al.* (2004). The RUC algorithm is adapted from Stoelinga and Warner (1999). The ETA model run at the US Weather Service also used the Stoelinga and Warner algorithm, and this was carried over into its successor the North American Mesoscale (NAM) Model.

Aerosol plays an important role in determining visibility. In HIRLAM-DMI, this is accounted for very simply by assuming everywhere a constant background concentration of aerosol, which is enhanced when the wind direction is likely to bring more polluted air towards Denmark. The RUC model neglects aerosol. The MetUM, however, carries aerosol as a prognostic variable and, amongst NWP models, this is a distinctive feature of its visibility prediction system. Other aerosol models exist for a variety of purposes in the fields of climate change, air quality and radiative transfer. Visibility is often one of the output parameters. Usually the aerosol model is driven by output fields from an atmospheric model, but the aerosol and atmosphere are uncoupled. An example is the NAAPS (Navy Aerosol Analysis and Prediction System) developed by the Naval Research Laboratory (NRL) in Monterey, California, which is driven by global meteorological fields from the Navy Operational Global Atmospheric Prediction System (NOGAPS). Another is the Northern Aerosol Regional Climate Model (NARCM) developed in Canada and applied by Munoz-Alpizar *et al.* (2003) to study visibility in the Mexico City area. Coupled (online) models also exist; for example, the Weather Research and Forecast (WRF) model has an online chemistry package (Grell, *et al.*, 2005), the MetUM itself has a sulphur cycle for climate use, and a more sophisticated chemistry and aerosols package, UKCA, is under development. However, such models are substantially more expesive to run than the pure meteorological forecast model.

Successful short-period forecasts of low visibility from NWP models are dependent on accurate initial conditions, and therefore on assimilation of all relevant observations. An early study which underlined this was by Ballard *et al.* (1991) who examined the sensitivity of coastal fog prediction to initial humidity and cloud water fields with a previous version of the Met Office Mesoscale Model. Macpherson *et al.* (1996) documented the benefit of assimilating screen-level relative humidity observations for fog prediction. In a later version of that system, visibility observations were introduced to update the prognostic aerosol field, with some additional benefit for visibility prediction on a 6–9-hour time-scale. Clark and Hopwood (2001b) incorporated local observations into the initial conditions for their one-dimensional model and found humidity data were the most important, though their influence on the forecast lasted no more than about 6 hours. Visibility observations have been used to validate the surface aerosol predictions of the NAAPS model (Lerner *et al.*, 2004) and there are plans to explore how they can be used to infer aerosol concentrations. This latter step is part of the novel variational assimilation of visibility observations described in this paper, which is believed to be the first scheme in which visibility data have a multivariate impact on both moisture and aerosol fields.

The paper is organised as follows. In section 2 we describe the general approach to visibility prediction and how we model the aerosol distribution and its role, along with humidity, in the diagnosis of visibility. In section 3 we give an account of the 3-dimensional variational assimilation algorithm and the treatment of observation and model errors. Results reported in section 4 are of several kinds: first a single-observation experiment, then a single analysis of many observations, followed by a batch of individual case-studies, and then a set of week-long sensitivity tests in which the relative contribution of different modelling features and assimilation is investigated. These are related to operational verification from a 3-month season. Finally in section 5 we conclude with a summary of future developments.

A companion paper, Haywood *et al.* (2008), henceforth referred to as Part II, reports aerosol data from flights of a research aircraft around the UK and special surface observations which have been analysed to validate aspects of the aerosol modelling system and reference will be made to these results where relevant, though the scheme reported here represents the situation prior to any improvements that might arise from this validation.

## 2. The visibility forecast model

### 2.1. Overall approach

Visibility is a complex issue, and at this stage it would be very difficult to capture all of its characteristics in a simple numerical model. Visibility is a non-local parameter, in that it represents the shortest horizontal distance visible considering all directions, so that a fog bank observed 500 m away from an observing site strictly means that the visibility at the site is 500 m, even if the site itself is clear. However, we have made the simplifying assumption that visibility can be related to the local scattering characteristics of the atmosphere. This is dominated, in general, by scattering from aerosol particles and, where present, cloud drops and other hydrometeors.




DOI: 10.1002/qj

Predicting aerosol content purely from sources and sinks would require the representation of complex atmospheric chemistry, using a substantial number of chemical species together with (probably highly uncertain) source and sink terms. While this is an approach which is becoming feasible, it adds very substantially to the cost of the model and may not be justified on the basis of improved visibility forecasts alone. Instead, the rather simple view is taken that, outside areas of precipitation, visibility is largely a function first of relative humidity and second of the characteristics of the air mass, which we shall describe as the aerosol content. 'Clean' air results in better visibility at a given relative humidity than 'dirty' air. The air-mass aerosol content, in the absence of sources and sinks, is assumed to be a conserved quantity on the timescale of a model forecast (1–2 days). A single prognostic aerosol variable is used to characterise the air mass.

A measure of aerosol concentration in the initial atmospheric state is derived using assimilation of surface (synoptic) measurements, in particular of relative humidity (via dry- and wet-bulb temperatures) and visibility. This presents a challenge, in particular since the relationship between aerosol concentration, relative humidity and visibility is very nonlinear. However, it also presents advantages for two reasons. First, the aerosol is derived from the inverse of the relationship used to predict visibility, which leads to considerable reduction in sensitivity to various model assumptions. Second, visibility observations represent a valuable source of information about humidity where the relative humidity is high. The presence of mist or fog can be a better indicator that air is close to saturation than a wet bulb temperature measurement.

At longer forecast times, aerosol sources and sinks must be accounted for, especially as the initial air may leave the domain of a limited-area model. This is extremely difficult using a single prognostic variable, and very simple approaches have been tried. In practice a surrogate source based upon air-pollution sources has been implemented. Once source terms are applied, similarly sinks are needed to ensure aerosol does not build up unrealistically. This has been confined to wet deposition by precipitation.

Visibility is further reduced by the precipitation, when present, so in principle the intensity and nature of precipitation is important. The current model includes parametrization of the reduction in visibility by convective and large-scale rain and snow. This uses the same underlying assumptions regarding hydrometeor size spectra as the related microphysical parametrizations to derive scattering coefficients and hence visibility in a very straightforward way. However, this part of the model is very distinct from the task of predicting visibility in clear air. Furthermore, validating the prediction of visibility in precipitation is dominated by the accuracy with which the model predicts precipitation, rather than the accuracy of the visibility predicted given the precipitation. This part of the model is therefore not considered further in this paper. In practice, observations made during precipitation are not used in the assimilation.

Sub-grid variability is a feature of any variable in any NWP system, but is of particular interest where visibility is concerned. Predicting the sub-grid variation of visibility may be difficult, but recognizing its existence is regarded as essential. Thus, we have taken an essentially stochastic approach, assuming that there is an underlying distribution of variables leading to our visibility diagnosis, and so a distribution of visibilities that may be observed at any location.

### 2.2. Treatment of aerosol

The scheme was first implemented operationally in the MetUM in 1995, with minor differences in formulation from that described here. The Analysis Correction assimilation scheme, operational at the time, used visibility observations to correct aerosol concentrations only (after standard assimilation of other data including moisture-related variables had been applied). Before data assimilation was introduced, the scheme was run continuously for a period of 3 months using only the source, sink and boundary terms described below within the 3-hourly cycled, continuous assimilation system. Comparison was made between predicted and observed visibility at synoptic stations using the predicted aerosol at analysis time and a rescaling applied (amounting to a change in $r_0$ and $N_0$, see below) to remove overall bias in the system. This also confirmed that, as expected, errors in $\log_{10}$(visibility) are much more normally distributed than errors in visibility. This reflects, in part, the quasi-logarithmic SYNOP code for visibility which allows more precision at lower visibility.

The scheme was then implemented in the NIMROD nowcasting system, where visibility observations were also used to modify surface humidity (Wright and Thomas, 1998) and further minor changes have been made since to take account of other physics changes (Radcliffe *et al.*, 1998).

We assume a spectrum of aerosol particles characterized by their dry radius $r_\mathrm{d}$ and a number density $n(r_\mathrm{d})$. The dry aerosol mass mixing ratio $m$ is given by:

$$\begin{aligned} m &= \int \frac{4}{3}\pi \rho r_\mathrm{d}^3 n(r_\mathrm{d})\,\mathrm{d}r_\mathrm{d}/\rho_a \\ &= \frac{4}{3}\pi\frac{\rho}{\rho_a} N\overline{r_\mathrm{d}^3} \\ &= \frac{4}{3}\pi\frac{\rho}{\rho_a} N r_\mathrm{md}^3, \end{aligned} \tag{1}$$

where $r_\mathrm{md}$ is the (dry) mean volume radius, $\rho$ is the density of the aerosol (taken to be $1700\,\mathrm{kg\,m^{-3}}$, roughly that of ammonium sulphate) and $\rho_\mathrm{a}$ the density of air. To recover $N$ and $r_\mathrm{md}$ from $m$ requires an assumption to be made about the aerosol particle size distribution; we make the simple assumption that $r_\mathrm{md}$ varies as a power of the aerosol mass concentration, $m$, i.e.

$$r_\mathrm{md} = r_0 \left(\frac{m}{m_o}\right)^p, \tag{2}$$

where $r_0$ is the radius of a 'standard' aerosol particle, $p$ is the power used to represent the variation in aerosol particle size with mixing ratio, and $m_0$, the standard mass mixing ratio of the aerosol, is given by Equation (1) with $N = N_0$, where $N_0$ is the standard number density of the aerosol.

We have taken $r_0 = 0.16 \times 10^{-6}$ m and $N_0 = 5.0 \times 10^8$ m$^{-3}$ based on the initial calibration described above. Where aerosol is derived from visibility assimilation, these choices are not crucial, in the sense that other parameters can be chosen to compensate for any changes, and remaining errors will cancel in the assimilation process.

The density of air is taken to be $1\,\mathrm{kg\,m^{-3}}$ for the purpose of defining $m_0$. Thus, the aerosol number density, $N$, is given by

$$N = N_0 \left(\frac{m}{m_o}\right)^{1-3p}. \tag{3}$$

If $p = 0$, the number density is proportional to mixing ratio, so the size spectrum of particles is independent of mixing ratio. On the other hand, if $p = 1/3$ the number density is fixed. We have chosen an intermediate value $p = 1/6$. Thus, as the mixing ratio of aerosol increases, the mean size of particles gradually increases as well as the number density. This is an arbitrary, though plausible, assumption that is supported by measurements reported in Part II of this paper, which show $p = 1/x$ with $x$ about 6 or 7 for different cases. The assumption appears to have little impact on results, and is included primarily so that sensitivity can be assessed.

The MetUM positive definite semi-Lagrangian tracer advection scheme is used to advect the aerosol. Furthermore, the MetUM boundary-layer vertical mixing scheme is used to mix it in precisely the same way as other scalar quantities (currently with a zero deposition flux surface boundary condition). The MetUM 'boundary-layer' scheme is only applied to the lowest 13 levels of the model (about 3.3 km). In practice, little is known about concentrations above the boundary layer, so advection has been restricted to the boundary layer, with fixed 'climatological' fields above. These have very little impact on surface values. The impact of vertical mixing of aerosol is very evident on the diurnal cycle of concentrations as enhanced mixing during the day tends to reduce surface concentrations (as well as reducing humidity). Similar effects are also evident around coasts.

Given a typical lifetime of aerosols of several days, it is anticipated that most information on aerosol distribution will be obtained from the data assimilation scheme. For a reasonably long forecast, some account of sources must be taken, as there is the likelihood of analysed distributions advecting out of the model. We have therefore adopted a very simple approach. The source terms represent primarily sulphur dioxide sources, using point and area data over the UK from Warren Spring Laboratory. Outside the UK, the 150 km inventory generated by the Co-operative Programme for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe (EMEP) has been used. The source data date from the initial implementation of the scheme (1993) and have not been updated. This is obviously a deficiency of the scheme. However, in practice the source terms primarily identify the major industrial areas of Europe.

Much atmospheric aerosol is not emitted, as such, but results from atmospheric chemistry. To represent this, a conversion factor is applied to the source terms representing the average conversion likely over the model domain. This factor is based on a simple estimate of conversion. If we consider transport over an average time of $t_{\max}$, then the average ratio of the aerosol concentration to the emission is

$$\begin{aligned}\overline{s(t)/c_0} &= \frac{1}{t_{\max}} \int_0^{t_{\max}} \frac{s(t)}{c_0}\,\mathrm{d}t \\ &= \frac{k}{k + \frac{v_\mathrm{d}}{h}} \left(1 - \frac{1 - \mathrm{e}^{-\left(k + \frac{v_\mathrm{d}}{h}\right) t_{\max}}}{\left(k + \frac{v_\mathrm{d}}{h}\right) t_{\max}}\right), \end{aligned} \tag{4}$$

where $s(t)$ is the amount of aerosol produced from a pollutant with initial concentration $c_0$, $k$ is a representative oxidation rate, $v_\mathrm{d}$ a representative dry deposition velocity and $h$ a typical boundary-layer depth. The implication of using this constant conversion factor is that we expect to overestimate aerosol concentration close to sources and underestimate some distance away (beyond $t_{\max}$). Over 18 hours for 'typical' parameter values associated with $SO_2$, this ratio has a value of order 0.1. A simple sinusoidal diurnal variation has been allowed in which the emissions peak during the day. An additional (small) surface term has been added everywhere to allow for natural sources. It should be emphasised that Equation (4) provides some justification for the conversion factor used but that this was also justified on the basis of the 3-month run without assimilation discussed above and subsequent verification which does not show strong biases developing with forecast time. The precise scaling of the source terms remains uncertain; their role is primarily to provide some discrimination between maritime and continental air.

Boundary conditions for aerosols are also highly uncertain. There is not the facility to run aerosols in the operational global model just to provide them, so we have to provide 'representative' boundary fluxes or concentrations. Zero or constant boundary concentrations are an option but, for specific UK use, an attempt has been made to recognise that Europe is a much more important source of aerosols than the Atlantic. The distribution of sources from the EMEP sulphur inventory outside the model was used to find a total 'upwind' source strength outside the model area by summing along lines radiating from the model domain centre. It was found that this angular distribution could be reproduced quite accurately by summing two Gaussian distributions, with centres a characteristic distance from the model area. At each boundary point the wind speed and direction is calculated, and this is assumed to apply to transport outside the area. This and

the source distribution and travel time are used to compute a boundary concentration according to

$$\frac{S}{uh}\frac{k}{(k+v_{\mathrm{d}}/h)}\left\{1-\exp\left(-\frac{kx}{u}\right)\right\}, \tag{5}$$

where $S$ is the source strength, $x$ the travel distance and $u$ the wind speed. The significance of this equation should not be overstated; its main purpose is to ensure an inverse windspeed relationship in the boundary concentration, somewhat modified by the exponential term. A small, and arbitrary, background was added to this to represent natural and very distant sources. This boundary-layer concentration is then assumed to decay exponentially in the vertical with pressure. Although very rough indeed, this mechanism at least allows for the influx of moderately realistic concentrations from Europe, particularly under light wind conditions. It should be borne in mind that the intention is to supply a very approximate influx which can be improved using the assimilation system. No attempt has been made to include diurnal or seasonal factors in either this relationship or the source strength scaling factor above.

Removal by precipitation is included using a simple linear scavenging proportional to precipitation rate. In the large-scale precipitation scheme, the precipitation rate at a level is used to scavenge from that level. Different rates may be used for rain and snow, but in practice we have used the same. In convection, the surface precipitation rate is used to scavenge all levels from the surface to the top of the boundary layer. The scavenging takes the form

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -k_{\mathrm{s}}(R)m\,, \tag{6}$$

where $k_{\mathrm{s}}$ is the scavenging rate, a function of precipitation rate $R$. For simplicity, we have used a form $k_{\mathrm{s}}(R) = 10^{-4}R$, where $R$ is the precipitation rate in $\mathrm{mm\,h^{-1}}$ and $k_{\mathrm{s}}(R)$ is measured in $\mathrm{s^{-1}}$. This is solved using a simple first-order fully implicit Euler scheme which has the advantage of being computationally cheap and applicable with arbitrary long time step.

### 2.3. The diagnosis of visibility

Visibility is diagnosed assuming a simple exponential scattering law, and a visual range defined by a fixed liminal contrast $\varepsilon$, given by 0.02 (Koschmeider, 1924). Other authors have used a liminal contrast of 0.05. A value of 0.02 was adopted for consistency with previous work in the Met Office forecast system (Ballard *et al.*, 1992). Thus, given a scattering coefficient $\beta$:

$$vis = -\frac{\ln\varepsilon}{\beta_{\mathrm{tot}}} = \frac{3.912023}{\beta_{\mathrm{tot}}}\,. \tag{7}$$

Here,

$$\beta_{\mathrm{tot}} = \beta_{\mathrm{air}} + \beta(RH, m)\,, \tag{8}$$

where the first r.h.s. term corresponds to the extinction coefficient of clean air, the second is due to aerosol, and $RH$ is the relative humidity. The extinction coefficient of clean air generally has no practical impact, and has been take to be equivalent to a visibility of 100 km to ensure that unrealistically high visibilities are never diagnosed.

With respect to visibility, the atmosphere can be divided into two distinct regimes, depending on whether or not the fog droplets are activated. The $RH$ used as input to derive the scattering coefficient in Equation (8) is generalized to include the condensed water, thus:

$$RH = \frac{q_{\mathrm{tot}}}{q_{\mathrm{sat}}}, \tag{9}$$

where $q_{\mathrm{tot}}$ is the total water and $q_{\mathrm{sat}}$ is the saturation specific humidity. In either regime we are fundamentally interested in the droplet radius, $r$, which determines the scattering.

To simplify analysis, the aerosol is assumed to be mono-disperse with uniform chemical constituents. Unactivated droplets are essentially aerosol particles with a small amount of liquid water, and their mean droplet radius, $r_{\mathrm{m}}$, can be related to the relative humidity, $RH$, by the equilibrium equation expressed in the familiar Köhler curve (e.g. Pruppacher and Klett, 1978):

$$RH = \exp\left\{\frac{A}{r_{\mathrm{m}}} - \frac{B}{(r_{\mathrm{m}}/r_{\mathrm{md}})^3 - 1}\right\}, \tag{10}$$

where $A$ is a constant related to the surface tension of water ($A = 1.2\times10^{-9}\,\mathrm{m}$) and $B$ is the activation parameter (taken here to be 0.5, roughly appropriate for ammonium sulphate aerosol). This curve has a peak at the activation radius. Where drops have grown to the activation radius, further growth can occur with reduced $RH$, so droplets are said to be activated.

To a very good approximation (given $r_{\mathrm{m}}$ is much greater than $r_{\mathrm{md}}$ close to saturation) the activation radius is given by

$$r_{\mathrm{act}} = \sqrt{\frac{3Br_{\mathrm{md}}^3}{A}}\,. \tag{11}$$

This can be substituted into Equation (10) to obtain the activation humidity, $RH_{\mathrm{act}}$.

When no activated particles are present, an analytic solution for $r_{\mathrm{m}}$ can be obtained if the term involving $A$ is ignored (which is a very good approximation away from activation):

$$r_{\mathrm{m}} = r_{\mathrm{md}}\left\{1 - \frac{B}{\ln(RH)}\right\}^{\frac{1}{3}}. \tag{12}$$

Close to activation, it is necessary to solve the full form of Equation (10), which accounts for the presence of unactivated droplets in supersaturated air. This is achieved as follows.

Activated fog droplets grow rapidly, and the cloud water content, $q_{\mathrm{L}}$, can be related to the volume mean droplet radius, $r_{\mathrm{m}}$:

$$q_{\mathrm{L}} = \frac{4}{3}\pi(r_{\mathrm{m}}^3 - r_{\mathrm{md}}^3)\rho_{\mathrm{w}}N\,, \tag{13}$$




DOI: 10.1002/qj

where $\rho_w$ is the density of water ($1000\,\mathrm{kg\,m^{-3}}$). This can be inverted to provide the droplet radius, $r_m$, but first the amount of liquid water must be diagnosed. Simply assuming all water above saturation is condensed is fairly accurate but leads to a discontinuous relationship when combined with the unactivated regime. A diagnosis which is continuous across regimes is needed, especially for data assimilation.

The droplet radius, $r_m$, is obtained by solving the equation

$$q_{tot} = RH\{\min(r_m, r_{act}), r_{md}\}\, q_s(T) + q_L(r_m)\,, \qquad (14)$$

where $RH(r_m, r_{md})$ is given by Equation (10) and $q_L(r_m)$ is given by Equation (13). The use of $\min(r_m, r_{act})$ ensures a single-valued function with negligible loss of accuracy. This is solved using Newton–Raphson iteration.

It is worth noting that, in assuming a single representative aerosol particle at this stage, we do not allow for partial activation. In practice, the smaller particles would not activate leading to somewhat higher visibilities in fog. We therefore anticipate a tendency to overforecast poor visibilities.

We now need to diagnose the visibility from the calculated droplet radius. The extinction coefficient due to particles, $\beta$, is given by:

$$\beta(RH, m) = \int \pi\, Q(r)\, r^2(RH, r_d)\, n(r_d)\, \mathrm{d}r_d, \qquad (15)$$

where $Q$ is the extinction efficiency, which may be derived from Mie theory. In practice we do not know the size distribution or how it varies with aerosol mass mixing ratio (apart from the assumption already made regarding mean volume radius in Equation (2)). Furthermore, we have not attempted to model the variation of particle growth with $RH$ as a function of size. We have made the simple assumption that $Q$ is independent of particle size, or equivalently regard $Q$ as the spectrally averaged value. If we also assume that $B$ is independent of particle size then we may simplify this to

$$\beta(RH, m) = \pi\, Q N \overline{r^2}\,. \qquad (16)$$

The result may be expressed in terms of the mean volume radius:

$$\beta(RH, m) = \pi\, Q \left(\frac{\overline{r^2}}{r_m^2}\right) N r_m^2 = \beta_0 N r_m^2\,, \qquad (17)$$

where

$$\beta_0 = \pi\, Q \eta \qquad (18)$$

and

$$\eta = \overline{r^2}/r_m^2\,. \qquad (19)$$

We have written the result in these terms to remind us that the parametrization depends upon average scattering properties ($Q$) and that these properties are weighted towards smaller particles ($\eta < 1$, given that aerosol size spectra are generally skewed towards small particles). However, the overall result is that scattering is proportional to $m^{1-p}$. For the constant of proportionality, we have assumed a value of $\eta$ of 0.75 and a value of $Q$ of 2.0 (essentially geometric scattering), so $\beta_0 = 1.5\pi$. Results reported in Part II suggest that this is probably too high, the amount depending on the nature of the air mass (between a factor of 1.22 for a heavily polluted air mass to 2 for a cleaner one), though, in a sense, the assumption can be seen as conservative, in that a given error in visibility thus results in a smaller increment to aerosol.

As a final statement of derivation of visibility outside precipitation, we can use Equation (17) in Equations (7) and (8) to give

$$vis = \frac{-\ln \epsilon}{N r_m^2 \beta_0 + \beta_{air}}. \qquad (20)$$

The complexities of size spectra and scattering are all hidden in the constant $\beta_0$; to keep within the philosophy of a single prognostic aerosol variable, improvements to this require more sophisticated parametrization of $\beta_0$ in terms of $m$.

### 2.4. Sub-grid variability and probabilistic diagnostics

The fundamental relationship Equation (14) determines droplet sizes and hence point visibility as a function of input $RH$. The $RH$ may vary significantly over a large grid box, leading to sub-grid variability in visibility. So may the aerosol concentration, but in most cases it is likely that the sub-grid variability of $RH$ is most important because of land use and orography. In the MetUM, we assume a sub-grid distribution of $RH$ based on the model cloud scheme (Smith, 1990), which assumes a triangular distribution of humidity. Thus, in principle, any point in the humidity distribution can be used as input to Equation (14). This can be inverted and the $RH$ corresponding to a given visibility derived – in practice the probabilities of visibility being less than given thresholds (200 m, 1 km and 5 km) are output routinely. Choice of the grid-box mean $RH$ corresponds to a median (*not mean*) visibility and, because of the inherent skewness of the visibility distribution, it has sometimes been found more useful to use a lower percentile (e.g. 40%) as a representative grid-box visibility. The Smith scheme has been used for consistency with the rest of the model, but, in principle, it may be possible to couple the sub-grid variability to known variables (based on land use and orography) to produce a more realistic variability. This has yet to be attempted in the MetUM, though the NIMROD scheme (Wright and Thomas, 1998) effectively does this through a downscaling algorithm.




DOI: 10.1002/qj

## 3. Assimilation of visibility

### 3.1. Development of the 3D-Var algorithm

The predecessor to the Met Office operational 3D-Var analysis was the Analysis Correction scheme (Lorenc *et al.*, 1991) in which variables were analysed in a sequential, univariate fashion. Impact of surface observations on visibility forecasts came through assimilation of screen-level *RH* data, as reported by Macpherson *et al.* (1996). A univariate visibility analysis step was added operationally in 1995, in which $\log_{10}$(aerosol) increments were proportional to errors in $\log_{10}$(visibility). This gave a further improvement in short-period visibility forecasts, with benefit from the observations lasting for 6–9 hours.

When 3D-Var was developed for the MetUM (Lorenc *et al.*, 2000), the potential was clear for an improved visibility analysis, since the variational analysis framework is ideally suited to the multivariate dependence of visibility on several model prognostic variables.

The problem of variational analysis is to find the model state $\mathbf{x}$ which minimises a penalty, $J$, made up from a background term $J^{\mathrm{b}}$ and an observational term $J^{\mathrm{o}}$:

$$J(x) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}}) + \frac{1}{2}(\mathbf{y}-\mathbf{y}^{\mathrm{o}})^{\mathrm{T}}(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{y}-\mathbf{y}^{\mathrm{o}}), \tag{21}$$

where $\mathbf{x}^{\mathrm{b}}$ is a prior (background) estimate of $\mathbf{x}$, with error covariance $\mathbf{B}$, $\mathbf{y}^{o}$ is a vector of observed values, with instrumental error covariance $\mathbf{E}$, and $\mathbf{y}$ is a prediction of the observed values:

$$\mathbf{y} = H(\mathbf{x}). \tag{22}$$

$\mathbf{F}$ is the error covariance of the 'generalised interpolation' $H$, also known as the 'observation operator', which may involve complex and nonlinear equations expressing the dependence of the observed parameter on the prognostic variables of the model.

Aside from its visibility analysis, the regional 3D-Var system is very similar to the global version described in Lorenc *et al.* (2000). One obvious difference is the presence of boundaries, which are treated by imposing the constraint that increments are zero on the boundary for scalar fields and the normal component of the wind. The tool for this purpose is the double-sine transform. For fields other than aerosol, forecast error covariances are based on several months of $(t+12)-(t+6)$ forecast differences for deriving vertical modes. We use an Incremental Analysis Update (IAU) initialisation scheme (Bloom *et al.*, 1996), in which the analysis increments are added over $\mathcal{N}$ time steps, $1/\mathcal{N}$th at a time. The IAU period was chosen to be two hours, centred on the nominal analysis time. Alongside the IAU, cloud and rainfall data are simultaneously assimilated by a nudging technique described in Macpherson *et al.* (1996) and Jones and Macpherson (1997).

Full details of the pre-operational trials and subsequent upgrades to the UK mesoscale 3D-Var system can be found in Macpherson *et al.* (2002) and Renshaw *et al.* (2004). The 3D-Var scheme has been run in several operational models: from October 1999 to October 2006 in a 12 km grid-length model covering the UK and near continent, from August 2003 in a larger 12 km North Atlantic and European (NAE) domain and since December 2005 in a 4 km resolution UK model, nested within the NAE. All these models have 38 levels in the vertical, with 13 in the boundary layer and the lowest theta level at 20 m.

### 3.2. Observation operator

In the case of visibility, the observation operator is highly nonlinear in the moisture variable, as demonstrated in Figure 1, showing the sudden drop in visibility with approach to saturation.

Within the descent algorithm to minimise the penalty function, the adjoint of the tangent linear version of the observation operator is required to compute the gradient of the penalty with respect to changes in the control variables. The tangent linear approximation to the visibility equation (20) can be written:

$$\log_{10}(vis) = \log_{10}(vis)_{\mathrm{g}} + q' \left.\frac{\partial \log_{10}(vis)}{\partial q}\right|_{\mathrm{g}} + \log_{10}(m)' \left.\frac{\partial \log_{10}(vis)}{\partial \log_{10} m}\right|_{\mathrm{g}} + T' \left.\frac{\partial \log_{10}(vis)}{\partial T}\right|_{\mathrm{g}}, \tag{23}$$

where the model perturbation variables within 3D-Var include specific humidity $q'$, temperature $T'$ and the log

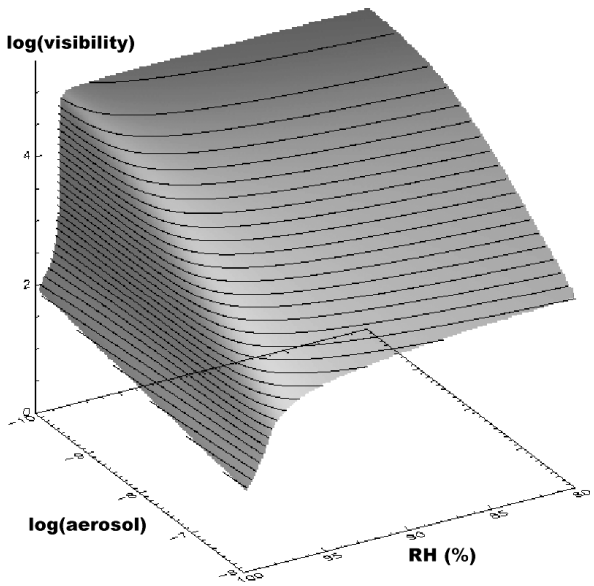


Figure 1. Graph showing the $\log_{10}$(visibility) surface as a function of relative humidity and $\log_{10}$(aerosol), where aerosol is defined as the mass mixing ratio in kg kg$^{-1}$. Temperature is held constant.

of the aerosol mass mixing ratio $\log_{10}(m)'$. The subscript g denotes a guess or background value. One may also use the tangent linear model to update the observed estimates **y** during the minimisation.

Tests were performed to confirm that the linearisation was correctly derived and coded. Taking Equation (23) and neglecting the humidity and temperature terms, we are left with an equation which can be rearranged to

$$\frac{\log_{10}(vis) - \log_{10}(vis)_{\mathrm{g}}}{\log_{10}(m)'} = \left(\frac{\partial \log_{10}(vis)}{\partial \log_{10} m}\right)_{\mathrm{g}}, \qquad (24)$$

where the left-hand side is a finite-difference estimate of the gradient and the right-hand side is the gradient as calculated from the linearisation. A correct linearisation implies that the ratio of these two expressions should tend to unity for small enough perturbations. This was tested over a range of $\log_{10}(m)'$ values and the results are shown in Figure 2.

At very low increment values, machine precision errors are present. The ratio remains close to 1.0 for a wide range of $\log_{10}(m)'$ values before the linearisation breaks down. Similar graphs confirmed that the humidity and temperature terms were also correct.

### 3.3. Observation errors

Visibility is typically measured by either forward-scattering devices or a human observer. The drive towards automation of observing networks in recent years means that the number of reports by human observers has declined appreciably. In the UK, the current ratio of automatic to manual reports can vary from about 50:50 to 70:30 depending on time of day and day of the week. The automatic measurement is based on an atmospheric sampling volume of order 0.1–10 litres, depending on the precise technology used. A human observer, on the other hand, will record the minimum visibility obtained by looking in various directions from the observing site for suitable markers which can be recognized against the horizon sky. This measurement is representative of a much wider area. Of course, the two kinds of measurements may have quite different error characteristics. One would expect visual observations to be less reliable at night, while automatic measurements should be insensitive to time of day. Automatic observations are less accurate and reliable when the visibility is good, due to the low density of scattering particles. Visiometer measurement quality is also dependent on effective maintenance, for example to ensure that a spider's web does not produce a spuriously low reading!

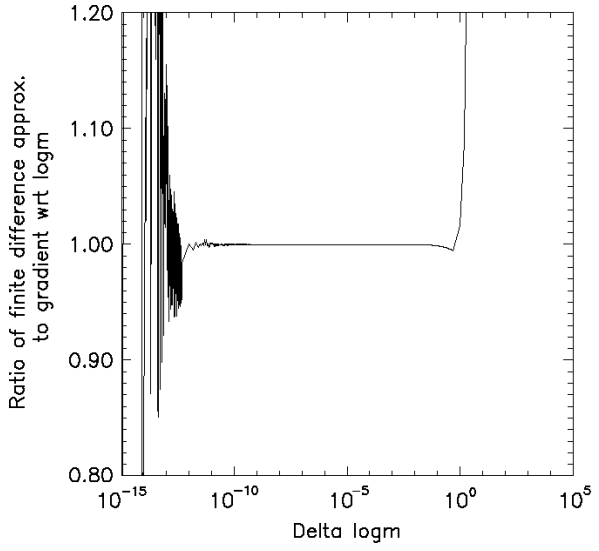


Figure 2. Graph showing the accuracy of the $\log_{10}$(aerosol) derivative.

Instrumental errors for forward-scattering devices as quoted by manufacturers are typically in the range 10–20%. Such errors are, however, dwarfed by the error of representativeness at scales typical of today's generation of operational NWP models with grid-lengths of $\sim 10$ km, for which subgrid-scale variability in visibility is high. Golding (personal communication) estimated the combined effect of instrumental errors and errors of representativeness by calculating the ratio of visibilities measured at nearby station pairs with separations varying from 5 to 30 km. Data were accumulated for 3 separate months (December, May and August) and included all observations, whether made during precipitation or not. There were 5 station pairs with separations less than 15 km. The root-mean-square visibility ratio for these station pairs was in the range 1.5–2, so the 'total' observation error (instrumental plus representativeness) for our 12 km model was set to 0.25 in terms of $\log_{10}$(visibility). This number determines the diagonal elements of the matrix $\mathbf{E} + \mathbf{F}$ in Equation (21).

### 3.4. Quality control

The patchy nature of fog and the relatively poor quality of model background visibility fields means that the conventional background and buddy checks used for other observation types are not useful. Instead, a consistency check is made between reported visibility and other collocated surface observations.

Observations of aerosol are not available, but we can check that the observed visibility lies within a realistic range. For this we assume either high or low aerosol content and diagnose visibility with the reported relative humidity and temperature. That is, we accept observations with visibility $vis_{\mathrm{obs}}$ only when the following two inequalities hold:

$$\begin{aligned} vis_{\min}(m_{\max}, RH_{\mathrm{obs}}, T_{\mathrm{obs}}) &< vis_{\mathrm{obs}}\,, \\ vis_{\mathrm{obs}} &< vis_{\max}(m_{\min}, RH_{\mathrm{obs}}, T_{\mathrm{obs}})\,, \end{aligned} \qquad (25)$$

where $m_{\max}$ and $m_{\min}$ are 'plausible' maximum and minimum aerosol values of 100 and 0.1 $\mu$g kg$^{-1}$, respectively. We note that the maximum value is of the same order as the 80 $\mu$g kg$^{-1}$ used in the HIRLAM-DMI visibility diagnosis (Petersen and Nielsen, 2000). In practice, the visibility is most sensitive to humidity near saturation when the range $vis_{\max} - vis_{\min}$ becomes small and variation in aerosol content has little impact. For this




DOI: 10.1002/qj

reason, and to account for the error of representativeness in reported visibility, we increase $vis_{\max}$ and reduce $vis_{\min}$ by a factor equal to the assumed observation error for visibility.

This 'visQC' scheme is supplemented by a step to flag visibility data where present weather reports indicate precipitation at the time of the observation. This is introduced because, although the model diagnosis of visibility accounts for precipitation, the algorithm for assimilation of visibility is not able to correct for reduced visibility due to precipitation in a sensible way. Assimilation increments are added only to the aerosol, moisture and temperature fields. Flagging visibility reports in rain prevents us, for example, adding a large quantity of aerosol to match low visibility in a heavy shower.

The basic consistency check was tested in case-studies and a real-time trial over 3 months in the 12 km UK Mesoscale model (Sharpe and Macpherson, 2005). The scheme flagged on average around 10% of data. The assimilation was found to converge around 5% faster, and to give aerosol increments around 10% smaller in r.m.s. terms. These reductions come essentially from rejection of visibility reports that contradict collocated humidity observations. The visibility observation is not flagged if the humidity report has itself been flagged.

The impact of visQC was assessed through the Equitable Threat Score (ETS) for visibility at three thresholds (200 m, 1000 m and 5000 m). ETS measures the skill of a forecast relative to a 'random' forecast, a score of unity being perfect, while a score of zero is for a forecast with no more skill than chance. The three visibility thresholds are those used within the 'UK NWP Index', an overall measure of skill for regional versions of the operational MetUM, which also includes skill scores for screen-level temperature, 10 m wind, total cloud cover, cloud base height and 6-hourly rainfall accumulation. Temperature and wind skill scores are a function of the ratio of the mean square model forecast error to that for a persistence forecast. Cloud and rain skill scores are ETS values for three thresholds of the relevant variable. Each variable contributes with equal weight to the final NWP index. Impact of visQC on the UK NWP index was neutral and the impact on the ETS visibility component alone (–1.5%) was of questionable significance. Objective assessment of a quality control scheme is complicated by the issue of whether or not to verify only against unflagged observations. For technical reasons, the visQC trial verification was against all observations.

As well as the real-time visQC scheme, a monthly reject list is compiled of stations which differ persistently from the model background field by a large amount. If the mean or standard deviation of the observation minus background (o–b) in $\log_{10}$(visibility) are greater than certain threshold multiples of the observation error, the station will be placed on the reject list. For the North Atlantic and European domain, around 5% of stations are listed.

### 3.5. Background errors

The background error model for the Met Office global 3D-Var system is described by Ingleby (2000). 3D-Var control variables have background-error statistics derived from a version of the so-called 'NMC method' (Parrish and Derber, 1992) of analysing accumulations of forecast difference statistics. The NMC method has recently been applied by Benedetti and Fisher (2007) to derive error covariances for aerosol within a global 4D-Var aerosol analysis and forecast system being developed as part of the Global Monitoring for Environment and Security (GMES) initiative. They point out that a variety of less sophisticated methods have been used to specify aerosol error covariances in other models.

During development of the visibility analysis, the background-error variance for $\log_{10}$(aerosol) was initially set to 1.0, or in other words, the error in the background aerosol was estimated to be an order of magnitude. Later, aerosol covariance statistics were calculated from $(t+6)-(t+12)$ forecast differences valid at 06 UTC, leading to an estimate of 0.25 in $\log_{10}$(aerosol). When this value was used in test analyses, however, the results were inferior to those with the initial default of 1.0, so this was retained. Recently, (o–b) statistics for visibility have been used to estimate a more realistic aerosol background-error variance of 0.4, and this value became operational in March 2007 in the NAE model, reducing aerosol increments appreciably.

In the UK Mesoscale 3D-Var system, horizontal background correlations are modelled by a second-order autoregressive (SOAR) function. The choice of scale parameter was guided by earlier statistical studies of (o–b) error correlation applied in the previous AC scheme (Macpherson *et al.*, 1996). These had led to a value of 90 km as the scale for $RH$. Aerosol observations are not readily available for statistical analysis, but since the aerosol field is mainly relevant, along with humidity, to visibility prediction, the same scale of 90 km was specified for $\log_{10}$(aerosol). A recent (o–b) error correlation study for $\log_{10}$(visibility) (Dow, personal communication) led to a larger value of 150 km becoming operational in the NAE model in April 2008, with further significant reduction in the typical size of aerosol increments.

The vertical covariances in our 3D-Var system are produced by decomposing the forecast differences into vertical modes. For the aerosol control variable, however, with observations available only at the surface, we take the simpler route of analysing increments to $\log_{10}$(aerosol) at the lowest model level, then applying a simple function to extrapolate them to all boundary-layer levels. The vertical correlation function is that used for analysis of surface data in the AC scheme:

$$f = \exp\left\{-\left(18\,\ln\frac{p}{p_*}\right)^2\right\}, \qquad (26)$$

which gives a value of $\sim 0.4$ at 950 hPa. Some tests were run with a value of 25 instead of 18, giving a value of

$f$ close to 0.2 at 950 hPa, but the model showed little sensitivity to this change.

### 3.6. Minimisation and convergence

Validation of the tangent linear observation operator confirmed the high degree of nonlinearity of the visibility equation since errors in the linearised approximation were seen at perturbation values typical of forecast errors. Because of this, the nonlinear observation operator for visibility was chosen as default during the minimisation, which uses a limited-memory quasi-Newton algorithm based on Gilbert and Lemaréchal (1989).

The convergence of the minimisation is defined by stopping criteria, for which there are various options. One is that the norm of the gradient of the penalty must decrease to a certain fraction of its original value. After experimentation, however, it was decided to apply the criterion that the penalty must decrease to within a chosen absolute margin of the estimated minimum penalty at 'full convergence'. Typically the assimilation may take about 40 iterations to reach convergence under this criterion, though this figure varies considerably according to the meteorological situation. Without visibility assimilation, convergence is achieved in around 20 iterations, and this figure varies little from case to case.

## 4. Results

### 4.1. Initial tests

Tests of the visibility assimilation began with investigations of the response to single observations. These were run with an observation error of 0.25 in $\log_{10}$(visibility), and background errors of 1.0 in $\log_{10}$(aerosol), 6.6% in screen-level *RH* and 1 K in screen temperature. To give an example, an observation of visibility 200 m and model background visibility of 2000 m gave rise to an order of magnitude increase in aerosol content from 140 to 1500 $\mu$g kg$^{-1}$, and a 2% rise in *RH* from 94.2% to 96.1%, while the temperature increment was a tiny reduction of 0.04 K.

Whilst the size of increments generated by low-visibility observations to aerosol, humidity and temperature depends sensitively on the background fields, the example quoted is not untypical. In a broad sense, the error statistics quoted give rise to aerosol increments of an order of magnitude, *RH* increments of a few per cent and very small temperature increments. Figure 3 shows the increments to $\log_{10}$(aerosol) following assimilation of all observations including visibility. Increments are concentrated over or near land since marine visibility and humidity observations were rejected. (Marine air temperature and humidity observations have recently been introduced to the operational assimilation in April 2008, and marine visibility data will be added later in 2008.) In this fairly representative example, increments to $\log_{10}$(aerosol) range from –1.3 to 1.4. The corresponding increments to low-level *RH* from all observations are shown in Figure 4(a), while in Figure 4(b) we have the *RH* increments when visibility observations are excluded from the assimilation. The difference is shown in Figure 4(c), from which we see that the impact of visibility assimilation can be up to around 10% or so, an appreciable fraction of the peak increments coming from other data types. Comparison of Figure 3 with Figure 4(c) reveals that, unlike the single-observation case, the signs of the analysis increments in aerosol and *RH* are not always correlated. This behaviour is a consequence of the presence of multiple observations and a very nonlinear relationship between visibility, aerosol and humidity.

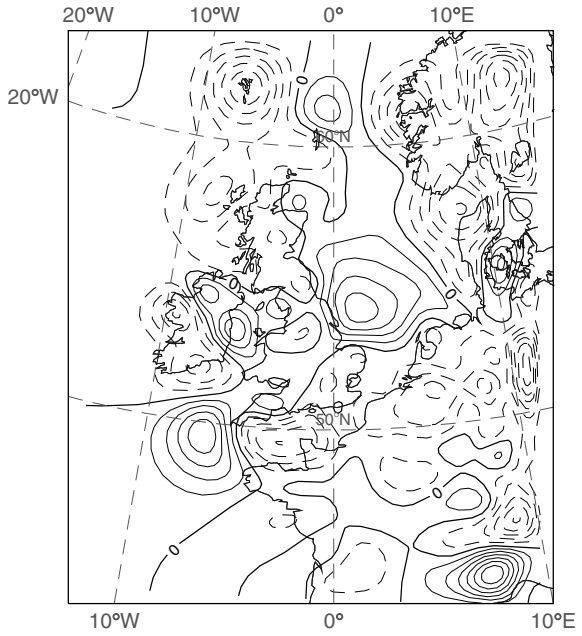


Figure 3. Analysis increments to $\log_{10}$(aerosol) at level one from 00 UTC on 25 January 2000. The contour interval is 0.2 and dashed contours show negative values.

### 4.2. Case-studies

To assess the performance of the visibility assimilation in 3D-Var, a set of eight foggy case-studies was run, each with 12 hours of assimilation followed by a forecast. Each data time was chosen for the presence of fog to be significant at a forecast range of 6 hours, for which we present in Table I the impact of visibility assimilation on the Equitable Threat Score (ETS) for visibility at three thresholds, averaged over all eight cases. A benefit from visibility assimilation is seen at each threshold. Rerunning the same cases with only 6 hours of assimilation and a 12-hour forecast, to verify at the same time, still showed a benefit but smaller than obtained at $t + 6$.

Other variables were slightly affected by including visibility assimilation. Averaged over all cases, the initial $(t + 0)$ r.m.s. errors in temperature and relative humidity were larger by 0.02 K and 0.1% respectively. At later forecast ranges, the results averaged over all cases showed little difference in r.m.s. errors between the two tests. However, visibility assimilation brought




DOI: 10.1002/qj

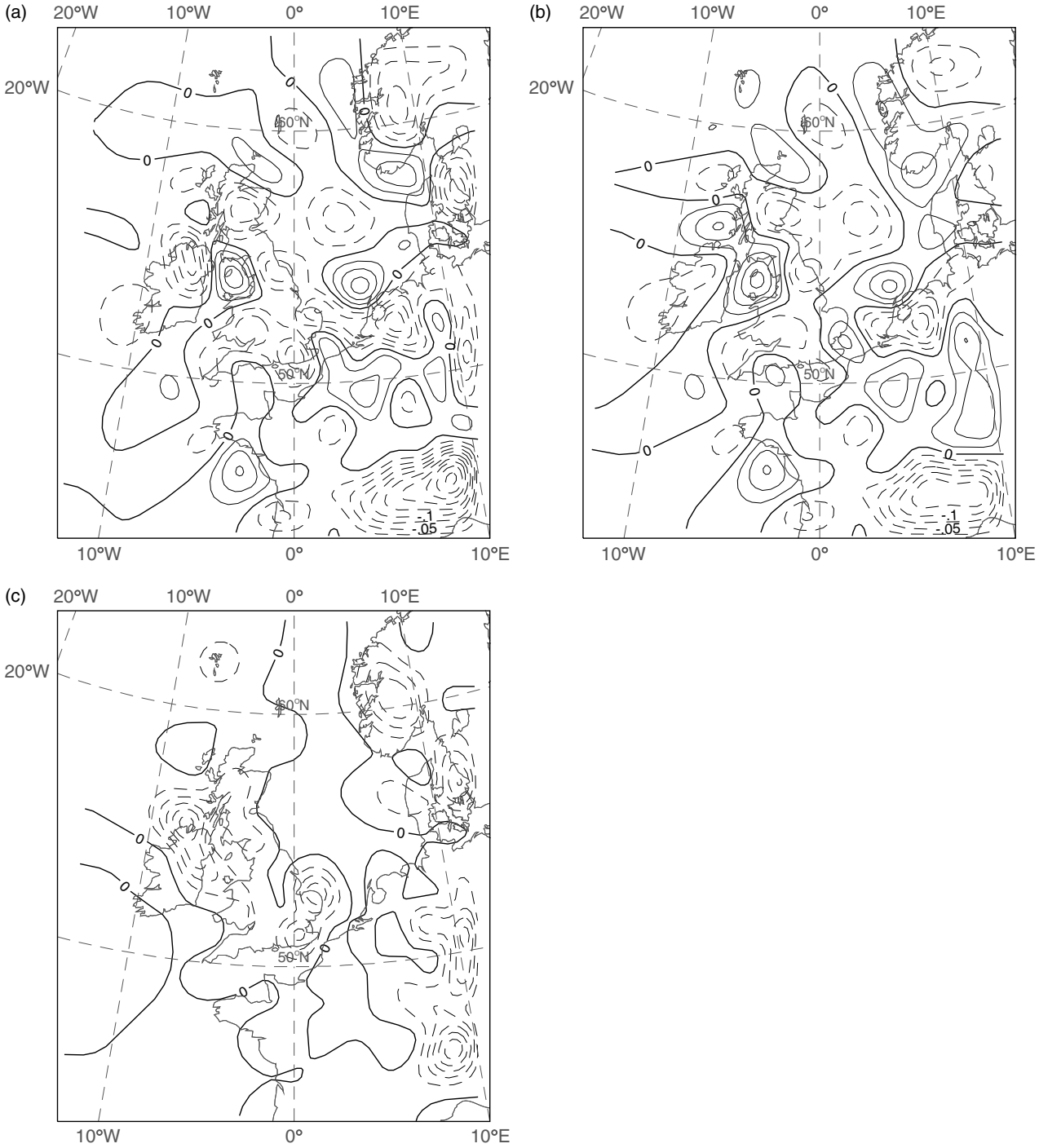


Figure 4. Analysis increments to relative humidity at lowest model level for 00 UTC on 25 January 2000. The contour interval is 2.5% and dashed contours show negative increments. (a) All observations assimilated including visibility. (b) All observations assimilated *except* visibility. (c) Difference (a) minus (b).

a reduction in the positive relative humidity bias at all forecast times (Figure 5).

A subjective evaluation of model visibility forecasts was also carried out. Runs with assimilation of visibility observations performed significantly better. Out of 24 pairs of visibility forecast fields compared over the 8 cases, the runs with visibility assimilation were assessed by a forecaster as better on 13 occasions and worse on 4. Figure 6(a) shows the $t + 6$ fog probability forecast for 12 UTC on 22 December 1998, from a trial using all observations including visibility. The 0.4 probability contour extends from south-west England through central England and over East Anglia. The trial without visibility (Figure 6(b)) shows a much reduced area of fog. Comparing with the observations valid at the time (Figure 6(c)) we see that the run with visibility assimilation performed quite well in predicting the fog over England. It also picked out the fog in western France which the run without visibility assimilation missed.




DOI: 10.1002/qj

Table I. Equitable Threat Score (ETS) at $t + 6$ hours for three visibility thresholds with visibility assimilation omitted (CONTROL) and included (TEST). Verification is averaged over eight foggy cases and all SYNOP stations in the model domain.

| Visibility threshold | ETS (CONTROL) | ETS (TEST) |
|---|---|---|
| 200 m | 0.18 | 0.21 |
| 1000 m | 0.20 | 0.25 |
| 5000 m | 0.20 | 0.23 |

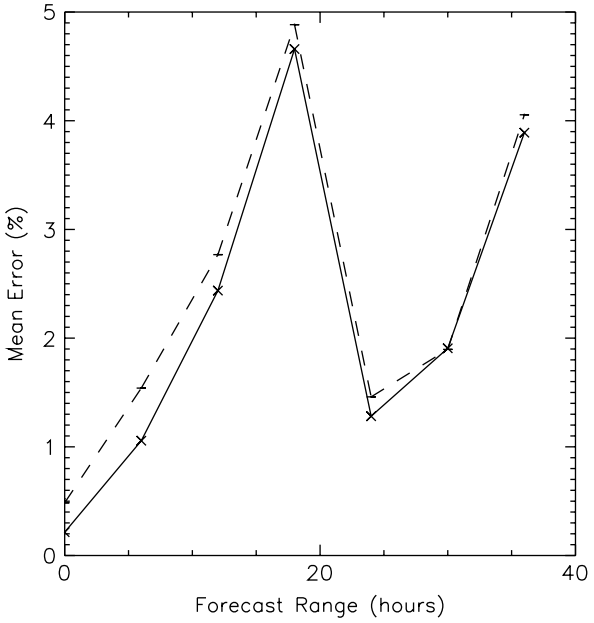


Figure 5. Verification of screen-level relative humidity (%). Mean bias over all 8 cases. Comparison of assimilating all observations including visibility (solid line) against all observations excluding visibility (dashed line).

### 4.3. Sensitivity experiments

The complete visibility modelling and assimilation system contains a number of components whose relative contributions to the performance of the system were investigated in a set of experiments run on the period 30 January to 6 February 2006. A continuous 3-hourly 3D-Var assimilation was run in the 12 km UK Mesoscale model, with four 36-hour forecasts per day. Weather over the UK was dominated by high pressure and a relatively high incidence of poor visibility. For example, around 4%, 9% and 48% of surface observations during the week had visibilities of less than 200, 1000, 5000 m respectively. Climatological frequencies of such low visibilities are more like 1%, 2% and 10–15% respectively. So, although the sample of verifying data from one week's observations is still small in statistical terms, it is a suitable period on which to test the system.

The control experiment (CNTL) contained all the standard operational features of the visibility system. In experiment NoVA, there was no visibility assimilation. In NoVASrce there was no visibility assimilation and aerosol sources were switched off. Finally, in NoVASAero, there was no visibility assimilation and all modelling of aerosol sources, advection, mixing and rainout was disabled. Visibility diagnosis was made with a fixed aerosol content of $10\,\mu g\,kg^{-1}$, as used in the global model.

In Figure 7 we see the impact of each component on the r.m.s. error in $\log_{10}$(visibility). The aerosol modelling (advection, mixing and removal) and sources are seen to be beneficial throughout the full forecast length, while visibility assimilation still gives some benefit as far as $t + 24$ hours. Early in the forecast, all three components bring comparable reductions in the r.m.s. error. It is noticeable that even the runs without visibility assimilation show a significant rise in error with forecast time. For the NoVA experiment, the assimilation of screen-level relative humidity data will give rise to smaller visibility errors early in the forecast, and the more accurate initial wind field will give better advection of moisture and aerosol from sources within the model domain. For experiment NoVASrce, the impact of the humidity data still applies, and there will be a contribution from more accurate advection of aerosol input from the boundary as well as from moisture advection. For NoVASAero, the humidity assimilation and moisture advection are still relevant.

Like NoVASAero, the global model has no prognostic aerosol or visibility assimilation, and until April 2008 it did not assimilate screen-level humidity data. Its r.m.s errors in $\log_{10}$(visibility) have a similar flat trend with forecast time (not shown) and values significantly higher than those in the mesoscale system equivalent to CNTL. For example, a 12-month mean value of the r.m.s errors in $\log_{10}$(visibility) at $t + 0$ is 0.45, compared with 0.32 for the NAE model equivalent to CNTL. The $t + 0$ value for CNTL in Figure 7 of just under 0.4 may reflect the difficulty in fitting observations with a greater proportion of low visibilities than normal. Smith *et al.* (2002) report corresponding values in the range 0.35–0.45 for visibility analyses from the RUC model over a 10-day period in the USA. Within long-term operational verification results for $\log_{10}$(visibility) there is a diurnal signal. Errors tend to be larger overnight (00 and 06 UTC) than during the day (12 and 18 UTC), with the peak at 06 UTC when the incidence of low visibility is greatest. Smith *et al.* (2002) also report lower errors during the day from the RUC20 model.

A slightly different view of the relative impact of different scheme components emerges when we focus on forecast skill at the important low visibility end of the observation distribution (Figures 8 and 9). For the 5 km threshold (Figure 8), the picture is similar to that in Figure 7, although this time the benefit of assimilation lasts throughout the 36-hour forecast, and the impact of aerosol sources is relatively larger than that of the aerosol advection, mixing and removal. For the 1000 m threshold (Figure 9), however, we find that visibility assimilation gives by far the largest benefit at $t + 6$ hours, but this impact is already much reduced by $t + 12$ hours. This signal is consistent with the very nonlinear dependence of visibility on relative humidity and the relative insensitivity of such low visibilities to aerosol content.



DOI: 10.1002/qj

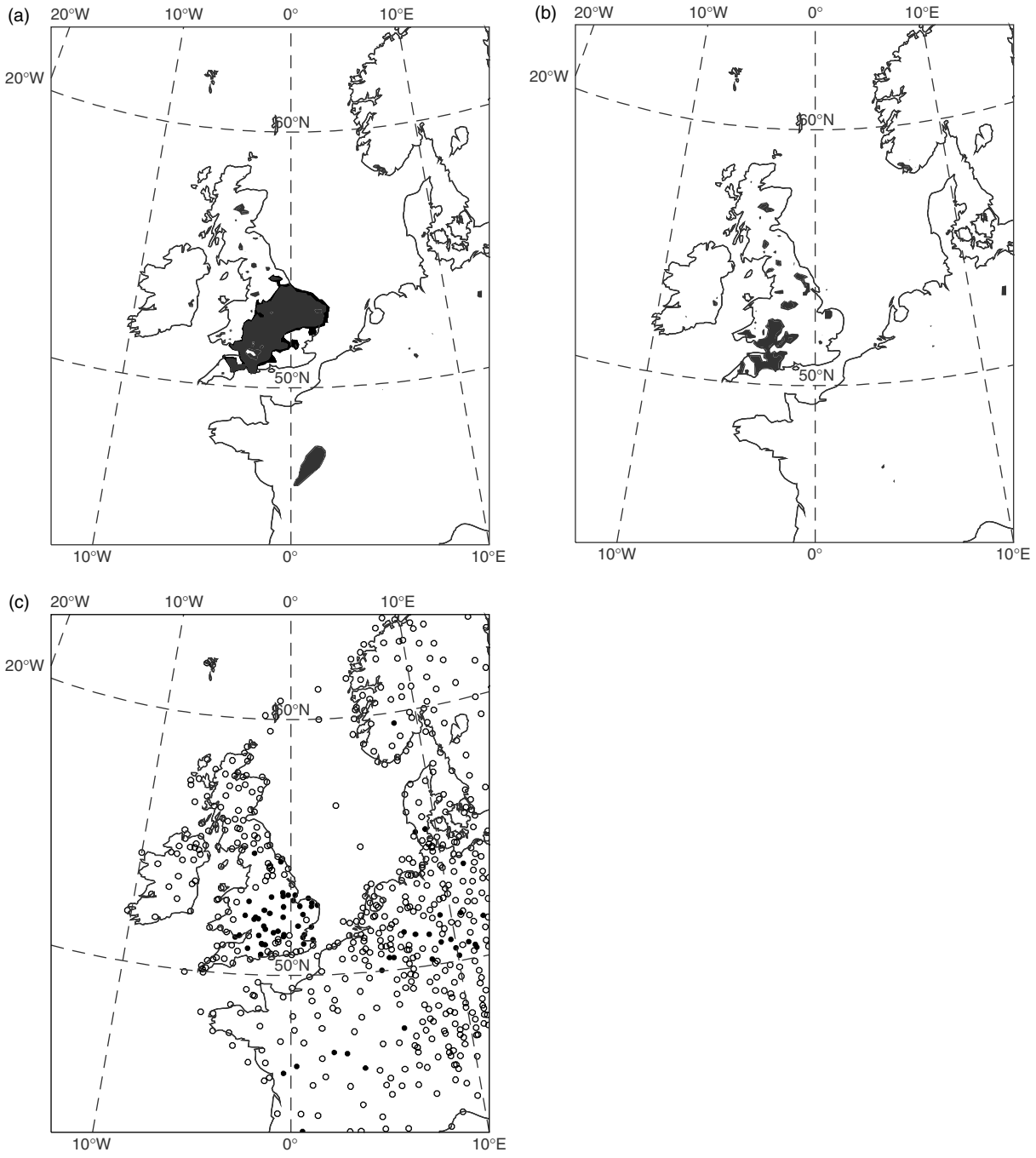


Figure 6. (a) $t + 6$ forecast of fog probability at 1.5 m valid at 12 UTC on 22 December 1998. The shaded region indicates a probability greater than 0.4. (b) is as (a) but visibility observations were excluded from the trial. (c) Observations of visibility extracted from stations for 12 UTC. Shaded (open) circles show visibility reports of less than (greater than) 1 km.

The total impact of aerosol modelling and visibility assimilation on long-term operational performance is hard to assess rigorously without long and expensive trials. A good indicator can, however, be obtained from a comparison of skill scores from the global model, with its visibility system as in NoVASAero, and the NAE model which contains all the features of the CNTL experiment of this section. A 3-month mean ETS score for the whole North Atlantic and European domain (Figure 10) shows a clear advantage at all forecast times for the NAE at each visibility threshold. There were around 7000 observations of visibility less than 200 m in the NAE domain during this period, so the sample size is good even at the lowest threshold. It should be noted that the NAE model assimilates screen-level humidity data, while for the relevant period in Figure 10 the global model did not, giving the NAE an advantage beyond visibility assimilation when it comes to prediction at the 200 m and 1000 m thresholds.

Regarding impact on other variables in the February 2006 sensitivity experiments, visibility assimilation gave a small increase (approximately 1%) in r.m.s. errors

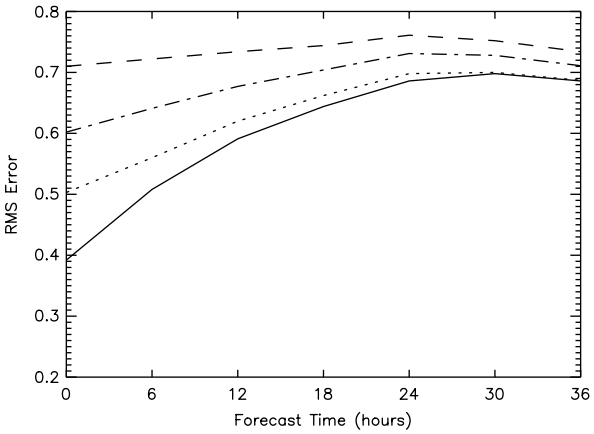


Figure 7. R.m.s. error in $\log_{10}$(visibility) as a function of forecast time for the four experiments CNTL (solid line), NoVA (dotted line), NoVASrce(dot-dash line) and NoVASAero (dashed line), averaged over 29 forecasts made during the period 30 January – 6 February 2006. All stations in the model domain were used. For explanation, see section 4.3.

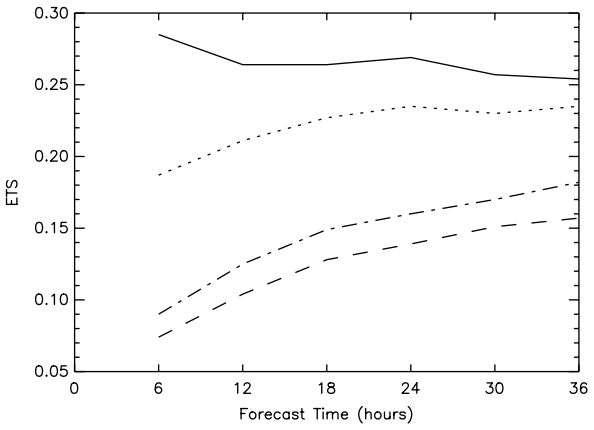


Figure 8. Equitable Threat Score (ETS) for the 5000 m threshold as a function of forecast time for the four experiments CNTL (solid line), NoVA (dotted line), NoVASrce(dot-dash line) and NoVASAero (dashed line), averaged over 29 forecasts made during the period 30 January – 6 February 2006. All stations in the 12 km Mesoscale Model domain were used. For interpretation, see section 4.3.

for screen temperature. The impact on screen relative humidity was neutral for r.m.s. errors, but showed a similar reduction in positive humidity bias to that noted in the case-study results shown in Figure 5.

## 5. Conclusions and future work

We have described and documented the performance of a visibility forecast and assimilation system. The model is based on a single prognostic aerosol variable to represent the air-mass characteristic, along with a simple description of aerosol sources within and at the boundaries of the limited-area model domain, plus a treatment of aerosol mixing in the boundary later and rainout. The assimilation is a 3-dimensional variational scheme in which the visibility observation operator is a very nonlinear function of humidity, aerosol and temperature. Visibility assimilation is most important

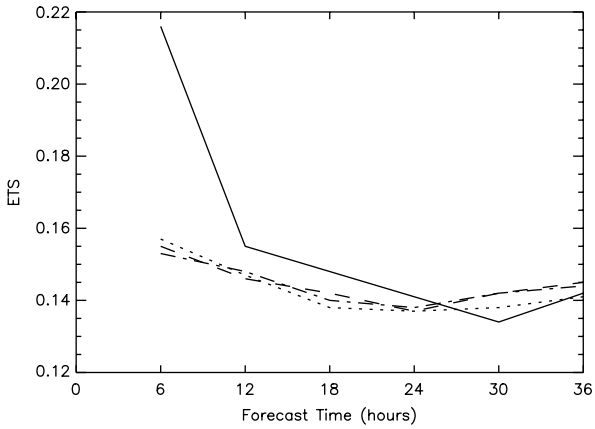


Figure 9. As Figure 8, but for the 1000 m threshold.

within the first 6–12 hours of the forecast and for the lowest visibilities, while aerosol modelling is important for slightly higher visibilities and is still significant at longer forecast times.

From the results of sensitivity experiments, a number of avenues for future work could lead to improvements in the system. Aerosol advection along with boundary-layer mixing and rainout are beneficial to visibility prediction, at least for the 5 km threshold. So one might expect that an extension of the 3D-Var system to 4D-Var would be helpful, with aerosol advection included within the linear 'Perturbation Forecast Model' of the Met Office 4D-Var system (Rawlins *et al.*, 2007). A 4D-Var scheme may also improve the analysis of humidity on which visibility depends so strongly. Indeed, the Met Office NAE model was upgraded operationally to 4D-Var in March 2006, and performance of the visibility analysis within 4D-Var will be reported in a future paper.

The industrial aerosol sources are also clearly beneficial to the scheme, yet they are known to be out of date, coming from figures released in 1993. An update based on more recent data is planned, and this must have the potential to improve the visibility predictions around the 5 km threshold still further. Not only the intensity and locations of industrial sources have changed over the years, but also the typical chemical composition of the aerosol. This has implications for the details of the parametrisation of visibility as a function of humidity and aerosol concentration. Aerosol data from flights of a research aircraft around the UK and special surface observations have been analysed to validate and improve such aspects of the aerosol modelling system. This work is reported in Part II of this paper. A more accurate aerosol distribution in the boundary layer might also allow benefit to be obtained from coupling the prognostic aerosol to the radiation scheme. At present, such coupling degrades the r.m.s. forecast errors for screen-level temperature by 1–2%. The degradation may result from inaccuracies in aerosol amounts originating in assimilation or the model of aerosol sources, transport and removal. It could also be partly due to the radiation scheme's characterisation of the proportions of water-soluble aerosol, soot and dust

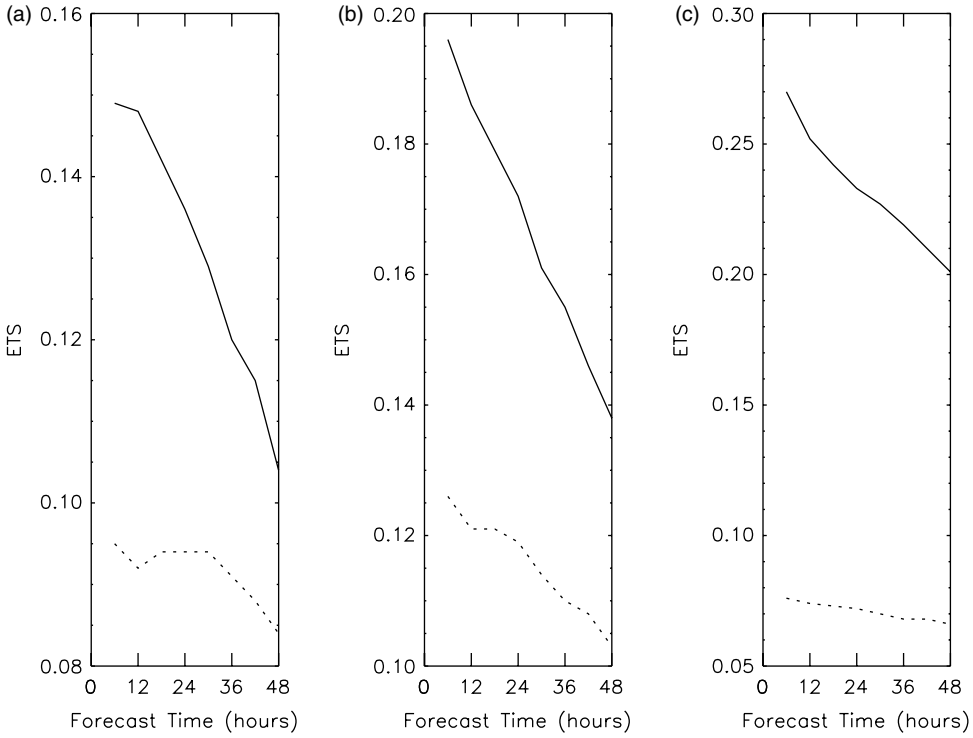


Figure 10. Equitable Threat Score (ETS) as a function of forecast time, for the operational North Atlantic and European Model (NAE; solid line) and for the global model (dotted line), averaged over the period December 2006 – February 2007. All stations in the NAE model domain were used. Curves show data for thresholds of (a) 200 m, (b) 1000 m and (c) 5000 m. For interpretation, see section 4.3.

that make up the aerosol. In the meantime, a climatological aerosol distribution is used within the radiation scheme.

## Acknowledgements

We thank Mike Bush and Richard Renshaw for support in performing some of the experiments in section 4.3, Tim Payne for advice on minimisation issues and Brian Golding for access to review material from within the COST 722 initiative on 'Short-range forecasting methods of fog, visibility and low clouds'.

## References

Ballard SP, Golding BW, Smith RNB. 1991. Mesoscale model experimental forecasts of the haar of northeast Scotland. *Mon. Weather Rev.* **119**: 2107–2123.

Ballard SP, Wright BJ, Golding BW. 1992. 'Diagnosis of visibility in the UK Met Office mesoscale model and the use of a visibility analysis to constrain initial conditions'. Forecasting Research Scientific Paper No. 4. Met Office: Exeter, UK.

Benedetti A, Fisher M. 2007. Background-error statistics for aerosols. *Q. J. R. Meteorol. Soc.* **133**: 391–405.

Benjamin SG, Weygandt SS, Brown JM, Smith TL, Smirnova T, Moninger WR, Schwartz B, Szoke EJ, Brundage K. 2004. 'Assimilation of METAR cloud and visibility observations observations in the RUC'. Preprints for 11th Conference on Aviation, Range, and Aerospace Meteorology, Hyannis, MA. Amer. Meteorol. Soc: Boston.

Bloom SC, Takaks LL, Da Silva AM, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon. Weather Rev.* **124**: 1256–1271.

Bott A, Trautmann T. 2002. PAFOG – A new efficient forecast model of radiation fog and low-level stratiform clouds. *Atmos. Res.* **64**: 191–203.

Clark PA, Hopwood WP. 2001a. One-dimensional site-specific forecasting of radiation fog. Part I: Model formulation and idealised sensitivity studies. *Meteorol. Appl.* **8**: 279–286.

Clark PA, Hopwood WP. 2001b. One-dimensional site-specific forecasting of radiation fog. Part II: Impact of site observations. *Meteorol. Appl.* **8**: 287–296.

Davies T, Cullen MJP, Malcolm A, Mawson M, Staniforth A, White A, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Gilbert JC, Lemaréchal C. 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical Programming* **45**: 407–435.

Golding BW. 1998. Nimrod: A system for generating automated very short range forecasts. *Meteorol. Appl.* **5**: 1–16.

Grell GA, Peckham SA, Schmitz R, McKeen SA, Frost G, Skamarock WC, Eder B. 2005. Fully coupled online chemistry within the WRF model. *Atmos. Environ.* **39**: 6957–6975.

Haywood JM, Bush M, Abel S, Claxton B, Coe H, Crosier J, Harrison M, Macpherson B, Naylor M, Osborne S. 2008. Prediction of visibility and aerosol within the operational Met Office Unified Model. II: Validation of model performance using observational data. *Q. J. R. Meteorol. Soc.* **134**: 000–000.

Ingleby NB. 2001. The statistical structure of forecast errors and its representation in the Met. Office global 3D-Var. *Q. J. R. Meteorol. Soc.* **127**: 209–231.

Jones CD, Macpherson B. 1997: A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorol. Appl.* **4**: 269–277.

Koschmeider H. 1924. Therie der horizontalen sichtweite. *Beitr. Phys. Freien Atmosph.* **12**: 33–53, 171–181.

Leigh RJ. 1995. Economic benefits of Terminal Aerodrome Forecasts (TAFs) for Sydney airport, Australia. *Meteorol. Appl.* **2**: 239–247.

Lerner JA, Westphal DL, Reid JS. 2004. 'Quality controlled surface visibility observations used to validate predicted surface aerosol concentration for Southwest Asia'. Proceedings of 16th Conference on Numerical Weather Prediction, 11–15 Jan. 2004, Seattle, WA. (CD-ROM, P4.3) Amer. Meteorol. Soc: Boston.

Lorenc AC, Bell RS, Macpherson B. 1991. The Meteorological Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.* **117**: 59–89.

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met Office global 3-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **126**: 2991–3012.

Macpherson B, Wright BW, Hand WH, Maycock AJ. 1996. The impact of MOPS moisture data in the UK Meteorological Office data assimilation scheme. *Mon. Weather Rev.* **124**: 1746–1766.

Macpherson B, Andrews PL, Harcourt SA, Ingleby NB, Chalcraft BV, Maycock AJ, Renshaw RJ, Parrett CA, Anderson SR, Sharpe M, Harrison DL, Gibson M. 2002. 'The Operational Mesoscale Data Assimilation System 1999–2001: Implementation of 3D-Var and later upgrades'. Forecasting Research Technical Report No. 374. Met Office: Exeter, UK.

Munoz-Alpizar R, Blanchet J-P, Quintanar AI. 2003. Application of the NARCM model to high-resolution aerosol simulations: Case study of Mexico City basin during the Investigación sobre Materia Particulada y Deterioro Atmosférico-Aerosol and Visibility Research measurements campaign. *J. Geophys. Res.* **108**: (D15) 4462.

Pagowski M, Gultepe I, King P. 2004. Analysis and modeling of an extremely dense fog event in southern Ontario. *J. Appl. Meteorol.* **43**: 3–16.

Parrish DF, Derber JC. 1992. The National Meteorological Center's Spectral Statistical Interpolation analysis system. *Mon. Weather Rev.* **120**: 1747–1763.

Pasini A, Pelino V, Potesta S. 2001. A neural network model for visibility nowcasting from surface observations: results and sensitivity to physical input variables. *J. Geophys. Res (Atmos)* **106**: 14951–14959.

Petersen C, Nielsen NW. 2000. 'Diagnosis of visibility in DMI-HIRLAM'. Scientific Report 00-11. DMI: Copenhagen, Denmark.

Pruppacher HR, Klett JD. 1978. *Microphysics of clouds and precipitation*. D. Reidel Publishing Company: Dordrecht, Netherlands.

Radcliffe J, Chalcraft BV, Wilson CA, Macpherson B, Wilson D, Clark PA. 1998. 'Report on pre-operational trials of the new microphysics scheme in the mesoscale model'. Forecasting Research Technical Report No. 259. Met Office: Exeter, UK.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.

Renshaw RJ, Candy B, Christidis N, Macpherson B, Chalcraft BV, Forsythe M. 2004. *'The 2002/03 data assimilation upgrades in the UK Mesoscale Model'*. Forecasting Research Technical Report No. 435. Met Office: Exeter, UK.

Rosenfeld J. 1996. Cars vs. the weather. A century of progress. *Weatherwise* **49**: 14–23.

Sharpe CT, Macpherson B. 2005. 'A quality control scheme for visibility observations'. Forecasting Research Technical Report No. 460. Met Office: Exeter, UK.

Smirnova TG, Benjamin SG, Brown JM. 2000. 'Case study verification of RUC/MAPS fog and visibility forecasts'. Pp. 31–36 in preprints for 9th Conference on Aviation, Range, and Aerospace Meteorology, Orlando, FL. Amer. Meteorol. Soc: Boston.

Smith RNB. 1990. A scheme for predicting layer clouds and their water content in a general circulation model. *Q. J. R. Meteorol. Soc.* **116**: 435–460.

Smith TL, Benjamin SG, Brown JM. 2002. 'Visibility forecasts from the RUC20'. Pp. 150–153 in preprints for 10th Conference on Aviation, Range, and Aerospace Meteorology, Portland, OR. Amer. Meteorol. Soc: Boston.

Stoelinga MT, Warner TT. 1999. Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event. *J. Appl. Meteorol.* **38**: 385–404.

Wright BJ, Thomas N. 1998. An objective visibility analysis and very-short-range forecasting system. *Meteorol. Appl.* **5**: 157–181.