University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Silver, JD;Christensen, JH;Kahnert, M;Robertson, L;Rayner, PJ;Brandt, J

Title:
Multi-species chemical data assimilation with the Danish Eulerian hemispheric model: system description and verification

Date:
2016-09-01

Citation:
Silver, J. D., Christensen, J. H., Kahnert, M., Robertson, L., Rayner, P. J. & Brandt, J. (2016). Multi-species chemical data assimilation with the Danish Eulerian hemispheric model: system description and verification. Journal of Atmospheric Chemistry, 73 (3), pp.261-302. https://doi.org/10.1007/s10874-015-9326-0.

Persistent Link:
https://hdl.handle.net/11343/282880

Noname manuscript No.
(will be inserted by the editor)

Multi-species chemical data assimilation with the Danish Eulerian Hemispheric Model: System description and verification

Jeremy D. Silver · Jesper H. Christensen · Michael Kahnert · Lennart Robertson · Peter J. Rayner · Jørgen Brandt

November 10, 2015

Abstract Satellite retrievals of atmospheric composition provide a wealth of data on a global scale. These complement results from atmospheric chemistry-transport models (CTMs), and can be combined using data assimilation. We present two assimilation schemes coupled to the Danish Eulerian Hemispheric Model (DEHM), a three-dimensional, off-line CTM with full photochemistry: a variant on the ensemble Kalman filter and the three-dimensional variational scheme. The aim of this paper is to describe the two schemes and present an initial assessment of their impacts on model skill.

Retrievals of multiple atmospheric trace gases are assimilated, namely: NO_2 tropospheric column densities, CH_4 total column densities, and partial column concentrations of O_3, CO and CH_4; these data are retrieved from four satellite sensors. Data for each species are assimilated independently of one another, and other species are only adjusted indirectly via the model's chemistry and dynamics. Assimilation results are compared with measurements from surface monitoring stations and other satellite retrievals, and preliminary validation results are presented.

Reference simulations (without assimilation) grossly underestimate surface CO concentrations, and both assimilation schemes eliminate this large and systematic model bias. The assimilation improves the spatial correlation of modelled CO with surface observations, and improves the spatial correlation between forecasts and retrievals for CO, NO_2 and O_3. Results for CH_4 show a loss of skill due to a mismatch in model bias between two assimilated CH_4 data-sets. Finally, we discuss differences in methodology and results between this paper and a recent study on multi-species chemical data assimilation. Joint optimisation of initial conditions and emission rates offers a promising direction for improving modelled boundary-layer concentrations.

Keywords Atmospheric Chemical-Transport Modelling · Data Assimilation · Satellite Retrievals · Ensemble Kalman Filter · Three-Dimensional Variational Analysis

1 Introduction

Concentrations of atmospheric constituents have important impacts on human and ecosystem health (Bytnerowicz et al 2007; Kampa and Castanas 2008), climate change (Seinfeld and Pandis 2012), and dynamic meteorology (Levin and Cotton 2009). Atmospheric chemistry-transport models (CTMs) are important tools for studying the effects of modified trace gas and aerosol concentrations (Jacobson 2005). In-situ observations and remotely-sensed retrievals of air pollution provide an independent source of information to complement modelling results (Hertel et al 2007; Martin 2008). Such data sources can be combined in a statistically coherent manner through data assimilation, in order to estimate atmospheric concentrations more accurately (Kalnay 2003).

The past two decades has seen a rapid development in the area of chemical data assimilation (DA), building on advances in computing resources, algorithms and remote-sensing data (Sandu and Chai 2011). Most of this development follows research in numerical weather prediction (NWP), however chemical DA faces a number of other challenges (Lahoz et al 2007). For instance, while the initial conditions in NWP are of crucial importance, in CTM simulations the source and sink terms contribute considerably to the uncertainties. Many of the chemical species of interest have an atmospheric lifetime in the order of a few days or even hours, thus rapidly erasing the impact of improved initial conditions, whereas other constituents remain in the atmosphere for months or years. The importance of uncertainties in source/sink terms can, to some extent, be accounted for by

Jeremy D. Silver · Jesper H. Christensen · Jørgen Brandt
Department of Environmental Science, Aarhus University, DK-4000 Roskilde, Denmark

Michael Kahnert · Lennart Robertson
Swedish Meteorological and Hydrological Institute, 60176 Norrköping, Sweden

Jeremy D. Silver · Peter J. Rayner
School of Earth Sciences, University of Melbourne, 3010, VIC, Australia. E-mail: jeremy.silver@unimelb.edu.au

inverse modelling to re-estimate parameters such as source terms from the lower and lateral boundaries (van Loon et al 2000; Elbern et al 2007; Constantinescu et al 2007a).

There are different motivations for DA in atmospheric chemistry-transport modelling. These include improved accuracy of forecasts or hindcasts of surface air pollution (Carmichael et al 2008; Curier et al 2012), forecasts of visibility for nautical or aeronautical purposes (Zhang et al 2008), better characterisation of the ozone distribution in the stratosphere to assist in global numerical weather prediction or climate simulations (Semane et al 2009; Milewski and Bourqui 2011), improved quantification of emission sources (Elbern et al 2007; Kopacz et al 2010), and estimating potential benefits of additional measurement equipment (Edwards et al 2009; Timmermans et al 2009).

In this study, we present results for assimilations performed in conjunction with the Danish Eulerian Hemispheric Model (DEHM; Brandt et al 2012), a three-dimensional, off-line CTM. Data assimilation schemes have in previous studies been coupled to DEHM or its predecessor DEOM (Brandt et al 2001): Frydendall et al (2009) implemented a two-dimensional optimal interpolation scheme to assimilate surface O_3 measurements into DEOM, and Silver et al (2013) presented a three-dimensional optimal interpolation algorithm for assimilation of tropospheric column NO_2 densities into the DEHM. In both of these studies the assimilation is restricted to a single species. The two assimilation schemes presented in this article allow for assimilation of multiple species, and direct adjustment of both observed and unobserved species during the assimilation.

The first of the two schemes considered is a three-dimensional variational scheme (3D-var), using the spectral covariance model described in Kahnert (2008), while the second is a variant on the ensemble Kalman filter (EnKF; Evensen 2009). Although the joint presentation and application of the 3D-var and EnKF schemes invites comparison of the two methods, this has been reported in detail in the NWP and CTM literature (Corazza et al 2007; Meng and Zhang 2008; Yang et al 2009; Buehner et al 2010a,b; Singh et al 2011), and is not one of the main focus areas of the present study.

This article has two aims: the first is to describe the implementation of two DA schemes that have been coupled with the DEHM, and the second is to present an initial assessment of the performance of the two schemes when assimilating satellite retrievals of trace gases. In particular, we examine how the assimilations affect model skill in the planetary boundary layer, since the DEHM has been designed and developed to model regional-scale, surface-level air quality. The data assimilated in this study are remote-sensing retrievals from four satellite sensors (OMI, MOPITT, TES, SCIAMACHY) of NO_2, CO, CH_4 and O_3 concentrations. To assess the impact of the assimilation on model skill, model results are compared with observational data from surface monitoring sites, as well as satellite retrievals of trace-gas concentrations. We also present validation results from a pair of observing system simulation experiments (OSSEs).

An OSSE involves two sets of simulations: the first simulation (the "nature" run, or NR) is used to generate "pseudo-observations" (PO), which are then assimilated into the second simulation (the "test" run). The resulting analysis or forecast from the test run can then be compared with the NR. The OSSE framework is often used to assess potential benefits of a new measurement platform, especially if the corresponding physical experiment would be technically difficult or very expensive to conduct (e.g. involving satellites). This approach has recently been used in several chemical DA systems to assess, for example, the potential benefits of additional satellite instrumentation for remote-sensing of aerosols or trace gases (Edwards et al 2009; Timmermans et al 2009), the observability of aerosol sub-types using LIDAR profiles (Kahnert 2009), and how joint assimilation of retrieved O_3 concentrations in a meteorological model can improve flow and temperature fields (Peuch et al 2000; Lahoz et al 2005). The use of the OSSE methodology in the context of air quality modelling is the subject of a recent review article (Timmermans et al 2015).

The OSSE framework is mainly used to address questions about the extra information content available from potential data sources. However this methodology has also been used to test the behaviour of a DA system, since the "true state" of the system is known; it is with this purpose that the OSSE framework will be used in this study. This usage of the OSSE methodology has been employed by others. For example, Errico et al (2007) designed an OSSE to estimate characteristics of the analysis errors for a NWP model (e.g. error variances, consistent biases, correlation length-scales), which are otherwise difficult to study. Privé et al (2013) performed a similar study, using one NWP model for the NR and a second NWP model for the test run; the authors found that the forecast skill of the OSSE was slightly higher than for real data, which they posited was due to insufficient model error in the OSSE. Kleist and Ide (2014) performed an OSSE (also using a separate NWP model to generate the NR) to compare forecast skill in two different DA schemes.

Each of the two OSSEs presented in this study (for verification purposes) uses a different NR. The first NR is a DEHM simulation using different forcing parameters (emissions, meteorology), while the second is from a different CTM (namely version 3 of the Model for Ozone And Related Tracers, or MOZART). Pseudo-observations are generated to mimic the satellite retrievals.

The two assimilation schemes described can either perform univariate or multivariate adjustments. The term *univariate* assimilation means adjusting only one (observed) species at a time. By contrast, *multivariate* assimilation refers to adjusting several species simultaneously (based on observations of one or multiple components). Univariate assimilation can be performed sequentially, meaning that independent adjustments are performed for each assimilated species. All assimilation results presented in this article are based on sequential univariate assimilation of retrieved NO_2, CO, CH_4 and O_3 concentrations.

2 Model and assimilation schemes

2.1 Chemical transport model

The Danish Eulerian Hemispheric Model (DEHM) is a three-dimensional, off-line CTM (Christensen 1997; Frohn et al 2002; Frohn 2004; Brandt et al 2012). The study domain used is a 96×96 horizontal grid on a polar-stereographic projection at roughly 150 km $\times$ 150 km resolution (true at 60° N), with 29 vertical layers using terrain-following σ-coordinates from the surface up to 100 hPa. In these simulations 58 reactive gases and 9 classes of aerosols are modelled. The DEHM's chemistry scheme is similar to that of the European Monitoring and Evaluation Programme (EMEP) model (Simpson et al 2003). A recent overview of the DEHM in terms of the physical and chemical processes represented, as well as parameterisations and numerical methods used, can be found in section 2 of Brandt et al (2012). In the following, we summarise features relevant for this study.

At the lateral and upper boundaries, fixed boundary conditions are imposed at in-flow edge grid-cells, and free boundary conditions are applied for out-flow edge-cells (Frohn et al 2002). For O_3, boundary condition concentrations depend on month, altitude, latitude and longitude, and are based on a climatology of ozonesonde data (Logan 1999). For all other species, the boundary condition concentrations are fixed in space and time, and for most species are set to a near-zero value. Only for CO, CH_4, H_2 (the reactive gases with the longest lifetimes among those modelled) are boundary condition concentrations set to hemispheric background concentrations (respectively 60, 1760 and 580 ppb). The effects of this treatment are discussed in section 3.3.

The meteorological inputs (e.g. wind speed, temperature, pressure) are extracted from MM5 (V3.7) NWP model runs (Grell et al 1995). Two MM5 simulations are run, both with the same horizontal and vertical grid as described above. The first (termed *M1*) involves a single, continuous run with grid-nudging every six hours above the boundary layer using the NCEP FNL Operational Global Analysis data set[1]. Grid-nudging (also termed "analysis nudging" or "four-dimensional data assimilation") is a form of Newtonian relaxation: the three-dimensional fields of horizontal wind, temperature and humidity mixing ratio are "nudged" towards the analysed fields by adding artificial tendency terms to the prognostic governing equations (Stauffer and Seaman 1990); the tendency terms depend on the difference between the modelled state and the analysed field. The second MM5 simulation (termed *M2*) is initialised with the same analyses, then run for 12 hours with nudging, then 60 hours with no nudging, of which the last 12 hours are stored. Both simulations are initialised using the aforementioned NCEP FNL analyses. Details of the various parameterisations used in these MM5 simulations are given in section 3 of Brandt et al (2012).

In order to couple the DEHM with the ensemble assimilation framework described below, a number of small modifications to the model are required. Typical DEHM simulations span individual months, whereas the simulations with assimilation ran for only 3 hours each (i.e. the assimilation is performed every 3 hours). At the beginning of a standard DEHM simulation a large amount of data is read and processed to calculate emission rates for the month-long simulation. The computing time required for this pre-processing accounts for only a small fraction of a month-long simulation, but significantly added to the computing time required for the short simulations. In the version used here, the resulting emission parameters are calculated for each month and stored, so they are simply read from file and the pre-processing step is thus avoided.

A month-long spin-up time is used for the DEHM simulations presented here. Generation of the ensemble initial conditions is described in section 2.4. Details of the emissions data used are given in section 2.2.

2.2 Reference simulations

Three reference CTM simulations are described here. The first (termed *ref*) provides a benchmark for what is achieved without data assimilation, while the other two (termed *NR1* and *NR2*) serve as the nature runs in the OSSEs described in section 3.2. The essential features of these reference simulations are summarised below.

Simulation *ref* used the same configuration of the DEHM model as in the runs with data assimilation. Meteorological parameters are obtained from the *M1* MM5 simulation (see section 2.1). The emissions databases used to calculate source terms at the lower boundary are as follows. Anthropogenic emissions are, over Europe, based on the EMEP emission inventory (Vestreng and Klein 2002), and elsewhere RCP 8.5 emissions are assumed (Lamarque et al 2010), which includes ship emissions. The GEIA inventory is used for biogenic emissions (Benkovitz et al 1996), including emissions of nitrogen oxides (NO_x = NO + NO_2) from lightning and soil. For wildfire emissions, the GFED (V3.1) database is used (van der Werf et al 2010). Aircraft emissions are not accounted for in DEHM. Emissions for each sector are distributed vertically depending on the respective SNAP category.

Simulation *NR1* is a modified version of simulation *ref*, using different inputs for the meteorology and emissions. Meteorological data are extracted from the *M2* MM5 simulation (see section 2.1). Global anthropogenic emissions are taken from the EDGAR database (van Aardenne et al 2001), wildfire emissions are derived from the RETRO inventory (Schultz et al 2007). No higher-resolution emissions are used over Europe. Emissions from shipping are obtained from the global database of Corbett and Koehler (2003). In the vertical, emissions are uniformly distributed between the surface and the top of the mixed

[1] Available at `http://dss.ucar.edu`

layer. Random perturbations are applied to the boundary condition concentrations. To do this, for each species and at each boundary point a random number is drawn from a normal distribution (with mean 0.0 and variance 1.0), which are smoothed using an exponential moving average smoother ($s_1 = z_1, s_k = \alpha z_k + (1-\alpha) z_{k-1}$, for smoothed values s_k and random draws z_k) both in the lateral and vertical dimensions (with $\alpha = 0.25$); these perturbations are shifted and re-scaled to have mean 1.0 and standard deviation 0.1, and boundary condition concentrations at each point are multiplied by the corresponding perturbation term. These are held constant for the one-month simulations, and re-sampled for each month.

Simulation *NR2* was obtained from the ECMWF, courtesy of the MACC project[2]. This simulation is described in detail by Stein (2009), and summarised briefly here. Version 3 of the Model for Ozone And Related Tracers (MOZART3; Horowitz et al 2003; Kinnison et al 2007) is coupled to the ECMWF Integrated Forecast system (IFS) as described by Flemming et al (2009). In the vertical, there are 60 model levels from the surface to 0.1 hPa, and the horizontal domain provides global coverage at $1.125^\circ \times 1.125^\circ$ resolution. The chemical mechanism incorporates 115 species, with most of the chemical reactions described in Kinnison et al (2007) and modifications detailed in Stein (2009). Anthropogenic and natural emissions are based on the RETRO inventory (Schultz et al 2007), for wildfire emissions the GFEDv2 database is used (van der Werf et al 2010), and ship emissions are based on Corbett and Koehler (2003).

A key feature of the *NR2* simulation is 4D-var assimilation of satellite retrievals within the GEMS/MACC framework (Inness et al 2009). In total, retrievals of five species (CO, O_3, CH_4, HCHO, SO_2) from nine satellite sensors are assimilated[3]. Only a small subset of the species modelled are available within the public version of this data-set, namely species that are observed and assimilated (or those with a close chemical coupling to observed species): NO_x, CO, O_3, CH_4, HCHO, SO_2 and SO_4^{2-}.

Section 4 in the Supplementary Material provides horizontal cross-sections of the mean concentrations and the temporal standard deviation, as well as differences in the mean, for these three reference simulations (shown for model levels 5 and 20, and for both simulation periods). It can be seen that the two DEHM simulations, *ref* and *NR1*, are much closer to one another than to *NR2*. Concentrations of NO_x, CO and CH_4 are consistently higher in *NR2*, compared to *ref* or *NR1*. For CO and CH_4, these differences are largest over highly industrialised areas at level 5 and over the poles at level 20, suggesting that uncertainties due to emissions are relatively more important in the lower model levels, whereas uncertainties due to transport and boundary conditions become more important in the upper model levels. For NO_x, these differences are largest over industrialised areas, which fits with "local" nature of NO_x, given its relatively short life-time (c.f. the other species under discussion). For O_3, there are large differences in concentration between *NR2* and the two DEHM runs both over and down-wind of industrialised areas; across most of the domain, O_3 concentrations predicted by *NR2* were higher except over the Arctic at level 5, which is likely due to differences in the meteorological representation of transport of air-masses to the pole (n.b. the MOZART-IFS runs used for *NR2* extend as high as 0.1 hPa, compared to 100 hPa for the MM5-DEHM simulations).

Although the differences in the seasonal averages between *ref* and *NR1* are smaller than between *NR2* and the DEHM simulations, it is the differences in three-hourly means that are used to calculate the background error standard deviations (see section 2.3). The pseudo-observations for the first of the two OSSEs (section 3.2) are also generated from these three-hourly values.

2.3 Three-dimensional variational assimilation

Three-dimensional variational assimilation (3D-var) is a commonly used technique for combining information from observations with a modelled field, weighted by their respective uncertainties (Talagrand 2010). Assuming that background and observational errors are unbiased, that background and observational errors follow a Gaussian distribution, and that background errors are independent of observation errors, then the maximum likelihood estimate can be found by minimising the cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y})^\top \mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}) \quad (1)$$

where $\mathbf{x}$ is a candidate value of the state vector, $\mathbf{x}_b$ is the initial estimate of the state vector (the "background"), $\mathbf{y}$ is the observation vector, H is the observation operator (which projects the modelled field into observation space), $\mathbf{B}$ is the background error covariance matrix, the super-script $^\top$ denotes the transpose of the matrix or column vector, and $\mathbf{R}$ is the observation error covariance matrix (e.g. Kalnay 2003). We note that in general (and in this application) H is non-linear.

The state vector is the three-dimensional concentration field for the species adjusted, arranged in a one-dimensional array. No forcing parameters (e.g. emissions, boundary-value concentrations) are adjusted via the assimilation. In this paper, adjustment of the four species studied (CO, NO_2, CH_4 and O_3) is done by via separate assimilations of the observations for that species. In other words four separate assimilations are performed (one each for CO, NO_2, CH_4 and O_3), each with $n_s = 1$ (where n_s is the number of species adjusted in the assimilation). The case of joint multi-species assimilation (i.e. $n_s > 1$) is the subject of a forthcoming article (Silver et al 2015). The foregoing comments apply for both the 3D-var and the EnKF.

[2] `http://macc-raq.gmes-atmosphere.eu/`

[3] For details, see `http://www.gmes-atmosphere.eu/about/project_structure/global/g_idas/g_idas_2/log_fbov/`

In this work, we have implemented the background error covariance model of Kahnert (2008), which in turn is based on the work of Berre (2000) and Gustafsson et al (2001). For brevity, we will use the same notation as Kahnert (2008), to which the reader is referred for detailed descriptions of each term. This model for **B** assumes that error correlations are homogeneous and isotropic in the horizontal plane (i.e. depending only on the separation distance), and non-separable in the horizontal, vertical and chemical dimensions. A consequence of the assumption of horizontally homogeneous and isotropic correlations is that the analysis increment around a single observation will be circular (n.b. this implies that information is propagated up-wind, down-wind and cross-wind of the observation). This assumption significantly reduces the number of parameters underlying **B** and improves the conditioning (as the inverse must be found). By contrast, the non-separability of covariance parameters in the horizontal, vertical and chemical dimensions is not an assumption, per se, but a lack of an assumption (as in a non-hydrostatic NWP model, which does not rely on the hydrostatic assumption).

The **B** matrix is decomposed as

$$\mathbf{B} = \mathbf{U}^{-1} \cdot \mathbf{U}^{-\dagger} \tag{2}$$

where † denotes the adjoint of the matrix. The matrix **U** is defined in equation 27 of Kahnert (2008). The control variable is

$$\chi = \mathbf{U}(\mathbf{x} - \mathbf{x}_b), \qquad \text{so that} \tag{3}$$

$$J(\chi) = \frac{1}{2}\chi^{\top}\chi + \frac{1}{2}(\mathbf{H}\mathbf{U}^{-1}\chi - \delta\mathbf{y})^{\top}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}^{-1}\chi - \delta\mathbf{y}), \tag{4}$$

where $\delta\mathbf{y} = \mathbf{y} - H(\mathbf{x}_b)$ and **H** is the linearised observation operator (i.e. the first derivative of the non-linear observation operator with respect to the state vector). The change of variables in equations 3 and 4 significantly improves the conditioning for the gradient minimisation. In practice it is not necessary to form any of the above matrices explicitly, and matrix-vector products (such as $\mathbf{U}^{-1}\chi$) are calculated algorithmically.

The starting point for these calculations is a climatological estimate of background covariances, which we we obtain by taking the synchronous differences between simulations *ref* and *NR1* (as in equation 8 Kahnert 2008), using a variant on the "NMC method" for background error estimation (Parrish and Derber 1992). The NMC method, developed for meteorological DA, involves comparing the 24- and 48-hour forecasts. However perturbations to the initial conditions of CTM simulations tend to decay with forecast outlook rather than grow (Constantinescu et al 2007b; Wu et al 2008), entailing that background errors from the original NMC method would be underestimated.

The modified NMC method used here follows the example of Kahnert (2008), in which three different methods of estimating statistics of the **B** matrix are compared. It was found that the modified NMC method provided an appropriate balance between large-scale and fine-scale horizontal and chemical cross-correlations. In the study of Kahnert (2008), the only difference between the two simulations used in the modified NMC is the NWP spin-up time, which one may expect would lead to error correlations representative of the features of the meteorological differences alone. However it was shown that the spectral properties of the synchronous differences in the individual species modelled are significantly different from one another (Kahnert 2008). It was also noted that the species modelled showed distinctly different statistical properties from the the meteorological variables, however these results are not shown. A similar approach is adopted in Silver et al (2013), by forcing the parallel simulations with output from two different NWP models.

Three-hourly mean concentrations, rather than instantaneous concentration fields, spanning the whole of 2008 for simulations *ref* and *NR1* are compared in this manner. As noted above, these two simulations differ in terms of their meteorology, emission rates, emission injection heights and boundary concentrations, hence the background errors account for the cumulative uncertainties from these factors.

There are three minor differences between our 3D-var configuration and that of Kahnert (2008). First, the observation increments used are interpolated from model space at observation time, rather than at analysis time – this is known as First Guess at Appropriate Time (FGAT) 3D-var. This has been shown to give a slight improvement over synoptic 3D-var (Lorenc and Rawlins 2005). This improvement is essentially due to a more informative innovation vector (i.e. $\mathbf{y} - \mathbf{H}\mathbf{x}_b$) than in standard 3D-var (Talagrand 2010); in other words, this method minimises errors introduced by timing inconsistencies between the observations and model field projected into observation space.

Second, the covariance parameters, as estimated using the framework outlined in Kahnert (2008), are not used directly. This is because we found that the resulting correlations did not decay to zero with increasing distance in all cases. For medium- to long-lived species spurious non-zero long-range correlations are evident in the estimated parameters. For this reason, the initial covariance parameters are converted from spectral space to physical space, replaced by fitted Gaussian decay functions, and then transformed back to spectral space. The updated spectral covariance parameters $\hat{\mathbf{D}}(k^*)$ are not necessarily positive-definite (PD) for each wave-number k^*, as should be the case. To enforce this condition, the non-PD $\hat{\mathbf{D}}(k^*)$ are replaced by the nearest PD matrix with the same diagonal entries using the method of Higham (2002), as implemented in Bates and Maechler (2013); in this context, the "nearest" PD matrix is the PD matrix for which the distance between the original and substitute matrices minimises the Frobenius norm (i.e. the square root of the sum of squares of all entries of the matrix). The resulting correlations in physical space are a good approximation for the fitted Gaussian decay curves mentioned above, and this procedure ensured that the resulting covariance matrix is PD and that diagonal elements of $\hat{\mathbf{D}}(k^*)$ are positive. The correlation between two variables (i.e.

of different species and different vertical levels) at zero horizontal distance is preserved during these corrections. Details of the transformation between the physical-space and spectral-space covariance parameters are given in section 3 of the Supplementary Material.

The long-range correlations that did not decay to zero at long separation distances appear to be related to boundary condition in-flow concentrations, most notably for long-lived species. This issue was much more pronounced for these species in a previous version of *NR1*, which used fixed and higher values of boundary condition concentrations, rather than the spatially-varying perturbations applied here (results not shown).

Third, we applied variational quality control to down-weight observations that deviated from the corresponding modelled quantities by several multiples of the observation error (Robertson and Langner 2000; Andersson and Thépaut 2010). This involves a reformulation of the cost function (equation 1) such that normalised observation errors (i.e. scaled by the inverse of observation error covariance matrix $\mathbf{R}$) are assumed to be drawn either from a normal distribution (with zero mean and variance) or from a uniform distribution (with mean 0 and width W). The latter case is intended to allow for "gross errors". This is because the Gaussian cost function is very sensitive to strong outliers (potentially leading to a very strong influence in the increment). Variational quality control reduces the need to make binary decisions about whether to assimilate a given observation, since distant outliers will not contaminate the analysis. This method has been used in cutting-edge variational data assimilation schemes for over a decade (Rabier et al 2000). We use an *a priori* probability of gross errors of 10% and a flat ("box-car") distribution for the gross errors with half-width of 5.0 standard deviations.

2.4 Ensemble assimilation scheme

Ensemble Transform Kalman Filter The second assimilation scheme implemented in this study is a local, asynchronous ensemble transform Kalman filter, which is a variant on the ensemble Kalman filter (EnKF; Evensen 2009). The EnKF solves the analysis equation

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^\top \left(\mathbf{H}\mathbf{B}\mathbf{H}^\top + \mathbf{R}\right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}_b) \tag{5}$$

where $\mathbf{x}_a$ is the analysis (i.e. the updated state vector). As mentioned in section 2.3, the state vector in this study is comprised of the three-dimensional concentration field of n_s species (where n_s is the number of species to be adjusted), rearranged as a column-vector. The background covariances are estimated as sample covariances in an ensemble of simulations $\mathbf{X}_b$, where the i^{th} column contains the background state $\mathbf{x}_b^i$, of the i^{th} ensemble member ($i = 1, \ldots, n_e$). This is given by

$$\mathbf{B} = \mathbf{A}_b^\top \mathbf{A}_b / (n_e - 1) \tag{6}$$

where the deviations from the mean ensemble vector $\overline{\mathbf{x}}_b$ are termed background ensemble anomalies, and defined as

$$\mathbf{A}_b = \mathbf{X}_b - \overline{\mathbf{x}}_b. \tag{7}$$

However the background error covariance need not be evaluated explicitly, and the largest system to be solved is of dimension $\min(n_e, n_o)$, where n_e and n_o are, respectively, the ensemble size and the number of observations (Evensen 2003).

There are a number of different formulations of the EnKF, which can be broadly classified as stochastic or deterministic filters (Evensen 2009). Stochastic filters treat observations as random variables, and each ensemble member is updated using perturbed observations with variances given by $\mathbf{R}$; deterministic filters avoid the need for perturbed observations by updating the ensemble mean and ensemble anomalies separately. The ensemble mean is updated using the standard EnKF analysis equations, and the following linear transformation is applied to the ensemble anomalies

$$\mathbf{A}_a = \mathbf{A}_b \mathbf{T}$$

where $\mathbf{T}$ is a square matrix of dimension n_e found such that the required variance properties of the analysis are satisfied (Bishop et al 2001). The ensemble transform Kalman filter is a deterministic filter, and has the advantage that the largest matrices that need to be solved are of dimension n_e, which in most applications is much smaller than n_o, thus resulting in significant computational savings (Tippett et al 2003).

Lawson and Hansen (2004) compared stochastic and deterministic formulations of the EnKF in several models with varying degrees of complexity, non-linearity and non-Gaussian errors. The deterministic filters are shown to maintain non-Gaussian moments, which are otherwise smoothed out by the stochastic filter. The accumulation of non-Gaussian structures can potentially lead to the occurrence of distant outliers among the ensemble members. However the problem of outliers can be addressed by the application of a random orthogonal rotation in the update step for ensemble anomalies (Sakov and Oke 2008). This solution, which also improves the performance of the filter and ensures that the row means of $\mathbf{A}_a$ are zero, is applied here.

Timing In order to account for discrepancies between observation time and assimilation time, we used so-called "asynchronous ensemble Kalman filter" (AEnKF) correction, which involves interpolating from model space to observation space at observation time rather than analysis time. This has been shown to improve performance over short assimilation windows (Hunt et al 2004). In experiments with a low-dimensional model it has been shown to outperform four-dimensional variational (4D-var) assimilation for short assimilation windows, since the AEnKF incorporates flow-dependent background errors, whereas these spin-up implicitly in 4D-var as the assimilation window increases (Fertig et al 2007). However for longer assimilation windows, the perfect-model and linearity assumptions underpinning the AEnKF lead to performance degradation and even filter divergence (Fertig et al 2007). We note that the correction required for the AEnKF is identical to the FGAT correction in the 3D-var case. The importance of such temporal treatment is mainly due to observations of diurnally-varying species such as NO_2 (section 3.1.1).

Perturbations In the EnKF, the ensemble spread represents the uncertainty of the corresponding variables. At each assimilation step the ensemble spread tends to contract, and over time risks becoming "overconfident" and thus "ignoring" information from new observations. This phenomenon, known as filter divergence (or ensemble collapse), arises because the EnKF does not explicitly account for model error. This is most commonly avoided by multiplying the background covariance matrix by an inflation factor (typically between 1.01 and 1.1), thus assuming that model errors proportional to background errors (Hamill and Whitaker 2005). Alternatively, one can explicitly add random perturbations to the ensemble to maintain ensemble spread, analogous to the way model error is treated in the Kalman filter (Evensen 2009). While multiplicative model error has been shown to be effective in relatively simple model systems, for highly multidimensional geophysical models, it has proven to be less successful than additive model error at improving the analysis and maintaining ensemble spread (Whitaker et al 2008). This is because in high-dimensional geophysical models, some parts of the state vector will be more strongly constrained by observations than others.

In this study, filter divergence is prevented by adding random Gaussian perturbations to the ensemble. The model error covariance matrix is assumed to be proportional to the climatological **B** matrix used in the 3D-var system, similar to the approach of Houtekamer et al (2005). These random perturbations are generated as follows. Let w be a vector of independent, normally distributed random variables with zero mean and unit variance. Let $\mathbf{C}$ be a symmetric positive definite matrix satisfying $\mathbf{C} = \mathbf{V}\mathbf{V}^{\dagger}$ for some $\mathbf{V}$; then $\mathbf{V}w$ will be normally distributed with zero mean and covariance matrix $\mathbf{C}$ (e.g. Gut 2009, pp. 132). In this study we set $\mathbf{V} := \mathbf{U}^{-1}$ from equation 31 of Kahnert (2008). The sampling is performed by computing the product $\mathbf{V}w$.

Although w is treated as a vector its entries correspond to particular eigenvalue-wavevector combinations. Further, to ensure that all elements of $\mathbf{V}w$ are real, w must satisfy

$$w_{I,-m,-n} = w^{*}_{I,m,n}$$

where I is the index of the eigenvector, $m = -K_x, ..., K_x$ and $n = -K_y, ..., K_y$ are the indices of the Fourier coefficients, and * denotes the complex conjugate. Beyond this constraint, entries of w are generated by independent draws from a standard normal distribution. Perturbations to generate the ensemble initial conditions are also generated in this manner.

The justification for this approach, as opposed to simpler alternative methods for generating additive model error (e.g. Evensen 2009, pp. 157-163), is that the perturbations capture the appropriate scales of variation in not only the horizontal plane, but also in the vertical and chemical dimensions, thus allowing for joint multi-species assimilation. Furthermore, it alleviates the need for manual calibration when assimilating different species. The model errors are treated as correlated in time, assuming a first-order auto-regressive model:

$$q_k = \alpha q_{k-1} + \sqrt{1-\alpha^2}\hat{q}_k \tag{8}$$

where q_k is the model error term for time t_k, of which $\hat{q}_k$ is generated independently, and α is the first-order autocovariance. A value of $\alpha = 0.5$ is used.

Model error is added to the analysis rather than to the background, as would be consistent with the classical Kalman filter formulation. This is primarily motivated by the use of the AEnKF correction, which means that the observation operator is applied at observation time, rather than at analysis time. To add model error to the background would substantially weaken correlation between the ensemble innovations and the background. The choice of whether to add model error to the background or the analysis is discussed by Houtekamer and Mitchell (2005), who write that the model error term represents several factors including "(i) errors in the forward interpolation operator, (ii) errors in the specification of the statistics of the observations, and (iii) errors due to the parametrisation of unresolved dynamical and physical processes". They note that the model error could instead be added to the analysis, rather than the background, such that the analysis error can also account for the "system error", which includes errors due to the non-optimality of the assimilation; such factors would include sampling errors due to the finite ensemble size, the use of localisation constraints to correct for this, and deviations from the underlying assumptions (e.g. Gaussian error statistics, correct specification of error covariances, bias). Applying model error to the analysis is not without precedent, for example Houtekamer et al (2009, pp. 2127) report that model error has been added to the analysis, rather than the background, in the Canadian Meteorological Centre's operational NWP EnKF system since 2005.

The procedure described above amounts to a means of simulating perturbations with the variance identical to the climatological background error variance. We instead chose to employ an adaptive scaling in order to keep $\bar{\sigma}^{e}_{i_z,i_s,j}$ (the mean ensemble

standard deviation at level i_z for species i_s at time-point j) approximately equal to $\bar{\sigma}^c_{i_z,i_s}$, the mean climatological standard deviation of level i_z for species i_s. Let $\bar{\sigma}^{e_b}_{i_z,i_s,j}$ and $\bar{\sigma}^{e_a}_{i_z,i_s,j}$ be the mean ensemble standard deviations for the background and analysis, respectively, at the j^{th} assimilation. Then the model error perturbations generated as described above are scaled in order to ensure that the mean standard deviation of the analysis ensemble is given by

$$\hat{\sigma}^{e_a}_{i_z,i_s,j} = \bar{\sigma}^c_{i_z,i_s} + \max(\bar{\sigma}^c_{i_z,i_s} - \bar{\sigma}^{e_a}_{i_z,i_s,j-1}, 0)/2. \tag{9}$$

The term $(\bar{\sigma}^c_{i_z,i_s} - \bar{\sigma}^{e_a}_{i_z,i_s,j-1})$ measures the decay in ensemble spread during the three-hour assimilation window (most substantial for shorter-lived species such as NO_2); this is divided by 2 so that the mean ensemble variance over the ensuing integration period would be $\bar{\sigma}^c_{i_z,i_s}$ assuming a linear decay in ensemble spread. Perturbations to the initial ensemble are scaled to have a mean ensemble standard deviation of $\sigma^c_{i_z,i_s}$.

Additive perturbations can lead to negative estimates of concentration, especially for highly variable species. Negative concentration estimates can also result from the assimilation. This could be prevented by log-transforming concentrations in the assimilation, and adding model error perturbations to log-transformed concentrations, yet this may cause the ensemble to diverge exponentially in regions with few observations. A simpler approach is taken here, namely to set negative estimates to a small non-zero value (10^{-6} multiplied by the mean concentration in the background for the corresponding level and species). This may, however, create a positive bias and thereby degrade the analysis. The same method is used to correct for negative concentrations in the analysis from the 3D-var. In practice, this is required for NO_2 but not the other species, due to the high spatial variability of NO_2 and very low global background concentrations.

In order to maintain some level of ensemble spread in regions downwind of in-flow edge-cells, normally distributed perturbation factors with mean 1 and standard deviation 0.1 are applied to the boundary condition concentrations of each ensemble member. These perturbation factors are held constant for a given ensemble member for the entire boundary for the entire simulation period, and each species had a different perturbation term. This only affected species with a non-zero in-flow concentration (see section 2.1).

Localisation An ensemble of 16 members is used. Although this number is rather small compared to many other applications, such small ensemble sizes have been found sufficient to provide satisfactory results in other CTM applications (Hanea et al 2004; Curier et al 2012).

The finite sample size of the ensemble results in "spurious" non-zero correlations (e.g. between distant grid-points). Furthermore, the analysis increment is restricted to the span of the ensemble members (e.g. see equations 5, 6), and for high-dimensional geophysical models, the ensemble size is typically far less than the number of effective degrees of freedom of the model state. If parametric assumptions are applied to damp long-range correlations, both of these issues are alleviated (Sakov and Bertino 2010). This is demonstrated by Ott et al (2004) and Hunt et al (2007) in what they termed the local ensemble transform Kalman filter (LETKF); we are unable to use the form of the LETKF outlined by Hunt et al (2007) due to implications (described below) from the use of the AEnKF.

We prevented spurious long-range correlations in the implied covariances by applying the assimilation grid-point by grid-point, known as "local analysis", with the approach described in equations 11-15 of Sakov and Bertino (2010). This approach effectively down-weights "distant" observations by inflating the associated observation errors. Using the AEnKF correction to account for timing discrepancies, the projection from observation space to model space takes place at observation time, so it is not possible to calculate explicitly the product $\overset{i}{\mathbf{H}}\overset{i}{\mathbf{A}}$ (as defined in Sakov and Bertino 2010), where the accent "i" denotes localised versions (of variables $\mathbf{A}$, $\mathbf{H}$) used in the update for grid-point i, as this would have required storing the entire ensemble at every model time-step. As only the product $\mathbf{HA}$ is available, the localised equivalent $(\overset{i}{\mathbf{HA}})$ had to be calculated.

The matrix $(\overset{i}{\mathbf{HA}})$ has dimension $n_o \times n_e$. Let us define the element

$$(\overset{i}{\mathbf{HA}})_{j,k} = d(i,j)(\mathbf{HA})_{j,k} \tag{10}$$

where j is the observation index, k is the ensemble index, and i is the grid-point index (Here, the term "grid-point" refers to the value of a single species, at some given point in the three-dimensional domain). The term $d(i,j)$ is a localisation weight between grid-point i and observation j. We next address the question of how to such construct localisation weights.

If the model domain were purely two-dimensional, these weights could be based on distance alone. However in this case, we must also take account of the vertical and chemical dimensions; this is especially relevant given the use of satellite retrievals, which provide information throughout a vertical column.

The localisation weight $d(i,j)$ between grid-point i and observation j should fulfill certain criteria. First, weights should fall in the interval 0 to 1. Second, the weight $d(i,j)$ should be a decreasing function of $h_{i,j}$, the horizontal distance between grid-point i and observation j. Third, $d(i,j)$ should reflect the strength of the relationship between grid-point i and observation j; one means of doing so is to sum, over all grid-points i', the relationship strength between grid-points i and i', multiplied by the relationship strength between grid-point i' and observation j. To make this more precise, the strength of the relationship between grid-points i and i' can be quantified based on the correlation $C_{h_{i,i'};i_z,i_s;i'_z,i'_s}$, where $h_{i,i'}$ is the horizontal distance between

grid-points i and i', i_z and i_s denote indices in the vertical and chemical dimensions, respectively, for grid-point i. One can then use the linearised observation operator $\mathbf{H}_{j,i}$ to quantify the relationship between grid-point i and observation j.

In light of these considerations, the distance $d(i,j)$ is defined as

$$\hat{d}(j,i) = \sum_{i'} |C_{0;i_z,i_s;i'_z,i'_s}| f(h_{i',j}/L_{i_z,i_s;i'_z,i'_s}) |\mathbf{H}_{j,i'}| \tag{11}$$

$$d(j,i) = \hat{d}(j,i) / \max_{i'}(\hat{d}(j,i')) \tag{12}$$

where $L_{i_z,i_s;i'_z,i'_s}$ denotes the fitted length-scale (described below) between vertical levels i_z, i'_z and chemical species i_s, i'_s. The operator $|\cdot|$ denotes the absolute value. Rather than using the correlation $|C_{h_{i,i'};i_z,i_s;i'_z,i'_s}|$ directly, it is modelled as the product of an intercept (representing the zero-distance correlation) and a decaying distance-dependent correlation function. The intercept term, $C_{0;i_z,i_s;i'_z,i'_s}$, denotes the zero-distance correlation between vertical levels and chemical species for grid-points i, i', as derived from the climatological background covariance decomposition used in the 3D-var. The term $h_{i,j}$ denotes the horizontal distance between grid-point i and observation j. A Gaussian distance function is used, as given in equation 4.12 of Gaspari and Cohn (1999)

$$f(r) = \exp\left(-r^2/2\right). \tag{13}$$

Note the use of absolute value signs around the correlation and the observation operator terms in equation 11. This is because correlations between species can be negative, and the averaging kernels used in the observation operators (section 3.1) can also be negative. The choice of scaling in equation 12 ensures that the maximum distance value of 1.0 is obtained by at some point in the column. In the hypothetical case of vertically constant averaging kernels, then the distance between the observation and grid-points underpinning this column is uniformly 1.0.

The derived correlations did not necessarily decay to zero at very large separation distances (e.g. PAN in the left panel of in figure 1). For this reason, we fitted parametric functions to each of the correlation-decay curves, and the fitted functions are used in equation 11. To model the decay in correlation, we compared six different correlation functions (including the popular fifth-order polynomial of equation 4.10 of Gaspari and Cohn 1999), and the Gaussian model proved to have the lowest MSE of the models considered. Rather than using the derived length-scale from the background error covariances (as calculated by equation 50 of Kahnert 2008), length-scale parameters are fitted to the derived correlations at different distances to ensure that the parameter values are consistent with the parametric localisation function; this corresponds, for example, with using correlations shown in the right panel of figure 1 rather than the left panel of figure 1.

3 Verification

In situ and remotely sensed observations from a range of different measuring platforms are included in this study, and these are described below. For the real-data experiments, as opposed to the OSSE simulations, satellite retrievals are assimilated and in situ measurements used for verification; verification with satellite retrievals is considered (section 3.3).

3.1 Data sources

The remote-sensing retrievals used in this study are chosen on the basis that they are freely available, valid for the study period (the year 2008) and that averaging kernels are supplied. These products differed substantially in their abundance and spatial coverage; they are illustrated in figure 2. Table 1 summarises key details of the remote-sensing products described below.

3.1.1 OMI tropospheric column NO_2

The Ozone Monitoring Instrument (OMI) sits aboard the sun-synchronous NASA satellite Aura, launched July, 2004. Aura crosses the equator near 13:45h local time. The OMI measures solar back-scatter in the visible and ultraviolet spectra. Pixels are 13 km $\times$ 24 km at nadir. Radiance measurements from the sensor are processed to obtain total or partial column concentrations for a range of atmospheric constituents including O_3, HCHO, NO_2, SO_2 and aerosols.

In this study, we used version 2.0.1 of the Dutch OMI NO_2 (DOMINO) retrievals of the tropospheric column NO_2 densities, made freely available through the web-site `www.temis.nl`. The retrieval scheme is described in Boersma et al (2002, 2007). Pixels are excluded if the cloud radiance fraction is greater than 50%, if the surface albedo is greater than 0.3, or if the tropospheric column flag is raised (e.g. due to row anomalies), as recommended by Boersma et al (2011). The DOMINO tropospheric column densities have been extensively validated and show good agreement with ground-based measurements (typically with Pearson correlation coefficient $R^2 > 0.6$), albeit with a slight underestimation of about 15-30% (see Celarier et al 2008, and references therein). Retrievals are only available during day-time over-passes; the spatial distribution is illustrated in figure 2 g, h, i.

3.1.2 TES partial column CH_4, CO, O_3

The Tropospheric Emission Spectrometer (TES), which like OMI, is mounted on the satellite Aura. The TES is a high-resolution infrared-imaging Fourier transform spectrometer, which performs global surveys and also makes "special observations" of, for example, volcano eruptions and biomass burning. Nadir pixels have a footprint of roughly 5 km $\times$ 8 km. Radiances are measured in the range 3.2-15.4 μ m, and retrievals have been derived from a large number of trace gases including HNO_3, CH_4, CO, CO_2, NH_3 and O_3.

In this study, we used version 5 (F06_09) nadir retrievals of CO, CH_4 and O_3 made available by NASA. Although these are provided as mixing ratios at 67 pressure levels (from 1212 to 0.1 hPa), and we assimilate only one layer: for CO and O_3 the 825 hPa layer is used while for CH_4 the 620 hPa layer is used. This selection is a compromise between choosing a level that is lower in the atmosphere (for better sensitivity to boundary layer concentrations), and a layer with lower retrieval error (n.b. the lowest retrieved layers are typically assigned the highest retrieval error). Retrieval algorithms are detailed in Osterman et al (2004). We followed the quality control guidelines for species-dependent data exclusion as specified in section 6.1.1.5 of Herman et al (2013). Compared to in situ measurements with sondes, TES O_3 profiles show a positive bias of about 15% in the vertical range 1000-100 hPa. Measurements from aircraft-mounted monitors match the TES CO profiles (in the range 700-200 hPa) within 5-10%, which is lower than the specified uncertainty of 10-20% (Luo et al 2007). Retrieved TES CH_4 columns show a positive bias of between 1-4%, with a dependence on latitude (Herman et al 2012). Retrievals are available during both day-time and night-time over-passes; the spatial distribution is illustrated in figure 2 j, k, l.

3.1.3 MOPITT partial column CO

The Measurements of Pollution in The Troposphere (MOPITT) sensor is mounted on the NASA satellite Terra, launched December 1999. Terra follows a sun-synchronous orbit with an equatorial crossing of 10:30h local time. Nadir pixels are roughly 22 km $\times$ 22 km. The MOPITT uses gas correlation spectroscopy and measures radiances in the near-infrared (NIR, 2.3 μ m) and thermal-infrared (TIR, 4.7 μ m) bands. From these radiances, total and partial column densities of CO and CH_4 are retrieved.

In this study, we used partial column densities of CO from version 5 based on the TIR radiances; retrievals are provided by NASA/NCAR[4]. In particular, we only used the 900 hPa layer of the retrieved column. At the time we downloaded and processed these data, the TIR product was recommended as more reliable than the NIR or NIR/TIR retrievals (Deeter 2011). Details of the retrieval scheme can be found in Deeter et al (2013) and references therein. We excluded pixels north of 65° N latitude, due to difficulties in cloud detection (as recommended by NCAR 2011). The V5 TIR surface retrievals show very close agreement with in situ measurements from aircraft profiles (bias $\sim 1.0\%$, $R^2 \geq 0.97$), however the bias appears to vary slightly over time and with latitude (Deeter et al 2013). In another study, V3 retrievals are found to have a positive bias of around 7-14% compared with in situ profiles (Emmons et al 2004). The TIR product includes both day and night retrievals, whereas the NIR or NIR/TIR products include only data for daytime scenes. The spatial distribution is illustrated in figure 2 a, b, c.

3.1.4 SCIAMACHY total column CH_4 and tropospheric column NO_2

The SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) instrument was launched aboard European Space Agency's ENVISAT satellite in March 2002. ENVISAT, which is no longer operational, had a sun-synchronous orbit with equatorial crossings at 10:00h local time. The pixel footprint is roughly 60 km $\times$ 30 km. Radiances recorded by SCIAMACHY, measured in the spectral region 214 nm to 1750 nm, have been used to retrieve concentrations of a range of species including CO, CO_2, CH_4, O_3, NO_2, BrO and SO_2.

We used the WFM-DOAS version 2.0.2 total-column CH_4 retrievals from the IUP/IFE, University of Bremen. For details of the retrieval algorithm see Schneising et al (2011) and references therein. Pixels flagged as having a poor quality CH_4 retrieval or with evidence of cloud contamination are excluded from processing. These retrievals are only available over land during day-time; the spatial distribution is illustrated in figure 2 d, e, f. Comparison with ground-based measurements shows an agreement within 20% (Kelder et al 2004).

We also used the TEMIS version 2.3 tropospheric column NO_2 retrievals, made freely available from the `www.temis.nl` web portal. Full details of the retrieval algorithm are available in Boersma et al (2004) and van der A et al (2010). When compared with an aircraft campaign (Heue et al 2005), these retrievals showed good agreement (regression slope of 0.93 ± 0.06). Ground-based retrievals from sites in East Asia showed a bias of around $-5 \pm 14\%$ (Irie et al 2012).

In this study, pixels are excluded if a) the tropospheric column quality flag is raised, b) they are not of nadir view, c) the cloud fraction or the cloud radiance fraction exceeded 0.5, or d) the surface albedo is greater than 0.3. This product is used for validation only, and thus no observation data thinning is applied (see section 3.1.5). These retrievals are available for day-time scenes; the spatial distribution is illustrated in figure 2 l, m, o.

[4] `https://www2.acd.ucar.edu/mopitt`

3.1.5 Observation processing

In this study, we only assimilated level-2 satellite products (derived geophysical quantities), rather than level-1 data (calibrated and collocated radiances). Retrieving level-2 information from level-1 data is generally an ill-posed problem, and typically can only be solved by the use of an *a priori* estimate of the atmospheric state. This has been shown to result in correlated observation errors (Kunzi et al 2011). The choice of assimilating retrivals rather than radiances was mainly due to practical reasons (e.g. expertise required, limited time), however we note that this choice is the norm in constituent DA (Lahoz et al 2007, pp. 5751). Furthermore, given the technical challenges involved (e.g. detecting and handling biases in the radiances, appropriate choice of micro-windows, advance radiative-transfer modelling), a recent review article does not recommend direct assimilation of radiances in chemical DA applications (Bocquet et al 2015, pp. 5341).

In principle, the error statistics used in the assimilation should be as accurate as possible. Here, like in many other assimilation applications, we have assumed a diagonal observation error covariance matrix $\mathbf{R}$, although this will not hold in practice due to the use of level-2 retrievals. It has been shown that the use of diagonal $\mathbf{R}$ matrix in the presence of correlated observation errors effectively leads to a loss of information and can degrade the analysis (Stewart et al 2008). Ideally, observation error correlations should be represented explicitly, however we have opted for a simpler approach. In the horizontal plane, observations are "thinned" to a lower density, thereby reducing the effects of correlated observation errors. After applying the quality control criteria to each of the satellite products, observations are sequentially discarded until no two retrievals are within retrieval 50 km of one another within a 30-minute interval. We could, alternatively, have created "super-observations", assimilating the average of a group of nearby observations, as in Miyazaki et al (2012), which would have had the added benefit of reducing representativeness errors. A recent comparison of data-thinning methods in NWP showed that there the super-observation approach had only marginal benefits above random thinning (Lazarus et al 2010); we note that it is difficult to generalise this finding to other applications (e.g. chemical data assimilation). The choice of thinning is motivated by the simplicity of the method, and that it would not need to be reconfigured for each different satellite data-set (each with its own vertical structure and set of averaging kernels).

Observation errors are estimated as the sum of two components: measurement error and representativity error. In principle, the observation errors should also account for errors in the observation operator, however this is not taken into account in this study. The measurement error for a given data-point is estimated as the median reported retrieval error for that data-type, for that month of the year, and for the appropriate latitude band. The choice of using median errors (as opposed to the errors provided with the product) is that for some of the satellite products, such as the OMI NO_2 tropsopheric columns, the observation error is found to be nearly proportional to the value of the observation itself; while this may seem appropriate from a measurement point of view, it can cause problems in an assimilation since observations that are assigned extremely small errors can have an unduly strong impact on the analysis. The representativity error for a given data point is estimated as the median standard deviation for the corresponding satellite product product within any DEHM grid-cell within any three-hour period, again segregated by month and by latitude band.

In the case of the MOPITT and TES data, we chose to assimilate retrieved partial-column mixing ratios at a single layer (namely, that closest to the surface), even though mixing ratios are retrieved throughout the vertical column. These vertical columns are provided with the corresponding error covariance matrices for the retrieval, so in principle it would have been possible to account for observation error correlations within the vertical profile. However, the assimilation schemes considered here cannot at present handle correlated observation errors. Such an extension may be considered in future, but went beyond the scope of the present study.

The sensitivity of the satellite retrievals varies throughout the vertical column, and also contains information from an *a priori* estimate of the concentrations throughout the column. In order to compare the retrieval with its modelled counterpart (properly accounting for the *a priori* and the issue of variable sensitivity), one must apply the appropriate averaging kernel to the retrieved column (Eskes and Boersma 2003). Assuming the averaging kernels are estimated accurately, this process smooths information from the model in the same way that information from different layers of the real atmosphere is smoothed in the retrievals. We note, however, that the pressure levels of model and the averaging kernels are, in most cases, different. Let us call the pressure levels of the averaging kernel the "target levels" for short. As such, it is necessary to transform the DEHM column to target levels.

In cases when the retrieved data are defined as a mixing ratio (or log-mixing ratio), this is done by calculating, for each target level, a weighted average over the DEHM column, weighted by the proportion of the target level covered by the DEHM level. In cases when the retrieved data are defined as density (e.g. with units of 10^{15} molecules per cm^2 over a defined vertical region), then the DEHM column defined at the target levels is calculated as a weighted sum of the densities of the individual layers, weighted by the proportion of the DEHM layer that overlapped the target layer.

The observation operator consisted of three parts. First, at each level the model field is bi-linearly interpolated from the four surrounding grid-points. Second, the resulting column is transformed vertically to the pressure levels at which the averaging kernels are defined, in the manner described above. Third, the averaging kernels provided in the satellite product are applied to ensure consistency in the vertical resolution of the modelled value and the retrieved quantity (Eskes and Boersma 2003). In the case of the TES and MOPITT partial column retrievals, the averaging kernels are applied to the logarithm of the volume mixing

ratio. Hence for these species the observation operator is non-linear, and the linearised observation operator for these retrievals is calculated as the derivative with respect to the components of the state vector (i.e. the concentration field). The same code is used for processing observations in both assimilation systems.

For some satellite products (e.g. the SCIAMACHY total-column CH_4) the layers of the retrieved column or the layers of the averaging kernels are defined higher than the DEHM's top vertical layer (at 100 hPa). At these levels, the interpolated column is set to be the same as the *a priori* column. The averaging kernels are generally applied in the observation operator in the following form:

$$\hat{\mathbf{x}} = \mathbf{x}_a + \hat{\mathbf{A}}(\mathbf{x}_m - \mathbf{x}_a)$$

where $\mathbf{x}_a$ is the *a priori* column, $\mathbf{x}_m$ is the modelled column, $\hat{\mathbf{A}}$ is the averaging kernel matrix and $\hat{\mathbf{x}}$ is the model-equivalent of the retrieved column. Thus above the model top, the modelled column does not contribute to $\hat{\mathbf{x}}$, and in particular, to the gradient of the cost function in the 3D-var.

3.1.6 Surface monitoring data

Hourly measurements of NO_x, O_3 and CO for the year 2008 are retrieved for European air-quality monitoring stations in the AirBase network[5]. Stations were selected for verification purposes if classified as "rural". For NO_x and O_3, the definition of "rural" is based on the objective classification developed by Joly and Peuch (2012) and implemented by Malherbe et al (2013). The classification provided by Malherbe et al (2013) assigns each station a category, ranging from 1 (rural background) to 10 (suburban traffic); we included stations in classifications 1 and 2. Given the very coarse horizontal resolution of the model, it can only be expected to describe rural background concentrations. Malherbe et al (2013) classify stations separately for individual species; for example, a station can be considered "rural" for O_3 but not for NO_x. For CO, no objective classification is available, the objective classification for O_3 is used in its place. For CH_4, there is only one rural measurement in AirBase site providing data to AirBase during the study period. Consequently, we downloaded measurements from the Global Atmosphere Watch monitoring network[6]. All such sites are deemed rural. As a second selection criterion, we only considered stations with less than 50% missing data for the given period.

Model data is interpolated horizontally with bi-linear interpolation from the four surrounding grid-points. For vertical interpolation, the reported site altitude is compared to the altitude given by interpolating the site coordinates on a global 5 km $\times$ 5 km digital elevation model (DEM) – let us call the difference $\mathrm{d}z$ (positive when the reported altitude is higher than the DEM altitude). Sites with an absolute value of $\mathrm{d}z$ more than 500m are excluded, as this is evidence of complex topography, which the model cannot be expected to describe with any skill due to its very coarse horizontal resolution. In cases where $\mathrm{d}z$ is higher than 100m, the model is interpolated from the layer corresponding to a height of $\mathrm{d}z$ metres above the surface; otherwise surface values are taken.

Since modelled data are stored as three-hourly averages, measured data are also averaged over corresponding three-hourly windows. In all verification, patterns of missing observation data are replicated in the modelled data to ensure an equitable comparison between the two data-sets.

3.2 Pseudo-observations and observing system experiments

The literature offers some suggestions for the design of OSSEs. Privé et al (2013) recommended (a) that different models should be used for the nature and test runs (n.b. as defined in the Introduction, the "test run" involves assimilation of pseudo-observations from the NR), (b) that the NR should be at a higher resolution that of the test run, (c) that the NR should accurately represent the processes that the test run is designed to capture, and (d) that pseudo-observations should be generated with a realistic spatio-temporal distribution. Halliwell Jr et al (2014) write that model errors can be attributed to errors in (a) the initial conditions, (b) the parameterisation of chemical and physical processes, (c) the numerical implementation, and (d) the applied forcing at the lower, upper and lateral boundaries; they note that error growth between the model state and the true atmosphere is due to contributions from each of these four factors, and that the NR should thus differ from forecast model (used in the assimilation and control runs) in ways that account for all four factors.

We conducted two OSSEs, which we shall refer to OSSE 1 and OSSE 2 as they rely on *NR1* and *NR2* (section 2.2), respectively, for their NRs. Recall that *NR1* is a DEHM simulation with various different input data (e.g. meteorology, emissions, boundary conditions), while *NR2* used a completely different CTM (a coupled IFS-MOZART simulation). This choice is largely governed by practical considerations of available computing and storage resources. The decision to use two different NRs is based on a concern that a NR from a different CTM could potentially make comparison difficult, whereas a NR from the same CTM may not yield sufficiently large differences to the reference run.

[5] `www.eea.europa.eu/themes/air/airbase`

[6] `http://www.wmo.int/pages/prog/arep/gaw/measurements.html`

The spatial resolution of *NR2* is slightly higher than that of the DEHM runs. Thus OSSE 1 passes only criterion (d) of Privé et al (2013), while OSSE 2 passes criteria (a), (b) and (d) – criterion (c) is debatable in both cases. The differences between *ref* and *NR1* entail that OSSE 1 satisfies criteria (a), (d) and partly (b) of Halliwell Jr et al (2014), while differences between the two CTMs result in OSSE 2 satisfying criteria (a), (b), (c) and (d). However the MOZART-IFS fields are only available at a lower temporal resolution (every six hours) than the DEHM fields.

It should be noted that simulation *NR1* is not only used as the nature run in OSSE 1, but also for estimation of background errors. This may lead to a positive bias in the potential impacts of the assimilation, since the background errors are well characterised (particularly in light of the fact, noted above, that differences between *ref* and *NR1* are smaller than between either run with *NR2*). However, the background errors will not be perfectly characterised for OSSE 1 for either the summer or winter simulation periods, since a full year's worth of data is used to estimate the background errors (averaging over diurnal and seasonal variation), and there can be strong seasonal variation in background errors (e.g. Silver et al 2013). On a balance, it is reasonable to assume that OSSE 1 will give an upper-bound on the skill gain from the assimilation, and OSSE 2 may be more realistic.

Pseudo-observations are generated for each of the retrieved data-sets used for assimilation, using exactly the same spatial and temporal distribution of these data as given in the satellite products. Interpolation from model space to observation space is based on the same code as in the observation operators used by both DA schemes. Using the same spatial and temporal distribution allowed us to use the same averaging kernel profiles as in the actual retrieved products. These synthetic observations are perturbed by adding independent, normally distributed random numbers added to each of the interpolated observations, with zero mean and standard deviation set to that of the prescribed observation error. The resulting observations are then truncated at zero to ensure positive-definite values.

3.3 Real data experiments

In this section, we assess the effects of the assimilation on the modelled concentrations. We ran four assimilation-coupled simulations: for each of the two simulation periods (January-February 2008, July-August 2008; henceforth "winter" and "summer", respectively) simulations are run for the two data assimilation schemes (EnKF, 3D-Var). This is summarised in table 2. The results are compared with simulation *ref*, which did not use data assimilation.

Surface validation Model output (interpolated vertically, horizontally and temporally as described) is compared with measured surface concentrations from monitoring stations. The measured time-series are verified against their modelled counterparts; if a data point is missing from the time-series, the corresponding modelled value is not used in the validation. The data are examined from three different angles:

1. Each station has a time-series of observed and modelled data. For each station we used these time-series to calculate the correlation coefficient (R^2), bias and root mean-squared error (RMSE). For each of these error measures (i.e. bias, R^2 and RMSE), we summarised the distribution of values (i.e. over stations) in terms of its mean and specific quantiles (5%, 25%, 50%, 75% and 95%). The correlation (R^2) referred to here is the *site-specific temporal* correlation.
2. For each three-hour period, we calculated the average concentration across sites for both the model and observations; this yielded a time-series, from which we calculated the bias, R^2 and RMSE. This is called the "mean-over-stations time-series". The correlation referred to here is the *mean-over-stations temporal* correlation.
3. For each of the stations, we calculated the mean modelled and observed concentration over the entire simulation period; this gave a scatter-plot of modelled versus observed values, from which we calculated the bias, R^2 and RMSE. The correlation referred to here is the *spatial* correlation.

The first and second of these methods gives some insight into the temporal skill of the model, while the third provides some information about the model's spatial accuracy. The bias, as estimated from the second and third methods, is identical. We note that the bias and correlation provide independent measures of the model skill, respectively measuring its ability to estimate the average and variation around the mean. The RMSE provides an aggregate measure of skill, and in the cases considered appears to be mainly dependent on the bias.

The three aforementioned validation methods are summarised in figure 4 for O_3, NO_x, CH_4 and CO for the summer and winter simulation periods, and for the EnKF and 3D-var. It is immediately apparent that the EnKF and 3D-var schemes give on average very similar verification results at the surface.

For O_3, both assimilation schemes show a slightly lower correlations (both temporal and spatial) compared to the correlation for the reference run (by about 0.03 in winter and 0.05-0.1 in summer), and both schemes are biased high ($\approx$ 5-8 ppb) whereas the *ref* has very little bias ($\approx$ 1 ppb winter, -3 ppb summer). The results for O_3 are sensitive to the location of the study sites (located in Europe, thus clustered within the hemispheric domain), as the corresponding plot from the GAW data-set (not shown) shows that both assimilation schemes have a slightly higher spatial correlation for O_3 than the reference run. The higher O_3 concentrations are largely due to the direct adjustment of the O_3 field, rather than due to altering concentrations of O_3 precursors (not shown). The skill of the 3D-var and the EnKF are similar for the winter simulation period, while for the summer simulations

the 3D-var shows slightly more skill for O_3 than the EnKF, for both seasons and all three skill scores (in winter the 3D-var had R^2 about 0.05 higher and bias about 2.5 ppb lower than the EnKF simulation).

This is consistent with the hypothesis that there is a limited potential for improving the accuracy of modelled O_3 in the PBL by assimilation of satellite-derived O_3 profiles, especially in light of the fact that the burden of O_3 in the vertical column lies is found in the stratosphere, thus limiting the observability of tropospheric O_3 from space. Nevertheless other studies have trialled this and found that assimilation of TES O_3 profiles can reduce discrepancies between the modelled field and O_3 profiles measured by sondes (Parrington et al 2008). In another study, it was found that these assimilations also increased the average value of modelled surface O_3 levels (by 2-9 ppb), and this lowered the bias at some monitoring sites and accentuated it at others, while showing no consistent improvement on the model-vs-observed temporal correlation at monitoring sites (Parrington et al 2009).

The assimilation has very little influence on the skill in simulated NO_x concentrations. Both assimilation schemes yield a marginally greater negative bias in winter compared to the reference simulation (-2.9 ppb vs. -2.6 ppb), and hence a larger RMSE (4.8 ppb vs. 4.5 ppb). It can be seen that assimilation of NO_2 does not notably improve the accuracy of the modelled NO_x concentrations, since the short life-time of this species entails that the spatial variation in concentrations is much higher than the model's coarse horizontal domain can resolve. The impacts of adjustment of NO_x are also short-lived, and due to the OMI's swath pattern, data is only available over any region at most once per day.

In a series of experiments involving univariate assimilation of individual species (not shown), it was found that univariate assimilation of NO_2 retrievals leads to increased correlation with surface measurements (by about 0.08 in winter), and little change to background O_3 concentrations (less than 1.0 ppb on average). In assimilation experiments that produced higher mean O_3 levels, there is limited night-time accumulation of modelled NO_2 (due to reaction with O_3), which reduced the amplitude of the diurnal variation and consequently reduced the correlation with surface measurements (as modelled and measured concentrations are compared based on three-hourly means, correlations are affected by the substantial diurnal variation in boundary-layer NO_2 levels). Thus the combined effect of assimilation of O_3 and NO_2 is that improvements model skill in the NO_2 field (from assimilation of NO_2) are counteracted by the higher background O_3 (from assimilation of O_3).

The assimilation of retrieved CO has a large influence on the modelled CO field. As mentioned above, the reference simulation shows a significant under-estimation of CO concentrations (in the order of 50-150 ppb). The underestimation is consistent with findings from a large multi-CTM comparison study, which showed that most (if not all) CTMs underestimate CO in the northern hemisphere (from 20-90° N), most markedly during the boreal spring, when concentrations are highest (Shindell et al 2006). This has been understood as a negative bias in emissions from biomass burning and fossil fuel combustion, especially in the East Asian region (Shindell et al 2006). In our results, the negative bias in the CO field is greatly reduced for both assimilation schemes especially in winter (falling from -110 ppb to -60 ppb in winter and -55 ppb to -25 ppb in summer), leading to a large reduction in RMSE (going from 120 ppb to 75 ppb in winter and 55 ppb to 30 ppb in summer). However although assimilation of retrieved CO leads to lower bias and higher spatial correlation (the R^2 goes from 0.2 to 0.5 in winter, and -0.3 to 0.15 in summer), the temporal correlation is substantially lower compared to the reference run (the R^2 falls from 0.45 in winter to 0.3 in winter, and from 0.3 to 0.1 in summer).

Three different factors may have a role to play in explaining this last result. First, there is limited sampling with data available at only four or six monitoring sites (after quality-control criteria are applied). Second, the MOPITT 900 hPa layer is sensitive to concentrations through a broad region of the free troposphere and the boundary layer (as quantified by the averaging kernels), and different processes affect temporal variation in the free troposphere (e.g. geostrophic winds, large-scale fronts, orographic effects) and at the surface (e.g. local emissions, boundary-layer depth, surface winds); adjustment of boundary-layer concentrations (via vertical correlations) based on information from the free troposphere may thus improve large-scale spatial bias if background errors are correctly specified, while at the same time weakening temporal correlations. Third, it can be seen that the DEHM CO field shows a strong negative bias, and thus the assimilation provides some level of bias-correction and effectively acts as a "source" of CO. As the assimilation schemes rely on a number of assumptions, including zero-bias in the model and observations, it would have been preferable to use an explicit bias correction scheme as part of the assimilation routine. Dee (2005, pp. 3325) illustrates that, to first order, any persistent biases in either the background or the observations are transferred linearly to the analysis, thus degrading its quality. Due to the frequent assimilation updates (every three hours), typically no more than 25% of the horizontal domain is effectively observed at any assimilation. As such, there is an imbalance between air-masses that are more or less recently observed and thus adjusted. As such, the temporal correlations are degraded by these sporadic disruptions, whereas nudging the CO modelled field towards the observations will reduce the spatial bias.

The modelled surface CH_4 concentrations showed a large negative bias (of around -150 ppb in *ref* for both seasons). This is likely to be due to a bias in the (spatially and temporally) constant boundary in-flow concentration, with a potential contribution from errors in the surface fluxes. As CH_4 is relatively long-lived (with an atmospheric lifetime of 9 years) and is thus well-mixed, the assumption of temporally constant in-flow concentrations is a better approximation than for the other species considered. However the assumption of spatially constant boundary condition concentrations appears to be less well-founded. This is illustrated in figure 3 (which shows the averaged profiles from *ref* and *NR2* simulations); CH_4 concentrations in *NR2* vary by up to 10% in the vertical extent. It is likely that the MOZART-IFS simulations, with a global domain, a much higher model

top and the assimilation of a wide range of satellite retrievals, will give a more realistic representation of vertical-meriodonal patterns.

This negative bias at the surface is substantially reduced by assimilation of CH_4 during winter (from -150 ppb to -100 ppb), but exacerbated by these assimilations during the summer simulation period (from -140 ppb to -180 ppb). The temporal correlations fell as a result of the assimilation (by about 0.25 in winter and 0.2 in summer), although the spatial correlations are marginally increased in winter (by about 0.05) and reduced substantially in winter (by about 0.5). The reason for the clear difference in performance between the two seasons relates to differences between the two retrieved CH_4 data-sets: the model field showed a negative bias with respect to the TES CH_4 mixing ratio at 620 hPa, and a positive bias with respect to the SCIAMACHY CH_4 total column retrievals (figure 5). Both the volume and geographic extent of the SCIAMACHY CH_4 retrievals is much greater in summer than in winter (due to the fact that it could could only be retrieved during day-time, cloud-free scenes), whereas the TES CH_4 retrievals are roughly equally distributed between the summer and winter simulation periods. The mismatch in sign in the bias for these two data-sets led to large increments of alternating sign, that led to an overall reduction in model skill, especially in the summer months.

Remote-sensing validation Validation at surface measurement sites provides point-wise information of model skill at specific locations. For NO_2, CO and O_3, model results are validated using AirBase monitoring data, whose geographical footprint is limited to Europe. By contrast, satellite retrievals (especially from instruments aboard orbiting satellites) have a much broader spatial coverage. A comparison of model results and retrieved concentrations can provide additional insight into the spatial information provided by the assimilation schemes.

However, these remote-sensing products are subject to various biases, and differ in terms of their precision. Furthermore, each of the retrieval products is sensitive to different levels within the vertical column. For example, MOPITT CO partial column concentrations are positively biased by 7-14% around 700 hPa (Emmons et al 2004), biases in the tropospheric NO_2 columns from SCIAMACHY and OMI have been estimated at -5% and -10%, respectively (Irie et al 2012). A summary of the errors in these products is given in section 3.1 and table 1.

Most of the satellite products used in this validation are also assimilated, and in order to avoid an over-optimistic assessment of the assimilation's impact we compared retrievals with the *forecast* concentrations (i.e. the background field), rather than the *analysis* (as done, for example, by Frydendall et al 2009). The forecasts ran for three hours, as part of the normal assimilation-forecast cycle. It is clear that although the verification data are distinct from the assimilated data, they are not truly independent, since they are both the product of a common retrieval scheme (with prior estimates of column concentration profiles, meteorological fields, etc., that are highly dependent). As such, comparison with these data sets represents more of a self-consistency test than an impartial assessment of the assimilation. In addition to this, the model data are also compared with the SCIAMACHY tropospheric column NO_2 densities (an independent data-set) for additional insight.

To facilitate the comparison, observations and the modelled counterparts are binned on the model's horizontal grid, and averaged over the simulation period. Examples of such gridded fields are illustrated in figure 2, which presents gridded fields for the retrievals (both those that are assimilated and those kept for verification alone). Summary statistics based on the comparison of these gridded fields are shown in figure 5.

Considering results presented in figure 5 for CO, it can be seen that both assimilation procedures reduce the bias in the modelled CO field (by 10-20 ppb), particularly in the winter months (figure 5 a, d); this is consistent with the comparison of modelled and measured CO at the surface. The spatial correlation for the MOPITT 900 hPa CO mixing ratio is already very high in the reference simulations (above 0.95) and its assimilation leads to an even higher correlation (of about 0.99), with the exception of the EnKF in winter. TES correlations are also increased (by 0.1-0.7 in winter and by 0.15-0.2 in summer).

For CH_4 (figure 5 b, e), it can be seen that the SCIAMACHY total column CH_4 densities suggest that the model is biased high (by about 4.0×10^{18} molecules CH_4/cm^2), whereas the TES partial column mixing ratios at 620 hPa suggest that the model is biased low (by about 70 ppb). This result is not necessarily contradictory; the total can well be biased in one direction while modelled values at an individual level are biased another level. The positive bias with respect to the SCIAMACHY data may be partly due to errors in the above model-top contribution (which contributes approximately 25% of the column CH_4 density, or 2.8×10^{18} molecules CH_4/cm^2); recall that above the model top, the modelled CH_4 is set to the *a priori* concentration profile provided with the satellite product. The discrepancy in sign in the estimated model bias for the two products results in CH_4 analysis increments with opposing sign, which leads to a reduction in spatial skill.

For NO_2 (figure 5 c, g), the model appears to have a small bias (less than 0.15×10^{15} molecules NO_2/cm^2 in magnitude) compared to tropospheric column concentrations retrieved for OMI and SCIAMACHY, suggesting slight underestimation of tropospheric NO_2 in winter and overestimation in summer. Assimilation of the OMI tropospheric columns resulted in an increase in spatial correlation for all cases considered (by about 0.02-0.05 for the OMI retrievals and 0.01-0.07 for the SCIAMACHY retrievals). The 3D-var scheme had little influence on the mean bias, while the EnKF simulation led to an increase in tropospheric NO_2 concentrations (of 0.2-0.5 $\times 10^{15}$ molecules NO_2/cm^2); this increase appears to be due to the positive-definite filter (i.e. forcing the analysed concentrations to be positive), rather than the assimilation itself. The positive-definite correction affected the NO_2 field almost exclusively, since the relatively short atmospheric lifetime of NO_x results in a low hemispheric background

concentration relative the additive model-error perturbations (based on the layer-average of the background error standard deviation). In three out of four results (comparison with two retrieved products, across two seasons), the 3D-var simulation showed the highest spatial correlation.

For O_3 (figure 5 d), the mean bias between the modelled and retrieved TES O_3 mixing ratios at 825 hPa is relatively small (less than 0.5 ppb in winter and around -1.8 ppb in summer). Assimilation leads to a clear increase in spatial correlation between the modelled and retrieved data (by 0.15 in winter and 0.05 in summer), and a reduction in the summer-time bias (from -1.7 ppb to -0.5 ppb). The EnKF and 3D-var schemes give very similar bias and spatial correlation results in both seasons.

3.4 Observation system simulation experiments

Knowing the "true" concentration field allows us to examine many different aspects: no DA vs. 3D-var vs. EnKF, two NRs, four different species, two different simulation periods, multiple verification statistics, averaged latitudinal-vertical cross-sections (i.e. zonal averages) or horizontal cross-sections (i.e. maps of surface concentrations). For brevity, we only present a small subset of the results (figures 6 to 9), and describe the main trends seen. Further, we focus on the results for the 3D-var scheme. As discussed in section 2.2, the results for *NR1* is in general a much closer match to the reference simulation compared to *NR2*, since *ref* and *NR1* are generated using the same forecast model. As discussed in section 3.2, the two OSSEs provide complementary information; OSSE 1 offers an upper bound on the assimilation impacts, while OSSE 2 can be expected to give a more realistic estimate of the skill gain from the assimilation scheme.

For CO, simulation *ref* showed a strong negative bias with respect to *NR2* throughout the model domain (of around 60 ppb in winter and 20 ppb in summer), and the 3D-var assimilation increased concentrations in the lower 15 model levels, thereby reducing this bias (less, in absolute value, than 20 ppb in winter and 10 ppb in summer). Compared to *NR1*, *ref* showed higher concentrations (by about 5 ppb) in most of the tropospheric domain (especially for levels 15 to 25, corresponding to 750 hPa to 300 hPa) with the exception of a band of slightly lower concentrations between 50° N and 65° N in winter and between 60° N and 65° N from the surface up to model level 10 (approx. 930 hPa). This band is due to lower CO emissions from Europe and Northern China. In both seasons and for both NRs, the 3D-var assimilation leads to a generally higher temporal correlation (by about 0.1) for CO with the corresponding NR. In winter the assimilation of CO pseudo-observations in OSSE 1 leads to a marked reduction in bias (to less than 2 ppb on average) in the free troposphere (levels 10-25) and at lower latitudes, while in summer the assimilation led to a negative bias instead of a positive bias (which, combined with the improvement in temporal correlation, led to a slight reduction in RMSE). For OSSE 2, univariate CO assimilation results in a reduced negative bias in winter, and a positive bias instead of negative bias in summer (again leading to an overall lower RMSE, due in part to higher temporal correlation).

Although we only assimilate synthetic retrievals representing MOPITT 900 hPa and TES 825 hPa partial column concentrations, the assimilation directly impacts concentrations throughout the vertical column for two reasons. First, the background error covariances show (positive) correlations between surface and upper model levels (figure 1b in the Supplementary Material). Second, the retrieved MOPITT 900 hPa layer partial column concentrations are sensitive to concentrations throughout the column (as quantified by the averaging kernels), and thus the observation operator is a function of the whole column. In addition to these are the indirect effects due to vertical exchange of air-masses.

For NO_x, the synthetic OMI NO_2 tropospheric column retrievals are the only data source assimilated, and these are least available at the high latitudes during winter (due to the polar night, as well as quality control exclusion criteria relating to surface albedo and cloud reflective fraction), as shown in figure 2 i. The atmospheric lifetime of NO_x is much shorter than CO ($\sim$ 1 day vs. 1-3 months in the lower troposphere), and thus adjustments to the concentration field are limited in their horizontal extent due to limited long-range transport and shorter correlation length-scales in the background errors (which entails that the increments are much more localised around the positions of the observations). A large portion of the vertical column is adjusted, for the same reasons as described above for CO.

A trend (across seasons and NRs) appears in the fractional bias plots: the 3D-var scheme leads to a general increase in NO_x concentrations (by roughly 10%) south of 55° N throughout most of the vertical column. In this region of the domain, negative biases are reduced and positive biases accentuated. On the balance, this leads to marginally higher RMSE for the 3D-var simulation than for the reference run. The changes in the correlation field (i.e. the temporal correlation between the reference run and NR compared to the temporal correlation between the 3D-var run and corresponding NR) show no consistent pattern.

As noted above, in the real-data experiments (section 3.3) the modelled CH_4 showed a significant positive bias with respect to the SCIAMACHY total column CH_4 retrievals, yet a negative bias with respect to the TES partial column CH_4 mixing ratios at 620 hPa. However in OSSE 1, the pseudo-observations are derived from *NR1* (a DEHM simulation) using the same observation operator code as used in the assimilation scheme; although the univariate 3D-var assimilation leads to an increase in the absolute bias in OSSE 1, the relative bias is less than 3%. In OSSE 2, as the vertical limit of the simulation *NR2* is 0.1 hPa there is no need to use the *a priori* profile to specify the above model-top contribution; the resulting SCIAMACHY total column CH_4 pseudo-observations are notably lower (despite simulating higher concentrations below 100 hPa – see figure 3), and this led to strong negative corrections throughout the model domain, a large increase in the negative bias (from about -3% to -10%), and a

reduction in the temporal correlation. The aforementioned negative bias is most prominent at lower latitudes (south of 50° N) in winter, and extended to the pole in summer, consistent with the seasonal distribution of the SCIAMACHY CH_4 retrievals.

As noted previously, the effect on the O_3 concentrations represents a mix of direct and indirect effects. In additional assimilation experiments (not shown), indirect effects (of NO_2 and CO assimilation) have larger impact on simulated O_3 in summer than in winter (due to differences in photochemical O_3 production), and that at the surface the direct effects (of O_3 assimilation) tend to be more influential than indirect effects (due to the influence of surface fluxes). For example, the higher CO concentrations (from assimilation of CO data) leads to overall higher O_3 concentrations. The net result of the direct and indirect effects is a slight decrease in O_3 concentrations throughout the model domain in winter for OSSE 1 (by about 10%, correcting a small positive bias), and an increase in concentrations for OSSE 2 in winter (by 5-10%, accentuating a small positive bias), OSSE 1 in summer (by about 5%, accentuating a positive bias) and OSSE 2 in summer (by about 5%, reducing the magnitude of a negative bias). As in the case of NO_x, there is no obvious pattern in the changes to the temporal correlation field.

4 Discussion

The assimilation of various types of satellite retrievals of atmospheric trace gases is, in principle, of great potential benefit for the modelled concentration fields. However it should not be taken for granted that it will result in more accurate estimates, especially at the surface. In many atmospheric chemistry-transport modelling studies, the lower troposphere is of key interest (due to impacts to human health and terrestrial ecosystems), and retrievals are not necessarily very sensitive to concentrations in this portion of the atmosphere. Furthermore, satellite measurements and retrieval processes are subject to systematic biases and correlated observation errors, and these are essentially overlooked by most chemical DA schemes (n.b. the present work is no exception).

As such, we have set out to test how well the two chemical DA schemes presented here affect estimates of atmospheric composition, with reference to surface monitoring data, satellite retrievals, and a pair of OSSEs. The comparison with surface measurements is of interest to establish potential benefits for forecasting and hindcasting in studies relating air pollution to human health or biodiversity. However the model grid used here is defined with a very coarse horizontal resolution (150 km × 150 km) and results do not necessarily reflect model skill for most modelling applications, in which the DEHM is run at a much higher resolution (down to 6.7 km × 6.7 km) in a series of nested domains over focus areas (Frohn et al 2002). Considering the spatial agreement between satellite retrievals and their modelled counterparts allows us to assess spatial skill of the model across the domain (as opposed to only at measurement sites), however it provides information about the model's skill in terms of a vertically-integrated function of the concentration field, which is not necessarily very useful if the aim is to improve the model skill at the surface. In another validation strategy, the OSSE gives us access to the "true" concentration field via the NR; a well designed OSSE will ensure that the differences between the NR and the forecast model are representative of the differences between the true atmosphere and the forecast model. This may be difficult to test in practice (especially given limitations on available concentration measurements, for example, in the upper atmosphere). The NRs underlying the two OSSEs presented both differed from the reference simulation, and as the differences for *NR2* are much more extensive it is reasonable to expect that OSSE 2 provides more realistic depiction of the assimilation impacts. Nevertheless OSSE 1 is still informative in providing an upper bound on the skill gain from the assimilation.

The comparison with surface modelled results showed that, on the whole, the differences between the 3D-var and EnKF results are relatively minor. The assimilation went a long way to correcting the strong negative bias in CO concentrations in the reference simulations (which is primarily due to underestimation of boundary in-flow concentrations), and led to a general increase in spatial correlation for CO yet a decrease in temporal correlation. The net result is an overall decrease in RMSE for CO. The impact of assimilation on NO_x verification statistics is minimal, which is partly due to the coarse horizontal resolution combined with the short lifetime of this species (and hence substantial sub grid-scale variability). In the boundary layer, changes in O_3 concentrations stem mainly from direct adjustment due to assimilation of TES O_3 partial column concentrations (particularly in winter), with a contribution from the indirect changes to precursor concentrations (e.g. the higher background CO concentrations leads to increased O_3 levels).

Errors in models and data typically have a systematic non-trivial component (Dee 2005), and neither of the assimilation schemes trialled here account for this. The use of bias-blind assimilation in the presence of biased observations or background will, to first order, lead to biases in the analysis. Simple treatment of boundary in-flow concentrations makes it reasonable to expect that concentrations of medium- to long-lived species will be biased, at least in and above the free troposphere. Improved specification of boundary-condition concentrations, either from climatological averages or global CTM results, is planned.

A recent study by Miyazaki et al (2012) also used a local ensemble transform Kalman filter, which was coupled to the global CHASER CTM (Sudo et al 2002) for simultaneous estimation of NO_2, O_3, CO and HNO_3 by joint assimilation of retrievals from multiple satellite instruments. The CTM was run at approximately 2.8° horizontal resolution with 32 vertical levels from the surface to 4 hPa. The entire chemical state vector (35 species) is perturbed for each ensemble member and updated during each assimilation cycle. Emissions of CO, surface NO_x and lightning NO_x were also appended to the state-vector, both to ensure ensemble spread and for joint state-parameter estimation. A constant horizontal localisation length-scale of 600 km was used for

all concentration and emission parameters, apart for surface NO_x emissions. Ensemble spread wass maintained by multiplicative covariance inflation, with a fixed inflation factor of 5%.

Miyazaki et al (2012) employed a multi-faceted validation strategy, considering further satellite retrievals, aircraft measurements and ozonesondes; however no surface measurements are used for assimilation or verification. The data assimilation resulted in substantial improvements in accuracy for various compounds, substantially reducing biases in NO_2, O_3 and CO. Miyazaki et al (2012) also presented a set of five "observing system experiments" (OSEs, as opposed to OSSEs) to examine the influence of each assimilated data-set; an OSE involves refraining from assimilating a given data-set in one experiment, and using it instead for validation purposes. Model results are then compared with non-assimilated observations, rather than the nature run (as would be done in an OSSE).

The OSEs of Miyazaki et al (2012) demonstrate the potential for joint multi-species assimilation; for example, assimilation of upper-tropospheric HNO_3 and O_3 Microwave Limb Sounder observations reduced the bias in the tropospheric NO_2 field by more than 30%; however it is unclear how much of this improvement is based on direct adjustment and how much is due to indirect effects via the model's chemistry and dynamics. It is shown that the estimation of emissions of O_3 precursors had greatest effect in the tropics and the northern mid-latitudes, and that direct estimation of concentrations is most important in the free-troposphere.

While the present study does not consider as many data-sources as Miyazaki et al (2012) for assimilation or validation, one factor that features in both studies is multi-species assimilation. In this study we present results from multiple single-species assimilations, and the comparison with the corresponding joint multi-species assimilations (as well as direct adjustment of unobserved components) is considered in a forthcoming article (Silver et al 2015).

Neither of the 3D-var or EnKF assimilation schemes presented here allowed for dynamic estimation of surface fluxes (e.g. as done by Miyazaki et al 2012). This is possible in a 4D-var framework, but not in the case of 3D-var. One consequence of using constant emission rates is that the analysis increment to the concentration field in the boundary layer tended to decay (more rapidly for the more reactive species). Another consequence is that the ensemble depended almost exclusively on variation in the ensemble initial conditions. Joint optimisation of emissions and concentrations (in either the variational or ensemble-based assimilation frameworks) is likely to be an important feature in advanced chemical DA schemes.

This fits with findings from recent work on O_3 assimilation. A range of studies have shown that assimilation of surface measurements of O_3 is effective at reducing forecast errors or errors at non-assimilated monitoring sites (e.g. Wu et al 2008; Frydendall et al 2009; Pagowski et al 2010; Curier et al 2012), however the benefits are short-lived due to forcings from the chemistry. Longer-lasting reductions in forecast errors are, however, possible by joint estimation of initial conditions and precursor emission rates (e.g. Elbern et al 2007).

The importance of developing systems for multi-species data assimilation is related to the importance of multi-species CTMs. The CTM represents a vital tool to gain a better understanding of atmospheric processes and composition, and to investigate and predict the outcome of different scenarios such as future climate change, large-scale geo-engineering or changes to emissions. Further development of these tools will help make the most of past, present and future atmospheric observing networks and remote-sensing instruments.

5 Conclusions

In this study, we describe for the first time multi-species chemical data assimilation in connection with the Danish Eulerian Hemispheric Model (DEHM); two different assimilation schemes are presented, respectively using the three-dimensional variational (3D-var) and ensemble Kalman filter (EnKF) methods. Both data assimilation (DA) schemes involved some development beyond the formulations on which they are based – for example, background error covariance matrix parameters in the 3D-var scheme are filtered to prevent long-range correlations, and the EnKF uses a specialised localisation scheme to account for features specific to each species and vertical level.

The influence of the assimilation on model skill is assessed in a set of simulations, in which retrievals of four trace gases and from four satellite sensors are assimilated. The assimilation schemes are evaluated by comparing model results to surface measurements and satellite retrievals, as well as with a pair of observing system simulation experiments (OSSEs). The reference DEHM simulation showed a large negative bias in the modelled surface CO, consistent with previous chemical transport model simulations in the Northern Hemisphere (Shindell et al 2006). This negative bias is substantially reduced by both DA schemes; the CO spatial correlation is higher in the assimilation results compared to the reference simulation, while while the assimilation resulted in a lower temporal correlation.

Comparison with satellite retrievals showed that the DA leads to an increase in spatial correlation in the NO_2 tropospheric column, a loss of spatial correlation for CH_4, a reduction in bias in CO and O_3, and an increase in the spatial correlation for CO. The OSSEs confirm a large decrease in CO bias, an increase in CO correlation, and suggest a general increase in NO_x levels throughout the column. The OSSEs agree with other verification methods in showing that the assimilation results in a general reduction in model skill in the CH_4 field, while for O_3 the results are mixed, with no consistent direction in the change model skill (as measured by the bias and temporal correlation) due to assimilation.

Further work needs to be done regarding how best to make use of available satellite data in order to improve modelled surface air quality. Joint optimisation of initial conditions and emission rates offers one promising direction. Another potential avenue of research is the combined assimilation of in situ and remotely-sensed measurement data. Finally, this study is continued in a forthcoming paper that investigates the effects of adjustment of non-assimilated species, as well as comparing sequential univariate assimilation and joint multi-species assimilation (Silver et al 2015).

Acknowledgements This study was funded by a program grant from the Danish Council for Technology and Innovation: *Effects of climate changes on ecosystems – a global comparative analysis* (ECOGLOBE). The final revisions to the manuscript were funded by a McKenzie Postdoctoral Fellowship at the University of Melbourne. P. Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). The authors thank the institutions providing measurements, modelling and retrieval data: NASA, TEMIS, AirBase, NCEP, GAW, NCAR and IUP/IFE at the University of Bremen. Henk Eskes of KNMI provided useful feedback on a draft of the manuscript. We thank an anonymous reviewer for a range of constructive criticism on the text.

References

A. Bytnerowicz, K. Omasa, E. Paoletti, Environmental Pollution **147**(3), 438 (2007). DOI 10.1016/J.ENVPOL.2006.08.028

M. Kampa, E. Castanas, Environmental Pollution **151**(2), 362 (2008). DOI 10.1016/J.ENVPOL.2007.06.012

J.H. Seinfeld, S.N. Pandis, *Atmospheric chemistry and physics: from air pollution to climate change* (John Wiley & Sons, 2012)

Z. Levin, W.R. Cotton (eds.), *Aerosol pollution impact on precipitation: a scientific review* (Springer, 2009)

M.Z. Jacobson, *Fundamentals of atmospheric modelling* (Cambridge University Press, 2005)

O. Hertel, T. Ellermann, F. Palmgren, R. Berkowicz, P. Løfstrøm, L.M. Frohn, C. Geels, C.A. Skjøth, J. Brandt, J. Christensen, K. Kemp, M. Ketzel, Environmental Chemistry **4**, 65 (2007)

R.V. Martin, Atmospheric Environment **42**(34), 7823 (2008)

E. Kalnay, *Atmospheric modelling, data assimilation and predictability* (Cambridge University Press, 2003)

A. Sandu, T. Chai, Atmosphere **2**(3), 426 (2011)

W.A. Lahoz, Q. Errera, R. Swinbank, D. Fonteyn, Atmospheric Chemistry and Physics **7**, 5745 (2007)

M. van Loon, P.J.H. Builtjes, A.J. Segers, Environmental Modelling and Software **15**, 603 (2000)

H. Elbern, A. Strunk, H. Schmidt, O. Talagrand, Atmospheric Chemistry and Physics **7**, 3749 (2007)

E.M. Constantinescu, A. Sandu, T. Chai, G.R. Carmichael, Atmospheric Environment **41**, 18 (2007)

G.R. Carmichael, A. Sandu, T. Chai, D.N. Daescu, E.M. Constantinescu, Y. Tang, Journal of Computational Physics **227**, 3540 (2008)

R. Curier, R. Timmermans, S. Calabretta-Jongen, H. Eskes, A. Segers, D. Swart, M. Schaap, Atmospheric Environment **60**, 217 (2012)

J. Zhang, J.S. Reid, D.L. Westphal, N.L. Baker, E.J. Hyer, Journal of Geophysical Research **113**, D10208 (2008). DOI 10.1029/2007JD009065

N. Semane, V.H. Peuch, S. Pradier, G. Desroziers, L.E. Amraoui, P. Brousseau, S. Massart, B. Chapnik, A. Peuch, Atmospheric Chemistry and Physics **9**, 4855 (2009)

T. Milewski, M.S. Bourqui, Monthly Weather Review **139**(11), 3389 (2011)

M. Kopacz, D. Jacob, J. Fisher, J. Logan, L. Zhang, I. Megretskaia, R. Yantosca, K. Singh, D. Henze, J. Burrows, M. Buchwitz, I. Khlystova, W. McMillan, J. Gille, D. Edwards, A. Eldering, V. Thouret, P. Nedelec, Atmospheric Chemistry and Physics **10**(3), 855 (2010)

D.P. Edwards, A.F. Arellano Jr, M.N. Deeter, Journal of Geophysical Research **114**(D14), D14304 (2009). Doi:10.1029/2008JD011375

R.M. Timmermans, M. Schaap, P. Builtjes, H. Elbern, R. Siddans, S. Tjemkes, R. Vautard, Journal of Atmospheric and Oceanic Technology **26**(12), 2673 (2009)

J. Brandt, J.D. Silver, L. Frohn, C. Geels, A. Gross, A.B. Hansen, K.M. Hansen, G. Hedegaard, C.A. Skjøth, H. Villadsen, J.H. Christensen, Atmospheric Environment **53**, 156 (2012)

J. Brandt, J.H. Christensen, L.M. Frohn, F. Palmgren, R. Berkowicz, Z. Zlatev, Atmospheric Environment **35, Supplement 1**, S91 (2001). DOI 10.1016/S1352-2310(00)00415-5. Selected Papers Presented at the Venice Conference

J. Frydendall, J. Brandt, J.H. Christensen, Atmospheric Chemistry and Physics **9**, 5475 (2009)

J.D. Silver, J. Brandt, M. Hvidberg, J. Frydendall, J.H. Christensen, Geoscientific Model Development **6**(1), 1 (2013). DOI 10.5194/gmd-6-1-2013

M. Kahnert, Tellus **60B**, 753 (2008)

G. Evensen, *Data assimilation: the ensemble Kalman filter*, 2nd edn. (Springer-Verlag, 2009)

M. Corazza, E. Kalnay, S.C. Yang, Nonlinear Processes in Geophysics **14**, 89 (2007)

Z. Meng, F. Zhang, Mon. Wea. Rev. **522-540**, 136 (2008). DOI 10.1175/2007MWR2106.1

S.C. Yang, M. Corazza, A. Carrassi, E. Kalnay, T. Miyoshi, Mon. Wea. Rev. **137**, 693 (2009). DOI 10.1175/2008MWR2396.1z

M. Buehner, P.L. Houtekamer, H.L.M. Cecilien Charette, B. He, Mon. Wea. Rev. **138**, 1550 (2010). DOI 10.1175/2009MWR3157.1

M. Buehner, P.L. Houtekamer, H.L.M. Cecilien Charette, B. He, Mon. Wea. Rev. **138**, 1567 (2010). DOI 10.1175/2009MWR3158.1

K. Singh, A. Sandu, K.W. Bowman, M. Parrington, D.B.A. Jones, M. Lee, Atmospheric Chemistry and Physics Discussions **11**, 22247 (2011). DOI 10.5194/acpd-11-22247-2011

M. Kahnert, Tellus B **61**, 747 (2009)

A. Peuch, J.N. Thépaut, J. Pailleux, Q. J. Roy. Meteor. Soc. **126**, 1641 (2000)

W. Lahoz, R. Brugge, D. Jackson, S. Migliorini, R. Swinbank, D. Lary, A. Lee, Quarterly Journal of the Royal Meteorological Society **131**(606), 503 (2005)

R. Timmermans, W.A. Lahoz, J.L. Attié, V.H. Peuch, L. Curier, D. Edwards, H. Eskes, P. Builtjes, Atmospheric Environment **115**, 199 (2015). DOI 10.1016/j.atmosenv.2015.05.032

R.M. Errico, R. Yang, M. Masutani, J.S. Woollen, Meteorologische Zeitschrift **16**, 695 (2007)

N.C. Privé, R.M. Errico, K.S. Tai, Quarterly Journal of the Royal Meteorological Society **139**, 1354 (2013). DOI 10.1002/qj.2029

D.T. Kleist, K. Ide, Monthly Weather Review (2014). DOI 10.1175/MWR-D-13-00351.1

J.D. Silver, J. Brandt, J.H. Christensen, M. Kahnert, L. Robertson. Multi-species chemical data assimilation with the Danish Eulerian Hemispheric Model: univariate and multivariate assimilation (2015). Submitted 9th of November, 2014

J.H. Christensen, Atmospheric Environment **31**(24), 4169 (1997)

L.M. Frohn, J.H. Christensen, J. Brandt, Journal of Computational Physics **179**, 68 (2002)

L.M. Frohn, A study of long-term high-resolution air pollution modelling. Ph.D. thesis, University of Copenhagen (2004)

D. Simpson, H. Fagerli, J.E. Jonson, S. Tsyro, P. Wind, J.P. Tuovinen, The EMEP Unified Eulerian Model. Model Description. EMEP MSC-W Technical Report 1/2003, Norwegian Meteorological Institute, Oslo, Norway (2003)
J.A. Logan, Journal of Geophysical Research **104**(D13), 16115 (1999)
G.A. Grell, J. Dudhia, D.R. Stauffer, A description of the fifth-generation Penn State/ NCAR mesoscale model (MM5). NCAR Technical Note NCAR/TN-398+STR, National Center for Atmospheric Research, Boulder, Colorado, USA (1995)
D.R. Stauffer, N.L. Seaman, Monthly Weather Review **118**(6), 1250 (1990)
V. Vestreng, H. Klein, Emission data reported to UNECE/EMEP: Quality assurance and trend analysis & Presentation of WebDab. MSC-W Status Report, Norwegian Meteorological Institute, Oslo Norway (2002)
J.F. Lamarque, T.C. Bond, V. Eyring, C. Granier, A. Heil, Z. Klimont, D. Lee, C. Liousse, A. Mieville, B. Owen, M.G. Schultz, D. Shindell, S.J. Smith, E. Stehfest, J. Van Aardenne, O.R. Cooper, M. Kainuma, N. Mahowald, J.R. McConnell, V. Naik, K. Riahi, D.P. van Vuuren, Atmospheric Chemistry and Physics **10**(15), 7017 (2010). DOI 10.5194/acp-10-7017-2010
C.M. Benkovitz, M.T. Scholtz, J. Pacyna, L. Tarrasón, J. Dignon, E.C. Voldner, P.A. Spiro, J.A. Logan, T.E. Graedel, Journal of Geophysical Research **101**(D22), 29239 (1996)
G.R. van der Werf, J.T. Randerson, L. Giglio, G.J. Collatz, M. Mu, P.S. Kasibhatla, D.C. Morton, R.S. DeFries, Y. Jin, T.T. van Leeuwen, Atmospheric Chemistry and Physics **10**(23), 11707 (2010). DOI 10.5194/acp-10-11707-2010
J. van Aardenne, F. Dentener, J. Olivier, C. Klein Goldewijk, J. Lelieveld, Global Biogeochemical Cycles **15**(4), 909 (2001)
M. Schultz, S. Rast, M. van het Bolscher, T. Pulles, R. Brand, J. Pereira, B. Mota, A. Spessa, S. Dalsøren, T. van Nojie, S. Szopa. Emission data sets and methodologies for estimating emissions. RETRO project report D1-6 (2007)
J.J. Corbett, H.W. Koehler, J. Geophys. Res. **108**(D20), 4650 (2003). DOI 10.1029/2003JD003751
O. Stein. Model documentation of the MOZART CTM as implemented in the GEMS system (2009). `http://gems.ecmwf.int/do/get/PublicDocuments/1531/1172?showfile=true`, accessed: 2013-06-11
L.W. Horowitz, S. Walters, D.L. Mauzerall, L.K. Emmons, P.J. Rasch, C. Granier, X. Tie, J.F. Lamarque, M.G. Schultz, G.S. Tyndall, J.J. Orlando, G.P. Brasseur, Journal of Geophysical Research: Atmospheres **108**(D24) (2003). DOI 10.1029/2002JD002853
D. Kinnison, G. Brasseur, S. Walters, R. Garcia, D. Marsh, F. Sassi, V. Harvey, C. Randall, L. Emmons, J.F. Lamarque, P. Hess, J. Orlando, X. Tie, W. Randel, L. Pan, A. Gettelman, C. Granier, T. Diehl, U. Niemeier, A. Simmons, J. Geophys. Res. **112**(D20) (2007). DOI 10.1029/2006JD007879
J. Flemming, A. Inness, H. Flentje, V. Huijnen, P. Moinat, M.G. Schultz, O. Stein, Geoscientific Model Development **2**(2), 253 (2009). DOI 10.5194/gmd-2-253-2009
A. Inness, J. Flemming, M. Suttie, L. Jones, GEMS data assimilation system for chemically reactive gases. Technical Memoradum No. 587, European Centre for Medium-Range Weather Forecasts (ECMWF) (2009)
O. Talagrand, in *Data assimilation: making sense of observations*, ed. by W. Lahoz, B. Khattatov, R. Ménard (Springer-Verlag, 2010), pp. 41–68
L. Berre, Mon. Wea. Rev. **128**, 644 (2000)
N. Gustafsson, L. Berre, S. Hörnquist, X.Y. Huang, M. Lindskog, B. Navascués, K.S. Mogensen, S. Thorsteinsson, Tellus A **53**, 425 (2001)
D.F. Parrish, J.D. Derber, Monthly Weather Review **120**, 1747 (1992)
E.M. Constantinescu, A. Sandu, T. Chai, G.R. Carmichael, Quarterly Journal of the Royal Meteorological Society **133**(626), 1229 (2007)
L. Wu, V. Mallet, M. Bocquet, B. Sportisse, Journal of Geophysical Research: Atmospheres **113**(D20310) (2008). DOI {10.1029/2008JD009991}
A.C. Lorenc, F. Rawlins, Q. J. R. Meteorol. Soc. **131**, 3247 (2005)
N. Higham, IMA Journal of Numerical Analysis **22**, 329 (2002)
D. Bates, M. Maechler, *Matrix: Sparse and Dense Matrix Classes and Methods* (2013). R package version 1.0-12. `http://CRAN.R-project.org/package=Matrix`
L. Robertson, J. Langner, in *Air Pollution Modeling and Its Application XIII*, ed. by S.E. Gryning, E. Batchvarova (Kluwer Academic/Plenum Publishers, 2000), pp. 299–309
E. Andersson, J.N. Thépaut, in *Data assimilation: making sense of observations*, ed. by W. Lahoz, B. Khattatov, R. Ménard (Springer-Verlag, 2010), pp. 41–68
F. Rabier, H. Järvinen, E. Klinker, J.F. Mahfouf, A. Simmons, Q. J. Roy. Meteor. Soc. **126**, 1143 (2000)
G. Evensen, Ocean Dynamics **53**, 343 (2003)
C.H. Bishop, B.J. Etherton, S.J. Majumdar, Monthly weather review **129**, 420 (2001)
M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, J.S. Whitaker, Monthly Weather Review **131**, 1485 (2003)
W.G. Lawson, J.A. Hansen, Monthly Weather Review **132**, 1966 (2004)
P. Sakov, P.R. Oke, Monthly Weather Review **136**, 1042 (2008)
B.R. Hunt, E. Kalnay, E.J. Kostelich, E. Ott, D.J. Patil, T. Sauer, I. Szunyogh, J.A. Yorke, A.V. Zimin, Tellus **56A**, 273 (2004)
E.J. Fertig, J. Harlim, B.R. Hunt, Tellus **59A**, 96 (2007)
T.M. Hamill, J.S. Whitaker, Monthly Weather Review **133**(11), 3132 (2005)
J.S. Whitaker, T.M. Hamill, X. Wei, Y. Song, Z. Toth, Monthly Weather Review **136**, 463 (2008)
P.L. Houtekamer, H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, B. Hansen, Monthly Weather Review **133**, 604 (2005)
A. Gut, *An intermediate course in probability* (Springer, Heidelberg, 2009)
P.L. Houtekamer, H.L. Mitchell, Quarterly Journal of the Royal Meteorological Society **131**, 3269 (2005)
P.L. Houtekamer, H.L. Mitchell, X. Deng, Monthly Weather Review **137**, 2126 (2009)
R.G. Hanea, G.J. Velders, A. Heemink, Journal of Geophysical Research: Atmospheres **109**(D10) (2004)
P. Sakov, L. Bertino, Comput. Geosci. **15**, 225 (2010)
E. Ott, B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D. Patil, J.A. Yorke, Tellus A **56**, 415 (2004)
B. Hunt, E. Kostelich, , I. Szunyogh, Physica D **230**, 112 (2007)
G. Gaspari, S.E. Cohn, Quarterly Journal of the Royal Meteorological Society **125**, 723 (1999)
F. Boersma, E. Bucsela, E. Brinksma, J.F. Gleason, in *OMI algorithm theoretical basis document, OMI trace gas algorithms*, *NASA technical report*, vol. IV, ed. by K. Chance (2002), *NASA technical report*, vol. IV. `http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/docs/OMI/ATBD-OMI-04.pdf`
K.F. Boersma, H.J. Eskes, J.P. Veefkind, E.J. Brinksma, R.J. van der A, M. Sneep, G.H.J. van den Oord, P.F. Levelt, P. Stammes, J.F. Gleason, E.J. Bucsela, Atmospheric Chemistry and Physics **7**, 2103 (2007)
K.F. Boersma, R. Braak, R.J. van der A, *Dutch OMI NO_2 (DOMINO) data product v2.0: HE5 data file user manual* (2011). `http://www.temis.nl/docs/OMI_NO2_HE5_2.0_2011.pdf`
E.A. Celarier, E.J. Brinksma, J.F. Gleason, J.P. Veefkind, A. Cede, J.R. Herman, D. Ionov, F. Goutail, J.P. Pommereau, J.C. Lambert, M. van Roozendael, G. Pinardi, F. Wittrock, A. Schönhardt, A. Richter, O.W. Ibrahim, T. Wagner, B. Bojkov, G. Mount, E. Spinei, C.M. Chen, T.J. Pongetti, S.P. Sander,

E.J. Bucsela, M.O. Wenig, D.P.J. Swart, H. Volten, M. Kroon, P.F. Levelt, Journal of Geophysical Research: Atmospheres **113**(D15S15) (2008). DOI 10.1029/2007JD008908

G.B. Osterman, R. Beer, K.W. Bowman, P.D. Brown, S.A. Clough, A. Eldering, A. Goldman, D.J. Jacob, M. Lampel, J.A. Logan, M. Luo, F.J. Murcray, D.M. Rider, C.P. Rinsland, C.D. Rodgers, S.P. Sander, M. Shepard, S.S. Kulawik, E. Ustinov, H.M. Worden, J. Worden, Tropospheric Emission Spectrometer (TES) level 2 algorithm theoretical basis document. Tech. Rep. D-16474, Version 1.99, Jet Propulsion Laboratory, California Institute of Technology (2004). `http://tes.jpl.nasa.gov/uploadedfiles/TES.ATBD.L2.V1.99.v10.pdf`, accessed: 2013-08-01

R. Herman, S. Kulawik, K. Bowman, K. Cady-Pereira, A. Eldering, B. Fisher, D. Fu, R. Herman, D. Jacob, L. Jourdain, S. Kulawik, M. Luo, R. Monarrez, G. Osterman, S. Paradise, V. Payne, S. Poosti, N. Richards, D. Rider, D. Shepard, M. Shephard, F. Vilnrotter, H. Worden, J. Worden, H. Yun, L. Zhang, Earth Observing System (EOS) Tropospheric Emission Spectrometer (TES): Level 2 (L2) Data User's Guide (Up to & including Version 5 data). Tech. Rep. D-38042, version 5.0, Jet Propulsion Laboratory, California Institute of Technology (2013). `http://tes.jpl.nasa.gov/uploadedfiles/TESDataUsersGuideV5_0_May8_FV-dhs-AcceptedwithEqs.pdf`, accessed: 2013-07-10

M. Luo, C. Rinsland, B. Fisher, G. Sachse, G. Diskin, J. Logan, H. Worden, S. Kulawik, G. Osterman, A. Eldering, R. Herman, M. Shephard, Journal of Geophysical Research: Atmospheres **112**(D24) (2007). DOI 10.1029/2007JD008803

R. Herman, G. Osterman, C. Boxe, K. Bowman, K. Cady-Pereira, T. Clough, A. Eldering, B. Fisher, D. Fu, R. Herman, D. Jacob, L. Jourdain, S. Kulawik, M. Lampel, Q. Li, J. Logan, M. Luo, I. Megretskaia, R. Nassar, G. Osterman, S. Paradise, V. Payne, H. Revercomb, N. Richards, M. Shephard, D. Tobin, S. Turquety, F. Vilnrotter, K. Wecht, H. Worden, J. Worden, L. Zhang, Earth Observing System (EOS) Tropospheric Emission Spectrometer (TES): Data Validation Report (Version F06_08, F06_09 data). Tech. Rep. D-33192, Version 5.0, Jet Propulsion Laboratory, California Institute of Technology (2012). `http://tes.jpl.nasa.gov/uploadedfiles/TES_Validation_Report_v50.pdf`, accessed: 2013-07-10

M.N. Deeter, MOPITT (Measurements of Pollution in the Troposphere) Version 5 Product User's Guide. Tech. rep. (2011). `http://www2.acd.ucar.edu/sites/default/files/mopitt/v5_users_guide_beta.pdf`

M.N. Deeter, S. Martínez-Alonso, D.P. Edwards, L.K. Emmons, J.C. Gille, H.M. Worden, J.V. Pittman, B.C. Daube, S.C. Wofsy, Journal of Geophysical Research: Atmospheres **118**(12), 6710 (2013). DOI 10.1002/jgrd.50272

NCAR. MOPITT V5 Level 2 Data Quality Summary (2011). `http://www.acd.ucar.edu/mopitt/mopitt_level2_ver5.shtml`, accessed: 30-07-2013

L.K. Emmons, M.N. Deeter, J.C. Gille, D.P. Edwards, J.L. Attié, J. Warner, D. Ziskin, G. Francis, B. Khattatov, V. Yudin, J.F. Lamarque, S.P. Ho, D. Mao, J.S. Chen, J. Drummond, P. Novelli, G. Sachse, M.T. Coffey, J.W. Hannigan, C. Gerbig, S. Kawakami, Y. Kondo, N. Takegawa, H. Schlager, J. Baehr, H. Ziereis, Journal of Geophysical Research: Atmospheres **109**(D3) (2004). DOI 10.1029/2003JD004101

O. Schneising, M. Buchwitz, M. Reuter, J. Heymann, H. Bovensmann, J.P. Burrows, Atmospheric Chemistry and Physics **11**(6), 2863 (2011). DOI 10.5194/acp-11-2863-2011

H. Kelder, A. Piters, R. Timmermans, K. Bramstedt, J.C. Lambert, in *Proceedings of the Second Workshop on the Atmospheric Chemistry Validation of ENVISAT (ACVE-2)* (ESA-ESRIN, Frascati, Italy, 2004), no. SP-562 in ESA

K. Boersma, H. Eskes, E. Brinksma, J. Geophys. Res. **109**(D04311) (2004). Doi:10.1029/2003JD003962

R. van der A, H. Eskes, M. Van Roozendael, I.D. Smedt, N. Blond, F. Boersma, A. Weiss, J. van Peet, Algorithm Document Tropospheric NO_2. Tech. rep., Tropospheric Emission Monitoring Internet Service (TEMIS) (2010). `http://temis.nl/docs/AD_NO2.pdf`, accessed: 04/12/2013

K.P. Heue, A. Richter, M. Bruns, J.P. Burrows, C. v. Friedeburg, U. Platt, I. Pundt, P. Wang, T. Wagner, Atmospheric Chemistry and Physics **5**, 1039 (2005). DOI 10.5194/acp-5-1039-2005

H. Irie, K.F. Boersma, Y. Kanaya, H. Takashima, X. Pan, Z.F. Wang, Atmospheric Measurement Techniques **5**, 2403 (2012). DOI 10.5194/amt-5-2403-2012

K. Kunzi, P. Bauer, R. Eresmaa, P. Eriksson, S.B. Healy, A. Mugnai, N. Livesey, C. Priget, E.A. Smith, G. Stephens, in *The remote sensing of tropospheric composition from space*, ed. by J.P. Burrows, U. Platt, P. Borell (Springer, Heidelberg, Germany, 2011), chap. 4, pp. 153–230

M. Bocquet, H. Elbern, H. Eskes, M. Hirtl, R. Žabkar, G.R. Carmichael, J. Flemming, A. Inness, M. Pagowski, J.L. Pérez Camaño, P.E. Saide, R. San Jose, M. Sofiev, J. Vira, A. Baklanov, C. Carnevale, G. Grell, C. Seigneur, Atmospheric Chemistry and Physics **15**, 5325 (2015). DOI 10.5194/acp-15-5325-2015

L.M. Stewart, S.L. Dance, N.K. Nichols, International Journal for Numerical Methods in Fluids **56**, 1521 (2008)

K. Miyazaki, H. Eskes, K. Sudo, M. Takigawa, M. van Weele, K. Boersma, Atmospheric Chemistry and Physics **12**(20), 9545 (2012)

S.M. Lazarus, M.E. Splitt, M.D. Lueken, R. Ramachandran, X. Li, S. Movva, S.J. Graves, B.T. Zavodsky, Weather and Forecasting **25**(3), 837 (2010)

H.J. Eskes, K.F. Boersma, Atmospheric Chemistry and Physics **3**, 1285 (2003)

M. Joly, V.H. Peuch, Atmospheric Environment **47**, 111 (2012)

L. Malherbe, A. Ung, P. Schneider, F. de Leeuw, Analysis of station classification. ETC/ACM Technical Paper 2012/17, European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM) (2013). `http://acm.eionet.europa.eu/reports/ETCACM_TP_2012_17_analysis_AQstation_classification`, accessed: 2013-11-25

G. Halliwell Jr, A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, R. Atlas, Journal of Atmospheric and Oceanic Technology **31**, 105 (2014)

M. Parrington, D. Jones, K. Bowman, L. Horowitz, A. Thompson, D. Tarasick, J. Witte, Journal of Geophysical Research: Atmospheres **113**(D18) (2008). DOI 10.1029/2007JD009341

M. Parrington, D. Jones, K. Bowman, A. Thompson, D. Tarasick, J. Merrill, S. Oltmans, T. Leblanc, J. Witte, D. Millet, Geophysical Research Letters **36**(4) (2009). DOI 10.1029/2008GL036935

D.T. Shindell, G. Faluvegi, D.S. Stevenson, M.C. Krol, L.K. Emmons, J.F. Lamarque, G. Pétron, F.J. Dentener, K. Ellingsen, M.G. Schultz, O. Wild, M. Amann, C.S. Atherton, D.J. Bergmann, I. Bey, T. Butler, J. Cofala, W.J. Collins, R.G. Derwent, R.M. Doherty, J. Drevet, H.J. Eskes, A.M. Fiore, M. Gauss, D.A. Hauglustaine, L.W. Horowitz, I.S.A. Isaksen, M.G. Lawrence, V. Montanaro, J.F. Müller, G. Pitari, M.J. Prather, J.A. Pyle, S. Rast, J.M. Rodriguez, M.G. Sanderson, N.H. Savage, S.E. Strahan, K. Sudo, S. Szopa, N. Unger, T.P.C. van Noije, G. Zeng, Journal of Geophysical Research: Atmospheres **111**(D19) (2006). DOI 10.1029/2006JD007100

D.P. Dee, Quarterly Journal of the Royal Meteorological Society **131**, 3323 (2005)

K. Sudo, M. Takahashi, , H. Akimoto, Journal of Geophysical Research **207**(4586) (2002). Doi:10.1029/2001JD001114

M. Pagowski, G.A. Grell, S.A. McKeen, S.E. Peckham, D. Devenyi, Quarterly Journal of the Royal Meteorological Society **136**, 2013 (2010). DOI 10.1002/qj.700

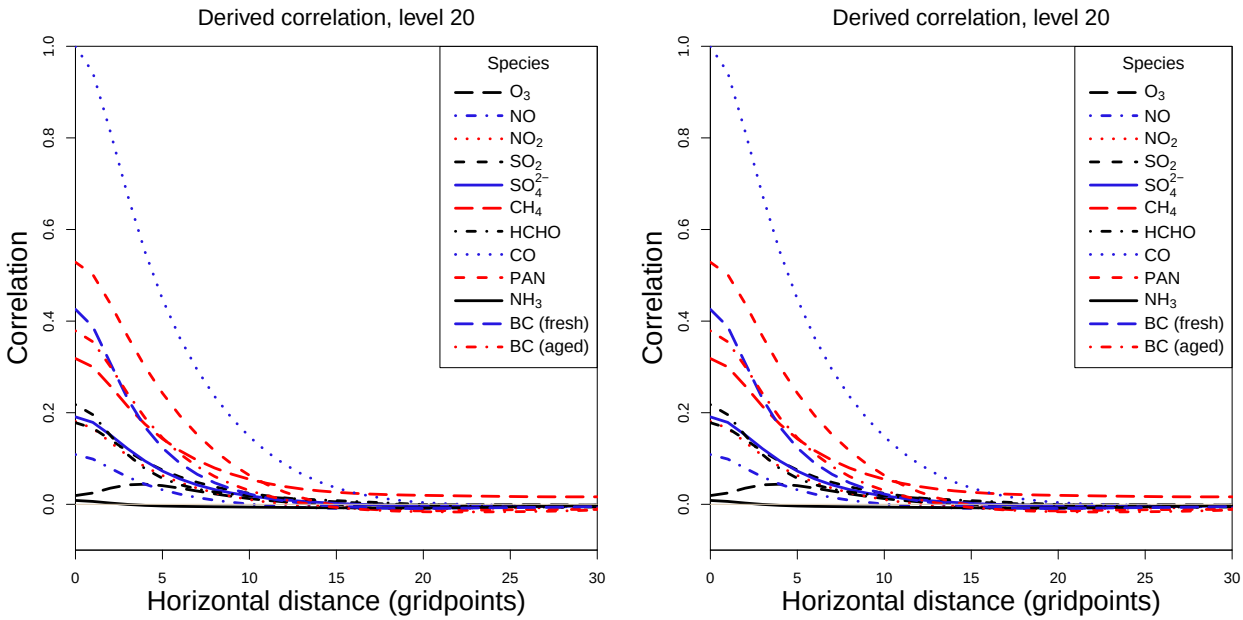


Fig. 1 Background error correlations between variables of the same species as a function of separation distance. Left panel: derived correlations from the background error decomposition. Right panel: fitted correlations modelled with the Gaussian function in equation 13.

6 Figures and tables

Instrument	Satellite	Column	Version	Provider	Species	Assim.	Error
MOPITT	Terra	Surf.	v5	NASA/NCAR	CO	Yes	Bias $\approx 7-14\%$ cf. profiles (700 hPa)
TES	Aura	Surf.	F06_09	NASA	CH_4	Yes	Pos. bias of $\approx$ 1-4%
—	—	—	—	—	CO	Yes	Error approx. 5-10% (700-200 hPa)
—	—	—	—	—	O_3	Yes	Pos. bias of $\approx$ 15% (1000-100 hPa)
OMI	Aura	Trop.	DOM. v2.0.1	TEMIS	NO_2	Yes	$R^2 > 0.6$, neg. bias of $\approx$ 15-30%
SCIAMACHY	ENVISAT	Tot.	v2.0.2	IUP/IFE	CH_4	Yes	Within 20%
—	—	—	v2.3	TEMIS	NO_2	No	Bias $\approx -5 \pm 14\%$

Table 1 Characteristics of the remote-sensing products used in this study. References for the error summary are given in section 3.1. Acronyms: "Assim" = Assimilated, "Surf." = Lowest level provided of a vertical profile, "Trop." = Tropospheric column, "Tot." = Total column, "DOM." = DOMINO, "Pos." = positive, "Neg." = negative. Notes: 1) Only a static averaging kernel profile is available

Period	Scheme	Datasets assimilated	Species adjusted
Winter	EnKF	OMI (NO_2), TES (O_3, CH_4, CO), MOPITT (CO), SCIAMACHY (CH_4)	NO_2, O_3, CH_4, CO
Summer	EnKF	—	—
Winter	3D-var	—	—
Summer	3D-var	—	—

Table 2 Summary of the assimilation-coupled simulations). The two simulation periods are January-February 2008 and July-August 2008, and are respectively termed "winter" and "summer" above. Two versions of each of the above experiments are run, one assimilating actual retreivals and the other assimilating pseudo-observations.

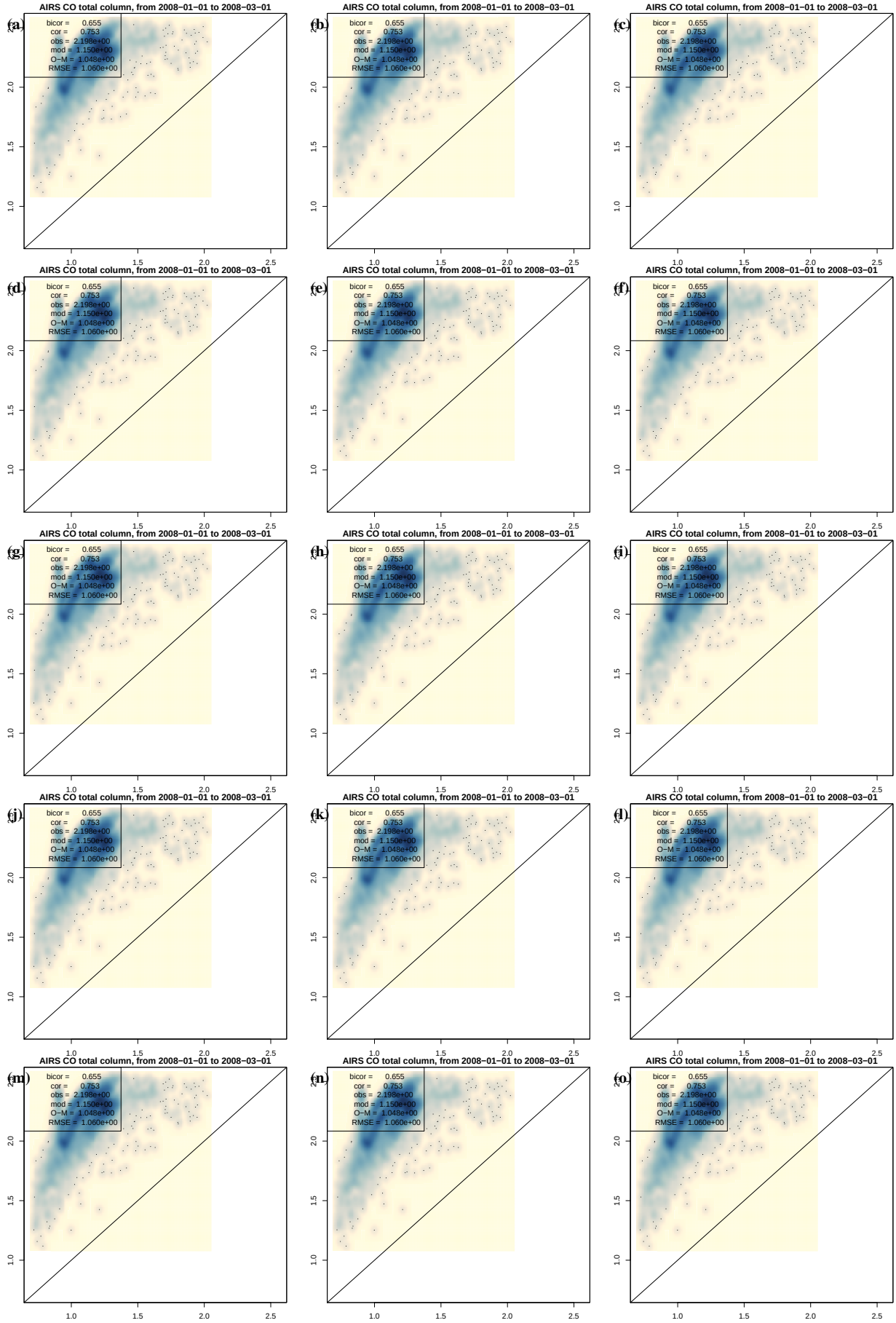


Fig. 2 Geographical distribution of mean satellite retrievals per grid-cell (left), mean modelled value per grid-cell (centre) and the total number of retrievals per grid-cell over the two-month winter period (right). Rows correspond to MOPITT CO 900 hPa partial column (top row), SCIAMACHY CH_4 total column (second row), OMI NO_2 tropospheric column (third row), TES O_3 825 hPa partial column (fourth row), and SCIAMACHY NO_2 tropospheric column (bottom row). Only pixels the used pixels are shown (i.e. retained after quality control checks and observation data thinning). Grey pixels (left and centre columns) indicate grid-cells without any retrievals during the simulation period. Modelled values are based on simulation *ref*, which used no data assimilation.

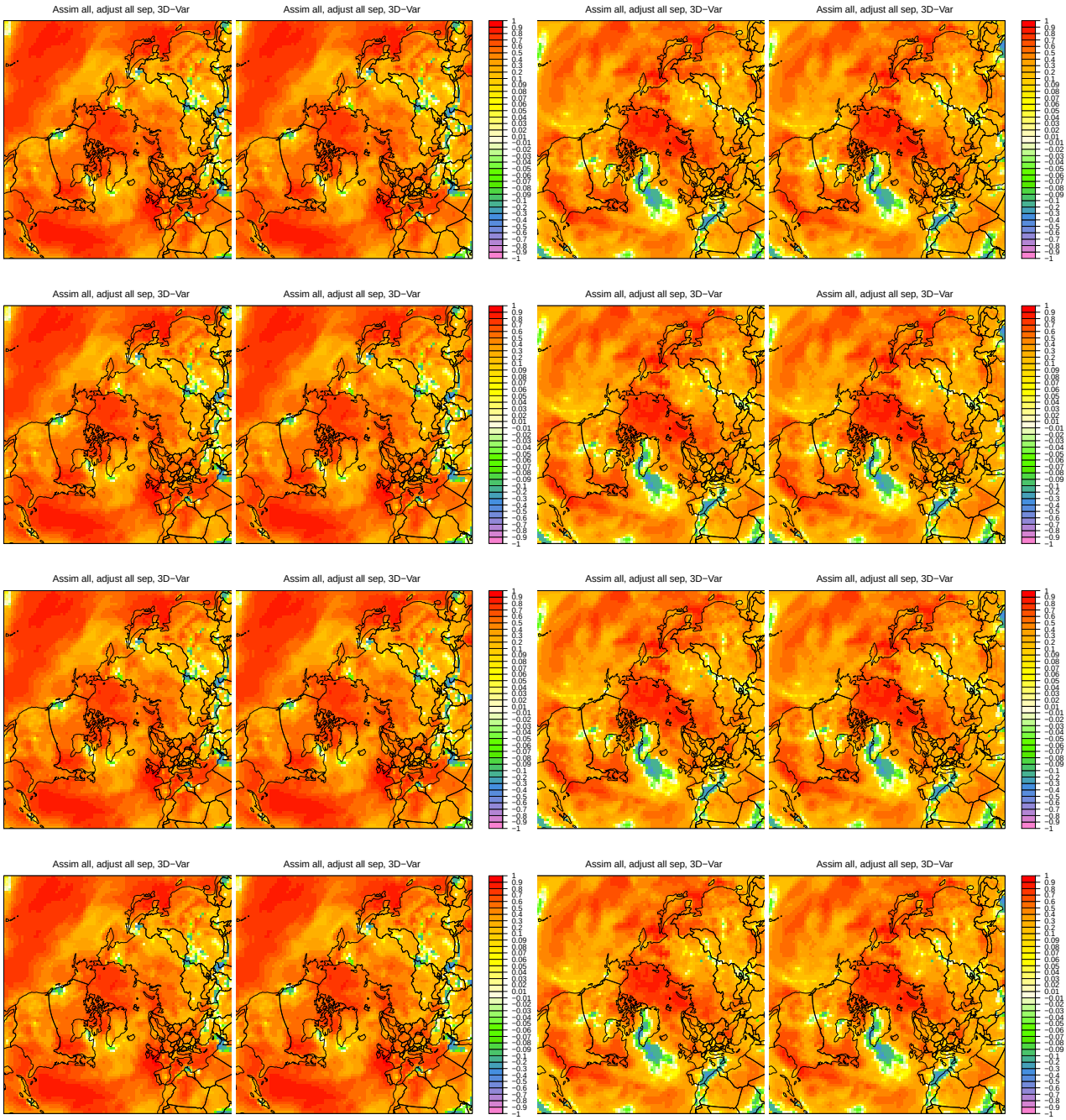


Fig. 3 Vertical profiles of each species (rows), for the DEHM and MOZART-IFS simulations for summer and winter (columns). The rows show, from top to bottom, zonal average concentrations for O_3, NO_x, CH_4 and CO (columns 1-4, respectively). The left-most two columns present January-February concentrations for the simulations *ref* and *NR2* (left and right, respectively), while the right-most two columns show the same data for July-August. The DEHM simulations are denoted "CN" in the panel title (columns 1 and 3), while the MOZART-IFS simulations are labelled "NR2" (columns 2 and 4). The colour scales differ between the left and right pairs of columns.

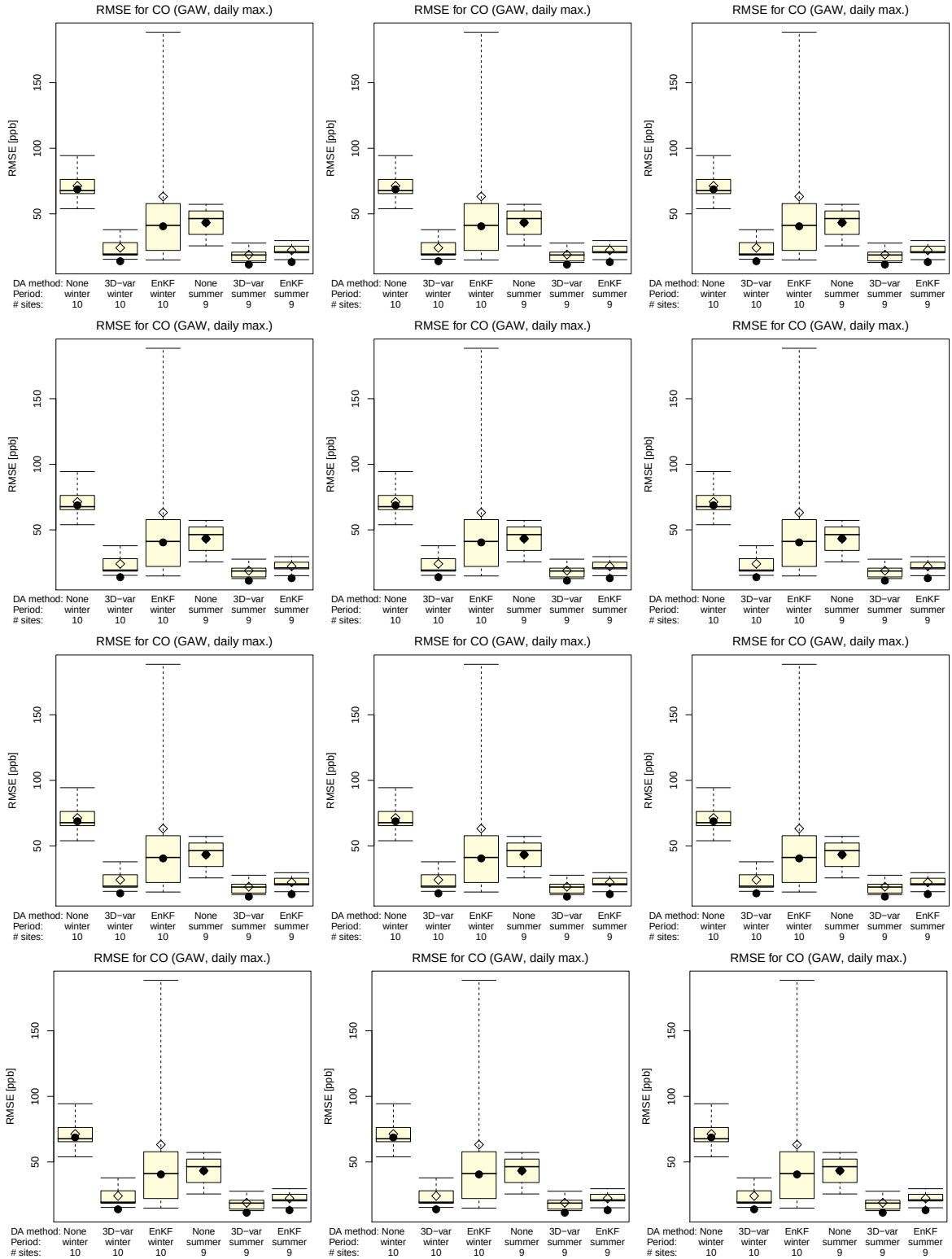


Fig. 4 Verification statistics for the four species considered. The correlation (R^2), bias and RMSE between the modelled and observed values are shown in the left, middle and right columns, respectively. Rows 1-4 display results for O_3, NO_x, CO and CH_4 respectively. Results for the winter and summer simulation periods are respectively presented in the left-most three and right-most three columns within each panel. The box-plot whiskers denote the 5% and 95% quantiles, the lower and upper edges of the box correspond to the 25% and 75% quantiles (respectively), and the median is indicated by a heavy horizontal black line. The mean is shown as an open diamond, and the statistic from the mean-over-stations time-series (i.e. the time-series of the geographical mean) is shown as a solid circle. The red stars show the value of the statistic calculated based on the station means alone.

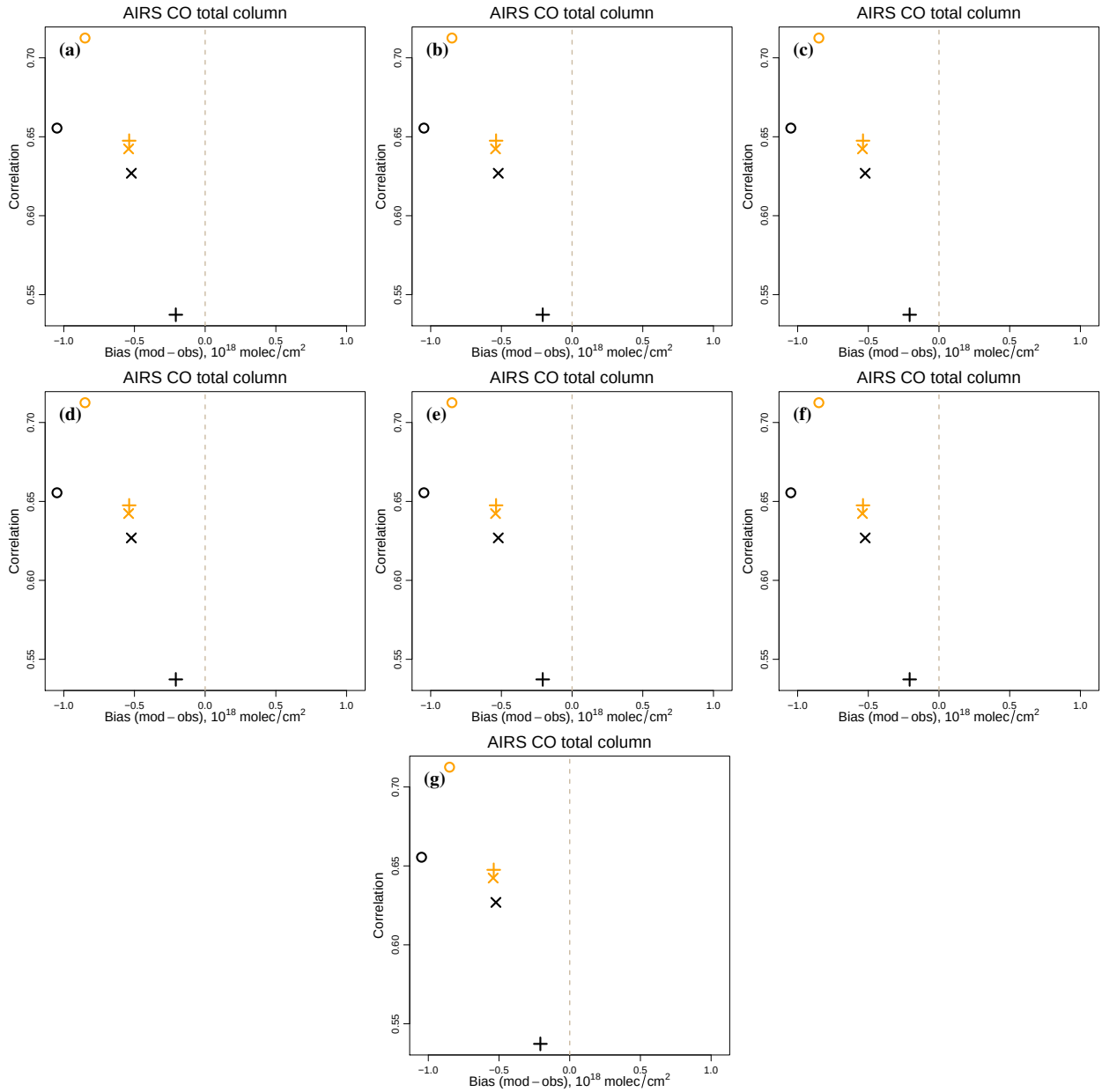


Fig. 5 Summaries of remote-sensing data, based on a comparison of the mean retrieval field and the corresponding mean modelled field (e.g. figures 2). The colour and shape of each point indicates, respectively, the simulation period and the assimilation treatment (see the caption in the upper-left panel). Note that the scales on the *x*- and *y*-axes are different for each figure, with the units for the bias indicated below the *x*-axis. The bias is computed as the difference model less observations. Each panel presents results for a different satellite product, as indicated in the panel headings.

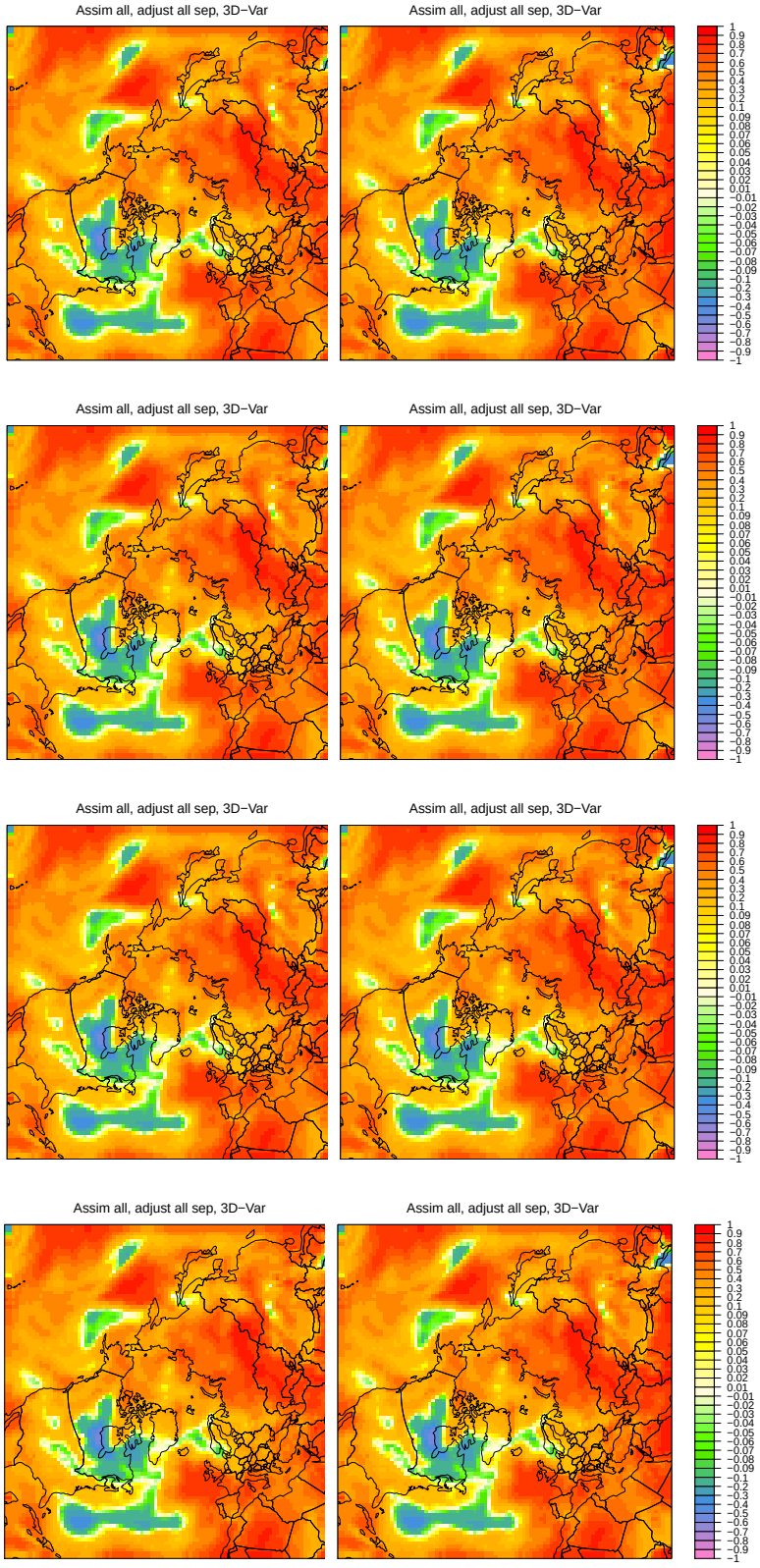


Fig. 6 Bias (from the OSSE) in the reference run (left column) and the 3D-var simulation (right column) with respect to *NR1*, normalised by the mean at each level in the NR. Rows 1-4 show results for species O_3, NO_x, CO and CH_4 are shown f respectively. Each cell shows the average at a given vertical model level, at a given latitude band (2.5° wide) for the winter simulation period. The colour-scale for each species (i.e. each row) is the same for both columns. Red/orange/yellow and purple/blue/green correspond to positive and negative values of the fractional bias (respectively).

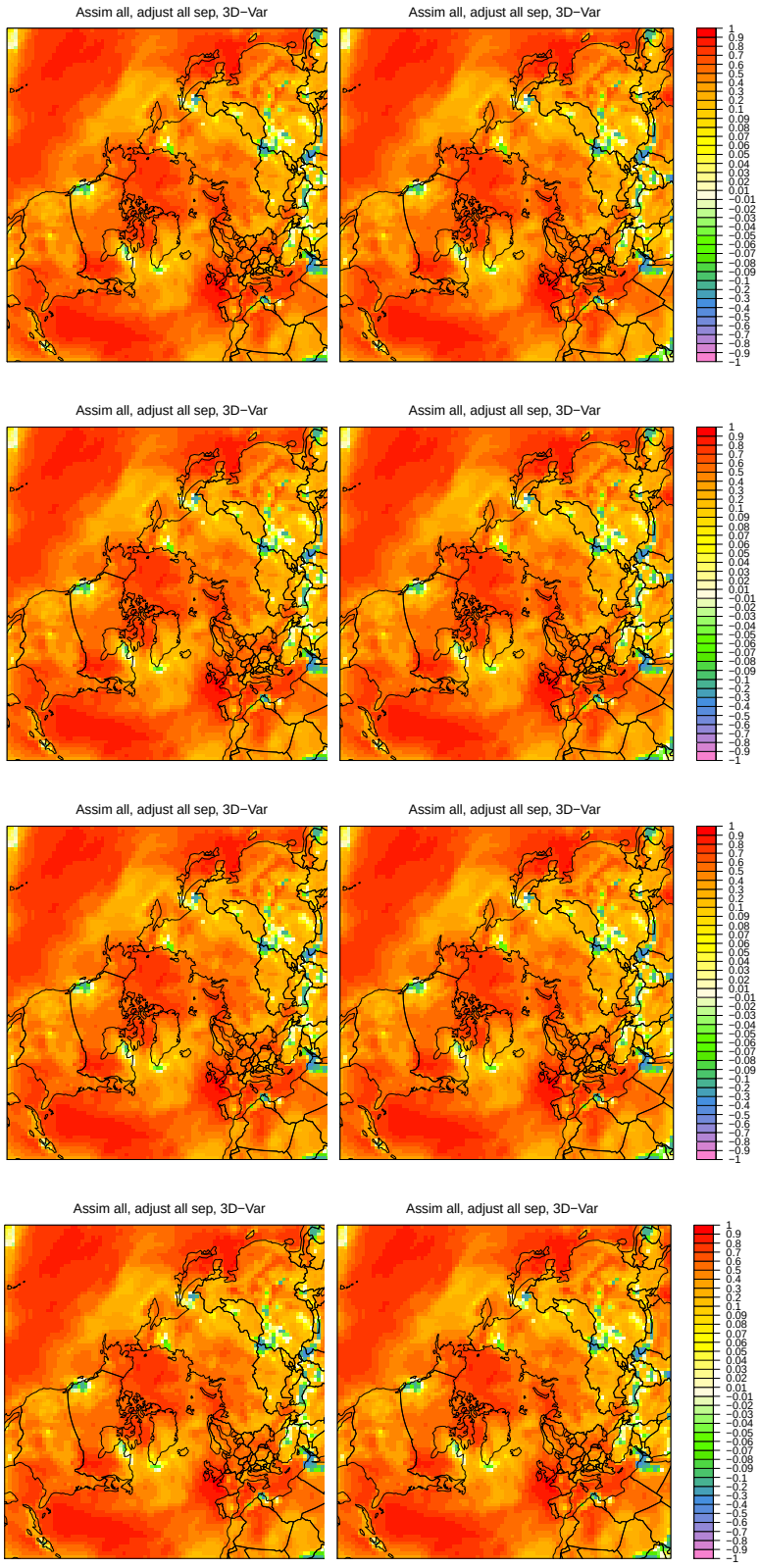


Fig. 7 Normalised bias from the OSSE compared to *NR2*, for the winter simulation period. See figure 6 for further details.

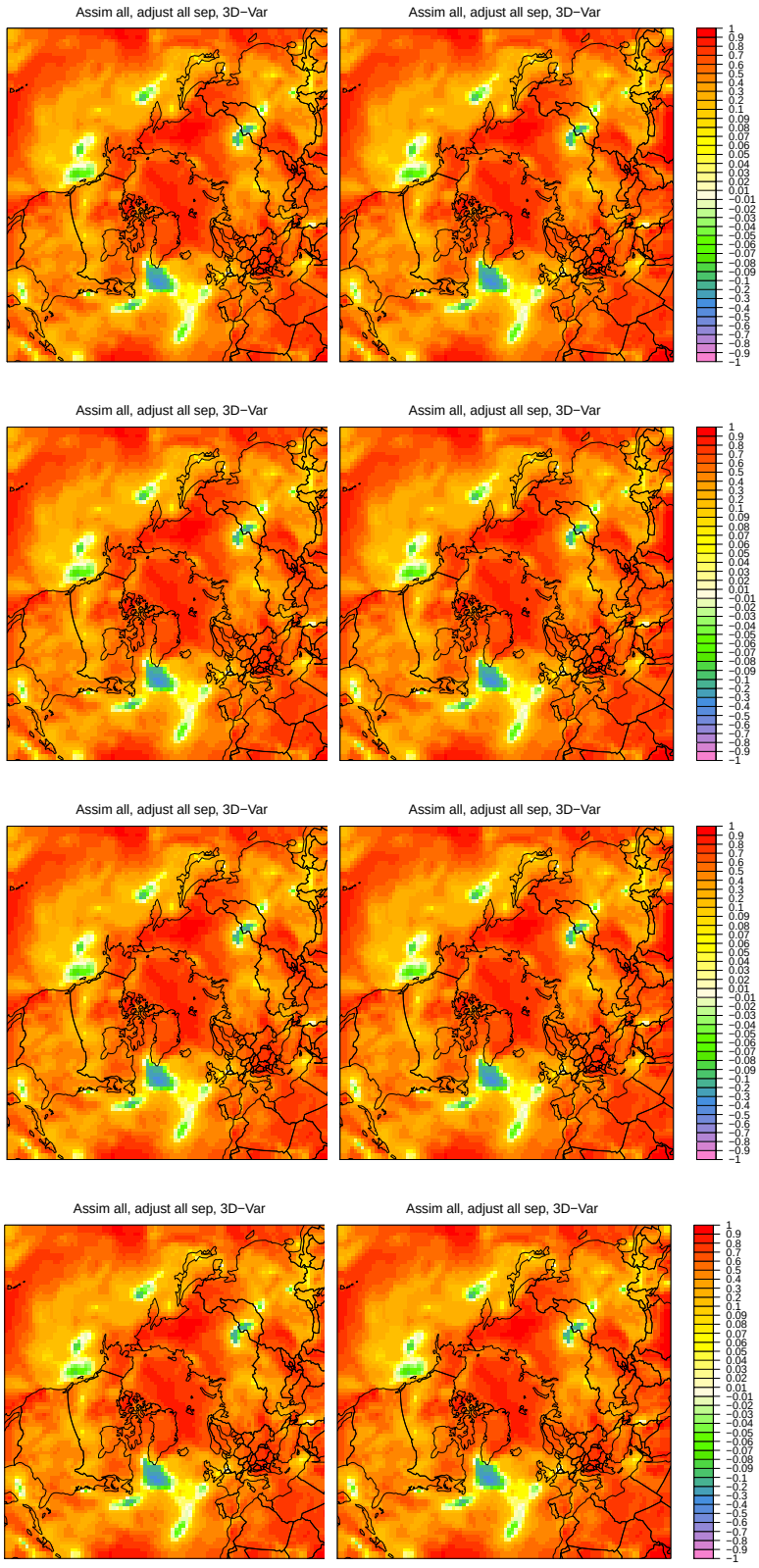


Fig. 8 Normalised bias from the OSSE compared to *NR1*, for the summer simulation period. See figure 6 for further details.

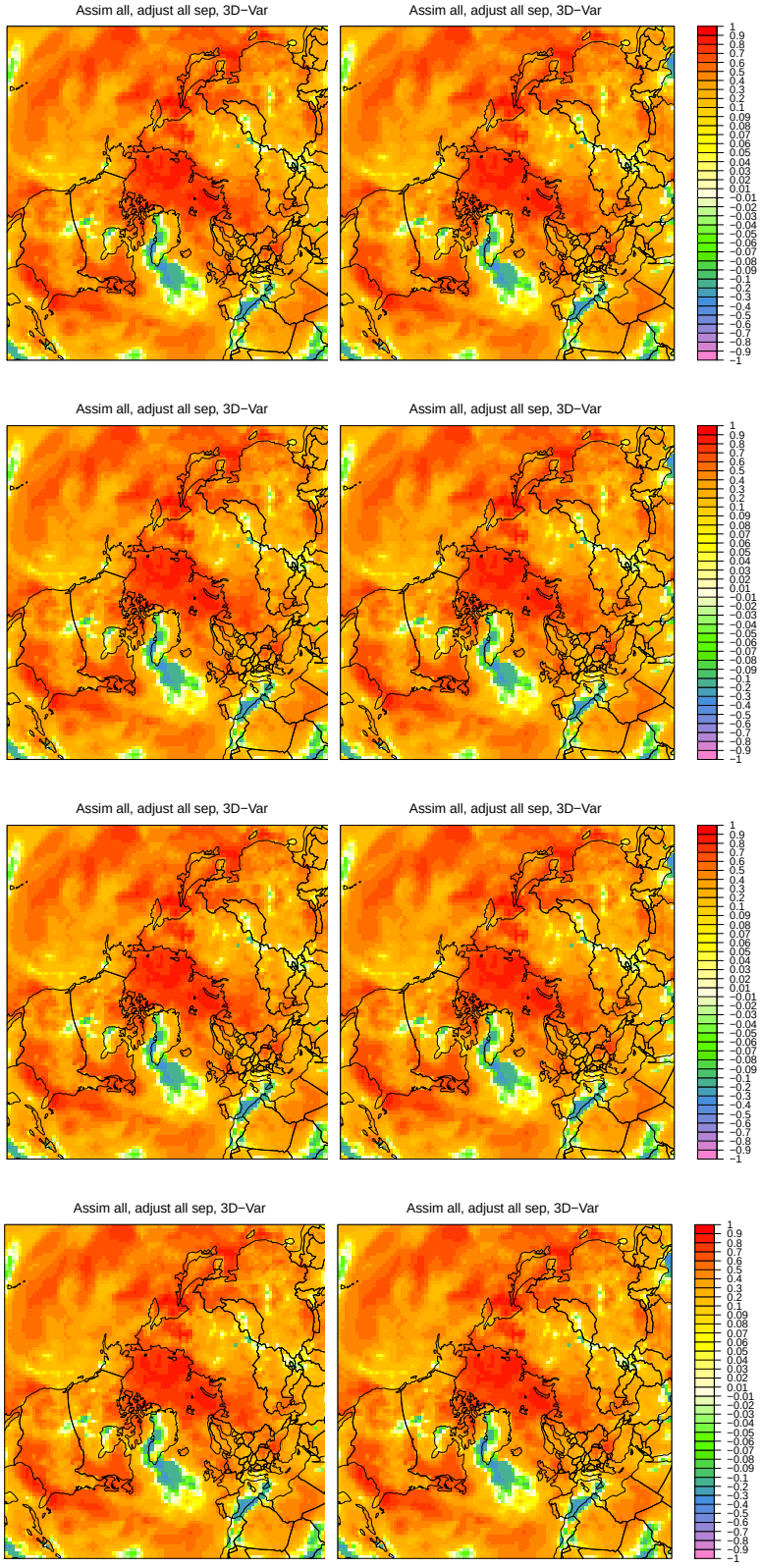


Fig. 9 Normalised bias from the OSSE compared to *NR2*, for the summer simulation period. See figure 6 for further details.