Available online at www.sciencedirect.com
SCIENCE DIRECT®
Atmospheric Science Letters 5 (2004) 35–41
ASL Atmospheric Science Letters
www.elsevier.com/locate/issn/1530261X

Representativeness uncertainty in chemical data assimilation highlight mixing barriers

David John Lary[a,b,c,*]

[a]*Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Code 910.3, Greenbelt, MD 20771, USA*
[b]*GEST, University of Maryland, Baltimore County, Baltimore, MD, USA*
[c]*Department of Chemistry, University of Cambridge, Cambridge, UK*

Received 14 February 2003; revised 9 July 2003; accepted 25 November 2003

Abstract

When performing chemical data assimilation the observational, representativeness, and theoretical uncertainties have very different characteristics. In this study, we have accurately characterized the representativeness uncertainty by studying the probability distribution function (PDF) of the observations. The average deviation has been used as a measure of the width of the PDF and of the variability (representativeness uncertainty) for the grid cell. It turns out that for long-lived tracers such as N_2O and CH_4 the representativeness uncertainty is markedly different from the observational uncertainty and clearly delineates mixing barriers such as the polar vortex edge, the tropical pipe and the tropopause.
Published by Elsevier Ltd on behalf of Royal Meteorological Society.

Keywords: Data assimilation; Representativeness uncertainty; Mixing barriers

1. Introduction

The key difference between conventional modeling and data assimilation is the use of observations and information on observational and other uncertainties The uncertainties normally considered are the observational uncertainty, the representativeness uncertainty (i.e. the spatial variability over an analysis grid cell), the background uncertainty, and the theoretical uncertainty. Accurately determining these uncertainties can be quite challenging. In this study we focus on the geophysical insights derived from examining

* Tel.: +1-301-614-6405.
E-mail address: david.j.lary@gsfc.nasa.gov (D.J. Lary).

1530-261X/$ - see front matter Published by Elsevier Ltd on behalf of Royal Meteorological Society.
doi:10.1016/j.atmoscilet.2003.11.002

the representativeness uncertainty of a chemical data assimilation system cast in flow tracking coordinates.

All data assimilation systems require reasonable estimates of the observation error statistics for each observation system used in the assimilation. Without such estimates, no assimilation system can extract all the available information from the observations (Daley, 1993).

In many data assimilation systems currently in use overly simplistic covariance models are used that cannot adequately describe state-dependent error components such as representativeness error (Dee et al., 1999). Cohn (1997) has highlighted the value of rigorous treatment of representativeness error and model error. The representativeness error is usually either ignored, treated as a constant, diagnosed by a time average of (observations − forecast, $O - F$), or sometimes taken as the variance of observations over a region (Stajner et al., 2001). Treatment of representativeness error is certainly an area which requires further attention in the future. Here, we show that actually using the data to define the representativeness uncertainty rather than using assumptions can provide interesting insights.

2. Flow tracking coordinates

Under adiabatic conditions, air parcels move along isentropic surfaces (surfaces of constant potential temperature, θ). So when considering tracer fields, θ is a suitable vertical coordinate, since it acknowledges the likely vertical motion of air parcels. McIntyre and Palmer (1983, 1984), Hoskins et al. (1985), and Hoskins (1991) have shown the value of isentropic maps of Ertel's potential vorticity (PV) in visualizing large scale dynamical processes. PV plays a central role in large-scale dynamics, where it behaves as an approximate material tracer (Hoskins et al., 1985).

As a result, PV can be used as the horizontal spatial coordinate instead of latitude and longitude (Norton, 1994; Lary et al., 1995). PV is sufficiently monotonic in latitude on an isentropic surface to act as a useful replacement coordinate for both latitude and longitude, reducing the tracer field from three dimensions to two. These ideas have already led to interesting studies correlating PV and chemical tracers such as N_2O and O_3 (Schoeberl et al., 1989; Proffitt et al., 1989, 1993; Lait et al., 1990; Douglass et al., 1990; Proffitt et al., 1989, 1993; Atkinson, 1993). A key result of these studies is that PV and ozone mixing ratios are correlated on isentropic surfaces in the lower stratosphere, as was first pointed out by Danielsen (1968).

Since the absolute values of PV depend strongly upon height and the meteorological condition, it is useful to normalize PV and use PV equivalent latitude (ϕ_e) as the horizontal coordinate instead of PV itself. ϕ_e is calculated by considering the area enclosed within a given PV contour on a given θ surface. The ϕ_e assigned to every point on this PV contour is the latitude of a latitude circle which encloses the same area as that PV contour. Therefore, for every level in the atmosphere ϕ_e has the same range of values, -90 to $90°$. This provides a vortex-tracking, and indeed a flow-tracking, stratospheric coordinate system.

We have taken these now well-established ideas and used them as a framework for our chemical data assimilation. This is certainly valid for our analysis interval of one day, and

often for up to 10 days or longer in the stratosphere. Because a major component of the variability of trace gases is due to the atmospheric motions it makes sense to use a co-ordinate system that 'moves' with the large-scale flow pattern to perform our data assimilation.

3. Representativeness uncertainty

In this study we seek to accurately characterize the representativeness uncertainty and improve the signal to noise by using all observations available within a grid cell and studying the probability distribution function (PDF) of these observations The width of the PDF is used as a measure of the variability or representativeness uncertainty for the grid cell. When more than one observation is available in a grid cell we take the median of the PDF as the observation for that cell, and the median of the observation uncertainty PDF as the observation uncertainty for that grid cell. So both the observation and observation uncertainty used are directly from the data but selected in a way that we are sure they are truly representative of that grid cell.

The criteria used to determine at what location we use an observation are equivalent PV latitude (ϕ_e), and potential temperature (θ). An observation is used in $\phi_e - \theta$ grid box where it lies. The grid used here has 21 potential temperature levels spaced equally in $\log(\theta)$ between 400 and 2000 K, and 32 equivalent PV latitudes spaced evenly between -90 and 90°.

The uncertainty of the reconstructed observation has two components. First, for each (ϕ_e, θ) grid box, we have a distribution of observed concentrations, and a distribution of observed concentration uncertainties. We take the observational uncertainty to be the median observed concentration uncertainty for the current distribution in the given (ϕ_e, θ) grid box. Second, the representativeness is taken to be the average deviation of the concentration distribution for the given (ϕ_e, θ) grid box. The average deviation, or mean absolute deviation, is a robust estimator of the width of the distribution (Press et al., 1992)

$$\sigma_{\text{rep}} = A\text{Dev}(\chi_1, \ldots, \chi_N) = \frac{1}{N}\sum_{j=1}^{N} |\chi_j - \bar{\chi}| \tag{1}$$

Let us now examine two examples of the representativeness uncertainty.

3.1. N_2O

Fig. 1 shows N_2O equivalent PV latitude–potential temperature cross-sections of (a) representativeness uncertainty (v.m.r.), (b) observational uncertainty (v.m.r.), (c) observation (v.m.r.), and (d) analyses uncertainty (v.m.r.) for January 1992. The data used is from the Upper Atmosphere Research Satellite (UARS) Cryogenic Limb Array Etalon Spectrometer (CLAES) version 9 for January 1992.

We notice that the representativeness uncertainty in Fig. 1(a) has more structure than in the observational uncertainty in Fig. 1(b). The mixing barriers associated with the vortex

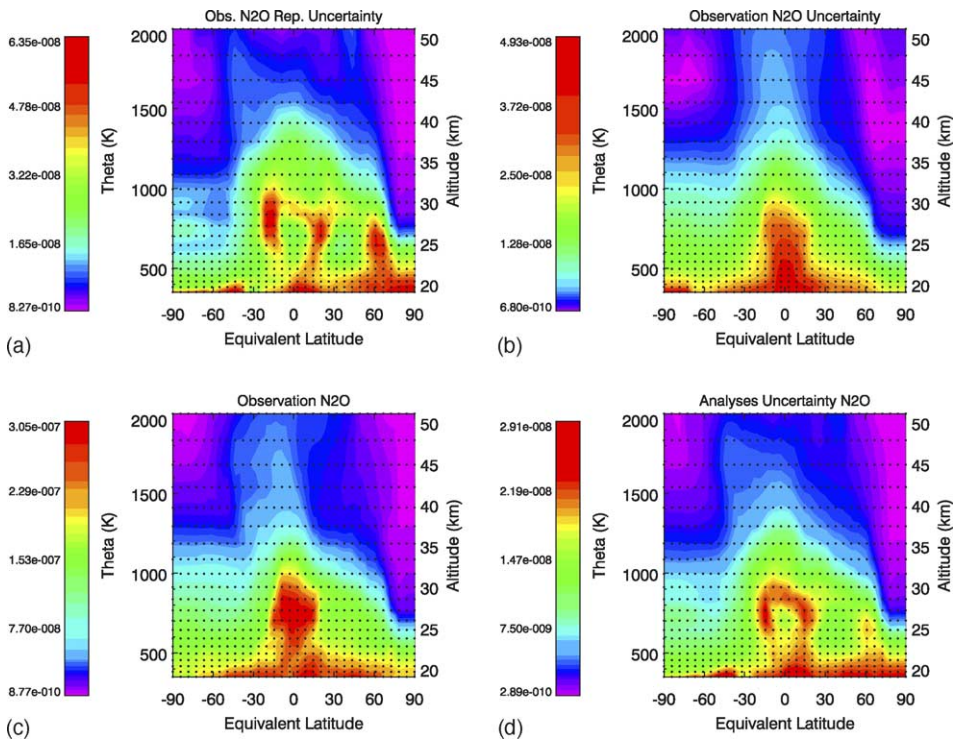

Fig. 1. N_2O equivalent PV latitude–potential temperature cross-sections of (a) representativeness uncertainty (v.m.r.), (b) observational uncertainty (v.m.r.), (c) observation (v.m.r.), and (d) analyses uncertainty (v.m.r.). The data used is from the Upper Atmosphere Research Satellite (UARS) Cryogenic Limb Array Etalon Spectrometer (CLAES) version 9 for January 1992.

edge (centered at around $\phi_e \approx 65°$) and the tropical pipe (at $\phi_e \approx \pm 20°$) are clearly visible. In each case this is close to sharp latitudinal gradients in N_2O, and that this uncertainty is reflected in the analyses uncertainty.

3.2. CH_4

Fig. 2 shows CH_4 equivalent PV latitude–potential temperature cross-sections of (a) representativeness uncertainty (v.m.r.), (b) observational uncertainty (v.m.r.), (c) observation (v.m.r.), and (d) analyses uncertainty (v.m.r.) for January 1992. The data used is from the Upper Atmosphere Research Satellite (UARS) Halogen Occultation Experiment (HALOE) version 19 for January 1992.

We notice that the representativeness uncertainty in Fig. 2(a) has more and very different structure than in the observational uncertainty in Fig. 2(b). The representativeness uncertainty beautifully captures the mixing barriers associated with the vortex edge (centered at around $\phi_e \approx 65°$) and the tropical pipe (at $\phi_e \approx \pm 20°$). In each case this is close to sharp latitudinal gradients in CH_4, and that this uncertainty is reflected in

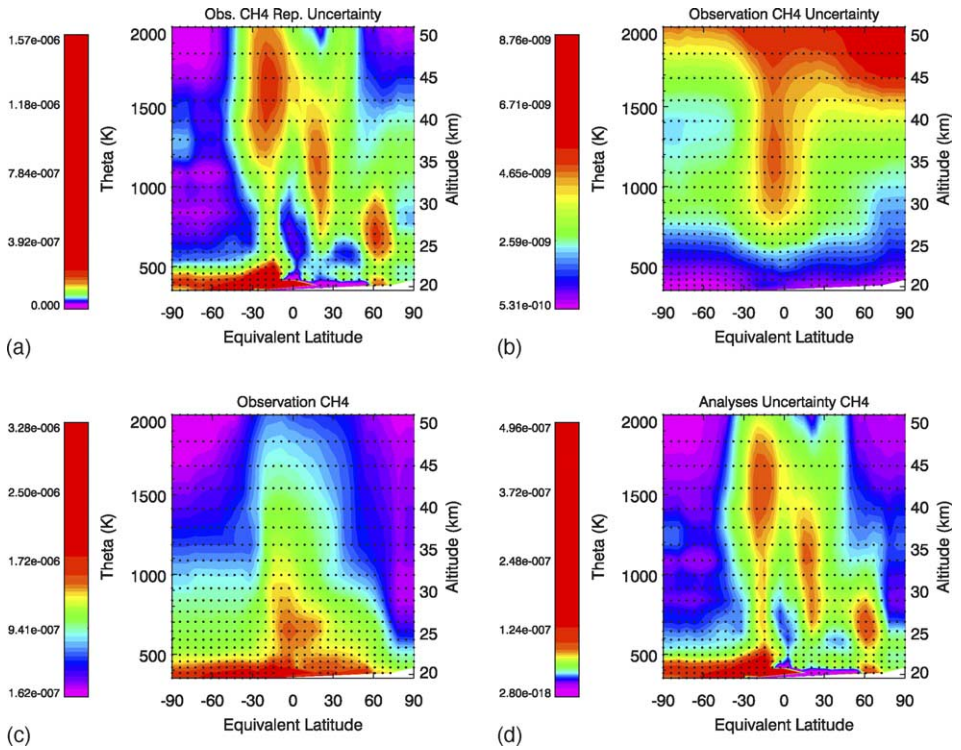

Fig. 2. CH_4 equivalent PV latitude–potential temperature cross-sections of (a) representativeness uncertainty (v.m.r.), (b) observational uncertainty (v.m.r.), (c) observation (v.m.r.), and (d) analyses uncertainty (v.m.r.). The data used is from the Upper Atmosphere Research Satellite (UARS) Halogen Occultation Experiment (HALOE) version 19 for January 1992.

the analyses uncertainty. Unlike N_2O, in the case of HALOE CH_4 the representativeness uncertainty also picks out the tropopause.

3.3. Other constituents

Shorter lived, chemically reactive constituents have also been examined, and have a different behavior to relatively inert tracers such as CH_4 and N_2O. Reactive species have large representativeness uncertainty in regions where we have considerable temperature, illumination, or other variations that impact the rates of reaction. For example, when using the Lagrangian flow-tracking coordinates described in the paper we have a large ozone representativeness uncertainty close to the summer pole. Such constituents were not considered here because of length constraints.

3.4. Other measures

The ‘width’ or ‘variability’ of the PDF can be characterized in several ways. Two other common measures are the variance, or its square root, the standard deviation. Both of these

measures have been tried and give essentially the same results. The reason for choosing the average deviation is that the variance and standard deviation depend on the second moment of the PDF. It is not uncommon to have a distribution whose second moment does not exist (i.e. is infinite). In this case, the variance or standard deviation is useless as a measure of the data's width about a central value. This can occur even when the width of the peak looks, by eye, perfectly finite. The average deviation is a more robust estimator that does not suffer from this problem (Press et al., 1992).

3.5. Sampling

In calculating the representativeness uncertainty, we are assuming that the observations adequately sample the atmosphere. With the relatively low data volume from instruments such as HALOE, and the UARS yaw cycle, we consider a month of data at a time in flow-tracking co-ordinates to give a statistically significant amount of data from which to calculate the average deviation. When flow-tracking coordinates are used in this way the representativeness uncertainty from both HALOE (a low data volume dataset) and CLAES (a high volume dataset) give similar results, suggesting that we have adequately sample the atmosphere.

The use of an entire months data will lead to a slight overestimate of the representativeness uncertainty due to the inevitable inclusion of some time variation. However, the choice of a Lagrangian coordinate system means that this is small, especially for constituents such as N_2O and CH_4. In addition, the slight overestimate that results here in the representativeness estimate based on the data itself should be compared to no estimates or crude estimates of the representativeness uncertainty used to date, often not based on the data itself.

4. Summary

When performing chemical data assimilation the observational, representativeness, and theoretical uncertainties have very different characteristics. In this study we have accurately characterized the representativeness uncertainty by studying the PDF of the observations. We have used average deviation as a measure of the width of the PDF and of the variability (representativeness uncertainty) for the grid cell. It turns out that for long-lived tracers such as N_2O and CH_4 the representativeness uncertainty is markedly different from the observational uncertainty and clearly delineates mixing barriers such as the polar vortex edge, the tropical pipe and the tropopause.

Acknowledgements

It is a pleasure to acknowledge: NASA for a distinguished Goddard Fellowship in Earth Science; The Royal Society for a Royal Society University Research Fellowship; The government of Israel for an Alon Fellowship; The NERC, EU, and ESA for research

support; Simon Hall of Cambridge University who has provided such excellent computational support.

References

Atkinson, R., 1993. An observational study of the Austral spring stratosphere: dynamics, Ozone transport and the Ozone dilution effect. Ph.D. Thesis, Massachusetts Institute of Technology.

Cohn, S., 1997. An introduction to estimation theory. J. Met. Soc. Jpn 75 (1B), 257–288.

Daley, R., 1993. Estimating observation error statistics for atmospheric data assimilation. Ann. Geophys.—Atmos. Hydro. Space Sci. 11 (7), 634–647.

Danielsen, 1968. Stratospheric–tropospheric exchange based on radioactivity, ozone and potential vorticity. J. Atmos. Sci. 25, 502–518.

Dee, D., Gaspari, G., Redder, C., Rukhovets, L., Da Silva, A., 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II. Applications. Mon. Wea. Rev. 127 (8), 1835–1849.

Douglass, A., Rood, R., Stolarski, R., Schoeberl, M., Proffitt, M., Margitan, J., Loewenstein, M., Podolske, J., Strahan, S., 1990. Global 3-dimensional constituent fields derived from profile data. Geophys. Res. Lett. 17 (4), 525–528.

Hoskins, B., 1991. Towards a pv-theta view of the general-circulation. Tellus, Ser. A 43 (4), 27–35.

Hoskins, B., McIntyre, M., Robertson, A., 1985. On the use and significance of isentropic potential vorticity maps. Q. J. R. Meteorol. Soc. 111 (470), 877–946.

Lait, L., Schoeberl, M., Newman, P., Proffitt, M., Loewenstein, M., Podolske, J., Strahan, S., Chan, K., Gary, B., Margitan, J., Browell, E., Mccormick, M., Torres, A., 1990. Reconstruction of O_3 and N_2O fields from ER-2, DC-8, and balloon observations. Geophys. Res. Lett. 17 (4), 521–524.

Lary, D., Chipperfield, M., Pyle, J., Norton, W., Riishojgaard, L., 1995. 3-Dimensional tracer initialization and general diagnostics using equivalent pv latitude–potential-temperature coordinates. Q. J. R. Meteorol. Soc. 121 (521), 187–210.

McIntyre, M., Palmer, T., 1983. Breaking planetary-waves in the stratosphere. Nature 305 (5935), 593–600.

McIntyre, M., Palmer, T., 1984. The surf zone in the stratosphere. J. Atmos. Terr. Phys. 46 (9), 825–849.

Norton, W., 1994. Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following technique for advecting material contours. J. Atmos. Sci. 51 (4), 654–673.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. Numerical Recipes in Fortran: The Art of Scientific Computing, second ed., Cambridge University Press, New York.

Proffitt, M., Steinkamp, M., Powell, J., Mclaughlin, R., Mills, O., Schmeltekopf, A., Thompson, T., Tuck, A., Tyler, T., Winkler, R., Chan, K., 1989. Insitu ozone measurements within the 1987 antarctic ozone hole from a high-altitude ER-2 aircraft. J. Geophys. Res. (Atmos.) 94 (D14), 16547–16555.

Proffitt, M., Aikin, K., Margitan, J., Loewenstein, M., Podolske, J., Weaver, A., Chan, K., Fast, H., Elkins, J., 1993. Ozone loss inside the northern polar vortex during the 1991–1992 winter. Science 261 (5125), 1150–1154.

Schoeberl, M., Lait, L., Newman, P., Martin, R., Proffitt, M., Hartmann, D., Loewenstein, M., Podolske, J., Strahan, S., Anderson, J., Chan, K., Gary, B., 1989. Reconstruction of the constituent distribution and trends in the antarctic polar vortex from ER-2 flight observations. J. Geophys. Res. (Atmos.) 94 (D14), 16815–16845.

Stajner, I., Riishojgaard, L., Rood, R., 2001. The GEOS ozone data assimilation system: specification of error statistics. Q. J. R. Meteorol. Soc. Part A 127 (573), 1069–1094.