Tellus (2001), 53B, 22–39
Printed in UK. All rights reserved

TELLUS
ISSN 0280–6509

Evaluating trace gas sampling strategies with assistance from a global 3D photochemical model: case studies for CEPEX and NARE O_3 profiles

By MARK G. LAWRENCE*, *Max-Planck-Institut für Chemie, Abteilung Luftchemie, Postfach 3060, 55020 Mainz, Germany*

(Manuscript received 24 November 1998; in final form 25 April 2000)

ABSTRACT

A technique is presented for evaluating sampling strategies based on *in situ* measurements and output from a 3D photochemical model. The technique quantifies the expected ability of various sampling strategies to give a representative picture of the actual mean trace gas concentration at the location being sampled. The representativeness is considered as a function of three parameters: (1) the number of samples, (2) the time between individual samples, and (3) the total time from the first to last sample. The technique is applied to analyzing the quality of various sampling strategies for ozone above five locations, ranging from the equatorial Pacific to the northern Atlantic. The analysis is based on data obtained by O_3 sondes during the CEPEX and NARE campaigns, along with the output from the global 3D photochemistry-transport model MATCH-MPIC. It is found that for all regions, the synoptic time scale of 3–5 days is a critical parameter for O_3. As long as the interval between soundings does not exceed this time scale, the number of soundings within a specific period does not strongly affect the representativeness of the sample mean profile; instead, the total period from the first to last sample is the most important parameter. For longer intervals between soundings, the behavior becomes more complex, with certain intervals resulting in better or worse representativeness than other intervals for the same total number of soundings. The model output is shown to generally be able to capture the synoptic time scale variability, and thus can be useful for establishing the relationship between synoptic meteorology and sampling strategies for various locations. However, the model underestimates the variability due to sub-gridscale processes (mainly convection), and thus must be used more cautiously in regions where this is critical, such as in the tropical upper troposphere. Future analyses such as this using previous observations and model output could assist field investigators in planning sampling strategies for campaigns or long-term monitoring.

1. Introduction

Trace gas mixing ratios vary substantially in time and space. Typical tropospheric O_3 mixing ratios range over an order of magnitude, from about 10 to 100 nmol/mol. At the extremes, O_3 in different regions of the upper troposphere has been observed to range from less than 5 nmol/mol over the equatorial Pacific (Kley et al., 1996), to as much as 500 nmol/mol over the tropical Atlantic (Suhre et al., 1997). The temporal variability in tropospheric O_3 observations is also relatively large, with a standard deviation equal to some 20–40% of the multi-year mean, depending on location (Logan, 1999). Furthermore, shifts in airmass origins can lead to rapid changes in concentrations; for instance, in an 18-h period, Oltmans et al. (1996) observed a change in ozone

* e-mail: lawrence@mpch-mainz.mpg.de

at 7 km above Cape Race from 150 to 50 nmol/mol, which they attributed to a shift in back trajectories from northern Canada to the Pacific Ocean. This large degree of variability makes it difficult to accurately determine the mean distribution of atmospheric trace gases. Nevertheless, there is an interest in accurately quantifying the average concentration of trace gases in various regions of the atmosphere over longer periods of time (e.g., seasonal or annual), for various purposes such as: determining the long term trends of O_3; simulating the climatic impact of key trace gases via absorption of solar and infrared radiation; and evaluating the ability of photochemical models to compute average trace gas concentrations.

Since the atmosphere cannot be sampled everywhere continuously, strategies must be devised for how best to obtain goals such as determining mean trace gas levels. In situ trace gas measurements are generally made within one of two frameworks: (1) during a field campaign over a limited time period that is focused on a particular region, or (2) at a long-term monitoring station. In either case, the appropriate sampling strategies will depend on the characteristics of the trace gas in question. In particular, Gibbs and Slinn (1973) and Junge (1974) established that an inverse relationship generally exists between the variability of trace gases and their tropospheric residence times; this implies that more measurements are needed of a shorter lived gas with greater variability to accurately establish its mean level (as well as other parameters of interest, such as seasonal cycle magnitudes).

A few studies have examined sampling strategy issues in more detail for various types of measurements. Gillette and Hanson (1983) considered flask sampling of CO_2, and used a simple analysis to show how one would expect the estimate of the seasonal cycle magnitude to improve if one were to increase the number of samples taken per year. Recently, Logan (1999) considered several decades of O_3 sonde data and found that, depending on the location, some 20–35 total soundings are needed to determine the multi-year mean profile to within $\pm 15\%$ at that location. The trace gas sampling strategies employed in specific field campaigns have also been examined. Ehhalt et al. (1997), using a 3D model with a set of tracers with various lifetimes, showed that the sampling applied during the PEM-West A and B campaigns likely provided a good picture of the actual mean trace gas distributions for the period sampled. Lawrence et al. (1999a) presented a similar analysis for O_3 measurements from CEPEX (the Central Equatorial Pacific Experiment) and found that while the single ship cruise which was made through the region was likely insufficient to capture the mean distribution of O_3 in the equatorial Pacific, on the other hand the ten O_3 sondes launched from Christmas Island were apparently sufficient to capture the mean profile there during the week of sampling.

Here, the issue of evaluating sampling strategies is addressed from a more general perspective. A technique is established for determining the expected departure from the mean (or any other target, such as the median value) for a large suite of possible sampling strategies. The intent is to provide supplemental information which can be used in choosing actual sampling strategies, in particular providing an estimate of a minimum level of sampling which is needed to determine the mean to a desired degree of accuracy over a certain period. This approach can be used with either actual observations or model output. Often insufficient observations are available to provide good statistics, in which case output from a model must be used as a surrogate. The best insights come when both field and model data can be used for the same location, with the observations providing a check on the model's accuracy, and with the model providing additional information on a level of detail not possible from the observations.

The technique is applied in this paper towards evaluating various sampling strategies for O_3 sondes at 5 different locations, ranging from the equator to 64°N. For this analysis, data are employed from two recent field campaigns, CEPEX and NARE (the North Atlantic Regional Experiment), along with corresponding output from a global photochemical model. The model, MATCH-MPIC, is described in an Appendix. The approach to assessing sampling strategies is first discussed in the next section, and O_3 sampling above each of the five chosen locations is then discussed in turn. Finally, a conclusions section considers implications of the results and possible further applications of this technique.

2. General description of the technique

Throughout this study, sampling strategies consisting of evenly-spaced O_3 soundings will be

considered. Such sampling strategies can be described in terms of 3 parameters, defined as follows:

N_s = the number of soundings in a sampling strategy
I_s = the interval between each sounding
T_s = the period being sampled (i.e., the time from the first to last sounding)

From these definitions, $T_s = I_s \times (N_s - 1)$. The technique presented here could also be applied to unevenly-spaced soundings, although the extra parameters required would make such an analysis much more complex.

The objective here is to use observations (where possible) and the output from MATCH-MPIC to see how well different sampling strategies are expected to represent the mean O_3 profile during a certain period (in this case, one month). The observations or model output are thus sub-sampled in the same way as the real atmosphere would be sampled for a given sampling strategy: one or more O_3 profiles, separated by a chosen time interval, are averaged together to obtain a mean profile for a specific sampling strategy. This will be called a "sample mean". This sample mean profile can then be compared to the "reference" mean O_3 profile for a month, which is calculated from all available observed or modeled profiles, as appropriate. It is important to note that this reference mean profile is itself a sample mean; here it will be assumed that the reference mean is close enough to the actual mean profile (which is unknown) that it will not make a difference to this analysis (this assumption should be reasonable for O_3 in this study as long as the reference mean consists of about daily or more frequent samples). In this section, the technique will be demonstrated using only the output from the model during March, 1993, above Christmas Island (2°N, 158°W). In the next section observations and four other locations will also be considered.

For a gas like O_3 with a large temporal variability, an arbitrarily chosen individual profile can differ significantly from the monthly mean; the same can apply to the sample mean of a set of profiles, especially when only a few profiles are included in the set. The expected difference can be characterized by various statistics. For this study, the statistic of the absolute deviation (d) will be used. This statistic is chosen because it applies regardless of how O_3 is distributed in time (e.g., normal, bimodal, log-normal, etc.). To obtain values which are readily comparable for different regions, d_n, the normalized absolute deviation (relative to the mean), will be used. For a time series of O_3 data during March at a particular vertical level k, d_n is defined as:

$$d_n(k) = \frac{\frac{1}{N}\sum_{i=1}^{N} |O_3(k,i) - \langle O_3 \rangle(k)|}{\langle O_3 \rangle(k)}, \quad (1)$$

where k is the vertical level index, i is the time-sample index, $O_3(k, i)$ is an individual sample or a sample mean, $\langle O_3 \rangle(k)$ is the March mean O_3 value at level k, and N is the total number of samples or sample means (note that this differs from N_s, which is the number of samples comprising each of the sample means). $d_n(k)$ thus indicates the average relative difference between an arbitrary O_3 value (or sample mean value) and the March mean value at level k.

The profile of d_n above Christmas Island based on the MATCH-MPIC output for March is plotted in Fig. 1. The corresponding O_3 profile is in Fig. 2, discussed in the next section. The model O_3 data were archived every 3 h for the month, so the total number of samples in this case is $N = 248$ ($= 31 \times 8$). The figure indicates that on average a single sounding would be expected to differ from the March mean by some 20–40%.

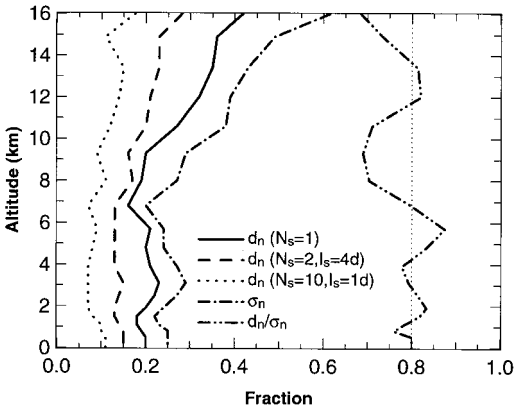

Fig. 1. Profiles of d_n and σ_n as computed from the MATCH-MPIC output above Christmas Island for selected sampling strategies. The dotted line at 0.8 indicates the value of d_n/σ_n expected for a normal distribution (from eq. (2)).

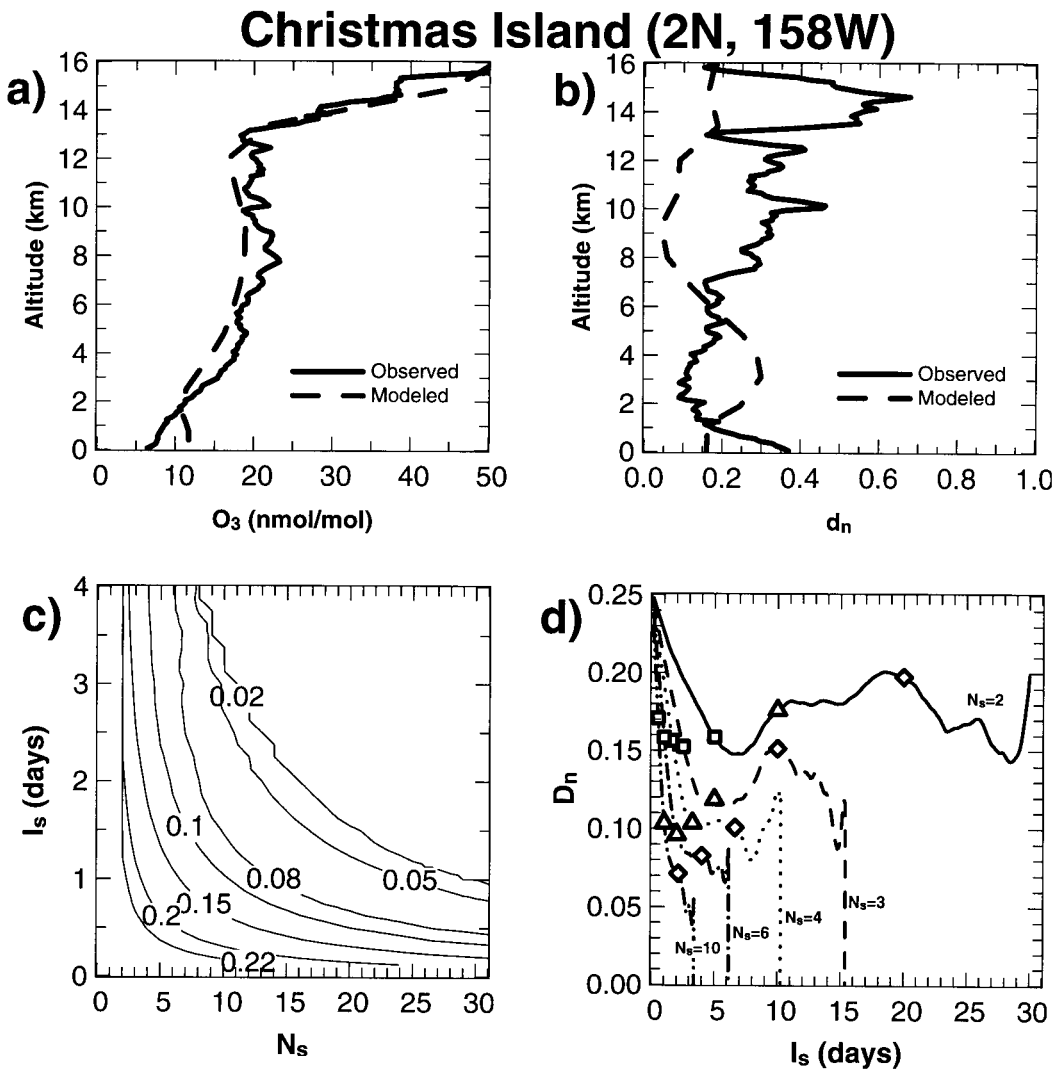

Fig. 2. Analyses for Christmas Island: (a) modeled and observed mean O_3 profiles; (b) modeled and observed normalized absolute deviations (d_n); (c) values of the column mean normalized absolute deviation (D_n) as a function of the number of soundings (N_s) and interval between soundings (I_s), focusing on scenarios mainly appropriate for a field campaign; (d) values of the column mean normalized absolute deviation (D_n) as a function of the interval between soundings (I_s) for selected values of number of soundings (N_s), focusing on scenarios mainly appropriate for long-term monitoring sites. For panels (a) and (b) the MATCH-MPIC values are for model output times corresponding to the CEPEX sonde launch times; for panels (c) and (d) the values are based on the entire month of 3-h model output for March, 1993, and the column mean is computed by first interpolating the model output onto 1 km thick layers from 0 to 16 km, then averaging the column (based on eq. (3)).

It is also possible for the $O_3(k, i)$ data in eq. (1) to consist of sample mean profiles instead of individual profiles. For instance, consider a sampling strategy with $N_s = 2$ and $I_s = 4d$ (i.e., each sample mean is the average of two soundings separated by 4 days). For this sampling strategy, there are $N = 216$ ($=[31 - 4] \times 8$) possible combinations for the March model output (less than

for $N_s = 1$ because 4 days must separate the first and last sample). Using these 216 sample mean profiles for $O_3(k, i)$ in eq. (1) results in the profile of d_n shown in Fig. 1. The $N_s = 2$ case shows a clear improvement over the $N_s = 1$ case, particularly in the upper troposphere (UT), leading one to expect that a sampling strategy with $N_s = 2$ and $I_s = 4d$ would result in a profile which on average differs by 15–30% from the actual March mean profile. Another example sampling strategy is $N_s = 10$, $I_s = 1d$; in this case, $N = 176$ ($= [31 - 9] \times 8$, since 9 days must separate the first and last samples). The d_n profile for this sampling strategy is also plotted in Fig. 1. On average, a sample mean ozone profile obtained using this sampling strategy would be expected to differ from the simulated March mean by about 10% over much of the tropospheric column, considerably closer than the $N_s = 2$ or $N_s = 1$ profiles discussed above.

The normalized absolute deviation can be related to the more familiar statistic of the (normalized) standard deviation, σ_n, by:

$$d_n = \sigma_n \sqrt{\frac{2}{\pi}\left(1 - \frac{1}{N}\right)} \approx 0.8\sigma_n, \tag{2}$$

provided the data are normally distributed; if not the relationship will deviate from eq. (2). The approximation on the far right applies for values of N larger than about 10. σ_n is also plotted in Fig. 1, along with the ratio d_n/σ_n. In general, the relationship in eq. (2) holds fairly well, particularly in the lower troposphere, though there are some notable deviations in the UT; this is consistent with the more extensive normality tests done in Logan (1999) (thus readers wishing to know the approximate equivalent value of σ_n for subsequent results in this study can generally multiply d_n by about 1.25).

To facilitate evaluating a wide range of possible scenarios, the average of d_n over the vertical column will be considered. This will be termed D_n, defined as:

$$D_n = \frac{1}{K'} \sum_{k=1}^{K'} d'_n(k), \tag{3}$$

where $d'_n(k)$ is the $d_n(k)$ profile interpolated onto an evenly-spaced vertical grid, and K' is the number of levels in the evenly-spaced grid. Since model output and observations are often on an unevenly spaced grid (e.g., with more grid points near the surface), such a vertical interpolation is necessary to reduce the dependence of D_n on the spacing of the vertical levels on which $d_n(k)$ is defined. For the model output at Christmas Island, the grid used here has $K' = 16$, with 1 km thick layers from the surface to near the tropopause at 16 km. For the three sampling strategies considered in Fig. 1, the D_n values computed from eq. (3) are listed in Table 1 (rightmost three columns). The variation of D_n with N_s, I_s, and T_s will be considered more generally in the following sections.

3. Sampling strategies for O_3 sondes at 5 locations

This section considers O_3 soundings at 5 locations: Christmas Island (2°N, 158°W), Tenerife, Izaña (28°N, 16°W), Bermuda (32°N, 64°W), Azores (38°N, 28°W), and Keflavik, Iceland (64°N, 24°W). The first location was sampled by CEPEX during March, 1993, while the latter four were sampled by NARE during August, 1993. An electrochemical concentration cell (ECC) detector was used during CEPEX, with an instrumental uncertainty of about 0.1 mPa O_3 (Smit et al., 1994), or about 1–2 nmol/mol near the surface, and about 5 nmol/mol in the UT. An ECC detector was also used during NARE, with an estimated accuracy of about 10% for normal tropospheric O_3 mixing ratios, and 15% for very low O_3 levels (Oltmans et al., 1996). The output from a corresponding run by the MATCH-MPIC model (see the Appendix), driven by meteorological data for 1993, is also considered.

3.1. Christmas Island

10 ozone soundings were made from Christmas Island during the CEPEX campaign in March, 1993 (Kley et al., 1996, 1997). The mean of the observed and modeled profiles are plotted in Fig. 2a. There is good agreement between the two mean profiles, and the model captures the main features of low O_3 near the surface, slowly increasing into the mid-troposphere, then falling again to a UT minimum before increasing rapidly near the tropopause (around 16 km). The model does not do as well at capturing the observed variability (Fig. 2b), notably overestimating it in the lower

Table 1. *Column mean values of D_n for selected scenarios*

			Modelled — full month[b]		
Location	Observed[a] $N_s = 1$	Modeled[a] $N_s = 1$	$N_s = 1$	$N_s = 2$, $I_s = 4d$	$N_s = 10$, $I_s = 1d$
Christmas Island (0–16 km)	0.27	0.16	0.25	0.17	0.10
Tenerife (0–15 km)	0.18	0.15	0.18	0.13	0.10
Bermuda (0–15 km)	0.26	0.16	0.15	0.11	0.07
Azores (0–14 km)	0.18	0.15	0.15	0.10	0.07
Keflavik (0–8 km)	0.15	0.14	0.15	0.11	0.08
Keflavik (8–12 km)	0.50	0.37	0.34	0.25	0.19

[a]The values correspond to the curves in Figs. 2b, 4b, 5b, 7b, and 9b, respectively; for the model they are computed from the set of model profiles corresponding to the launch times for the observed soundings; the number of soundings made from each location is given in the text.
[b]The values are computed from the full month's worth of 3-h model output data, for March, 1993, for Christmas Island, and for August, 1993, for all other locations.

troposphere (LT), and underestimating it in the UT. The underestimate in the UT is largely due to the finite model resolution, which smears out the effects of convective events, which stochastically pump O_3-poor boundary layer air into the UT and enhance the variability. This same feature will also be seen later for other locations. It is interesting to note that the variability in O_3 is greater in the UT than near the surface, although its residence time would be expected to be shorter in the tropical marine boundary layer (MBL). This is in contrast to the general conclusions of Gibbs and Slinn (1973) and Junge (1974) noted earlier. However, Junge (1974) makes clear that the relationship should only hold under certain assumptions which can often be applied to the troposphere as a whole, for instance that the variability of the sinks is much smaller than the variability of the sources; separating out the subdomains of the UT and MBL results in a failure to fulfill these assumptions.

The rest of this section considers how D_n at Christmas Island is expected to vary with N_s, I_s, and T_s. For this, the 3-h model output dataset will be used, since the 10 observed soundings are insufficient to produce the statistics required for this analysis. Fig. 2b indicates that one should expect that the mean deviations computed by the model will likely underestimate the deviations that one would expect in reality for the UT, and overestimate those for the LT; in the UT in particular this means that more sampling would need to be done in actuality than indicated by the model to obtain a certain desired D_n.

Fig. 2c shows D_n as a function of N_s and I_s, focusing on the range of values primarily applicable to a campaign scenario, that is, at most a few days between the soundings, and up to 31 soundings in all. On the whole, D_n decreases fairly smoothly as N_s and I_s (and thus T_s) are increased (the small fluctuations are due to random variations in the statistics of the model dataset). It is somewhat counterintuitive that sampling less frequently (i.e., increasing I_s) actually leads to a better representativeness (smaller D_n). This is because increasing I_s while holding N_s constant means that a longer part of the month is being sampled.

On the whole, among the three parameters affecting D_n, T_s is the most important. For a given T_s (recall that $T_s = I_s \times [N_s - 1]$), there is generally not a significant improvement when more profiles are included in the sample mean (within the range of possibilities shown in Fig. 2c). For example, suppose ten days are available for sampling. During this period, six soundings could be made, once every two days, yielding a sample-mean profile which, according to the model, is expected to differ from the mean by about 10%. Launching 21 sondes during a 10-day period (once every 12 h) would provide no notable improvement in the representativeness expected based on the model: D_n would still be about 10%. This is because the diurnal cycle of O_3 over the equatorial Pacific is

small (e.g., Thompson, et al., 1993), and the main variations in the O_3 mixing ratios in this region occur on time scales of several days. Thus, as long as a sample is taken at least once every 4 days, the mean distribution over the limited period being sampled will already be fairly well represented, and adding more soundings during the same period (i.e., launching more frequently) will do little to improve this; the main factor controlling D_n in this case will be how representative the period being sampled is of the entire month. In other regions where a large diurnal cycle has been observed, such as the Indian Ocean (Rhoads et al., 1998), this conclusion would likely not hold as well; in that case daytime and nighttime sampling would be required to accurately estimate the mean.

The behavior of D_n becomes much more complex when longer intervals between soundings are considered, as shown in Fig. 2d. This figure provides information which is generally more relevant to long-term monitoring scenarios. For all cases considered in the figure, the value of D_n decreases monotonically as the interval is increased up to some certain point, then enters a regime of notable fluctuations around roughly constant values until the maximum number of samples which can be made in a month for a certain I_s is reached (i.e., until $T_s = 31d$). The particularly large fluctuations immediately prior to the $T_s = 31d$ cutoff are due to poor statistics (since N in eq. (1) becomes small).

Fig. 2d shows that the point made earlier about T_s being the most important parameter is limited; here it is seen that this only holds up to a certain maximum interval between soundings, which is approximately 4–5 days for this location. This is demonstrated by the extra markings on the figure: the squares indicate the points along each line ($N_s = 1$ to $N_s = 10$) where $T_s = 5d$, while the triangles indicate $T_s = 10d$, and the diamonds indicate $T_s = 20d$. When $T_s = 5d$, D_n is about the same value (0.15–0.17) regardless of the value of N_s (and thus I_s), similar to what was seen in Fig. 2c. However, for the markings indicating $T_s = 10d$ or $20d$ (triangles and diamonds), this only holds if I_s is less than about 4 days. If I_s is increased much beyond this (e.g., the diamonds on the $N_s = 2$ or 3 lines), then D_n increases. Thus, a minimum criterion for obtaining a good estimate of the mean O_3 profile in the equatorial Pacific appears to be that a sounding should be made at least once every ≈ 4 days (i.e., the synoptic time scale).

The synoptic time scale of ≈ 4 days can also be ascertained from the autocorrelation of the model dataset, that is the correlation of the dataset with itself, but lagged by a certain amount of time. The autocorrelation (averaged over the column) as a function of the lag time is plotted in Fig. 3, which shows that the values for r^2 drop steadily from 1.0 at zero lag time to a minimum at about 4 days lag time, indicating the synoptic time scale. In Fig. 3 there is also a peak around 20 days, similar to the peak seen in Fig. 2d in the $N_s = 2$ and $N_s = 3$ lines. This is due to the occurrence of strong convective activity during the first few days of March and around 20 March (Lawrence, 1996), which leads to similar patterns in the O_3 profiles for these two periods, in particular in the lowest few kilometers where subsidence associated with the convection drives the O_3 towards higher mixing ratios. It is unknown whether such a 20-day cycle is a typical feature of the region; a multi-year analysis of this nature would reveal this, and may thus provide valuable information for planning future measurements in this region.

3.2. *Tenerife*

During the August 1993, component of NARE, O_3 soundings were made from four islands in the

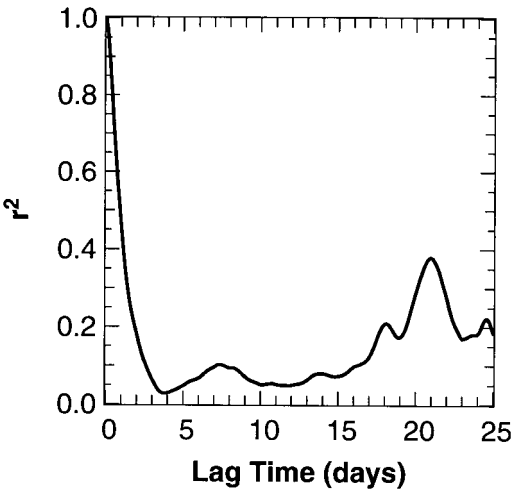

Fig. 3. The column mean autocorrelation (r^2) of the MATCH-MPIC output above Christmas Island; the column mean is computed by first interpolating the model output onto 1 km thick layers from 0 to 16 km, then averaging the column (based on eq. (3)).

North Atlantic. The southernmost of these was Tenerife, Izaña, from which 9 sondes were launched. The observed and simulated means and normalized absolute deviations are plotted in Fig. 4. On the whole, the model reproduces the observed quantities quite well, with the only notable discrepancy being an underestimated variability from 6–11 km. The mean O_3 level is considerably higher than above Christmas Island, while the variability is much less, indicating a lesser influence of maritime convection and greater role of nearby precursor emissions and synoptic meteorology. In this case, more confidence can be placed in the model's predictions than in the previous section.

Fig. 4 also shows the values of D_n computed by the model for various sampling strategies (again only model output is used here, in lieu of sufficient observed soundings). The main points which were brought out in the previous section also hold here.

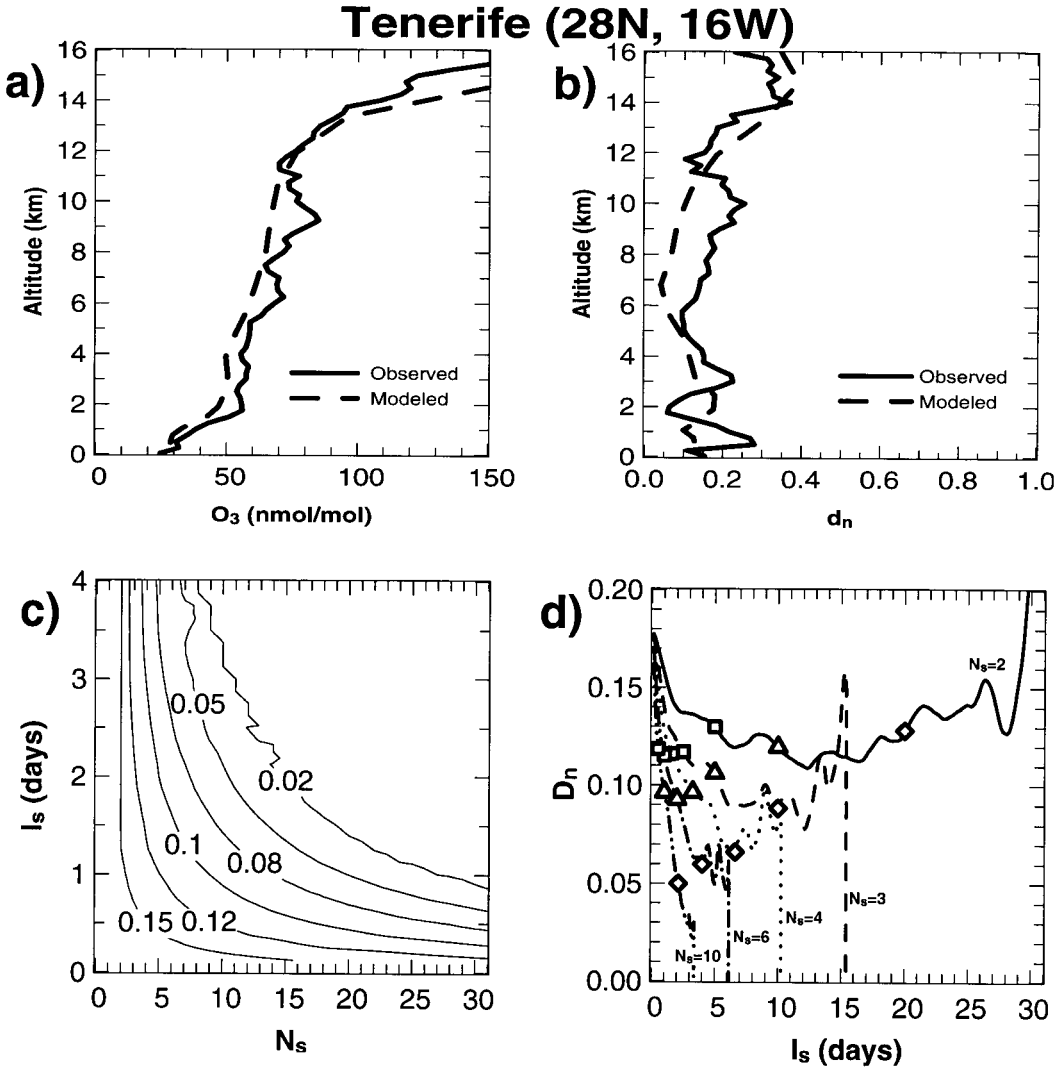

Fig. 4. As Fig. 2, except for Tenerife; panels (a) and (b) use data from the NARE campaign and corresponding model profiles; panels (c) and (d) use only model output from August, 1993, and consider the column from 0 to 15 km for computing D_n.

In Fig. 4c, it is seen that for scenarios with $I_s < 4d$, T_s is the critical parameter. Fig. 4d shows the same behavior of a decrease in D_n with increasing I_s out to about the synoptic time scale (about 4 days), then reverting to fluctuations around a base value which depends on N_s (e.g., ≈ 0.13 for $N_s = 2$, or ≈ 0.10 for $N_s = 3$). Also similar to Christmas Island, the extra markings on Fig. 4d for $T_s = 5$, 10 and 20d show relatively constant values when $I_s < 4d$, and a sharp increase for $I_s > 4d$. As was done in the previous section, the synoptic time scale of 4 days for this location can also be confirmed by the autocorrelation (not shown). On the whole D_n at Tenerife is considerably lower than at Christmas Island, particularly when only one or a few profiles are included in the sample mean (best seen in Table 1). This is likely in large part due to the lesser intensity of convective activity at this subtropical site, and indicates that it is not as difficult to obtain a good estimate of the mean profile here as it is over the equatorial Pacific.

3.3. Bermuda

The next location to the north which was sampled during NARE in August, 1993, was Bermuda, on the other side of the Atlantic Ocean from Tenerife; 19 soundings were made. As for the previous sites, Fig. 5 shows that the mean profile is simulated well, while the variability is generally underestimated, especially in the UT. Bermuda lies in the path of major tropical storms and hurricanes which cross the Atlantic, and is prone to convection due to the warm waters of the Gulf Stream. Furthermore, Bermuda is closer to the subtropical tropopause break than Tenerife, and is thus more prone to the influence of downward mixed stratospheric air in the UT. Specific events reflecting both of these features are documented in Oltmans et al. (1996). These processes lead to an enhanced UT variability, as seen in Fig. 5b. As with Christmas Island, the model significantly underestimates this variability; it also underestimates the variability near the surface, although it does show the feature of greater variability there than in the mid troposphere. The main features of the sampling strategies analysis (Fig. 5c,d) for this location are in good agreement with the previous two sections, with the synoptic time scale of 4 days again delineating a break in the expected quality of the mean profile estimated from possible sampling strategies.

Above Bermuda, unlike the previous stations, sufficient soundings were made to allow the same analysis to be applied directly to the observations. The analog of Fig. 5d, but instead using observed data, is plotted in Fig. 6 (the analog to Fig. 5c is not shown because 19 soundings are still insufficient to fill in the contour space well). The basic character of the model prediction is confirmed in this figure, with the values falling off from the D_n for $N_s = 1$ (≈ 0.25) to a minimum near the synoptic time scale before the onset of fluctuating values for longer I_s. The first minimum in the observations is reached at $I_s = 5d$, later than in the model ($I_s = 4d$). Whether this is due to a deficiency in the model, or instead to statistical variations in the relatively small observed dataset, is unclear. Furthermore, the model underestimates the D_n values, which could be anticipated based on Fig. 5b. The observations show a notable peak in D_n at $I_s = 7$–$9d$; this peak is also present, though weaker, in the model curves for $N_s = 2$ and $N_s = 4$. Similarly, the minimum in the $N_s = 2$ line at 20 days is also reproduced by the model. Though this agreement could certainly be fortuitous, similar agreement is also seen in later sections, and thus is probably due to the model capturing gross synoptic features which influence the variability.

3.4. Azores

The next NARE station to the north is Azores, from which 23 soundings were made during August, 1993. Like at the previous stations, the modeled and observed mean profiles are in good agreement (Fig. 7a), while the simulation also captures the observed variability very well except near the tropopause. Oltmans et al. (1996) concluded that the same stratospheric influx events which were noted at Bermuda also influenced O_3 in the UT above the Azores, resulting in the enhanced variability. Since the model clearly captures the mean level here, but not the variability, it is likely able to simulate the overall O_3 flux to this region, but more smeared out than in reality. The results for the sampling strategy analysis in Fig. 7c and Table 1 are very similar to those computed for Bermuda, consistent with the connection between these two locations suggested by Oltmans et al. (1996). Based on Fig. 7d (as well as the autocorrel-

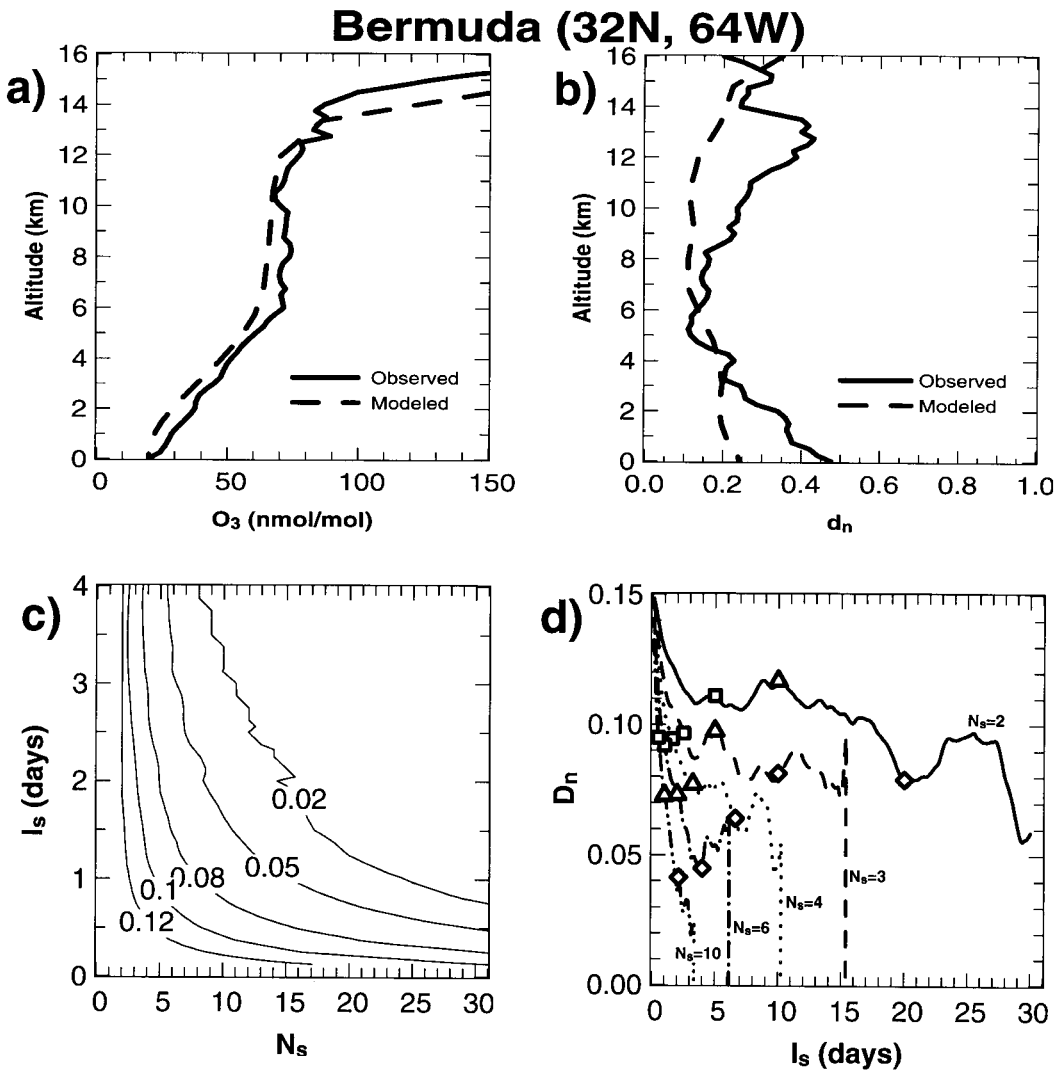

Fig. 5. As Fig. 4, except for Bermuda.

ation, not shown) it appears that the synoptic time scale is about three days, shorter than at the previously-discussed locations.

The observations (Fig. 8) also indicate a shorter synoptic time scale, with the first minimum in D_n occurring at 4 days (compared to 5 days at Bermuda). The values of D_n based on the observations are, not surprisingly, somewhat higher than those predicted by the model. However, as for Bermuda, the model captures certain key features: on the $N_s = 2$ curves, there are peaks at $I_s = 17$ and 24*d* (15 and 23*d* for the model) surrounding a wide minimum, while on the $N_s = 3$ curves there is a minimum at $I_s = 8$–9 days surrounded by peaks at 7 and 12 days (model: 6 and 12*d*).

3.5. *Keflavik*

The final NARE location considered in this analysis is Keflavik, Iceland, from which 22 sondes were launched during August, 1993. Here, as before, the model closely simulates the mean pro-

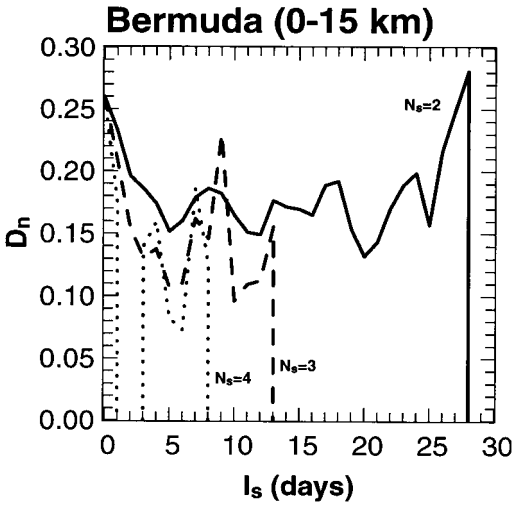

Fig. 6. Sampling strategies analysis based on NARE O_3 soundings from Bermuda: values of the column mean normalized absolute deviation (D_n) as a function of the interval between soundings (I_s) for selected values of number of soundings (N_s). This is analogous to Fig. 5d, except for observations instead of model output. The column mean is computed by first interpolating the soundings onto 100-m thick layers from 0 to 15 km, then averaging the column (based on eq. (3)). (Note that no value could be computed for $N_s = 4$, $I_s = 2d$; no combination of the actual sonde launch times met this requirement.)

file (Fig. 9). It also captures the variability better than at the other locations, in particular the peak near the tropopause, though this is still underestimated. In this case the variability is controlled mainly by synoptic variations (see below), rather than convection, so that one might expect the model to provide a better simulation. Because there are two distinct regimes apparent in the profiles in Fig. 9, the analysis of sampling strategies will be split into two columns, one extending from the surface to 8 km, the other from 8 km to 12 km. For the lower regime, the values of D_n for various sampling strategies with $I_s < 4d$ (Fig. 9c and Table 1) are consistent with the other North Atlantic locations, while the values near the tropopause (8–12 km) are higher than those for any of the other locations in this study.

The sampling strategies analysis plots for the observed data (Fig. 10) are also split into the same two vertical regimes. In the 0–8 km regime, unlike in the previous cases, the D_n values are similar in magnitude to those computed by the model (Fig. 9d). Like for the Azores, both the model and observations (as well as the autocorrelations, not shown) indicate a synoptic time scale of about three days for the 0–8 km regime.

For the 8–12 km regime, the model somewhat underestimates the values of D_n based on the observations, though it captures some of the main features, such as the peaks at $I_s = 8$, 15–19, and 24 days, and the dip around 6–7 days. Both figures show evidence of regular oscillations on the 3–4 day synoptic time scale (the observations more clearly than the model). The very high D_n values and the pattern of fluctuations are due to the nature of this region. Keflavik is located at a latitude that is alternately inside or outside the polar front, which influences the tropopause height. South of the front, the tropopause is generally 12–14 km, whereas north of the front it is generally lower, around 8–10 km. Thus, depending on whether the front is north or south of Keflavik, O_3 levels in the 8–12 km regime are either low or high; north-south movement of the front leads to the very large variability observed here. A similar feature has also been seen at other high-latitude O_3 sounding sites (Kyrö et al., 1992). Fig. 11 shows the variation during August of the "chemical tropopause" height, defined here as the altitude at which O_3 first exceeds 150 nmol/mol (the same qualitative result is obtained when a threshold of either 100 or 200 nmol/mol is used); the heights based on both the model and observed profiles are plotted. The strong oscillations on about a 3-day (peak-to-valley) time scale due to the polar front motion are simulated remarkably well by the model, which lends considerable confidence to using a model like MATCH-MPIC to evaluate sampling strategies in regimes where well-modeled synoptic processes such as this are the key factors controlling the O_3 variability.

4. Conclusions

This study addressed the issue of using a 3D photochemical model along with previous in situ observations to provide quantitative information on the representativeness which can be expected from various sampling strategies. The technique was demonstrated with the example of O_3 soundings made from a stationary launch point; how-

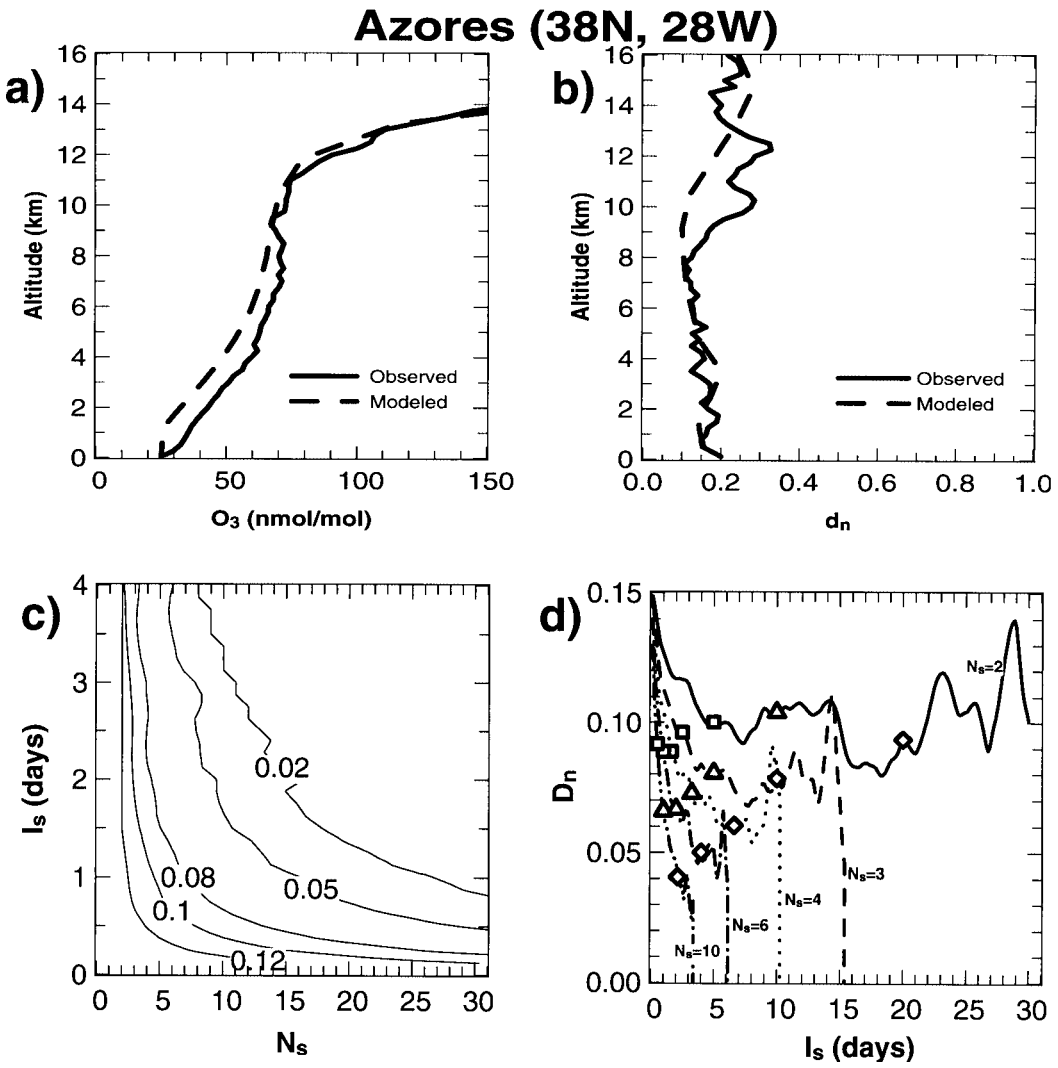

Fig. 7. As Fig. 4, except for Azores, and for a column only extending from 0 to 14 km.

ever, it can also be applied to other types of sampling. Lawrence (1996) considered its use for O_3 sondes launched during a research ship cruise. Other possible applications are to long-term monitoring with flask samples, flight trajectories (especially for commercial aircraft programs such as MOZAIC, Marenco et al., 1998), and perhaps even satellite remote sensing, in terms of choosing orbital parameters. Furthermore, the technique can also be used to estimate the representativeness of other parameters besides the mean, such as the median, quartiles, and the standard deviation; an example of this was given in Lawrence (1996).

In this study, ozone soundings were considered from Christmas Island (March, 1993) and from Tenerife, Bermuda, Azores, and Keflavik (August, 1993). Scenarios appropriate both to field campaigns and to long term monitoring were examined. The model was found to reproduce the observed mean well at all of the locations, and also to mostly capture the variability, except for generally underestimating it in the UT and occa-

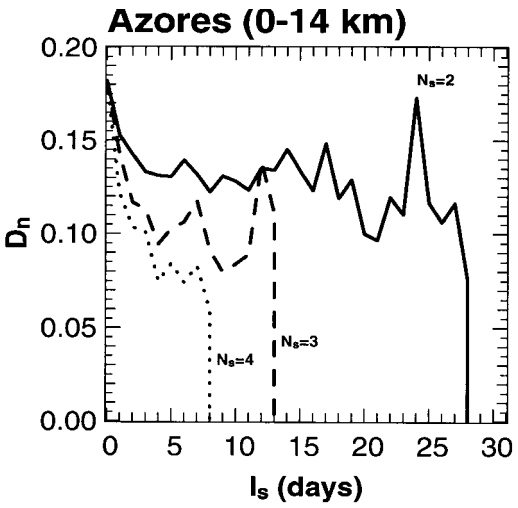

Fig. 8. As Fig. 6, except for Azores observations, and for a column only extending from 0 to 14 km.

sionally near the surface. The underestimated variability is in large part due to smoothing out of sub-gridscale processes such as convection. It may also point towards missing O_3 sinks, such as halogen reactions (Dickerson et al., 1999) or reactions on ice (Kley et al., 1996), which would reduce the lifetime and thus enhance the variability. In any case, for most locations the model appears to be an appropriate surrogate where insufficient observations are available for the analysis.

A key feature found for all the locations was the importance of the synoptic time scale of 3–5 days, which was generally shorter for the higher latitude sites considered here. As long as the interval between soundings (I_s) does not exceed this time scale, then the representativeness depends almost entirely on T_s, the total period from the first to last sample, which is related to the number of soundings (N_s) and I_s by $T_s = I_s \times (N_s - 1)$; for a given value of T_s, changing the values of N_s and I_s does little to change the representativeness of the mean. Once I_s exceeds the synoptic time scale, however, the representativeness becomes worse for increasing I_s (assuming a constant T_s). Furthermore, certain clear structures in the analyses curves indicated principle O_3 variations on various longer time scales (e.g., 20 days at Christmas Island), due to large-scale meteorological patterns; in several cases, the structures evident in the observations were also reproduced by the model. One such structure that was surprisingly well reproduced was the variability near the tropopause above Keflavik due to motions of the polar front. This indicates that the meteorological observations network in and around the North Atlantic is able to constrain a model like MATCH-MPIC to a relatively high degree of accuracy, at least for a synoptically-based process and diagnostic such as this. This also suggests that establishing a network of this density in other regions could significantly improve global atmospheric tracer transport simulations.

The analysis presented here suggests that an "optimal" sampling frequency for O_3 sondes from long term monitoring stations in marine locations with small diurnal cycles is once every 3–5 days (since sampling more frequently would not do much to help improve the estimate of the mean). At many stations, sampling this frequently is presently unrealistic. According to the global O_3 sounding compilation of Fortuin and Kelder (1998), on average about 3 soundings/month are made at long term monitoring stations, though the number varies from less than 1/month at some stations to more than 10/month at others. In reality, this comes from various sampling procedures; at many stations, soundings are made once per week (usually on Wednesdays, the international geophysical observations day), while at others soundings are made only every other week, or in some instances up to three times per week (Jennifer Logan, personal communication, 2000). For the stations considered here, 4 soundings/month (assuming one every 7 days) results in a D_n in the range of 6–13%, whereas 2 soundings/month (one every 15 days) gives a D_n of 10–23%. Although this level of uncertainty is generally quite sufficient for evaluating global chemistry models, it is more difficult to apply towards analyzing trends in tropospheric O_3, which are generally in the 1–2%/year range (Logan, 1994). On the other hand, if a station were to launch 8 soundings/month (one every 4 days), this would result in $D_n < 2\%$, according to the locations considered here. The considerable improvement in the latter scenario is because sampling every 4 days means that the sampling frequency has been reduced to about the synoptic time scale, seen earlier to be a critical factor in long-term sampling accuracy. Note that these

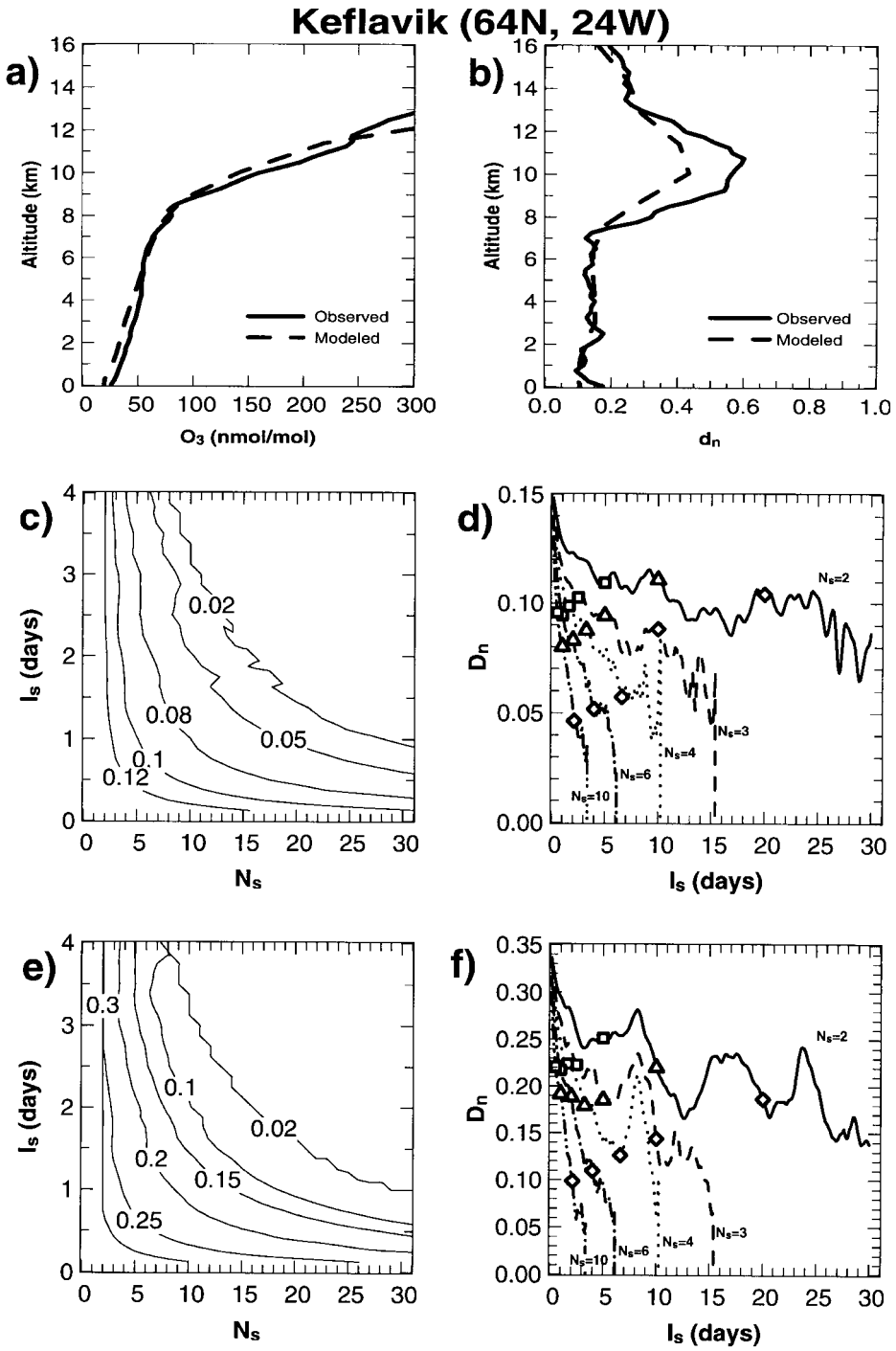

Fig. 9. As Fig. 4, except for Keflavik, and for columns extending from 0 to 8 km (panels c and (d) and from 8 to 12 km (panels e and f).

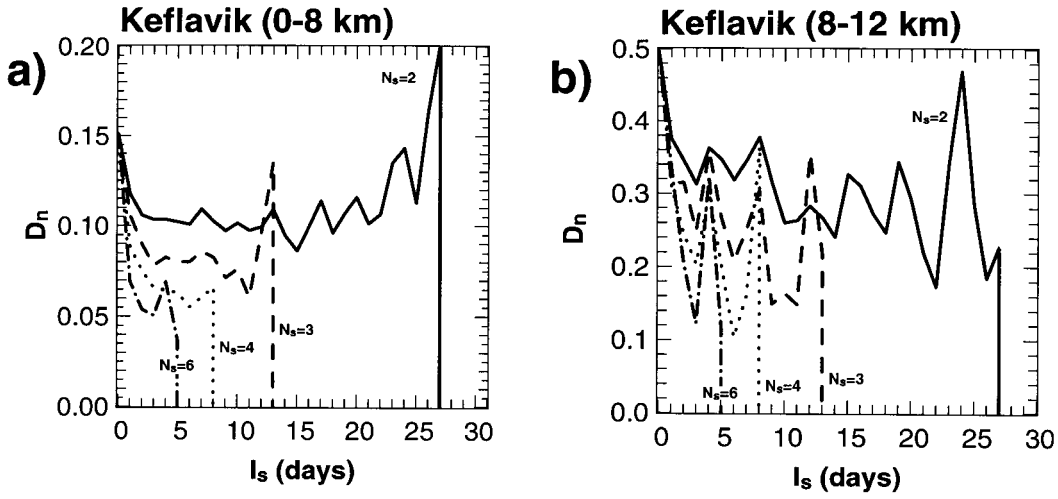

Fig. 10. As Fig. 6, except for Keflavik observations, and for columns extending from 0 to 8 km (panel (a) and from 8 to 12 km (panel (b)).

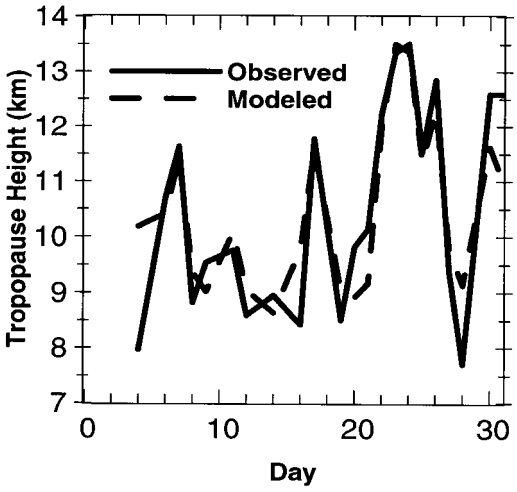

Fig. 11. Observed and modeled values of the "chemical tropopause" height (see text) above Keflavik during August 1993.

results only apply to a single month in a particular year. Interannual variability can also be considerable for O_3. Logan (1999) found that some 20–35 soundings in a given calendar month are needed to establish the interannual mean to within 15%. According to the present analysis, more than 20 (evenly-spaced) soundings in a single month would be expected to give an estimate of the mean profile which is within a few percent of the actual mean for that specific month; thus most of the larger uncertainty indicated by the analysis of Logan (1999) is due to interannual variability.

The main value of this technique is in establishing a minimum amount of sampling which is expected to achieve a desired degree of accuracy. On the other hand, one should be cautious about using the results of this study to argue for limiting sampling to only a moderate rate (e.g., 8 soundings per month), even where the means may exist for more intense sampling. It is important to bear in mind that any decision to limit the amount of sampling, for whatever reason (financial, model-based, or otherwise), comes with the risk of missing rare phenomena and surprises which are always in store in the atmosphere.

5. Acknowledgements

My appreciation goes to my thesis committee, in particular Paul Crutzen, Philip Rasch, and William Chameides, for their advice in the development of MATCH-MPIC and encouragement in finding creative ways to employ it for atmospheric analysis. Thanks also to Geert-Jan Roelofs, Jos Lelieveld, Frank Dentener, Carl Brenninkmeijer, Andreas Zahn, and Jonathan Williams for valuable discussions at various stages of this work, and to Sam Oltmans for insights into the NARE

data. The very thoughtful reviews of Jennifer Logan and an anonymous referee were highly valuable in expanding and improving this study. This study was supported by the Center for Clouds, Chemistry, and Climate (C4), and is C4 publication #214; during this work I was also supported by an NSF Graduate Fellowship, a stipend from the Max Planck Society, and the EU SINDICATE project.

6. Appendix

Model description

The model used in this study, MATCH-MPIC (Model of Atmospheric Transport and Chemistry), is briefly described in this section. The model is described in detail in several other works: the meteorological component is introduced in Rasch et al. (1997), and evaluated with tracer studies in Mahowald et al. (1997a,b); the photochemical component is described and evaluated in Lawrence (1996) and Lawrence et al. (1999a,b). (For reference, this study employs the version MATCH-MPIC 1.2 as discussed in Lawrence et al., 1999b.)

MATCH-MPIC is an offline model. The version employed in this work is driven by wind, temperature, pressure, water vapor, and surface heat flux and wind stress data from the National Centers for Environmental Prediction (NCEP) reanalysis (Kalnay et al., 1996), at T63 horizontal resolution (roughly 1.9 degrees in latitude and longitude), and with 28 vertical levels (from the surface to about 2 hPa); a 30-min time step was employed for the simulation considered here. The model version considered here uses the Semi-Lagrangian Transport algorithm for advection (Rasch and Williamson, 1990); convection is diagnosed online based on the Pan-Wu scheme (which is native to the NCEP model; Pan and Wu, 1995); cloud fractions are diagnosed based on a modified version of the Slingo scheme (Slingo, 1987); and a semi-explicit cloud microphysics scheme (Rasch and Kristjansson, 1998) is used to obtain cloud water amounts and large-scale precipitation.

The MATCH-MPIC tropospheric photochemistry module includes explicit source terms for the major NO_x and CO emissions (e.g., fossil fuel burning, biomass burning, etc.); methane is fixed at the surface based on observations (Fung et al., 1991), and O_3 and NO_y are fixed in the stratosphere based on satellite observations (Keating et al., 1987) and observed ratios between NO_y and O_3 (Murphy et al., 1993), respectively. The chemical transformation processes are handled by employing a quasi-steady-state approach. The reaction rates (from DeMore et al., 1994) are reevaluated each time step. The photolysis rates are also recomputed each time step, based on the method of Landgraf and Crutzen (1998). The dry deposition sink is based on monthly gridded deposition velocities (Ganzeveld and Lelieveld, 1995). Wet scavenging is applied based on the cloud water and precipitation rate parameters from the convection and microphysics parameterizations, with Henry's Law constants reevaluated each time step. Finally, heterogeneous loss of N_2O_5 on aerosols and cloud droplets is included based on Dentener and Crutzen (1993).

REFERENCES

DeMore, W. B., Sander, S. P., Howard, C. J., Ravishankara, A. R., Golden, D. M., Kolb, C. E., Hampson, R. F., Kurylo, M. J. and Molina, M. J. 1994. *Chemical kinetics and photochemical data for use in stratospheric modeling*. NASA JPL (Jet Propulsion Laboratory), Pasadena, California.

Dentener, F. J. and Crutzen, P. J. 1993. Reaction of N_2O_5 on tropospheric aerosols: impact on the global distributions of NO_x, O_3, and OH. *J. Geophys. Res.* **98**, 7149–7163.

Dickerson, R. R., Rhoads, K. P., Carsey, T. P., Oltmans, S. J., Burrows, J. P. and Crutzen, P. J. 1999. Ozone in the remote marine boundary layer: a possible role for halogens. *J. Geophys. Res.* **104**, 21385–21395.

Ehhalt, D. H., Rohrer, F., Kraus, A. B., Prather, M. J., Blake, D. R. and Rowland. F. S. 1997. On the significance of regional trace gas distributions as derived from aircraft campaigns in PEM-West A and B. *J. Geophys. Res.* **102**, 28,333–28,351.

Fortuin, J. P. F. and Kelder, H. 1998. An ozone climatology based on ozonesonde and satellite measurements. *J. Geophys. Res.* **103**, 31709-31734.

Fung, I., John, J., Lerner, J., Matthews, E., Prather, M., Steele, L. P. and Fraser, P. J. 1991. Three-dimensional model synthesis of the global methane cycle. *J. Geophys. Res.* **96**, 13,033–13,065.

Ganzeveld, L. and Lelieveld, J. 1995. Dry deposition parameterization in a chemistry general circulation model and its influence on the distribution of reactive trace gases. *J. Geophys. Res.* **100**, 20,999–21,012.

Gibbs, A. G. and Slinn, W. G. N. 1973. Fluctuations in trace gas concentrations in troposphere. *J. Geophys. Res.* **78**, 574–576.

Gillette, D. A. and Hanson, K. J. 1983. Sampling strategy to obtain data used in models of global annual CO_2 increase and global carbon cycle. *J. Geophys. Res.* **88**, 1345–1348.

Junge, C. E. 1974. Residence time and variability of tropospheric trace gases. *Tellus* **26**, 477–488.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Met. Soc.***77**, 437–471.

Keating, G. M., Young, D. F. and Pitts, M. C. 1987. Ozone reference models for CIRA. *Adv. Space Res.* **7**, 105–115.

Kley, D., Crutzen, P. J., Smit, H. G. J., Vömel, H., Oltmans, S. J., Grassl, H. and Ramanathan, V. 1996. Observations of near-zero ozone concentrations over the convective Pacific: effects on air chemistry. *Science* **274**, 230–233.

Kley, D., Smit, H. G. J., Vömel, H., Grassl, H., Ramanathan, V., Crutzen, P. J., Williams, S., Meywerk, J. and Oltmans, S. 1997. Tropospheric water vapor and ozone cross-sections in a zonal plane over the central equatorial Pacific. *Quart. J. R. Met. Soc.* **123**, 2009–20040.

Kyrö, E., Taalas, P., Jorgensen, T. S., Knudsen, B., Stordahl, F., Braathen, G., Dahlback, A., Neuber, R., Krüger, B. C., Dorokhov, V., Yuskov, V. A., Rudakov, V. V. and Torres, A. 1992. Analysis of the ozone soundings made during the first quarter of 1989 in the Arctic. *J. Geophys. Res.* **97**, 8083–8091.

Landgraf, J. and Crutzen, P. J. 1998. An efficient method for online calculations of photolysis and heating rates. *J. Atmos. Sci.* **55**, 863–878.

Lawrence, M. G. 1996. *Photochemistry in the tropical Pacific troposphere: studies with a global 3D chemistry–meteorology model.* Doctoral Thesis, Georgia Institute of Technology, Atlanta, Georgia, USA, pp. 520.

Lawrence, M. G., Crutzen, P. J. and Rasch, P. J. 1999a. Analysis of the CEPEX ozone data using a 3D chemistry–meteorology model. *Quart. J. Roy. Met. Soc.* **125**, 2987–3009.

Lawrence, M. G., Crutzen, P. J., Rasch, P. J., Eaton, B. E. and Mahowald, N. M. 1999b. A model for studies of tropospheric photochemistry: description, global distributions, and evaluation. *J. Geophys. Res.* **104**, 26,245–26,278.

Logan, J. A. 1994. Trends in the vertical distribution of ozone: an analysis of ozonesonde data. *J. Geophys. Res.* **99**, 22,553–25,585.

Logan, J. A. 1999. An analysis of ozonesonde data for the troposphere: recommendations for testing 3D models and development of a gridded climatology for tropospheric ozone. *J. Geophys. Res.* **99**, 22,553–25,585.

Mahowald, N. M., Rasch, P. J., Eaton, B. E., Whittlestone, B. and Prinn, R. G. 1997a. Transport of ^{222}Radon to the remote troposphere using MATCH and assimilated winds from ECMWF and NCEP/NCAR. *J. Geophys. Res.* **102**, 28139–28152.

Mahowald, N. M., Prinn, R. G. and Rasch, P. J. 1997b. Deducing CCl_3F emissions using an inverse method and chemical transport models with assimilated winds. *J. Geophys. Res.* **102**, 28,153–28,168.

Marenco, A., Thouret, V., Nedelec, P., Smit, H. G., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C. and Cook, T. 1998. Measurement of O_3 and water vapor by Airbus in-service aircraft: the MOZAIC airborne program, an overview. *J. Geophys. Res.* **103**, 25631–25642.

Murphy, D. M., Fahey, D. W., Proffitt, M. H., Liu, S. C., Chan, K. R., Eubank, C. S., Kawa, S. R. and Kelly, K. K. 1993. Reactive nitrogen and its correlation with ozone in the lower stratosphere and upper troposphere. *J. Geophys. Res.* **98**, 8751–8773.

Oltmans, S. J., Levy II, H., Harris, J. M., Merrill, J. T., Moody, J. L., Lathrop, J. A., Cuevas, E., Trainer, M., O'Neill, M. S., Prospero, J. M., Vömel, H. and Johnson, B. J. 1996. Summer and spring ozone profiles over the North Atlantic from ozonesonde measurements. *J. Geophys. Res.* **101**, 29179–29200.

Pan, H.-L. and Wu, W.-S. 1995. Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NMC Office Note, No. 409, 40 pp.

Rasch, P. J. and Kristjansson, J. E. 1998. A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations. *J. Clim.* **11**, 1587–1614.

Rasch, P. J. and Williamson, D. L. 1990. Computational aspects of moisture transport in global models of the atmosphere. *Quart. J. R. Met. Soc.* **116**, 1071–1090.

Rasch, P. J., Mahowald, N. M. and Eaton, B. E. 1997. Representations of transport, convection, and the hydrologic cycle in chemical transport models: implications for the modeling of short lived and soluble species. *J. Geophys. Res.* **102**, 28,127–28,138.

Rhoads, K. P., Kelly, P., Dickerson, R. R., Carsey, T., Farmer, M., Oltmans, S. J., Savoie, D. and Prospero, J. M. 1998. The composition of the troposphere over the Indian Ocean during the monsoonal transition. *J. Geophys. Res.* **102**, 18,981–18,995.

Slingo, J. M. 1987. The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. R. Met. Soc.* **113**, 899–927.

Smit, H. G. J., Sträter, W., Kley, D. and Proffitt, M. H. 1994. The evaluation of ECC ozone sondes under quasi flight conditions in the environmental chamber at Jülich. In: *Transport and transformation of pollutants in the troposphere*, P. M. Borrell, P. Borrell, T. Cvitas

and W. Seiler (eds). SPB Academic Publishing, Den Haag.

Suhre, K., Cammas, J.-P., Nedelec, P., Rosset, R., Marenco, A. and Smit, H. G. J. 1997. Ozone-rich transients in the upper equatorial Atlantic troposphere. *Nature* **388**, 661–663.

Thompson, A. M., Johnson, J. E., Torres, A. L., Bates, T. S., Kelley, K. C., Atlas, E., Greenberg, J. P., Donahue, N. M., Yvon, S. A., Saltzman, E. S., Heikes, B. G., Mosher, B. W., Shashkov, A. A. and Yegerov, V. I., 1993. Ozone observations and a model of marine boundary layer photochemistry during SAGA 3. *J. Geophys. Res.* **98**, 16,955–16,968.