Estimation of surface O$_3$ from lower-troposphere partial-column information: Vertical correlations and covariances in ozonesonde profiles

Robert B. Chatfield$^{a,*}$, Robert F. Esswein$^{a,b}$

$^a$ NASA Ames Research Center, Moffett Field, CA 94035, USA
$^b$ Bay Area Environmental Research, 560 Third Street West, Sonoma, CA 95476, USA

ABSTRACT

Analysis of the spatial correlation of ozone mixing ratio in the vertical provides information useful for several purposes: (a) it aids description of the degree of regionality of the ozone transport-transformation processes, (b) the information provided in the form of a priori covariance matrices for remote retrieval algorithms can simplify and sharpen accuracy of the resulting estimates, and most importantly, (c) it allows a first evaluation of the improvement that remote retrievals can give over boundary-layer climatology. Vertical profiles of mean, variance, and vertical autocovariance, and vertical autocorrelation of ozone mixing ratios were estimated and given parameterizations. The WOUDC ozonesonde network database was used. During the years 2004–2006, these were considerably augmented by sondes taken by NASA, NOAA, and Canadian agencies during recent summertime intensive periods in North America. There are large differences across the North American continent in the patterns and magnitudes of correlation, especially in the lowest 2–3 km of the troposphere. This is especially significant for the near-surface layers (100 s of meters deep) which determine actual surface O$_3$ variability. Sources and mixing processes explain lower-troposphere O$_3$ variability. Near-surface ozone maps can be inferred from 0 to 3 km averages. New satellite remote-retrieval instruments can map near-surface ozone.

1. Introduction

The measurement of atmospheric ozone from space has a long history and retrievals of the total column have become so precise (Liu et al., 2010) that it is natural to use retrievals to map pollutant
ozone near the surface. The characterization of near-surface ozone is important for human health and plant growth (US EPA Ozone (O₃) Standards: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr.html). Promising newly refined methods for remote retrieval of tropospheric ozone (Worden et al., 2007; Landgraf and Hasekamp, 2007; Natraj et al., 2011) allow samples for more appropriate regions, i.e., partial column estimates with 2–3 pieces of information for the whole troposphere. The lowest layer can be about 3 km deep. Description of the vertical profile in higher layers of the troposphere is also relevant. Middle and upper tropospheric ozone need to be mapped for the interest of several agencies (e.g., in the USA, EPA, NOAA, and NASA) in understanding hemispheric and global pollution, particularly as it progressively has increasing effects on the radiation budget and global climate changes. For these reasons, it is important to quantify ozone by altitude and by concentration as accurately as possible, and to understand how relevant partial-column estimates are to the determination of vertical profiles.

One goal of this report is to aid retrieval of the whole profile of tropospheric ozone by describing the sounding-to-sounding means and variances of ozone at a level as they vary—significantly—by geographical location and in the vertical. These aid ozone retrievals. It is easy for the convergence search of estimation procedures to stray to clearly unreasonable solutions since the retrieval problem is poorly posed and susceptible to solutions far from the norm (solutions that are mathematically allowable but not likely in terms of our prior experience) (Backus and Gilbert, 1970). The range of reasonable estimates e.g., mean, standard deviations, and the likelihood that these change rapidly in the vertical, also aids the algorithm by suggesting soft limits for parameter estimates as well as first-guess values. The expected covarying structure of estimated parameters like χ₀ manufacturers have shown how cloud-slicing can enhance retrievals of boundary layer. Nevertheless, statistical relationships of TOR and geographical location and in the vertical. These aid ozone retrievals. It is easy for the convergence search of estimation procedures to stray to clearly unreasonable solutions since the retrieval problem is poorly posed and susceptible to solutions far from the norm (solutions that are mathematically allowable but not likely in terms of our prior experience) (Backus and Gilbert, 1970). The range of reasonable estimates e.g., mean, standard deviations, and the likelihood that these change rapidly in the vertical, also aids the algorithm by suggesting soft limits for parameter estimates as well as first-guess values. The expected covarying structure of estimated parameters like χ₀, between differers levels is also useful in estimation algorithms (Rogers, 2000; Maddy and Barnet, 2008).

Understanding the vertical covariance structure of ozone mixing ratio, χ₀, level to level, helps to prescribe the smoothness of solutions.

A more immediate use of a better description of vertical covariances is to assess the correspondence of the retrievable (a layer average of tropospheric ozone) to the relevant (maps of ozone as it determines daily human exposure). Satellite mapping can serve several purposes to the extent that a retrieval of a partial column can relate strongly to what we will describe as “near-surface ozone,” θₘ. We wish θₘ to have dimensions of a mixing ratio, ppb, and to be a useful description for the ozone budget over many hours of the day, not necessarily a mixing ratio measured at a single site, e.g., a particular existing measurement station. More discussion of the averaging implied by this desire appears below. We used a pressure coordinate system for the vertical, which simplifies averaging; however, the near-surface layer had to respect surface pressure variation. We chose the bottom of the layer to be the 80th percentile surface pressure level p₈₀₆₈₀ and the top of the layer to be p₈₀₆₈₀ − 50, i.e., 50 hPa lower pressure, ca. 500 m higher altitude. The p₈₀₆₈₀ varies from 0 to ca. 125 m off the surface. This allows a pressure coordinate but also describes a region ca. 500 m deep, sometimes slightly less deep. The reasons for this choice will become more evident in a later section on origins of vertical variation of ozone.

2. Background

From early in the era of ozone-remote-sensing instruments, the advent of TOMS (Total Ozone Mapping Spectrometer, launched 1978), there have been repeated observations of features identifiable as regional smog in total ozone imagery (Fishman et al., 1987). Early methods concentrated on removing the large partial column of stratospheric ozone from the TOMS signal, leaving a much smaller partial column of ozone, a Tropospheric Ozone Residual or “TOR” (Fishman et al., 1990). The noisy difference, the TOR, represents at best ozone in the whole the troposphere; correlations between total tropospheric ozone and surface ozone may therefore be sporadic, depending on the variance in ozone above the polluted boundary layer. Nevertheless, statistical relationships of TOR and geographical patterns of surface ozone and surface effects have been noted repeatedly, although surface concentrations are not typically estimated (Hudson et al., 1995; Fishman et al., 2010; Wang et al., 2011). Accuracy of algorithms has improved, and the successor to TOMS, the OMI (Ozone Monitoring Instrument), provides retrievals of total ozone column to a very few percent. A more recent TES (Tropospheric Emission Spectrometer) technique using thermal IR appears to give good resolution of two layers (partial column estimates of ozone 2–5 km, and 5–12 km (Nassar et al., 2008). As a conspicuous example, as Nassar’s Section 6 indicates, in the first release of TES estimates, a large number of retrievals which minimized error also differed greatly in shape and 0–1 km magnitude from results that conform to prior experience and also the initial guess. A common feature of retrieval sets to contain some widely unrealistic appearing profiles. This is probably due to a minimization finding a physically unrealistic minimum (Rogers, 2000), can be ameliorated when there is better knowledge of a mean expectable profile and the typical (We will address this feature of retrievals again in the latter portion of our report.). IASI or future instrumentation may do slightly better than TES (Natraj et al., 2011).

Thermal techniques have nearly zero information at the surface (see Fig. 1), where air and surface temperatures are nearly equal, denying any distinction of a vertical scale. Techniques based on UV alone also have markedly decreased sensitivity near sea level due to rapidly increasing competition of Rayleigh scattering with O₃ absorption (Natraj et al., 2011). However, several retrieval situations, e.g., with low clouds or scattering aerosol can improve the discrimination of near-surface ozone. Ziemke et al. (2001, 2009) have shown how cloud-slicing can enhance retrievals of boundary

Fig. 1. Ability of several instrumental combinations to resolve lower tropospheric ozone, using data from Natraj et al. (2011). In this presentation, the degrees of freedom for signal are integrated from the surface upwards. Green: IR only; Orange: UV + Vis, Red: UV + IR, black dots: UV + Vis + IR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
layer ozone when there is moderate cover by low clouds. Joiner et al. (2009) extend this methodology while emphasizing the need to understand the scattering and absorbing properties of the cloud layer on retrievals. When there are appropriately distributed low-level clouds one may estimate a layer height above the surface and the ozone within that layer. A very different approach uses neural network techniques developed in land-surface sensing to estimate near-surface ozone directly, but the geographical regions and conditions in which this will give reproducible results are not yet clear (Sellitto et al., 2011). The network selected without guidance to employ wavelength regions thought to be uninformative, the Chappuis bands near 600 nm. While such methods provide little physical insight, the result promises that detailed attention to the physics of retrieval will be rewarding.

Data assimilation of ozone is an extremely promising approach to the characterization of near-surface ozone. The technique’s success in the stratosphere the uppermost troposphere and tropospheric column (Stajner et al., 2008; and references) has worked its way down in the troposphere (Doughty et al., 2011) and to the boundary layer (Zoogman et al., 2012). Three aspects of assimilation are worth noting; these suggest that it can be augmented by non-assimilation methods. First, the technique combines observations and modeling (simulations based on “bottom-up” or physical and chemical theory) in ways that can be difficult to disentangle. It adds in information from other times, but these additions necessarily require theoretical understandings of reaction, conservation, transport, and sources (a model, or the synthesis of several models) that may not always apply. Data assimilation faces a persistent criticism that it is just simulation that has been perversely edited so as only not to be contradicted by data, and only when there is data (The authors’ outlook is that this view is oversimplified, but contains some truth which could be emphasized in legal or political discourse.). Secondly, in the case of ozone, assimilation does well on the upper troposphere where chemical timescales are long and flow is predominantly isentropic, broad-scale, and accurate. The lowermost troposphere is more strongly and locally affected by isolated but widely distributed sources of ozone precursors and destroyers. Thirdly, evaluation of assimilation’s success rests on statistical theory that is broadly similar to kinds of inference that we use below. A main difference is that with assimilation, physical theory and statistics are complexly intermixed. Nevertheless, assimilation will play a role in the description of air pollution; we guess that it will work especially well when subsidence (responding to data), and local influences (responding to strong local sources and sinks of ozone) play similar roles. This publication will suggest that more clearly defined modeling, e.g. statistical models of geophysical correlation both have their independent role and also cannot be ignored in understanding the origins of the success of assimilation.

Following this background, our work begins with a description of methods and the quantification of level-to-level means, variance, and vertical-coordinate covariances. In practical terms, this is a portrayal of correlations between pairs of 50-hPa thick layers all throughout the troposphere. This description moves to the mathematical descriptions of the variance-covariance matrices themselves. We publish spline-function descriptions in an attempt to report them systematically for use in retrieval calculations. This method is relatively free of dependence on our sampling grid, and so should be easy to apply on other vertical grids.

Details of the variation seen at more polluted stations differ from the others and have implications for the practical usefulness of retrievable ozone data. Consequently, we outline a review/synthesis of current knowledge about sources, meteorology and chemistry of the lowest kilometers of the troposphere that may simplify the reader’s interpretation of the different stations.

Finally, we examine our central question, the relationship of near-surface ozone to layer-average ozone in the lower troposphere, lower-tropospheric partial column O$_3$, or O3pC. Following recent results on retrievability (Natraj et al., 2011), we take O3pC to measure mean O$_3$ for the region 0–3 km, expressed as an equivalent mean mixing ratio. These lower-tropospheric correlations appear to exhibit broad regional patterns; we speculate on effects that forthcoming local studies may confirm.

3. Variance and vertical autocorrelation in the troposphere and stratosphere

In our analysis, ozonesondes were used to describe the structure of the atmosphere for the troposphere and up to approximately 25 km in the stratosphere. The ozonesondes launches chosen for analysis were sponsored or coordinated by three main groups, the NOAA CMD (Global Monitoring Division, Samuel Oltmans), the Canadian Atmospheric Environment Service (David Tarasick, Environment Canada, Air Quality Research Division), or NASA’s continuing IONS project (Thompson et al., 2010). Most of the stations are in remote or lightly populated sites. Exceptional stations, in polluted areas, were in Houston, TX, Beltsville, MD, and Boulder, CO. These stations were in the center of a large highly industrialized city, in a suburb of Washington DC, and in a smaller city near a larger metropolis, Denver, CO. A summary of the measurement program through 2008 may be found in Tarasick et al. (2010) and Thompson et al. (2010). A broader summary of ozonesonde statistics may be found in Filmes et al. (2011). Data was assembled from the World Ozone and Ultraviolet Radiation Data Center (WOUDC) and the NASA Goddard sites for the IONS projects.

Examination of the variations of sondes profiles in the stratosphere suggests that the statistics on the important amount of ozone above 22–26 km is now better described by satellite retrievals exploiting limb scanning such as the Microwave Limb Sounder than by sondes in the region ∼22–30 km, where demands on pumps in the ozonesonde instrument become extreme. The statistical description of ozone’s probability distribution of values has two aspects which we attempt to harmonize. From the viewpoint of retrieval, the Rodgers outlook stresses an understanding of a priori (current knowledge) mean, variance, and vertical covariance. The former is a vector of means, the latter two are combined into a covariance matrix (or the related error-covariance matrix). More detailed information on the distribution is desirable but not often available (Rodgers, 2000, chapter 4). The natural measures are related directly to the effects of O$_3$ on radiance, i.e., partial columns measured in mole cm$^{-2}$ or in Dobson units ($2.69 \times 10^{16}$ mole cm$^{-2}$). These are also additive to a total O$_3$ column. However, for human comprehension, it makes sense to use O$_3$ mixing ratios in ppb (γ as indexed by vertical level above the surface $\gamma_i$; the specifier O$_3$ used above, X$_O$, will be omitted from here on). These work more intuitively in considering transport processes and chemical transformation. We will harmonize these views by describing the mass-averaged mixing ratio within a characterized level $\gamma_i$ where the averaging relationship for layer $i$ bounded by pressure levels $p_b$ and $p_l$ (in hPa) is

$$x_i = (p_b - p_l)^{-1} \int_{p_b}^{p_l} x(p) dp$$

and

one DU of ozone $= 789.14 \int_{p_b}^{p_l} x(p) dp$.

The use of mixing ratios also facilitates the adaptation to ones own most suitable set of levels. Logarithms of ozone are also used.
for iterative retrieval calculations; for these additivity does not hold
and near-conservation of column estimates might be considered as
one test of accuracy.

Means and standard deviation estimates over many samples for
\[ \text{Mean}(\chi_i) = \chi_i \] and \[ \sqrt{\text{Var}(\chi_i)} \]
of a selection of the sondes that we
analyzed are graphed in Fig. 2. These describe the atmosphere up to
approximately 55 hPa. The mean and variances were derived and fit
with spline curves to this level, as shown. These statistics are useful
in remote sensing studies, as they can define a priori values for the
retrieval. The variance estimates associated with the standard
deviations shown are also useful in defining the strictness or laxity
that should be accorded to estimation at each level. For example,
wide variations allow the algorithm to define a high-tropopause
situation (low ozone mixing ratio) or a low-tropopause situation
(high mixing ratio).

These regions of high variance are functions of both latitude and
longitude, as the Fig. 2 shows. Stations to the west of Valparaiso in
the Midwest and north of 35 N show a region with wide variation
and a defined “knee,” or bend, at 175–250 hPa, that is transitional
to the stratosphere.

Fig. 3 describes the spatial ozone autocorrelation structure in
the troposphere. While autocovariances determine the mathem-
atics of retrieval, autocorrelations are more easily understood by
the human eye. The autocovariance matrix is easily regained by
multiplying the autocorrelation matrix by a diagonal matrix con-
sisting of the variances at each level (Incidentally, all calculations
describing autocorrelations were derived by reversing this process.
All operations used the statistics of the autocovariances of layer
partial-column ozone, since variances add but correlations do not.)
The variation above and below the reference level is not symmetric
in Fig. 3, since these portrayals of the autocorrelation describe the
columns of the autocorrelation matrix, not the cross-diagonals
(Symmetrical graphs may be obtained by a measure of “off diagonal
distance” between pressure levels. This portrayal is the basis of the
numerical fits we describe below in the Supplementary Material.
No graphs of these are shown.) The figures show the relationships
centered on a single level and the autocorrelation with levels above
or below. For several of the stations and reference altitudes, the
correlation dies off rapidly with altitude, for others not.

Glancing across the first row of local autocorrelation plots in
Fig. 3, one can see that the breadth (vertical separation) of the
autocorrelation traces tend to become broader as the stations
progress from west to east. This is most obvious for the traces
describing the near-surface reference levels (long trace line above,
short one below the reference point), by 500 hPa, mid-troposphere,
the broadening is much less. Indeed, the jaggedness of the auto-
correlation estimates down near 0.3 or less suggests that sampling
error is becoming important.

![Fig. 2. Profiles of mean ozone mixing ratio (ppmv) with standard deviation indicated to the left and right of means, summertime six months 2004–2006. The solid lines indicate spline fits made for the mean and variance in ppmv terms. Tarasick et al. (2010) and Thompson et al. (2010) for description of stations and dates of soundings.](image-url)
4. Parameterized covariance estimates for remote sensing

The Supplementary Material gives spline fit coefficients for the profiles of mean and (square root of) variance for the North American sondes set some of which are shown in the graphs of Fig. 2. Variance estimates were obtained from layer partial-column ozone and normalized by layer depth to form an equivalent mixing ratio. Then \(\log_{10}(c_{ij})\) the variance diagonal of the autocovariance matrix was spline fit in terms of \(p = \log_{10}(\text{Pressure})\): a sample fitted function according to this formula is shown in Fig. 4 in terms of the standard deviation. Fig. 4 illustrates the judgments we made regarding the need to spline-fit sampling variability and structure evident from several sounding sites. The spline fit parameters are listed in the Supplementary Material.

We took care to provide estimates that could be easy to move to another grid. For example, separation between levels can be described in differences between the center pressures of the layers that we chose, layers that fit our dataset. Variance and autocorrelation expressed this way can be adapted to other grids. One note: although we find it easy to graph and describe correlations and plot standard deviations, all formal estimation is based on estimates of means, variances, and autocovariances. The statistics of means and variances add, and therefore they average correctly (Graphs showing a satisfying symmetry of autocorrelations and autocovariances around a center diagonal are obtainable using the user’s measure of “off diagonal distance” measured by difference of pressure levels. This portrayal is the basis of the numerical fits we describe below in the Supplementary Material).

It is a frequent practice to use \(\log(\chi)\) rather than \(\chi\) in retrieval practice (Natraj et al., 2011). More properly, the log of partial optical path (e.g., Dobson units) is the fundamental quantity, but this can be related to the appropriately pressure-averaged mixing ratio units. One good feature of logs is that the distribution of ozone amounts in the region of high variability at and above the tropopause is better described: high standard deviation in \(\chi\) is compensated by high mean. Some stations shown in Fig. 2 have extreme standard deviations that do not describe the distribution well, implying at worst negative concentrations. However, we found that use of logarithms required us to loosen the strict additivity used for non-log estimates. Still, all this estimation was

Fig. 3. Vertical autocorrelation scales in the troposphere. These describe the columns of the autocorrelation matrix \(r_{ij}\). Legends indicate the “reference level” for each column of the correlation matrix (see text). Similarities between plots help suggest statistical variation; variations in the progression to lower values also indicate sampling variance of the estimates. Error bars are somewhat elusive: neighboring estimates have more combined certainty than individual estimates.
To avoid repetition in discussing several stations and features, is

carried out using the basic rules that pressure-layer depths, and
means and variances have simple summation rules, and so produce
appropriate averages. Since the estimation procedure converges on
a state vector described by $\log(\chi)$, these are taken as additive once
the narrow-layer averages are established. These estimates are
available from the authors.

5. Origins of the vertical variations of lower tropospheric $O_3$

Our focus now turns from broad perspectives of process of
estimation of atmospheric ozone with maximum accuracy to the
topic of how well retrievals can characterize smog ozone for
exposure and regulation. In the following sections, we will discuss
some recurrent but perhaps unexpected trends in mean, variance,
and covariance in the first few thousand meters above the surface.
To avoid repetition in discussing several stations and features, is
useful to provide a review synthesizing aspects of lower-
tropospheric boundary layer meteorology, reaction, and sources
that have proven useful in our current analysis. It is particularly
appropriate to an analysis of urbanized regions like Baltimore-
Washington region (the Beltsville, MD, sondes, also, Chatfield
et al., manuscript in preparation, 2012). Fig. 5 shows mixing in the
lowest kilometers of the atmosphere over a land surface and over
a period of three days of fair weather without frontal passages or
thunderstorm activity. These conditions appear to be useful in our
analysis of several features of the ozone mean and variance
statistics. We will suggest that stations with other behavior have
plausible differences usefully describable as contrasts or additions
to the picture.

During each of the three days shown, starting soon after dawn,
air from the surface to the mixing layer top is carried by thermals
and other eddies that also generate smaller eddies which complete

![Diagram](image)

**Fig. 4.** Sample plots of spline fits to autocorrelation data for Beltsville. (a) Autocorrelation in the lower troposphere (centered on 780 hPa) is maintained for several hundred hPa and then drops precipitously. (b) Autocorrelation near the mean tropopause (238 hPa) has very short length scales. Green dots show the spline fits described in the Supplementary Material. Red dots show the individual autocorrelation estimates. Three adjacent pressure levels served as the basis for each spline fit; hence the large number and scatter of the dots. The central pressure level (e.g., 350 and 238 hPa) had significantly stronger weight. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

![Diagram](image)

**Fig. 5.** "Retrievable" and "relevant". A cartoon depiction of the daily cycles (slightly variable) of PBL-inversion height, convective eddies and appearance of full mixing vs layering of air pollutants. Vertical axis is distance from the surface, while horizontal axis shows time progressing through three days and also some horizontal features such as eddy-turnover. The region marked with yellow shows a ca. 3-km region that can be reliably characterized from space, while the regions highlighted in green shows a region strongly useful the characterization of air pollution on scales of several kilometers wide, i.e., the retrievable and the relevant. The afternoon mixed layer reaches various heights on the different days, as the planetary boundary layer height responds to surface heating and also to possible large-scale air subsidence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
the diffusive mixing. Mixing progresses over 0.4 h (initial plume transit bottom to top) to <2 h (substantial homogenization if chemical reaction does not interfere). Several factors make the behavior similar and generalizable among days with more or less surface heating and subsidence at the mixing layer top. 20–45% of mixing-layer air is in the updraft, as the thermals expand and dilute during their rise (Stull, 1987; Chatfield and Brost, 1987, and references). Typical dimensions of whole thermal regions are ~1.5 × broader than their height (Stull, 1987; Lothon et al., 2006); the general size is therefore determined by the boundary layer top. As thermals penetrate slightly into the inversion at the planetary boundary layer top, they gradually raise the boundary layer depth over hours, as shown. The boundary layer top in Fig. 4 is portrayed as a single thin line, but occasional small pockets of previously lofted air are trapped in a region of stable air just above the line in the boundary layer inversion: these quiet regions preserve concentrations of air mixed at a previous time or, occasionally, bits of relatively unmixed near-surface air. Often the inversion continues to rise with further thermals, and these regions are entrained into the underlying mixed layer. Transport above the mixing layer top is indicated very simply in the figure. Four important processes require mention: (i) oftentimes the stability to overturning at the mixed layer top allows differential advection, so that, e.g., high concentrations of pollutant ozone are replaced by low concentrations from regions exposed to a cleaner boundary layer; (ii) small fair-weather clouds can mix pollution upward from the top, forming a semi-mixed region with some character of the mixed layer; and (iii) synoptic weather patterns can either act to lift the air pollution to 2 or 3 km, typically involving clouds of only hundreds of meters deep, or (iv) compensating subsidence regions produced by synoptic weather can push down the mixed layer top, as shown in the second day of Fig. 4.

How do ozone variations relate to all this? Ozone is a secondary pollutant species, so its production occurs very broadly throughout the mixed layer and above. Production is due to chemistry that varies in timescale from ~6 h with particularly concentrated emissions of VOC and NOx pollutant in optimal proportions to ~10 days with little pollution; Ozone-determining pollution enters the boundary layer at very low levels, 0–3 m for nitrogen oxides and VOCs associated with traffic, and a few hundred meters for industrial power plants (which are widely scattered but significant sources). Both industrial and traffic emissions of NO act to destroy O₃ very locally, via the reaction NO + O₃ → NO₂ + O₂, diminishing it from a small percentage all the way to 0 ppb (complete titration of ozone). Regions of significant titration are limited both horizontally and vertically during the day. Unfortunately, most O₃ monitors are placed at a few meters altitude, and a significant fraction can be under the strong influence of one or several nearby NO sources. Some monitors are more free of these effects, but we know of no systematic discussion or characterization of monitors. All are located at the bottom of the mixed layer, an ozone destruction region which responds to the full mixed layer ozone budget. Thermals and other eddies carry NO aloft and its effects in titrating O₃ decrease as the thermals expand and dilute.

Above this region, variations in tropospheric ozone derive from a mixture of effects: older smoggy or clean areas lofted by mid-latitude storms in warm fronts, wisps and airmasses of clean and dirty air from far upwind (Europe, Asia, the Gulf of Mexico, ...), and streamers of O₃ from the stratosphere that have had variable amounts of dilution (Newchurch et al., 2003; Li et al., 2005; Doughty et al., 2006).

Some implications of this discussion are that (i) the vertical region that determines ozone levels upon which daily and hourly effects build is conveniently approximated as at least the depth of the highest afternoon mixed layer, ~1–2 km, the region that is fully mixed in a day’s time and (ii) variability in ozone should commonly decrease in the vertical from the surface to several km, as the destruction and production effects of precursor emissions die away; (iii) since the regions in the first km of the Earth’s surface tend to influence regions above with diminishing effect, satellite resolution of concentrations characterizing a several-km deep region may be quite informative about the first kilometers, where the budget of ozone affecting surface conditions is determined. All of these generalities deserve detailed analysis for particular sites; in the following sections they help to form an impression on the extent to which such a narrative may be useful.

6. Correlation near surface and individual levels aloft

How well might we characterize the lower troposphere? Evidence from simulation studies regarding methods to estimate lower tropospheric column ozone using the different information available from several wavelength regions gives us reason to expect that remote retrieval of ozone can be relevant to smog pollution, but also sets limits on how well we can estimate surface ozone. Fig. 1 represents from a new perspective the information on some estimated capabilities of remote sensing of (Natraj et al., 2011) to frame the question. It portrays a current understanding of our ability to resolve ozone mixing ratio in layers of the atmosphere extending from the surface up, under the conditions of clear skies and low aerosol loading (Cloud and aerosol effects can help as well as hinder retrievals of lowermost-troposphere ozone, as current work by Natraj and the GEO-CAPE remote-sensing sensitivity team will show). A common situation of stronger aerosol influence, where there is a scattering, not too absorbing, layer near the surface, can significantly enhance resolution near the surface. Photons returning to the satellite sensor, have a longer path length in the layer (V. Natraj, X. Liu, personal communications, 2011, manuscripts in progress).

The vertical lines of Fig. 1 are of particular significance. The line showing “Degrees of Freedom for Signal” (DoFS) = 1 roughly indicates a region that is fully resolved. For the most informative combinations of wavelengths and instruments this occurs with layers extending from the surface to 650–690 hPa, or about 3 km. This is a characterization of the radiation physics and retrieval methodology examined, and is a quantification of a transition behavior; i.e., retrievals can still have some sensitivity to large ozone changes somewhat above the layer, and retrievals with only DoFS = 0.8 can still be primarily sensitive to ozone below 3 km. Since DoFS depends on radiational physics, instrumental noise, and retrieval numerics, it is not sensitive to the actual variability of ozone in the vertical due to geophysical processes. Variation in a retrieved layer can be predominantly due to the most variable sub-regions; we shall see that these regions are often near the surface.

Consider now the correlation of surface ozone with ozone at various levels aloft. The red lines of Fig. 6 show the correlation coefficient r(χ₀, χₜ) of near-surface ozone mixing ratio (nsO₃): χ₀ with ozone aloft, χₜ, at various 50-hPa layers for three ozonesonde stations. Note the large variation in the plots. The vertical spatial autocorrelation is quite small at the westernmost region, Trinidad Head, CA. In mid-country, the correlation remains high until 850 hPa but also sets limits on how well we can estimate surface ozone. The budget of ozone affecting surface conditions is determined. All of these generalities deserve detailed analysis for particular sites; in the following sections they help to form an impression on the extent to which such a narrative may be useful.

The black lines in Fig. 6 describe a linear “transfer function” b = Cov(χ₀, χₜ)/Var(χₜ) which can be used to estimate variations aloft in
Fig. 6. (Red lines) Correlations of surface ozone, \( r(\text{ns}O_3) \) with ozone at the indicated levels in the lower atmosphere. (Black lines) “Amplification factors,” or transfer function coefficients, \( b \), by which nsO\(_3\) concentrations are to be multiplied so as to estimate ozone at higher altitudes. Green dots: Standard deviation of ozone at the indicated levels, ppb, with a scale factor of 0.1. Notice the strong differences between the extreme West Coast station, Trinidad Head, CA (a), affected by frequent subsidence and strongly laminated composition, as compared to the stronger correlation with the East Coast site at Beltsville, MD (c), which experience frequent episodes of airmasses with similar ozone chemistry up to \( \sim 3 \) km (700 hPa). The Huntsville, AL (b), trends are intermediate, possibly indicating an alternating influence of recirculating East Coast and Southeast Central air pollution with an influence of cleaner Gulf Coast air, particularly above \( \sim 2 \) km, 800 hPa. Values of \( b \) greater than 1 may be attributed to the greater productivity on O\(_3\) with rising and dilution of precurors (Chatfield and Delany, 1990), as well as more local production/loss processes near the ground. The latter could moderate and dilute larger scale regional variations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Inference of nsO\(_3\) from retrievable average lower tropospheric ozone

While level-to-level correlations describe the climatology of ozone correlations in some detail, we can examine a more basic question on the relevance of remotely sensed ozone to the analysis of smog. Satellite retrievals, especially from nadir-facing instruments, retrieve layer averages with sensitivity functions weighting some individual levels more heavily. These sensitivity functions (or layer-oriented averaging kernels) may vary with atmospheric conditions such as aerosol scattering or temperature lapse rate. Consequently we will describe a simple general relationship of nsO\(_3\) with these layers as if the weighting were uniform with the number of molecules sampled throughout a layer. We will simplify the view presented by Fig. 1 by assuming sensitivities can be represented by straight lines approximating the curves over the appropriate interval. This approximation is reasonable for the curves marked in orange, red, and black, as they vary between DoFS from 0 to 1. For the green curve, describing the use of thermal radiation only (instruments such as TES or IASI), the linear 0-to-1 DoFS approximation might be better taken as extending between ca. 900 hPa and 450 hPa.

Fig. 7 shows the correlation of near-surface (thin-layer) to lower tropospheric partial-column ozone directly for twelve North American stations. Our goal is to predict \( x_0 \), near-surface ozone mixing ratio (nsO\(_3\)), using quantities that can be retrieved using satellite radiances. We want to understand the best linear-regression predictor of \( x_0 \), and the confidence we have in the relationship. The annotation “Predict this layer” thick short red double-headed arrow describes \( x_0 \), and all the graphed quantities refer to aspects of its prediction.

We expect to use various remote retrievals of partial column lower tropospheric average mixing ratio \( \text{prctO}_3 \) or \( \text{prctO}_3 \), the overbar in \( \text{prctO}_3 \) emphasizes the thickness of the averaging layer. “Optimistic” estimates: the long red arrows describe a region with DoFS = 1 using the best estimators, i.e., the multi-instrument retrievals (red and orange lines and dotted black line in Fig. 1).
Another region above, not marked by an arrow, can also be estimated with DoFS = 1 (These regions are also marked in Fig. 1.). “Conservative” estimates: the region marked by the green arrow is an approximation to a region using very low noise thermal sensing, slightly better than are available from TES or IASI today (Natraj et al., 2011). Conservative estimates are available with current technique. Early versions of the Optimistic estimates have been outlined in theory (Worden et al., 2007) and satellite retrievals are now being improved (Xiong Liu, Kelly Chance, personal communication, 2011).

Using terminology similar to the previous section, we graph the several quantities. The correlation coefficients $r_{(10,17)}$ shown by the thick angled red and green lines deserve first attention. If they are high enough, use of satellite data deserves our attention. The actual
estimates of nSO\textsubscript{3} variation are given by $\chi_0 = a + b\tau_7$, where $b = \text{Cov}(\chi_0, \tau_7)/\text{Var}(\tau_7)$. The intercept can be estimated by linear regression over the sonde samples or equivalently as $a = \text{Mean}(\chi_0 - b\tau_7)$. The first six diagrams describe correlations ranging across the middle of the continent from Pacific to Atlantic.

The next three sondes diagrams include examples extending to the northeast from the US Midwest, by Walsingham southeast of Toronto, up to Sable Island in the Maritimes, a station commonly under strong influence of outflow of North American pollution in the summertime. The last three sondes describe the southern United States. Holtville, CA had sondes in only one month, August, 2006, although Houston, TX, and Huntsville, AL, had repeated measurement campaigns. We do not venture into mathematical statistics describing confidence intervals in this work, since the underlying source of “error” variation has temporal autocorrelation mostly described by weather, which has autocorrelation structure over hours, days, months, and year-to-year (Sample calculations we made which added 5% or 10% sample-to-sample completely random error, $\sigma/\mu \approx 0.05$ to 0.10, suggested only minor changes in the estimates.).

A broader view of these correlations is mapped out over temperate North America in Fig. 8, using the complete set of sondes over mid-latitude North America, and using a “sensed” layer that extends from 0.5 to 4 km. Fig. 8 shows broad similarities among sites geographically near to each other with the most general variation extending from northwest to southeast across North America. However, regions near bays, oceans or lake shores, particularly in the Great Lakes region, show smaller correlation than other regions.

The patterns we see in Fig. 8 suggest a narrative of regional meteorology on ozone correlations involving near surface layers. We expect that this description can be tested and expanded by case studies and further statistics. Jensen et al. (2012) provide an example of such a statistical approach. Rather than establish particular descriptions of correlations at individual stations, we wish to motivate a sense of weight of evidence: correlations appear to vary geographically with reasonable patterns. If this is so, neighboring estimates tend to provide mutual reinforcement of our confidence. For example, Trinidad Head is frequently under the influence of the eastern, subsiding side of the permanent Pacific Anticyclone. The effect of the subsidence is to produce a stable atmosphere in which lamina remain unmixed but spread out horizontally (thus becoming thinner) as they descend. Very little vertical correlation of ozone producing or destroying regions remains. The Eastern Seaboard from the Mid-Atlantic northwards is characterized by migratory anticyclones and cyclones. Air from the region and points to the west can form ozone in regions 1–3 km thick (adding to a 1–1.5-km boundary layer produced by dry convection is common mixing upwards by hundreds of meters) due to fair-weather cumulus clouds. The anticyclone may remain and recirculate in some circumstances. When the migratory anticyclone is displaced by a warm frontal region, polluted air will move generally toward the northeast on the western side of the high and rise, often toward a cyclonic system. Fair-weather clouds can add additional lofting. These motions produce a vertical structure in which different altitudes have ozone concentrations produced under broadly similar conditions. The interplay of ozone production and mixing produces variations within a larger scale general rise of ozone levels. When cool air-masses move in, concentrations can initially lower as relatively unpolluted air from the Great Plains and the Central and Prairie Provinces moves in. Alternatively, less polluted air can move in from the Atlantic or the Gulf of Mexico, with the greatest effect near the coasts (Such maritime air may also be polluted in recirculation that extends offshore.). Similar versions of this explanation obtain across continental North America. There are stations that exhibit lower correlations, 0.77–0.79. Commonly, these are regions in which lower tropospheric winds are likely to take air over land-sea or land–Great Lake trajectories within the past few hours, presumably creating internal boundary layers and structures. When it is likely that near-surface air is exposed to multiple land–water transitions (e.g., Walsingham, ON, Egbert, ON, Pellston, MI, the ship RMV Ron H. Brown, when stationed near New Hampshire in the Gulf of Maine, Sable Island, NS, Narragansett, RI) correlations seem particularly low, 0.7–0.77. On the other hand, Boulder, CO, is well known to exhibit frequent sharp transitions in at layers where westerlies flow unpumped past the high wall of the Front Range. Downtown Houston, TX, and Valparaiso, IN, near Chicago exhibit somewhat lower correlations, 0.84–0.85. Although the statistics of the sampling variance of the correlation coefficient can be complex, the geographical patterns generally suggest that correlation patterns will reproduce no worse than ±0.03.

8. Conclusions

We presented a first attempt to relate satellite “retrievable” quantities to measures of O\textsubscript{3} that are “relevant” to air pollution assessment, mitigation measures, and forecasting. Our statistics (Figs. 6 and 7) reinforce the conclusion that full-troposphere-residual (TOR) estimates rarely are quantitatively useful in describing near-surface ozone measures that are useful for air pollution surveys; when middle and upper troposphere variations are minimal or averaged away, they can be useful (Fishman et al., 1990). The evidence of the ozonesondes we have analyzed shows that ozone mixing ratios in our analysis exhibited considerable correlation in the vertical, especially in the lowest kilometers of the atmosphere. Since it appears possible to retrieve partial column lower tropospheric ozone (pctO\textsubscript{3}) in regions approximately 3 km above the Earth’s surface in clear skies, this analysis suggests very substantial skill in describing near-surface ozone nSO\textsubscript{3}, defined as the layer average approximately 500 m above the surface. This skill
is sometimes better than simple vertical autocorrelation function graphs seem to suggest (see Fig. 4), due to the interaction of sources, chemistry, and meteorology which we cartooned in Fig. 5. This approximate description merits additional study concentrating on individual stations and their weather. There are two contributing reasons (a) the layer averaging tends to smooth variations in autocorrelation seen at finer scales, and (b) the frequent decrease in variance with distance from the surface (Fig. 2, notably the stations with stronger precursor emissions) tends to emphasize the contribution of the NO_{3} layer and other layers relatively near the surface. Locations with high variance near the surface provide considerable information for covariance analyses, and this bodes well for good statistics in regions with strong precursor emissions.

The geographic pattern of correlation of \text{pcl}_O to \text{NO}_{3} seems generally reasonable and will reward more detailed analyses of the effects of general lower tropospheric subsidence, frequency of lifting of surface ozone and precursors by clouds, and decorrelation when frequent land–water transitions produce internal boundary layers, etc.

Besides emphasizing the usefulness of remote retrieval of \text{pcl}_O by methods now being evaluated (Natraj et al., 2011) and implemented (Worden et al., 2007), the analysis may be useful in constraining remote retrievals, providing estimates of a priori mean \chi and cross–covariance \text{C}(\chi_{1},\chi_{2}). We present in the figures and in the Supplementary Material details which allow a relatively simple adaptation to any user’s own vertical grid.

Acknowledgments

We particularly appreciate the contributions of the many ozonegeben operators and the industry of the organizers (especially D.W. Tarasick, S.J. Oltmans, and A.M. Thompson) of the sonden networks, funded by Environment Canada and allied organizations, NOAA, and NASA. Funding for this effort was provided by the Earth Science Division of the NASA Science Mission Directorate. Thanks to Laura Iraci and Michael Newchurch who reviewed an earlier draft.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.atmosenv.2012.06.033.

References


Worden, H., 2007. Improved tropospheric ozone pro-

Please cite this article in press as: Chatfield, R.B., Esswein, R.F., Estimation of surface \text{O}_3 from lower-troposphere partial-column information: Vertical correlations and covariances in ozonesonde profiles, Atmospheric Environment (2012), http://dx.doi.org/10.1016/j.atmosenv.2012.06.033.