

Study of the processes in the ozonosphere using MS Excel Tables: archiving, analysis, and simulation

I.L. Galkina,¹ V.V. Zuev,² and G.M. Kruchenitskii¹

¹ Central Aerological Observatory, Dolgoprudnyi, Moscow Region

² Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk

Received April 4, 2003

Potentialities of the MS Excel software in archiving, analysis, and presentation of the experimental data on various ozone characteristics are discussed. It is shown that the use of MS Excel enables archiving, comparison, and visualization of the results of diverse ground-based, balloon- and satellite-borne measurements, ozonometer network quality control, and solving various problems of ozone anomaly detection and diagnosing within a single approach. Some issues of the interaction with other application packages and database management tools, as well as of construction of regression models are considered.

Study of the processes in the ozonosphere, as well as atmospheric processes in general, has some peculiarities as compared with the traditional approach used in natural science when studying one or another phenomenon. This is caused, first of all, by the following specific, but traditional for geophysics, reasons:

- a large number of different processes affecting the process under study and related difficulties in identification of the occurrence;
- low (as compared with the majority of other physical fields) accuracy in measurement of parameters of the process under study;
- a large number of feedbacks relating the process under study and other processes that strongly complicates both correct determination of the list of parameters to be measured during field experiments and the cause-and-effect diagnostics of the observed phenomena.

The features listed above lead to the situation that the need in various databases increases strongly, as well as the role of statistical simulation, and the bulk of experimental data needed for testing one or another regularity revealed in studying some processes in atmospheric investigations (as compared with other physical fields).

Advent and wide spread of personal computers and diverse software for them in the last decades has made the geophysical investigations much more correct. However, this potential is often used for construction of the so-called numerical models including hundreds, thousands, and even tens of thousands of poorly determined constants (chemical, optical, gas-dynamic, and others). These models do not allow qualitative analysis and are very computationally expensive. In our opinion, an alternative to this approach is the direct use of computers, that is, for solution of the following problems:

- compiling vast and convenient-in-use databases;
- enhancing the signal-to-noise ratio in series of geophysical observations;

- automation of routine computational and information retrieval operations;

- invoking voluminous (comprehensive whenever possible) measurement data to testing the analytical results or results obtained for deficient experimental data.

In this paper, we aim at demonstrating the capabilities of one of the most popular software (MS Excel Tables¹) for realization of this approach as applied to studying the ozone behavior in the atmosphere. It should be mentioned that this paper does not pretend to be a comprehensive description of the Excel capabilities as applied to the ozone problem. Such capabilities are if not infinite, then so wide that their detailed description would call for quite a big book. Our goal is only to attract attention of our colleagues to the use of powerful and efficient software in everyday practical work.

First of all, it should be noted that databases (DB) organized at the current level of requirements to convenience of the user should provide for the following possibilities:

- easy access to the available data and easy way of adding new data;
- possibility of using structured QSL queries;
- possibility of visualizing (two- and three-dimensional) the information stored in DB and/or obtained from analysis;
- possibility of applying both standard analytical tools and user's algorithms for information processing.

Generally speaking, now all these requirements are easy to meet with standard program packages and DBMS aids, such as Access, FoxPro, Mathcad, etc. However, to become familiar with this software at the level sufficient for implementation of user's algorithm, the user often has to spend inadequately long time. At the same time, fulfillment of all the above requirements forces the user to master at least one DBMS and one package of analytical programs. To minimize the efforts, the user has to exert to make all the above possibilities accessible to him/her, it seems

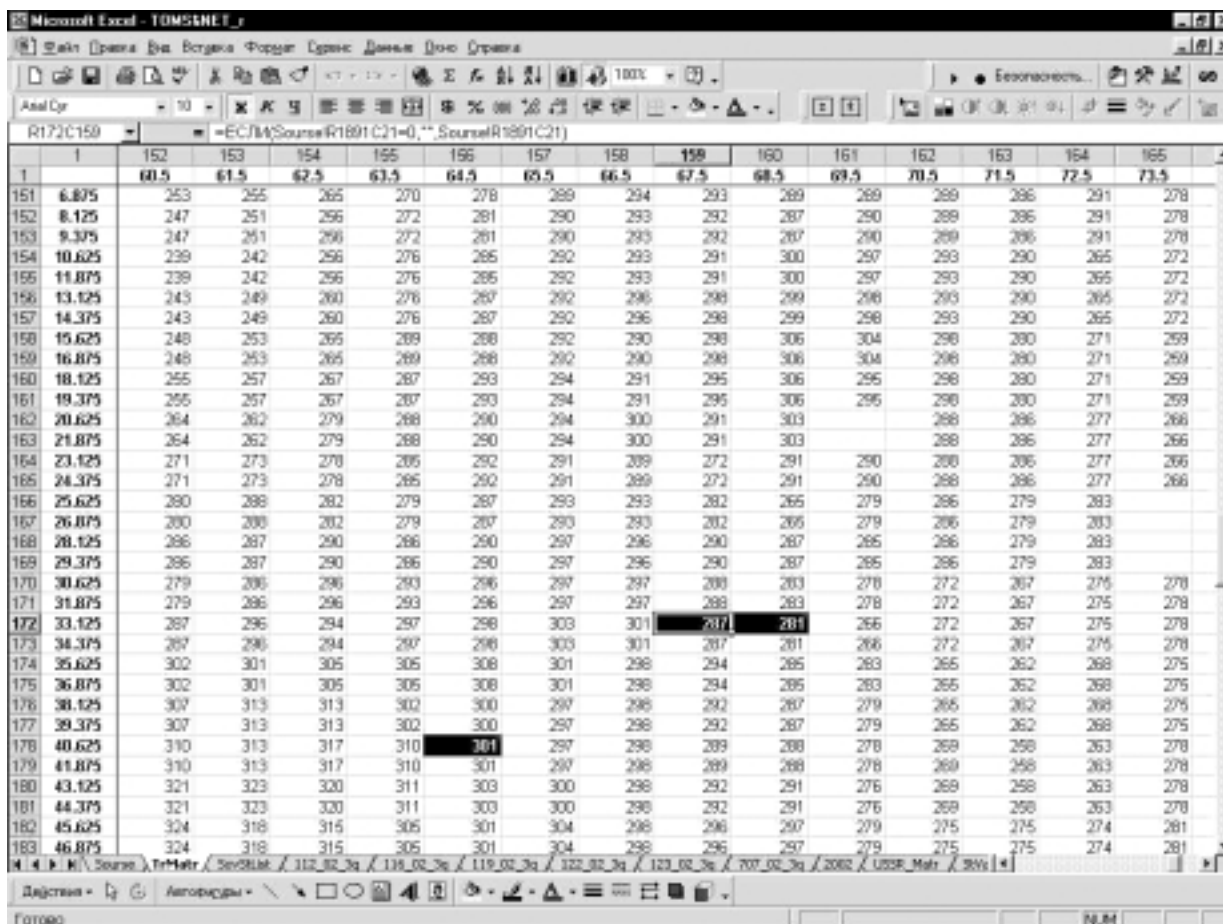


Fig. 2. TOMS data presented as transposed Mercator array.

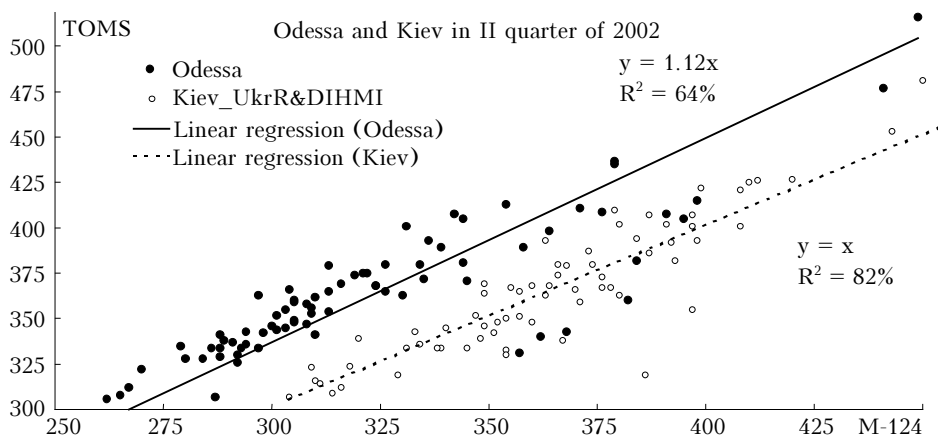


Fig. 4. Comparison of the quality of operation of the ground-based stations Kiev and Odessa.

Figure 4 and the calculated parameters presented that characterize the station performance are obtained using the Excel tools in the macro mode, that is, the list of stations (Odessa, Kiev) can be automatically replaced with any other, and the results of comparison are obtained almost immediately. It can be seen from Fig. 4 that the quality of the data from the station Kiev is much higher, because the presented regression coefficient of the ground-based and satellite data is almost ideal (equal to unity), and the spread of the data determined by the

difference of the regression determination coefficient R^2 from unity is much more narrow. (Generally speaking, one must not require identity of the ground-based and satellite data for any season and any station not only because of errors present in the satellite data, but also because different parameters are measured: midday (local time) TOC value from a satellite and daily average TOC value from the ground. However, the fact that the regression coefficient differs by 12% from unity is indicative of obvious underestimation of the ground-based data.)

As applied to Excel capabilities of archiving and visualizing the satellite TOC data, it is easy to perform spatial and temporal averaging of primary data, which is needed because of the following circumstances:

- low altitude of the Earth Probe station bearing TOMS, the TOC data in the equatorial zone and, partly, in mid-latitudes have some gaps (see Fig. 3);
- readouts in neighboring TOMS Grid cells in the middle and polar latitudes duplicate each other;
- for analysis of long-term changes in the global TOC distribution the spatial resolution corresponding to the TOMS Grid cell is certainly redundant.

The Excel capabilities on averaging of the TOMS data are illustrated in Fig. 5, which was drawn for the monthly mean TOC data averaged in space with the 3° step in latitude and the 5° step in longitude. Averaging was performed by linking worksheets using the Excel function, which is exemplified below.

```
IF(COUNTERGAPS(TrMatr!R[183]
C[102]:R[186]C[104])>7,"",
AVERAGE(TrMatr!R[183]C[102]:R[186]
C[104]))
```

This function averages a TOMS Grid 3 × 4 array, if it includes no less than eight readouts. The worksheets are linked by a code similarly to the macro, whose code is given in Appendix 1. The topographic map in Fig. 5 was generated by the macro, whose code is given in Appendix 2. Whenever necessary (for example, for the following publication), a gray-scale map with high resolution all over the dynamic range can be drawn, for example, in Mathcad by importing the data from Excel. An example of such a map is depicted in Fig. 6.

Consider now archiving and analysis of the data on the vertical ozone distribution (VOD) performed with the MS Excel. These data are also the results of balloon-borne and satellite measurements. There are three formats used for archiving the data on VOD: the

WOUDC new and old formats and the NASA format. The data in these formats can be found on the CD.³ All the three formats can be easily assimilated by Excel, which has significant advantages in the search of particular VOD profiles or retrieval some additional information about them. Explain this with some examples. Since radiosonde measurement systems used by the global network of balloon-borne VOD measurements do not use radars, the altitude reference is absent in data of two of three formats mentioned above. The Excel's intrinsic functions allow this reference to be restored using the recursion equation following from the barometric formula:

$$h_{i+1} = h_i + \frac{(T_{i+1} + T_i)R}{2\mu g} \ln \frac{p_i}{p_{i+1}},$$

where h_i is the height of the i th readout; $T_i = 273.15 + t_i^0$; t_i^0 is the i th temperature readout; p_i is the i th pressure readout; R is the absolute gas constant; μ is the molecular weight of air; g is the acceleration due to gravity. Besides, the user can calculate the ozone mixing ratio and concentration profiles, integrated ozone profile (Integrated Ozone), potential temperature, etc., as well as perform model calculations for the whole data array by specifying the needed formula in a single cell of the corresponding column and extending it to the entire workbook using the Autofill method. The results of modeling can be easily visualized using the standard Excel tools. For example, in Ref. 4 we showed that the probability of direct destruction of the ozone molecule in collisions at the altitude h with a particle of polar stratospheric clouds (PSC) W is connected with the ozone chemical potential v as:

$$W = 1 - \operatorname{erf}\left(\sqrt{\frac{v}{RT}}\right) + 2\left(\sqrt{\frac{v}{\pi RT}}\right) \exp\left(-\frac{v}{RT}\right),$$

where erf is the error function⁵; T is the absolute temperature at the altitude h ;

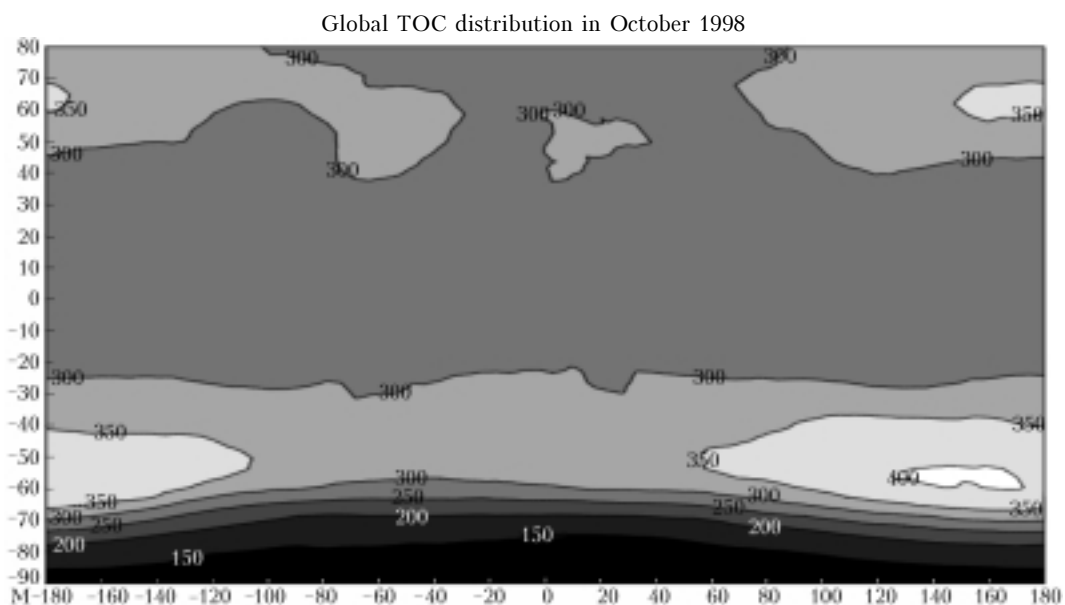


Fig. 6. Example of visualization of TOMS data imported to Mathcad from Excel.

$$v = \mu_3gh + R \int_0^h \frac{T}{p_3} dp_3 ;$$

μ_3 is the molecular weight of ozone; p_3 is the ozone partial pressure.

It is seen from Fig. 7 that on the Excel worksheet it is convenient to calculate the profiles of the chemical potential, conditional probability W , and the intensity of the ozone sink to PSC particles, which is proportional to W multiplied by the surface area of PSC particles per unit volume. Assuming that this area, in its turn, is proportional to the difference between the absolute temperature and the temperature threshold of PSC existence ~ 200 K, we can easily calculate the vertical profiles of the intensity of ozone sink accurate to a constant factor. One of such profiles is shown in Fig. 8 along with the concentration profile; Figure 8 has been plotted using the standard Excel tools.

It is seen from Fig. 8 that anticorrelation between the ozone concentration and the intensity of the ozone sink to the surface of PSC particles is no less obvious than that between the ozone concentration and the chlorine oxide concentration, which was depicted in the well-known figure from Ref. 6, published in its time in almost all magazines and newspapers, intimidating readers with mythic fear threat.

Here it is appropriate to emphasize one important advantage of the MS Excel more. All calculations performed using Excel can be readily checked from the viewpoint of formulas specified and exclude computational errors. This means that the results based on modeling by the Excel tools automatically meet the requirement of *repeatability* commonly accepted in natural sciences that is now almost forgotten in the current models of atmospheric processes.

Another significant facility given to the user by the Excel is the intrinsic function LINEST, which calculates the coefficients of linear regression of a modeled series from a set of regressors including up to 15 series. (This latter restriction is not significant, one of the ways to overcome it is considered, for example, in Ref. 8). Besides the linear regression coefficients, the LINEST function gives the errors of their calculation (as well as the model determination coefficient, the Fisher statistics, and the number of degrees of freedom). In combination with the function giving the Student's t -statistics that is also intrinsic feature of the Excel, this allows a user to determine the level of statistical confidence for each of the regressors and to select those of them, which are confident at the user-specified level. These operations can be easily automated by VBA macros. As an example, Fig. 9 depicts the latitude-time slice of the TOC seasonal behavior as judged from the TOMS data.

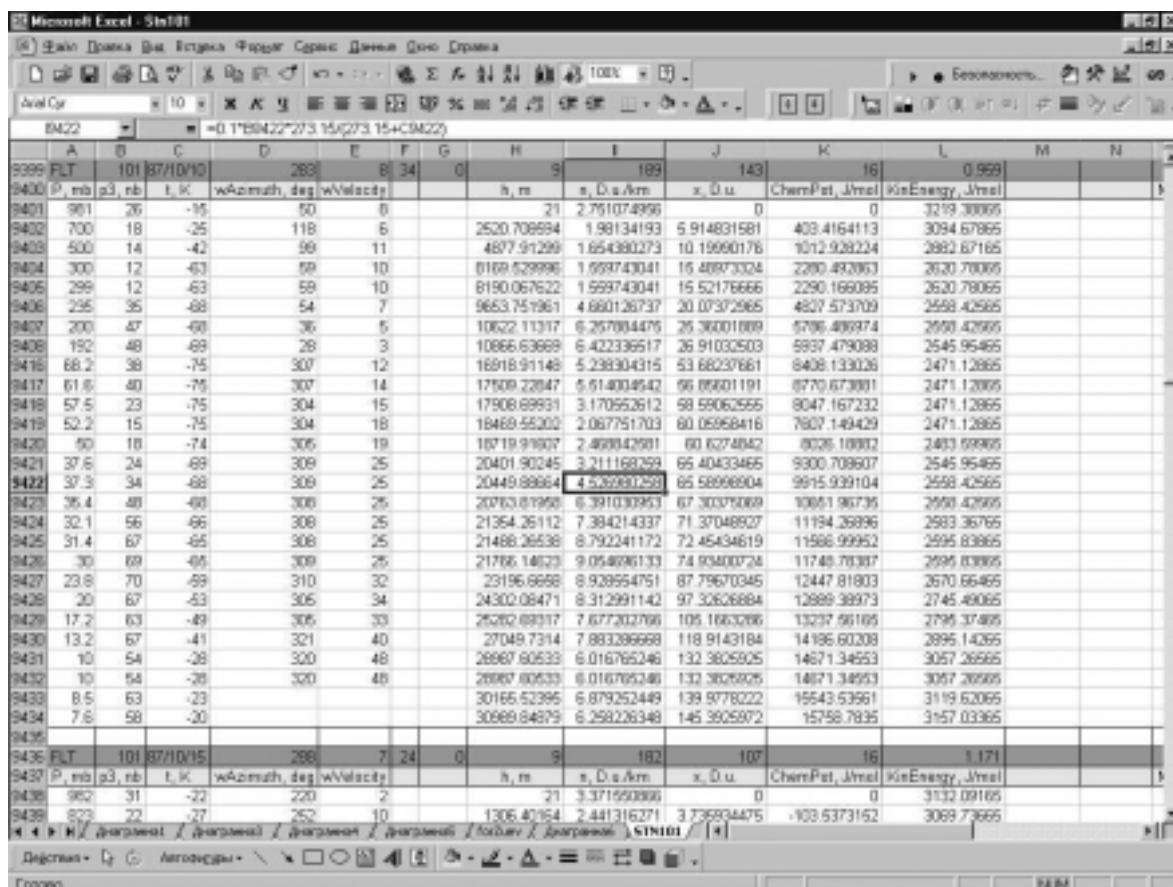


Fig. 7. Excel worksheet with the archive of balloon-borne ozone data.

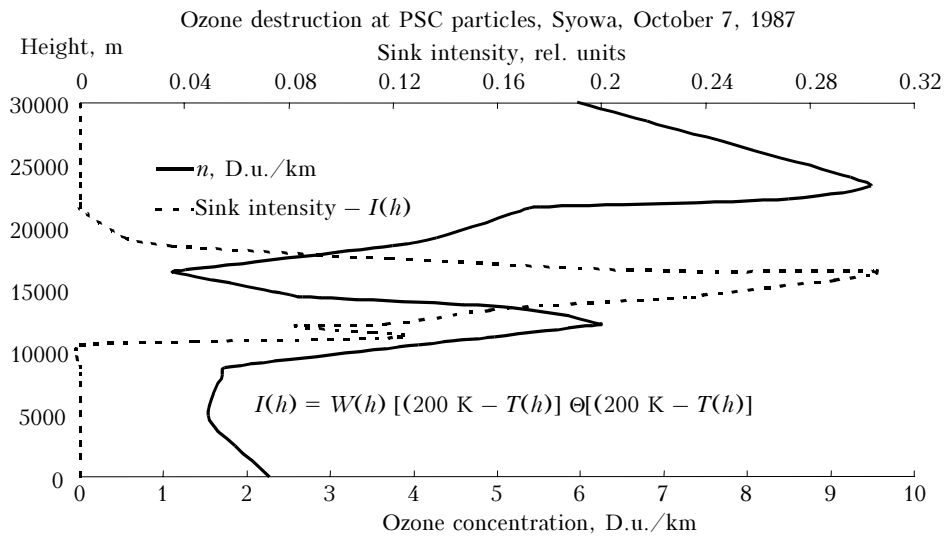


Fig. 8. Anticorrelation between the ozone concentration and the intensity of ozone sink at the surface of PSC particles.

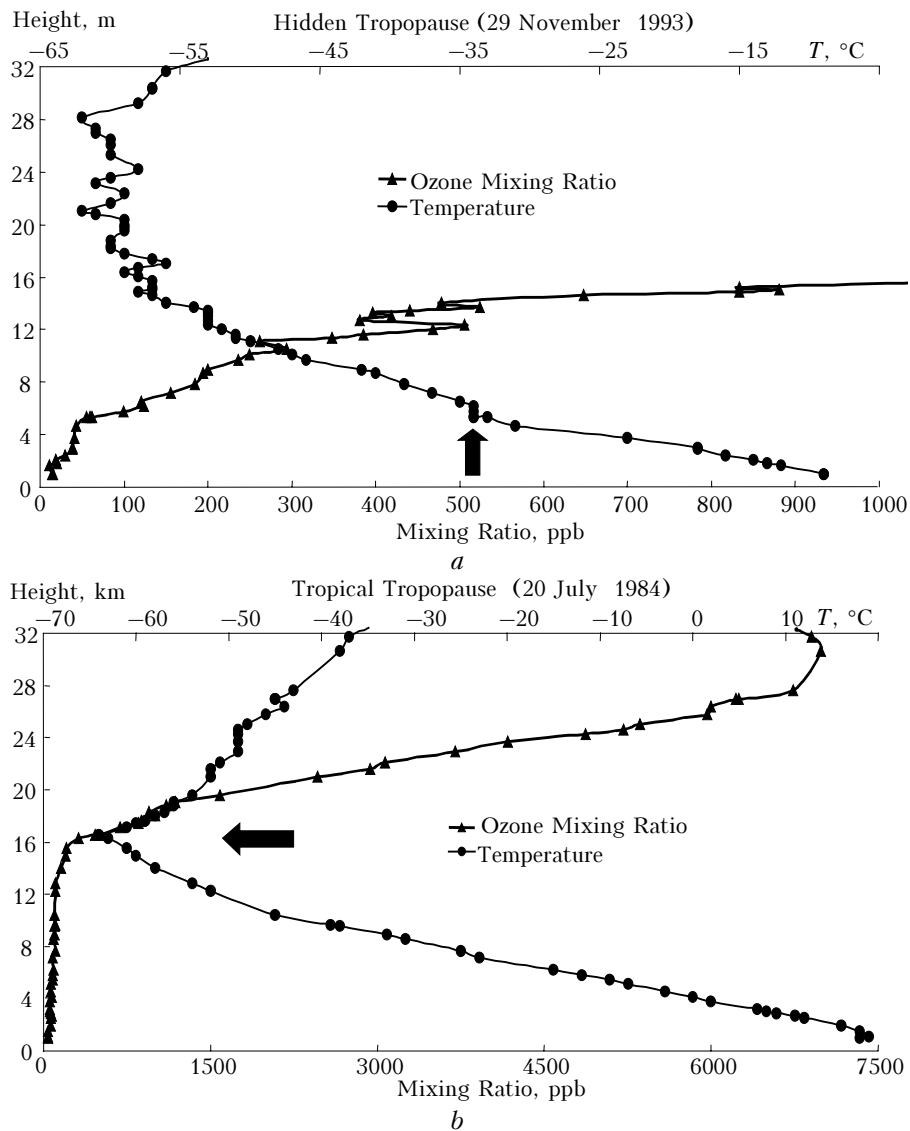


Fig. 10. Smallest (a) and largest (b) tropopause height for Hohenpeissenberg station (48°N, 11°E) according to the results of statistical diagnostics of the archive of tropopause observations by Excel macro.

The results depicted were obtained through regression of the zonally and monthly mean series to statistically confident (at the level of 95%) harmonics of the annual behavior.

Figure 9 was plotted in Mathcad with the data imported from Excel in order to use the wider Mathcad capabilities in gray-scale and color isoline drawing.

Excel allows efficient multi-criterion search. For example, when analyzing VOD or the vertical temperature (VT) profiles, it is important to find the tropopause height to separate tropospheric ozone that attracts now the increased interest of investigators from other data, as well as to determine the type of an air mass (tropic, polar, etc.).

The corresponding algorithm based on the definition from Ref. 9 can be easily realized as a macro capable of processing the entire archive. (Tropopause height is determined as a first level lying no lower than the 500 mbar isobaric surface, the mean vertical temperature gradient between which and the upper levels in the 2-km layer does not exceed 2°/km.) The results of automatic determination of the tropopause height can be easily tested by the peer review method, by selecting the smallest and largest values of the tropopause height using Excel intrinsic functions. (For this purpose, it is very convenient that Excel allows the user to find not only the largest parameter, but also the parameters second, third, and so on in value.)

Figure 10 shows the largest and smallest values of the tropopause height determined by Excel for the data obtained at the Hohenpeissenberg station. It is seen from Fig. 10 that Excel successfully copes with

this task even in the nontrivial case of hidden tropopause (Fig. 10a) and its diagnostics is confirmed by a jump of the ozone mixing ratio.

However, the Excel searching capabilities show themselves most efficiently when operating with arrays of satellite data on the vertical distribution of weather parameters and minor gaseous constituents of the atmosphere (MGCA). Using the Excel filters **Autofilter** and **Custom Filter**, as well as **Sort**, the user can easily organize the data in the form convenient for analysis and visualization. Filters allow the user to have on the worksheet only the data meeting the user's criteria (in both values of the parameters and time or geographic reference), while sorting orders the filtered array in the ascending or descending order. As an example, we present the latitude-altitude slice of the ozone mixing ratio as measured by the CRISTA-2 multichannel radiometer from onboard the space Shuttle in August 1998. Figure 11 depicts these data, which were obtained with the use of Excel **Filter** and **Sort** functions, as well as the **Average** function, which gave the zonally average values of the mixing ratio.

Certainly, the above examples do not exhaust the Excel capabilities as applied to archiving, analysis, and visualization of the ozonometric data. However, we hope that they convince our colleagues to widely use this powerful and efficient instrument in their everyday work. The use of MS Excel tables not only improves the efficiency of research, but also significantly facilitates the exchange of data between specialists and mutual check of the obtained scientific results.

Appendix 1

Codes for referencing between Excel worksheets to convert TOMS Grid data into array of readouts

```
Global Const TGFCOLUMNS = 25'   Number of rows in latitude block of TOMS Grid format
Global Const TGLATES = 180'     Number of latitude blocks
Global Const TGLONES = 288'     Number of readouts (resolved elements) in a latitude block
```

```
Public Sub ToTrMatr()
```

```
'Converts TOMS Grid format on the first worksheet into transposed array on the second
'worksheet
```

```
Dim lat As Integer, lon As Integer, ro As Integer, co As Integer, f As String
```

```
For lat = 1 To TGLATES
```

```
For lon = 1 To TGLONES
```

```
ro = 12 * lat - 11 + (lon - 1) \ TGFCOLUMNS
```

```
co = (lon - 1) Mod TGFCOLUMNS + 1
```

```
f = Worksheets(1).name & "!R" & Trim(ro) & "C" & Trim(co)
```

```
f = "=" & "IF(" & f & "=0,""""" & f & ")"
```

```
Worksheets(2).Cells(lon + 1, lat + 1).FormulaR1C1 = f
```

```
Next lon
```

```
Next lat
```

```
End Sub
```

Appendix 2

Public Sub GeogrBG()

'Draws topographic map on the Map diagram; latitude and longitude of points are given in the 'first two columns of the worksheet ("GBG")

Dim intRow As Integer, sngX As Single, sngY As Single, intI As Integer

Charts("Map").Activate

intRow = 1

sngSLat = Worksheets("GBG").Cells(1, 4).Value

sngNLat = Worksheets("GBG").Cells(1, 5).Value

sngKlon = Charts("Map").Axes(xlCategory).Width / 355

sngBlon = Charts("Map").Axes(xlCategory).Left + 177.5 * sngKlon

sngKlat = Charts("Map").Axes(xlSeries).Height / (sngSLat - sngNLat)

sngBlat = Charts("Map").Axes(xlSeries).Top - sngNLat * sngKlat

Do While Worksheets("GBG").Cells(intRow, 1) <> ""

intI = intI + 1

If Abs(Worksheets("GBG").Cells(intRow, 1).Value) < 177.5 And _

Worksheets("GBG").Cells(intRow, 2).Value > sngSLat And _

Worksheets("GBG").Cells(intRow, 2).Value < sngNLat Then

sngX = Worksheets("GBG").Cells(intRow, 1).Value * sngKlon + sngBlon - 0.5

sngY = Worksheets("GBG").Cells(intRow, 2).Value * sngKlat + sngBlat - 0.5

ActiveChart.Shapes.AddShape(msoShapeRectangle, sngX, sngY, 0.1, 0.1).Select

If intI Mod 400 = 0 Then DoEvents

End If

intRow = intRow + 1

Loop

'sngX = sngBlat

ActiveChart.Shapes.AddLine(-177.5 * sngKlon + sngBlon, sngBlat, _
177.5 * sngKlon + sngBlon, sngBlat).Select

Selection.ShapeRange.Line.Weight = 0.5

Selection.ShapeRange.Line.DashStyle = msoLineDash

ActiveChart.Shapes.AddLine(sngBlon, sngBlat + Worksheets("GBG").Cells(1, 4).Value *
sngKlat, _

sngBlon, sngBlat + Worksheets("GBG").Cells(1, 5).Value * sngKlat).Select

Selection.ShapeRange.Line.Weight = 0.5

Selection.ShapeRange.Line.DashStyle = msoLineDash

Worksheets("GBG").Cells(1, 8).Value = 1

End Sub

References

1. P. Blattner, L. Ulrich, K. Cook, T. Dyck, *Using Microsoft Excel 2000* (1999).
2. A. GarnaeV, *Teach Yourself in VBA* (BHV, St. Petersburg, 1999), 503 pp.
3. *Ozone Data for the World*. ODW CD#1 (Environment, Canada, 1999).
4. A.M. Zvyagintsev, V.V. Zuev, G.M. Kruchenitskii, and T.V. Skorobogatyi, *Issled. Zemli iz Kosmosa*, No. 3, 29–34 (2002).
5. E. Jahnke and F. Emde, *Tables of Functions with Formulae and Curves* (Dover Publications, New York, 1945).
6. J.G. Anderson, W.H. Brune, S.A. Lloyd, et al., *J. Geophys. Res.* **94**, 11480 (1989)
7. I.L. Galkina and G.M. Kruchenitskii, *Atmos. Oceanic Opt.* (in print).
8. A.M. Zvyagintsev, N.E. Kadygrov, and G.M. Kruchenitskii, *Issled. Zemli iz Kosmosa*, No. 3 (in print; 2003).
9. *Instructions for Weather Stations and Posts. Issue 4. Upper-air Observations at Stations*. Part IIIa (Gidrometeoizdat, Leningrad, 1973), 113 pp.