NWP SAF

Satellite Application Facility
for Numerical Weather Prediction

Sampled databases of 60-level atmospheric profiles from the ECMWF analyses

Frédéric Chevallier

European Centre for Medium–Range Weather Forecasts

Document No. NWPSAF–EC–TR–004

Version 1.0

January 2002

Sampled databases of 60-level atmospheric profiles from the ECMWF analyses

Frédéric Chevallier

ECMWF

This documentation was developed within the context of the EUMETSAT satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 25 November 1998, between EUMETSAT and the Met Office, UK, by one or more partners within the NWP SAF. The partners in the NWP SAF are the Met Office, ECMWF, KNMI and Météo France.

Change record			
Version	Date	Author / changed by	Remarks

Abstract

This report summarises the characteristics of two databases that sample the temperature, water vapour and ozone profiles simulated by the European Centre for Medium-Range Weather Forecasts system in the version used for the 40-year re-analysis project. The first database contains 13495 atmospheric situations, of which the second is an 80-profile subset. Their potential applications include statistical regressions and the validation of various models, in particular in the field of radiation.

1 Introduction

The availability of large datasets, like the day-to-day observations of the Earth-atmosphere system or the recent re-analyses of the Numerical Weather Prediction (NWP) centres have made it necessary to develop methods for the extraction/ reduction of information. First-order statistical moments or variance-component analysis provide essential tools that significantly reduce the number of freedom degrees. For some applications, the information can not be projected on a new variable space, but needs to be simply sampled. This is the case when one wants to validate an algorithm on a reduced, but representative, number of cases, or for statistical regressions. In the latter applications, a regular spread of the cases on the variable space is desirable, so that equal weight be given to frequent as to seldom situations.

A major attempt to sample atmospheric profiles on a global scale, despite the high dimensionality of the problem, has been the constitution of successive versions of the Thermodynamic Initial Guess Retrieval database (TIGR) (Chédin *et al.* 1985). Each version groups together hundreds of soundings sampled from larger databanks of observations of the atmosphere. The analyses of the NWP centres provide an interesting alternative to the use of radiosonde reports, because they are homogeneous and cover all latitudes, longitudes and days of the year. Moreover they provide a whole set of variables consistent with each other for each profile, like layer temperature, water vapour, cloud cover and condensate, and surface characteristics. Recent developments at European Centre for Medium-Range Weather Forecasts (ECMWF) add prognostic and assimilated ozone to the variable list. This report summarises the characteristics of a diverse profile dataset from the ECMWF 40-year re-analysis (ERA-40) (Simmons and Gibson 2001). The sampling method takes temperature, specific humidity and ozone into account. The database is described in section 2. Section 3 presents a further sampling of the database, so as to reduce it from 13495 situations to 80. Both sets are compared to four other databases that are used in the atmospheric research community: the previous version of the ECMWF sampled database, the third version of TIGR and the two databases on which the linear regressions of the Radiative Transfer for Tiros Operational Vertical Sounder fast radiative transfer model (RTTOV-5: Eyre 1991; Saunders *et al.* 1999) rely. Section 5 provides an overall summary.

2 Characteristics of the database

2.1 The sampling technique

As for the TIGR databases, the sampling strategy is a two-step method. The first step consists in filtering the infinity of possible profiles in the atmosphere, by gathering a much reduced but diverse sample of them. Let us call S this initial database. The sampling of S with a topological approach is the second step of the method. It relies on an index I, that measures the dissimilarity between two atmospheric situations. The process is iterative. At step one, a first atmospheric situation from S is randomly drawn and archived in a new set E. At step n, a n^{th} atmospheric situation is randomly drawn and archived in E if it is different enough from the already selected situations, relative to index I. With that approach, the distribution of E over the space of the various atmospheric variables is smoother than that of S. In practice, restriction to some variables has to be made. In the following, temperature, water vapour and ozone profiles are taken into account along the lines of Chevallier *et al.* (2000a). Those authors do not discuss the introduction of ozone. Its treatment is similar to that of specific humidity.

2.2 The model profiles

The model profiles used here are generated by the ERA-40 assimilation system, that relies on the 3-dimensional variational scheme described by Courtier *et al.* (1998). Conventional and satellite observations provide it with information on pressure, temperature, humidity, ozone and wind. Analyses are performed at 00, 06, 12 and 18 UTC, with the first-guess interpolated at appropriate time for comparison with the observations. The forecast model is a global spectral T_L159L60 model. The reduced horizontal grid corresponds to a regular grid size of about 125 km from the equator to the poles. In the vertical, a hybrid coordinate of 60 levels between the surface and the top of the atmosphere (0.1 hPa) is used. The physics package is an improved version of that described by Gregory *et al.* (2000), with the main modifications given by Jakob *et al.* (2000) and Morcrette *et al.* (2001).

In the present study, the initial database S results from the aggregation of 48 days over two years (namely the first and the 15th of each month between January 1992 and December 1993) of profiles from the ERA-40 analyses. Each day includes a description of the atmosphere on the corresponding 60-level vertical grid and on the model reduced Gaussian grid (35718 grid points at a horizontal resolution of about 125 km) every six hours. Doing so, S consists of about seven million profiles. It is divided into seven subgroups differing by the total precipitable water vapour content of the profiles: the first group ranges from 0 to 0.5 $kg.m^{-2}$, the second from 0.5 to 1.5 $kg.m^{-2}$, the third from 1.5 to 2.5 $kg.m^{-2}$, and so on, until the seventh one that goes from 5.5 $kg.m^{-2}$ up to the highest values.

The sampling approach described above is used for the extraction of about the same number (N) of samples from each class, except for the first one, where twice as many ($2 \times N$) profiles are extracted, in consideration of the higher temperature variability: this class includes all types

of situations from polar to tropical. N determines the density of the sampled database and $N \simeq 1700$ was chosen. The whole sampled database includes 13495 profiles.

2.3 Available variables

Each situation in the 60-level sampled database, hereafter referred to as 60L-SD, is indexed by its space-time location:

- the longitude, between 0° and 360°, eastward counted
- the latitude, between -90° and 90°
- the date, as *yyyymmddhh*, where *yyyy* is the year, *mm* the month, *dd* the day, and *hh* the synoptic hour

As said before, the sampled variables are:

- the atmospheric temperature, in K, on the 60-level grid
- the atmospheric specific humidity, in kg/kg, on the 60-level grid
- the atmospheric specific ozone, in kg/kg, on the 60-level grid

The vertical pressure grid is a linear function of the surface pressure P_s. Indeed for each level l, the pressure $P(l)$ is expressed as: $P(l) = a_l + b_l P_s$. The pressure grid is illustrated in Table 1. The minimum pressure is 0.1 hPa.

Other variables of the sampled situations have been extracted from the ECMWF archive and complete the database:

- the surface pressure (hPa)
- the surface geometric height (m)
- the surface skin temperature (K)
- the 2-meter temperature (K)
- the 2-meter specific humidity (kg/kg)
- the 10-meter u and v components of the wind (m/s)
- the land fraction (0 corresponds to sea-only points)
- the cloud cover, on the 60-level grid
- the cloud liquid water content, in kg/kg, on the 60-level grid

- the cloud ice water content, in kg/kg, on the 60-level grid
- the vertical velocity, in Pa/s, on the 60-level grid
- the type (see Table 2) and cover of low vegetation
- the type (see Table 2) and cover of high vegetation
- the temperature (K) and volumetric water (m^3/m^3) in four soil layers. Downward from the surface, the depth of the layers is successively: 7, 21, 72 and 189 cm.
- the ice cover and its temperature (K) in four layers. Downward from the surface, the depth of the layers is successively: 7, 21, 72 and 50 cm.
- the snow temperature (K), depth (m), density ($kg.m^{-3}$) and albedo (0-1)
- the surface albedo (0-1)
- the surface roughness (m)

The sampling is performed on the ECMWF model vertical layers and not on fixed pressure layers. As a consequence, the sampled database gathers profiles corresponding to various ocean conditions as well as to land conditions, including high elevated grounds. The lowest surface pressure in the database is 520 hPa and the highest 1049 hPa.

2.4 Distribution of the variables

The histograms of the 60L-SD are presented on Figures 1 to 5 for each geotype (sea and land) as a function of the following variables: the total water vapour content, the total ozone content, the skin temperature, the date (month and local time), the location (longitude and latitude), the surface pressure, the temperature and the specific humidity in model layer 47, the specific ozone in model layer 10, the 500 hPa vertical velocity, the total liquid water content, the total ice water content. Layers 47 and 10 correspond to pressure levels of 787 hPa and 4 hPa respectively when the surface pressure is 1000 hPa (see Table 1) and have been chosen as examples of the layer histograms. The distribution of the vegetation type of the land profiles is also presented on Figure 6.

An ideal sampling would lead to a regular distribution of the variable values, at least for temperature, humidity and ozone, but is impossible because of the constraints imposed by the physical laws. From the various histogram shapes it is clear that the 60L-SD results from compromises. If the distribution of the situations as a function of month, local time and longitude is regular for each geotype, the other histograms are more irregular due to physical constraints. As an example, the wing in the temperature (respectively specific humidity) in layer 47 histograms between 240 and 280 K (respectively 0.002 and 0.015 kg/m^2) illustrates the weak variability of specific humidity (respectively temperature) in this temperature (respectively specific humidity) range in this layer and in the initial set S. The difference between the specific

humidity histogram for sea and that one for land also stems from the different occurrence of each type of situation, even if the natural distribution has been strongly smoothed. Since the representation of high water vapour contents has been forced in the sampling (see section 2.2), the wing in the water vapour histogram is more regular than that of the temperature histogram.

2.5 Statistical characteristics

The main statistical characteristics (minimum and maximum values, mean, standard deviation and median per pressure level) of the 60L-SD temperature, humidity and ozone profiles are illustrated in Figure 7. For comparison, the characteristics of two other databases are presented in Figures 8 and 9.

The first database (Figure 8) is the previous version of the ECMWF diverse profile set, that is described on 50-levels (Chevallier 1999). It gathers 13766 profiles of forecasts from the ECMWF system in the operational set-up with a resolution of about 60 km and will be referred to as 50L-SD in the following. The ozone profiles were not available as model variables when it was set up, and therefore climatological values from Fortuin and Langematz (1994) were used. From Figures 7 and 8, the two ECMWF sampled datasets share very similar statistical features for temperature and water vapour. This emphasises the coherence and robustness of the forecast system despite large number of improvements in the model and differences in the observation network. The 60L-SD prognostic and assimilated ozone and the 50L-SD climatological ozone are also similar in the stratosphere in terms of mean and standard deviations, but the extrema span a larger range in the 60L-SD, illustrating the advantage of instantaneous over climatological values for the present kind of application. In the troposphere, the 60L-SD has much more variability than the 50L-SD, which may be caused by an inappropriate side effect of the ozone assimilation in ERA-40 (A. Dethof, 2001, personal communication).

The TIGR-3 database (Chevallier *et al.* 1998) from Laboratoire de Météorologie Dynamique is the second database in the comparison (Figure 9). It is made of 2311 40-level profiles, that include both radiosonde reports and tropical-type satellite retrievals. The ozone data come from selected radiosonde reports archived at the World Ozone and Ultraviolet radiation Data Centre (F. Karcher, 1997, personal communication) and have been added to the temperature and water vapour profiles with respect to the date (month and day) and latitude. Quality controls have ensured that the highest measured temperature in TIGR-3 reaches 30 hPa at minimum (Escobar-Munoz 1993). Above 30 hPa, the temperature profiles have been extrapolated above the highest measured pressure level using a statistical procedure (Moulinier 1983). The specific humidity profiles reach at least 300 hPa, before any extrapolation. For ozone, TIGR-3 use extrapolated values above about 10 hPa. From Figures 7 and 9, the 60L-SD has colder minima than TIGR-3 in most pressure levels, warmer maxima below 500 hPa and colder maxima above 500 hPa. The colder tropospheric minima in the 60L-SD are due to the presence of profiles from the Antarctic plateau, that is not represented in TIGR-3. As said before, in the middle and in the high stratosphere, TIGR-3 may suffer from extrapolation artifacts. The temperature standard deviations are rather similar between TIGR-3 and the 60L-SD, even though they are slightly larger for TIGR-3 in the troposphere. The mean temperature profiles, that stem from

the relative spread between the various air masses, are very different. The specific humidity maxima show wetter values of the 60L-SD maxima between 450 and 750 hPa, and drier in the other tropospheric pressure levels. The standard deviations are slightly larger for the 60L-SD. The ozone statistics are very different, with a much smaller spread of the 60L-SD in the stratosphere. This tends to indicate that despite a good agreement between ERA-40 and the assimilated Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultraviolet Radiometer (SBUV) ozone measurements (A. Dethof and E. Holm, 2001, personal communication), the extreme profiles may still not be well captured in the stratosphere. The opposite situation happens in the troposphere, where the 60L-SD has a much larger spread, but as mentioned above, some of the 60L-SD values may be too high.

As complementary results, Principal Component Analyses have been performed on the temperature, humidity and ozone fields, in order to compare the vertical resolution of the three above-mentioned datasets. The cumulated variance as a function of the number of leading eigenvectors is presented in Figure 10. The temperature plot (Figure 10a) illustrates the increasing resolution obtained when increasing the number of levels from 40 (TIGR-3), to 50 (the 50L-SD), and to 60 (the 60L-SD). The ozone plot (Figure 10c) shows a similar result, knowing that the ozone climatology in the 50L-SD was originally given on 19 levels. However the opposite feature appears for the humidity plot (Figure 10b), with the 60L-SD having the smallest vertical resolution and TIGR-3 having the highest. The reason why the observation-based TIGR-3 performs better than the two ECMWF sampled datasets for humidity is likely to be found in the fact that the representation of water vapour in forecast systems is still not as satisfactory as that of other variables like temperature and winds. Comparing the two ECMWF sampled datasets, the difference in resolution appears for profiles located over land. Over sea, the 60L-SD has a slightly higher resolution than the 50L-SD (not shown). As a consequence, the change of orography between the 125 km-resolution ERA-40 (that the 60L-SD use) and the 60 km-resolution operational archives (that the 50L-SD use) shall explain the behaviour over land.

3 Strategy for a further reduction of the database

3.1 Description of the 80 profile database

For some computationally expensive applications, such as radiative "line-by-line" computations, the number of profiles of the 60L-SD may still be too high. The International ATOVS Working Group suggested that about 80 profiles only are needed for the creation of line-by-line transmittance datasets (ITWG 2000).

This further reduction of the size of the database could be achieved by re-sampling the initial 7,000,000 profile initial set (the set S of section 2.2) with a lower value for N. A simpler approach, that was already used for the 50L-SD (Chevallier 1999), is used here. Under the restrictions examined in section 2, the 60L-SD is a regular mesh of the 7,000,000 profile initial set. A random sampling of it enlarges the mesh without modifying its distribution.

This approach is used to select a reduced set of 77 profiles out of the 13495 profile database. A drawback of the random sampling is that the extreme values have few chances of being selected. As a consequence, two synthetic additional profiles have been added that represent the extremes. They are both given a standard surface pressure of 1013.25 hPa. The first profile is made of the minimum values of temperature, water vapour and ozone at each pressure level. The second one contains the maximum values. Finally the mean profile of the 60L-SD is added to the set since it can be a relevant information as well. Its surface pressure is also set to 1013.25 hPa. This reduced dataset of 80 profiles is referred to as 60L-SDr in the following.

The histograms of the 60L-SDr are shown in Figures 11 to 14. They reproduce the shapes of those of the 60L-SD (Figures 1 to 4), except that due to the small number of profiles, more bins are empty.

Figure 15 presents the statistics of the 60L-SDr. As expected, they are very close to the 60L-SD ones, even though one can notice a smoothing of the extrema, due to vertical interpolations for the computation of the extrema on pressure levels. For comparison, the statistics of two small diverse profile datasets are shown in Figure 16. The ozone statistics (Figures 16d and 16e) are presented for the 33-profile dataset that is used for the ozone regressions in RTTOV (Saunders *et al.* 1999). The temperature and water vapour plots (Figures 16a to 16c) correspond to the 43-profile set that are at the basis of the other RTTOV regressions. The latter is a modified subset of the TIGR-2 database (Achard 1991; Escobar-Munoz 1993). These two sets contain radiosonde reports only.

The temperature extrema of the RTTOV set are very close to those of TIGR-3 and share similarities to those of the 60L-SDr (see the comments in section 2.5 on the differences between TIGR-3 and the 60L-SD). The RTTOV standard deviations for temperature are much larger than the other datasets. For specific humidity, the RTTOV set has the driest maxima, means and medians in the troposphere, because the sampling method used in TIGR-2 took only temperature into account in the choice of the profiles. This has been improved in the TIGR-3 dataset, as shown in Figure 9. The RTTOV ozone statistics clearly show the weakness of the extrapolation of the ozone radiosondes in the high stratosphere. Between 30 and 400 hPa, the statistics are close to those of the 60L-SDr. Below 400 hPa the 60L-SDr has much more, and likely excessive (cf. section 2.5), variability.

4 Summary and future developments

Two datasets have been sampled from the 60-level ECMWF model outputs. They may be used for a wide range of applications, depending on the computational expense. The sampling method used allows for a regular distribution of physically consistent atmospheric temperature, water vapour and ozone profiles in each set. As illustrated by Chevallier *et al.* (2000b), this kind of database is suitable for regression applications. It can also serve as an independent validation set for various algorithms (e.g., Chevallier and Mahfouf 2001).

Both datasets are available from the NWP-SAF [1]. All comments or questions should be sent to the author [2].

The sampled databases presented here should not be considered as final ones. They carry both qualities and weaknesses from the ECMWF assimilation-forecast system. Compared to the previous ECMWF diverse profile datasets, the present versions have a better description of the boundary layer and of ozone. Further improvements of the system will enable further improvements of the databases, for instance in the description of cloudiness, for the representation of the atmosphere above 0.1 hPa, or for the introduction of other gases.

Acknowledgments

The author thanks R. Saunders (Met Office) and P. Brunel (MétéoFrance) for encouraging support of this work, and P. Viterbo (ECMWF) for his help with the ERA-40 archive.

References

Chédin, A., N. A. Scott, C. Wahiche and P. Moulinier, 1985: The Improved Initialization Inversion method : a high resolution physical method for temperature retrievals from satellites of the TIROS-N series. *J. Climate Appl. Meteor.*, **24**, 128-143.

Chevallier, F., F. Chéruy, N. A. Scott, and A. Chédin, 1998: A neural network approach for a fast and accurate computation of longwave radiative budget. *J. Appl. Meteor.,* **37**, *1385-1397.*

Chevallier, F., 1999: TIGR-like sampled databases of atmospheric profiles from the ECMWF 50-level forecast model. *NWP SAF Research Report No. 1*, 18 pp. [Available from the librarian at ECMWF].

Chevallier, F., A. Chédin, F. Chéruy, J.-J. Morcrette, 2000a: TIGR-like atmospheric profile databases for accurate radiative flux computation. *Quart. J. Roy. Meteor. Soc.*, **126**, 777-785.

Chevallier, F., J.-J Morcrette, F. Chéruy, and N. A. Scott, 1999b: Use of a neural network-based LW radiative transfer model in the ECMWF atmospheric model. *Quart. J. Roy. Meteor. Soc.*, **126**, 761-776.

Chevallier, F., and J.-F. Mahfouf, 2001: Evaluation of the Jacobians of infrared radiation models for variational data assimilation. *J. Appl. Meteor*, **40**, 1445-1461.

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljević, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I.: formulation. *Q. J. Roy. Meteor. Soc.*, **124**, 1783-1808.

[1] http://www.metoffice.com/research/interproj/nwpsaf/rtm

[2] f.chevallier@ecmwf.int

Escobar-Munoz, J., 1993 : Base de données pour la restitution de variables atmosphériques à l'échelle globale. Étude sur l'inversion par réseaux de neurones des données des sondeurs verticaux atmosphériques satellitaires présents et à venir. PhD thesis, Univ. Paris VII, 190 pp. [Available from LMD, Ecole Polytechnique, 91128 Palaiseau cedex, France].

Eyre, J. R., 1991: A fast radiative transfer model for satellite sounding systems. ECMWF Technical Memorandum No. 176, 28 pp. [Available from the librarian at ECMWF].

Fortuin, J. P. F. and Langematz, U., 1994: An update on the global ozone climatology and on concurrent ozone and temperature trends. *Proceedings SPIE*, **2311**, 207-216.

Gregory, D., J.-J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale, 2000: Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. *Q. J. Roy. Meteor. Soc.*, **126**, 1685-1710.

ITWG 2000: Working group report on radiative transfer and surface property modelling. In *International ATOVS Working Group report on the Eleventh International TOVS Study Conference, Budapest, Hungary, 20-26 September 2000,* 9-14.

Jakob, C., E. Andersson, A. Beljaars, R. Buizza, M. Fisher, E. Gérard, A. Ghelli, P. Janssen, G. Kelly, A. P. McNally, M. Miller, A. Simmons, J. Teixeira, and P. Viterbo, 2000: The IFS cycle CY21r4 made operational in October 1999. *ECMWF Newsletter*, **87**, 2-9.

Morcrette, J.-J., E. J. Mlawer, M. J. Iacono, and S. A. Clough, 2001: Impact of the radiation-transfer scheme RRTM in the ECMWF forecasting system. *ECMWF Newsletter*, **91**, 2-9.

Moulinier, P., 1983: Analyse statistique d'un vaste échantillonnage de situations atmosphériques sur l'ensemble du globe. *LMD Internal note 123*, 30 pp., in French [Available from LMD, Ecole Polytechnique, 91128 Palaiseau cedex, France].

Saunders, R., M. Matricardi, and P. Brunel, 1999: An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quart. J. Roy. Meteor. Soc.*, **125**, 1407-1425.

Simmons, A. J., and J. K. Gibson (Eds), 2000: The ERA-40 project plan. *ERA-40 Project Report Series No. 1*, 62 pp.

White, P., 2001: IFS Documentation Part IV: Physical processes (CY23R4). In press. [available from ECMWF, Shinfield Park, Reading, Berks. RG2 9AX, UK].

level	pressure (hPa)	level	pressure (hPa)	level	pressure (hPa)	level	pressure (hPa)
1	0.10	18	23.31	35	351.28	52	918.53
2	0.29	19	28.88	36	385.84	53	937.16
3	0.51	20	35.78	37	421.61	54	953.09
4	0.80	21	44.33	38	458.37	55	966.35
5	1.15	22	54.62	39	495.87	56	977.04
6	1.58	23	66.62	40	533.83	57	985.35
7	2.08	24	80.40	41	571.96	58	991.51
8	2.67	25	95.97	42	609.96	59	995.86
9	3.36	26	113.41	43	647.51	60	998.82
10	4.19	27	132.71	44	684.32		
11	5.20	28	153.89	45	720.07		
12	6.44	29	176.91	46	754.47		
13	7.98	30	201.72	47	787.23		
14	9.89	31	228.27	48	818.11		
15	12.26	32	256.53	49	846.86		
16	15.19	33	286.49	50	873.28		
17	18.81	34	318.11	51	897.21		

Table 1: 60-level vertical grid of the ECMWF model, when the surface pressure equals 1000 hPa. The general formulation depends on the surface pressure.

Index	Vegetation Type	High/Low ground
1	Crops, Mixed Farming	L
2	Short Grass	L
3	Evergreen Needleleaf Trees	H
4	Deciduous Needleleaf Trees	H
5	Deciduous Broadleaf Trees	H
6	Evergreen Broadleaf Trees	H
7	Tall Grass	L
9	Tundra	L
10	Irrigated Crops	L
11	Semidesert	L
13	Bogs and Marshes	L
16	Evergreen Shrubs	L
17	Deciduous Shrubs	L
18	Mixed Forest	H
19	Interrupted Forest	H

Table 2: Definition of the vegetation types in ERA-40 (White 2001).

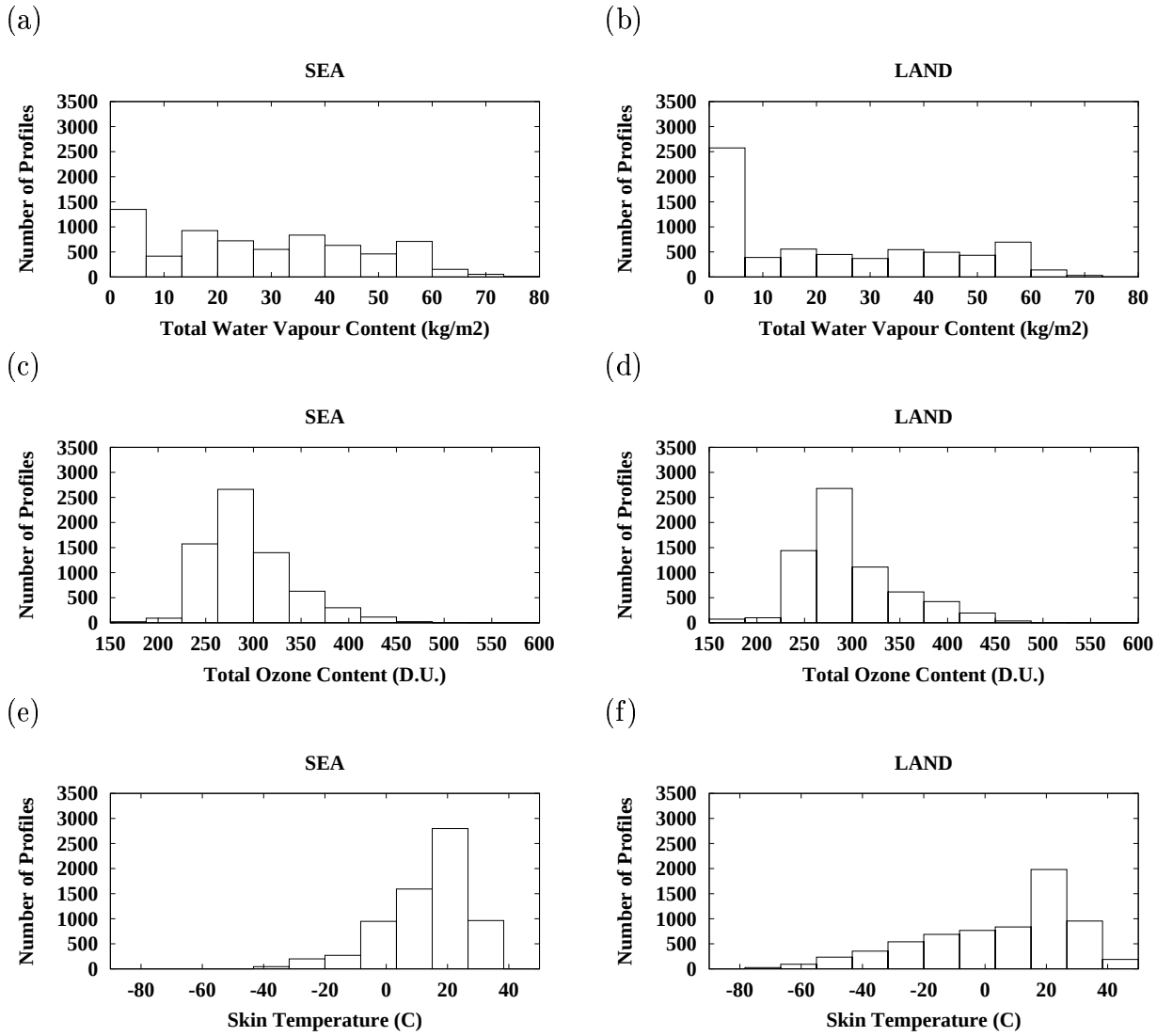

Figure 1: Distribution of the situations in the 60-level sampled database (60L-SD, 13495 profiles) as a function of some variables and for each geotype. The total ozone content is given in Dobson Units.

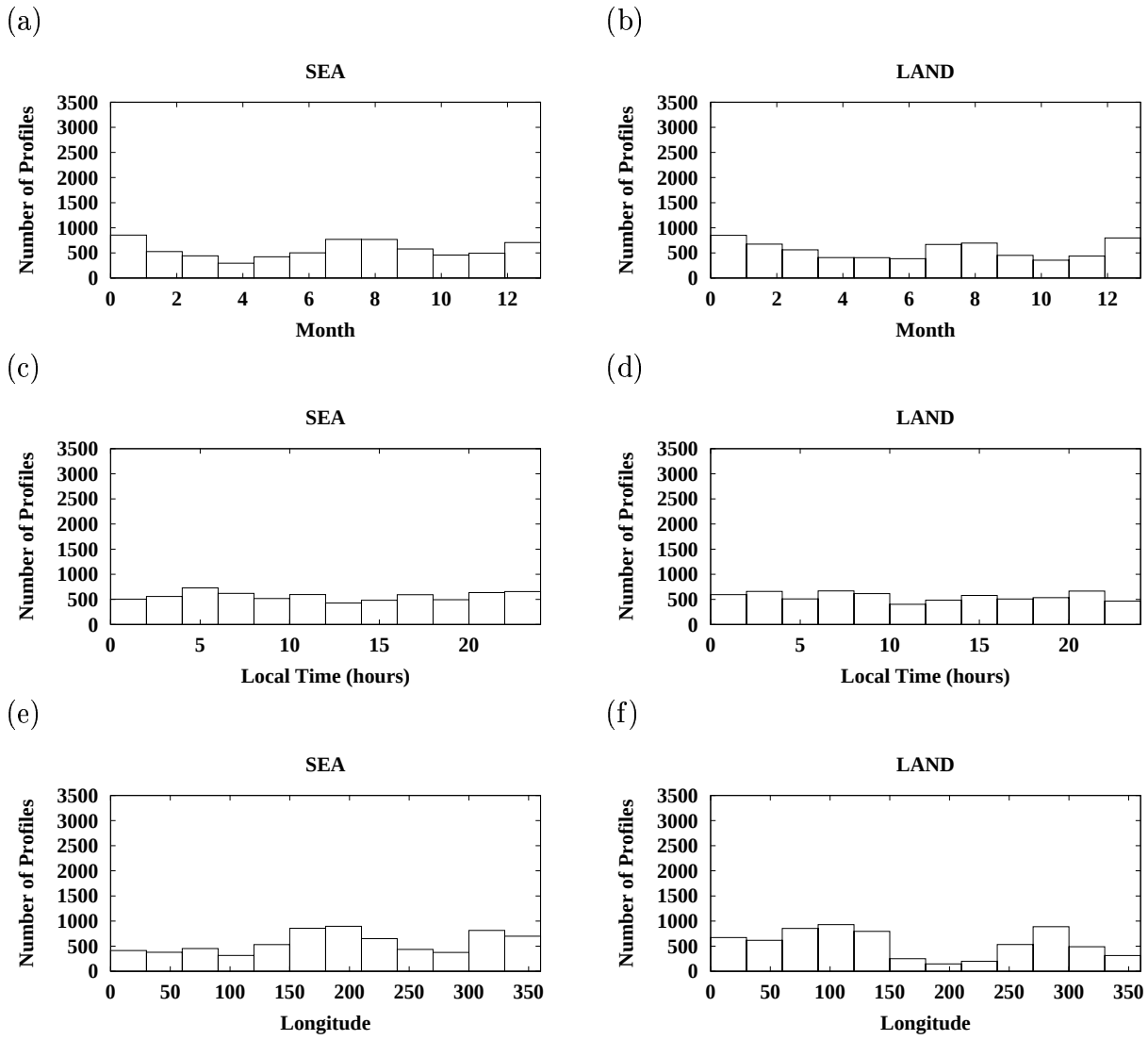

Figure 2: Same as previous.

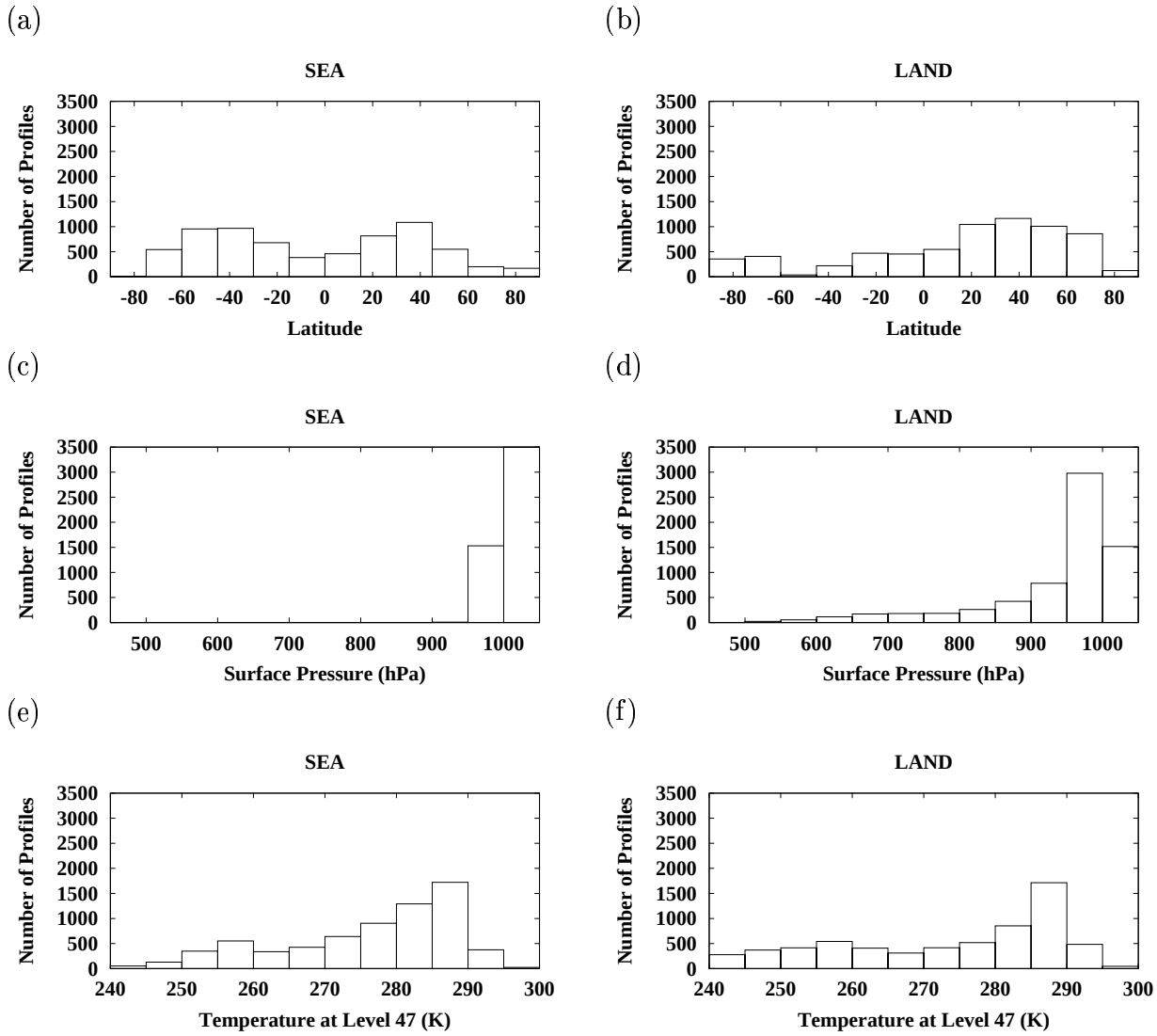

Figure 3: Same as previous. Layer 47 corresponds to a pressure level of 787 hPa when the surface pressure is 1000 hPa (see Table 2) and has been chosen as an example of the temperature and humidity layer histograms.

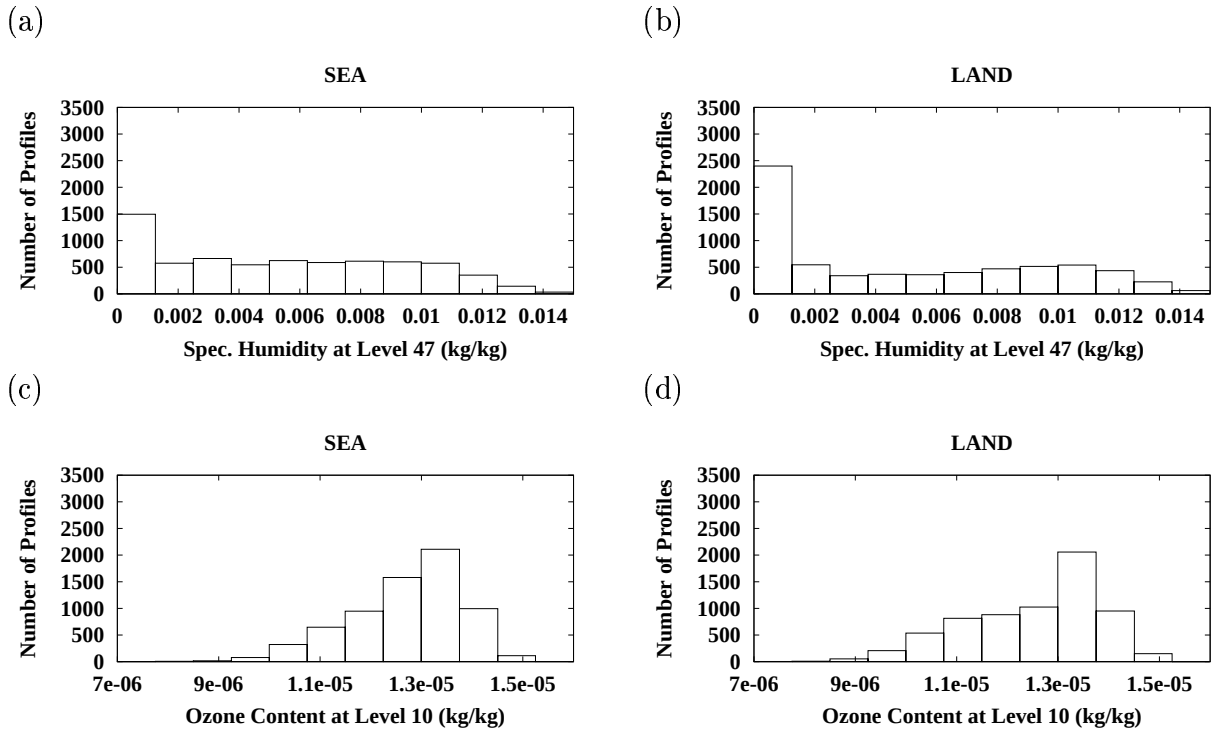

Figure 4: Same as previous. Layer 10 corresponds to a pressure level of 4 hPa when the surface pressure is 1000 hPa (see Table 1) and has been chosen as an example of the ozone layer histograms.

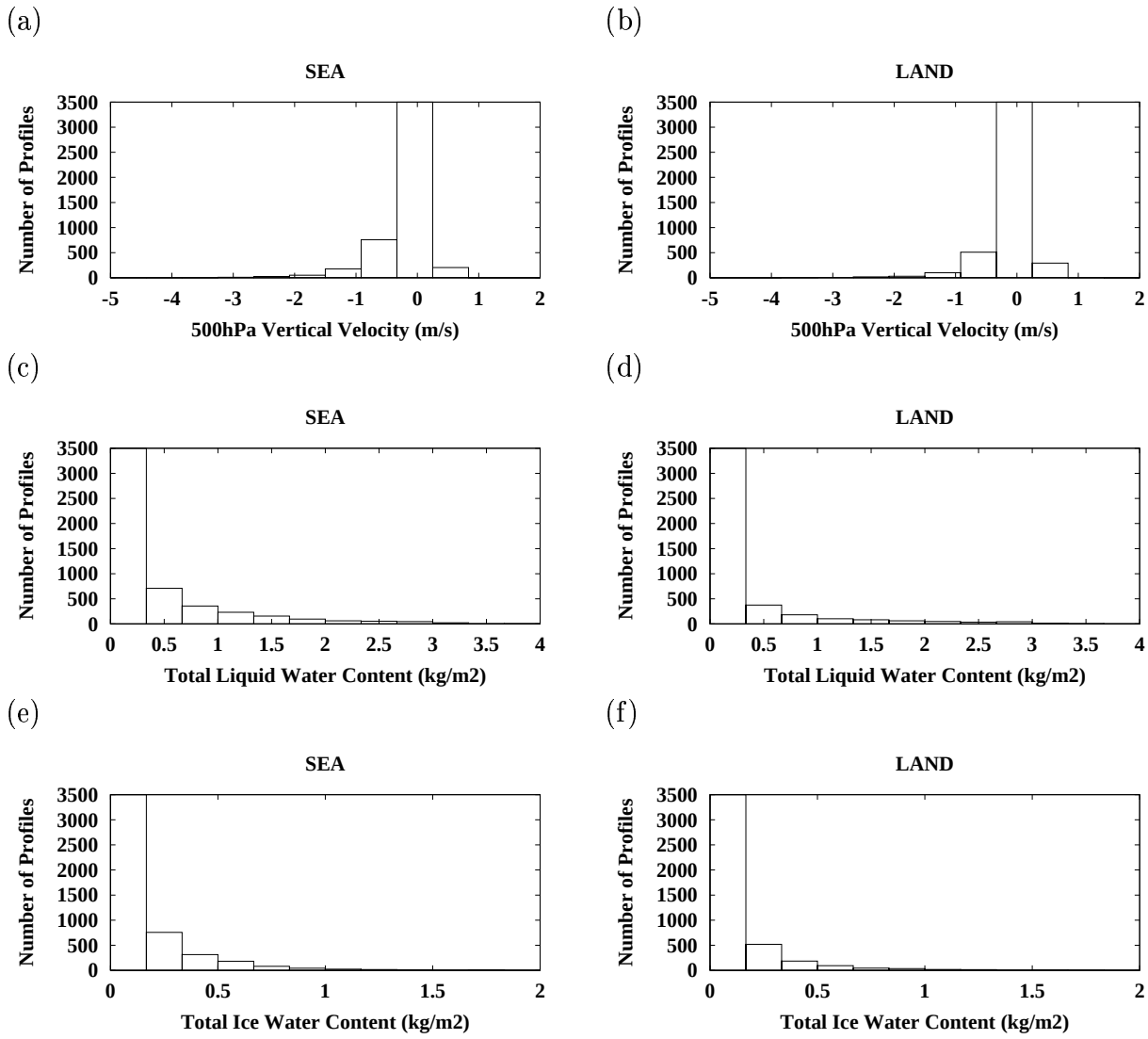

Figure 5: Same as previous.

(a) (b)

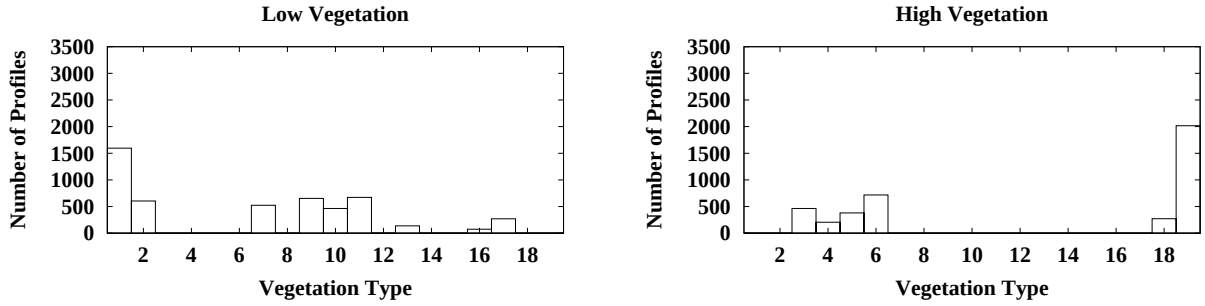

Figure 6: Distribution of the land profiles in the 60-level sampled database as a function of vegetation type (see Table 2) for high and low vegetation.

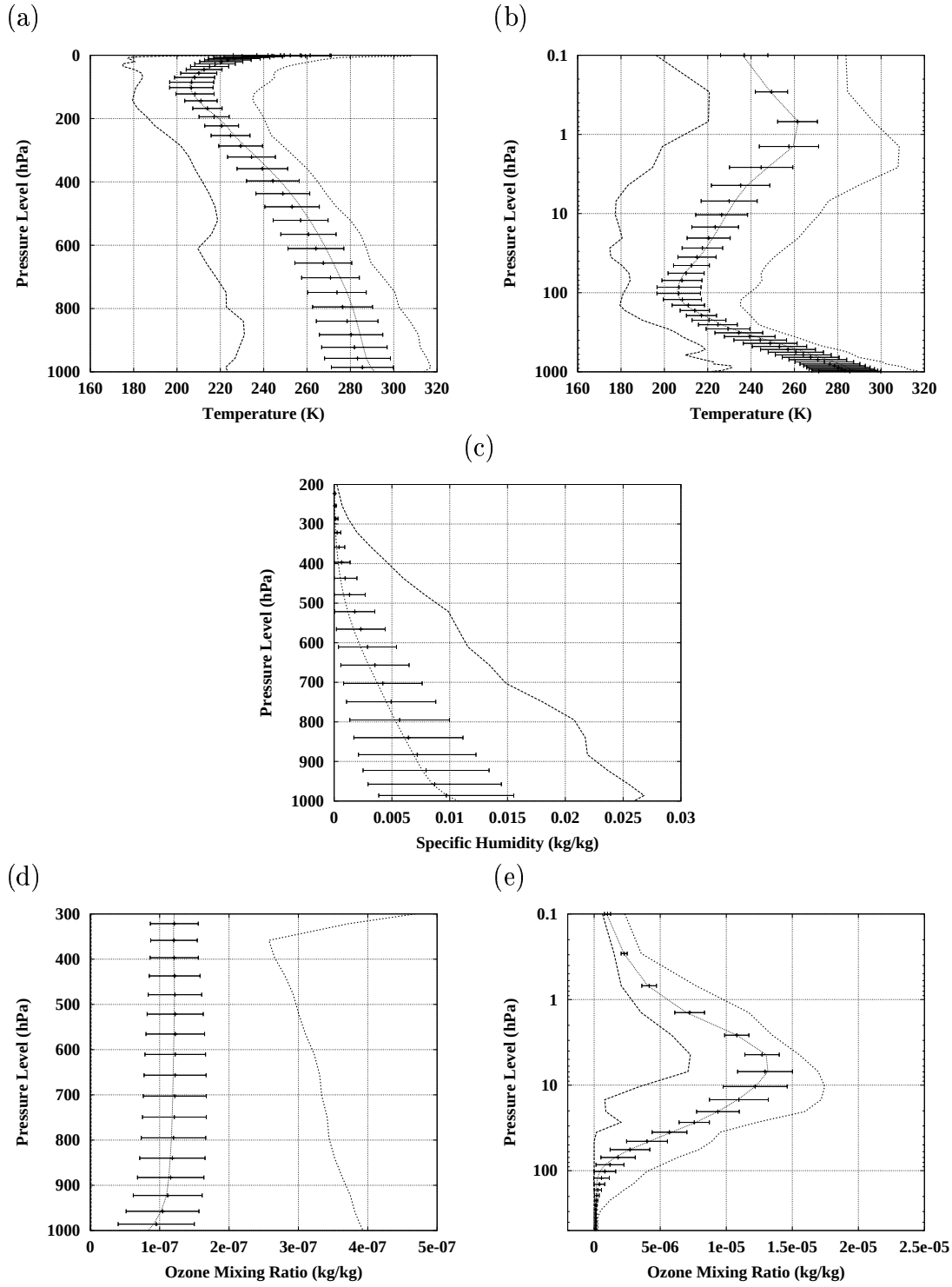

Figure 7: Statistics of the 60-level sampled database (noted 60L-SD). The outer curves show the minimum and maximum values. The horizontal bars have a length of twice the standard deviation and are centered at the mean. The inner curve is the median. The dataset is interpolated on a fixed pressure level grid: the 43-level RTTOV grid. Note that the 60-level sampled dataset has values below 1013 hPa, that are not shown here.

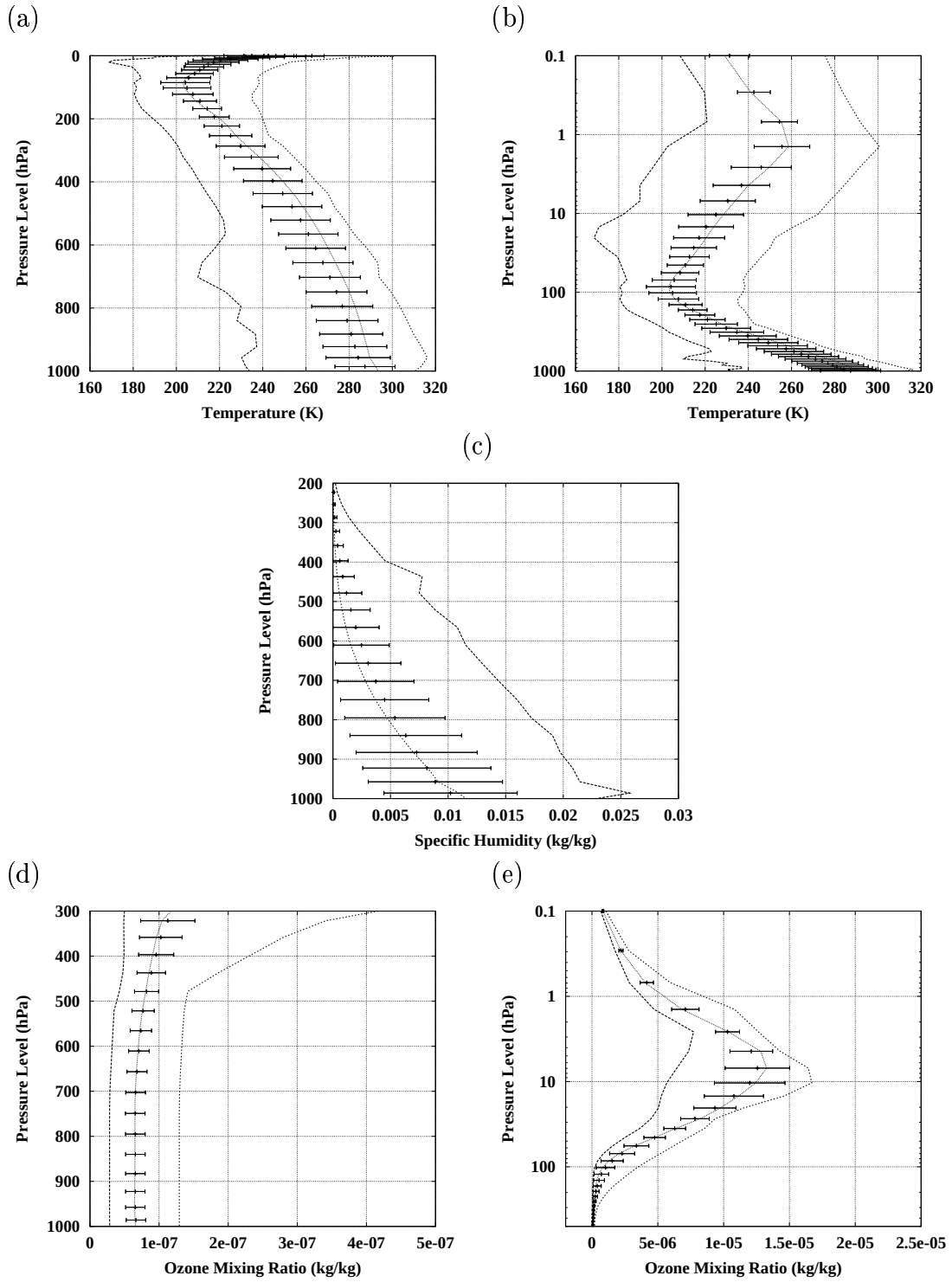

Figure 8: Same as previous, but for the 50L-SD.

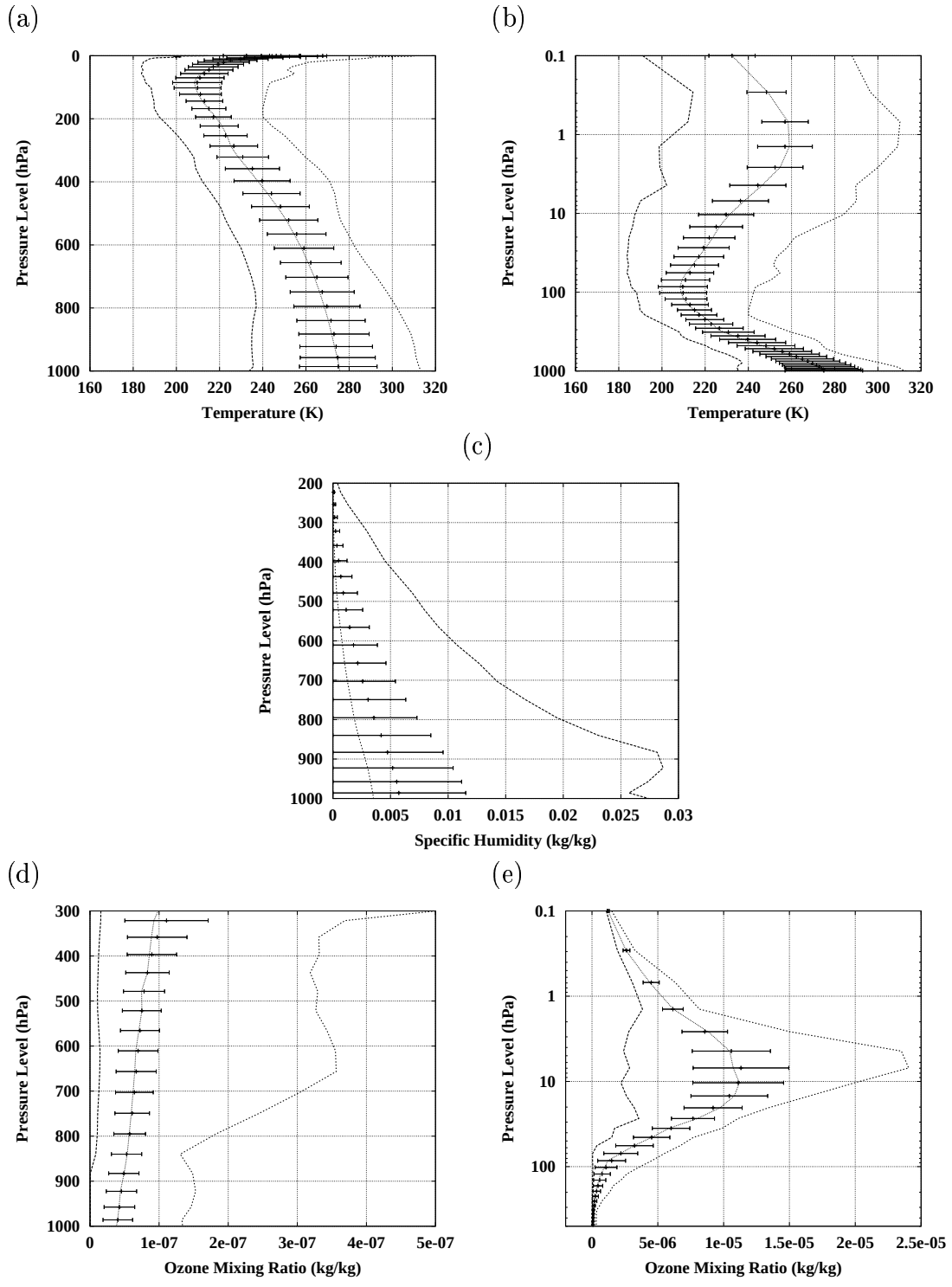

Figure 9: Same as previous, but for the TIGR-3 database.

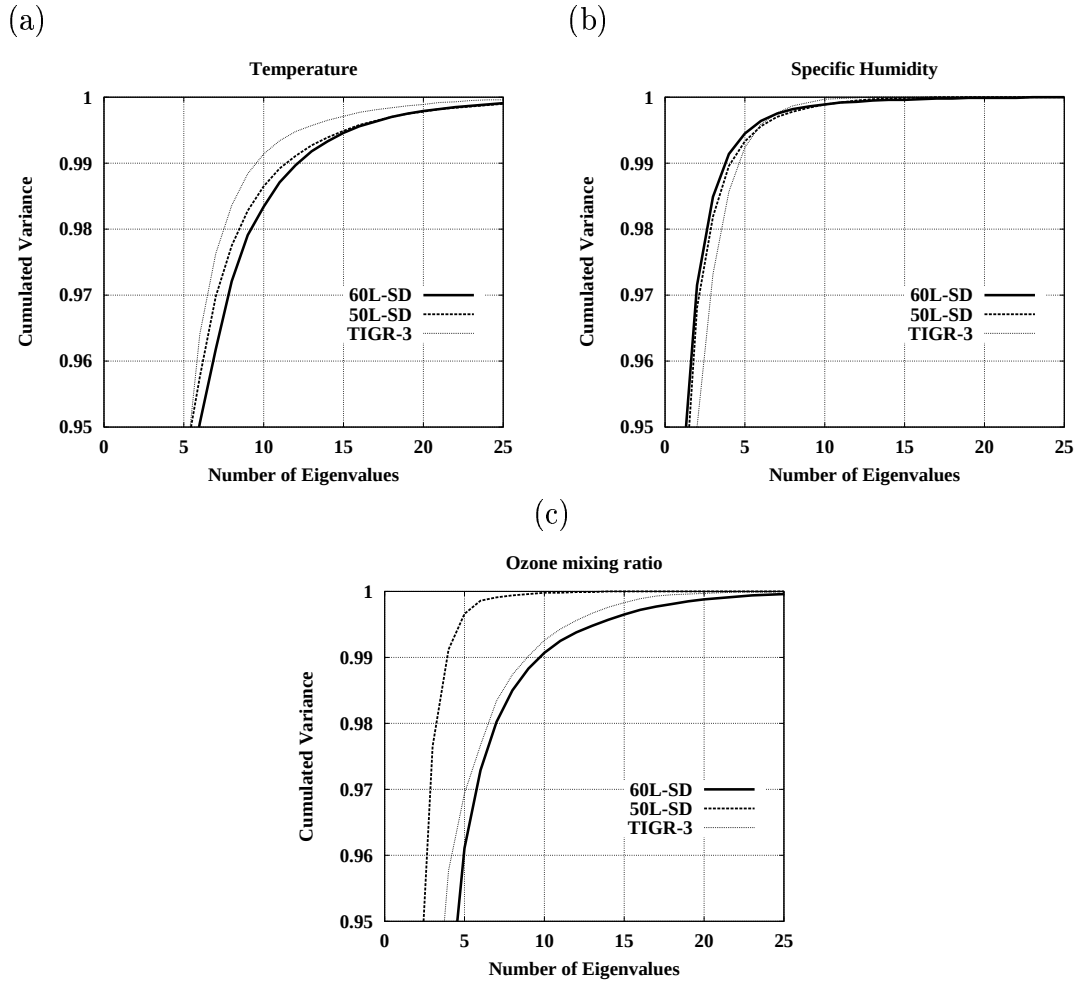

Figure 10: Cumulated variance as a function of the number of leading eigenvalues in the Principal Component Analysis of the temperature (a), specific humidity (b) and specific ozone (c) fields for the 60L-SD, the 50L-SD, and TIGR-3.

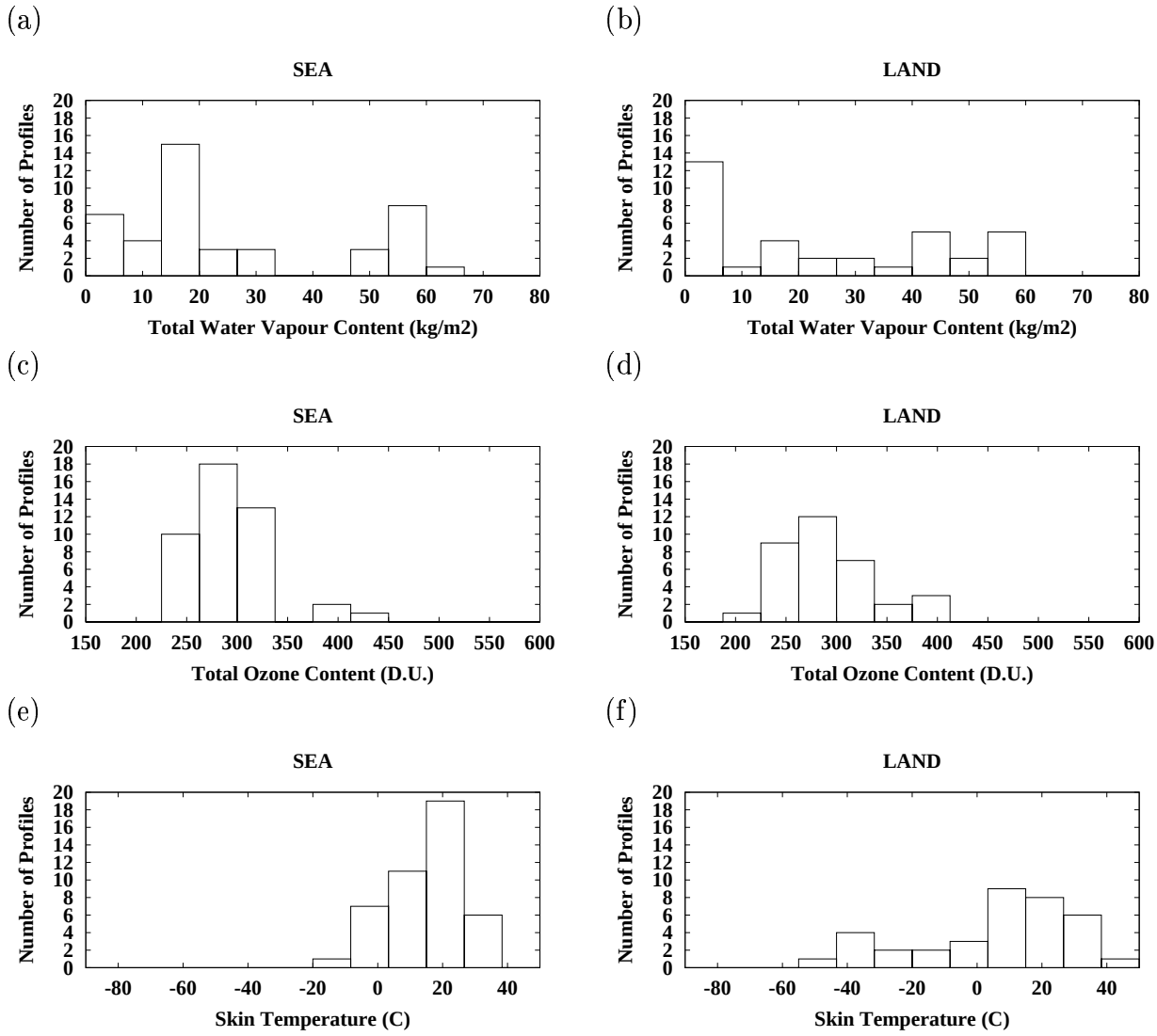

Figure 11: Distribution of the situations in the reduced 60-level sampled database (60L-SDr, 80 profiles) as a function of some variables and for each geotype. The total ozone content is given in Dobson Units.

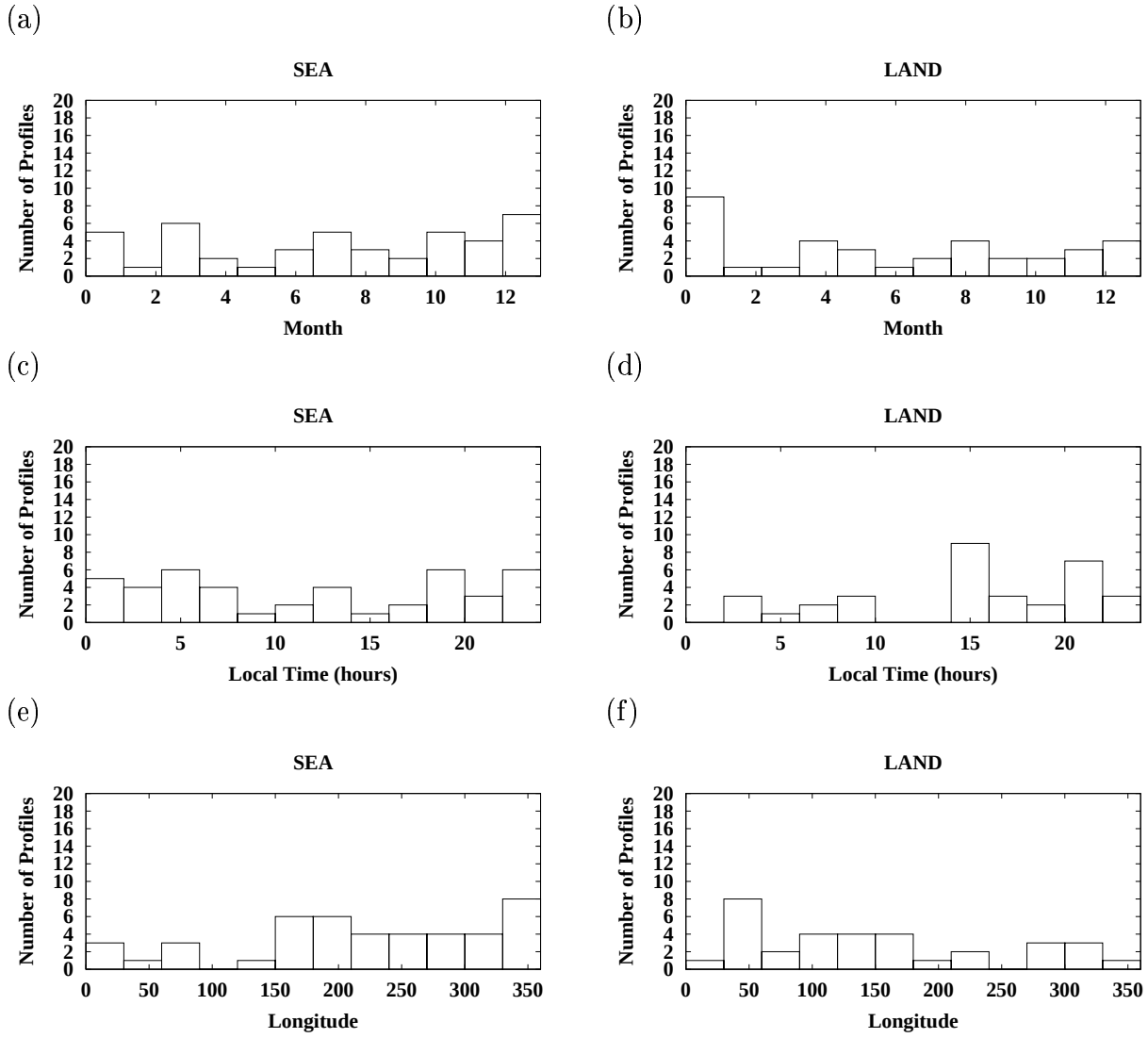

Figure 12: Same as previous.

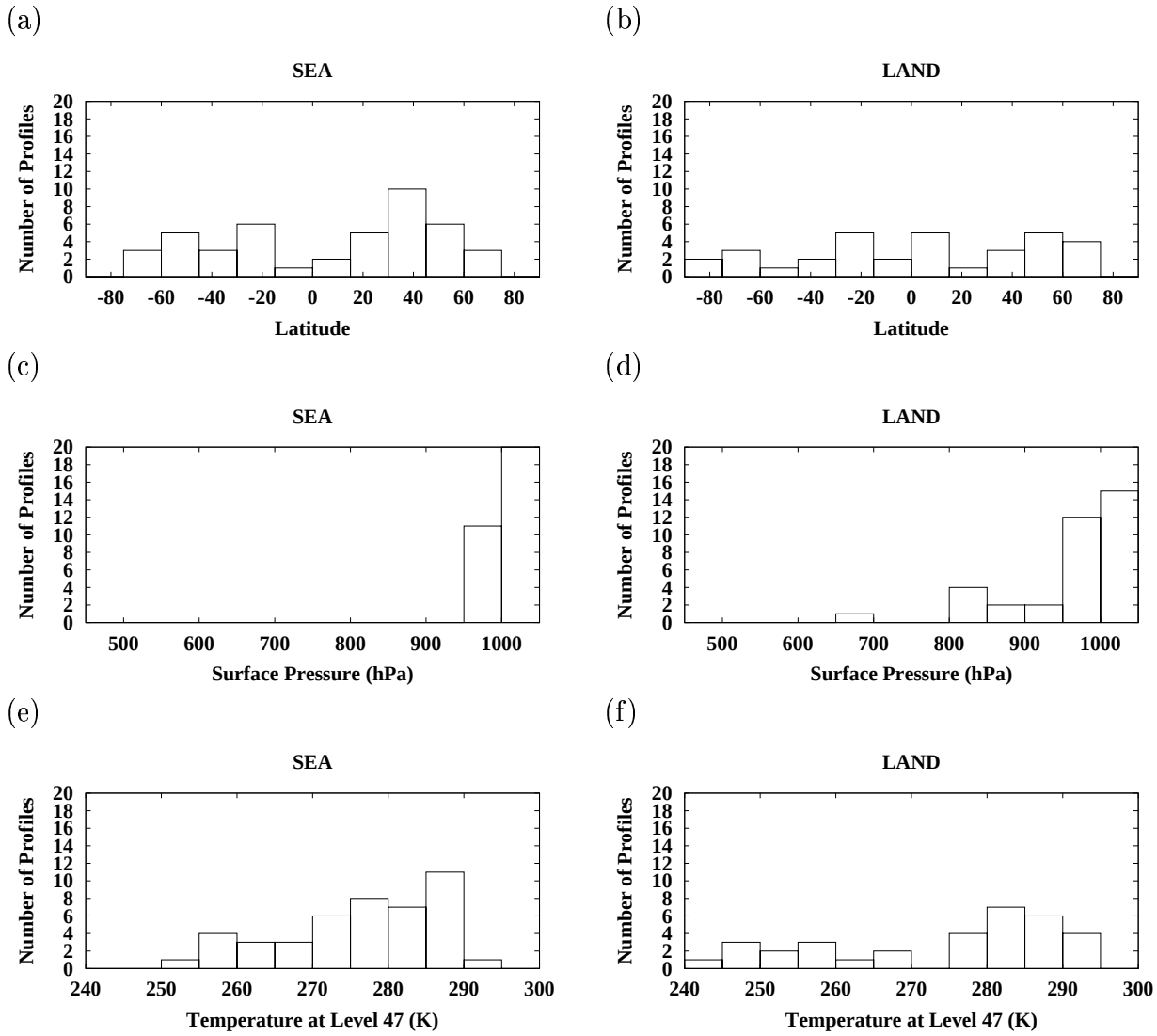

Figure 13: Same as previous. Layer 47 corresponds to a pressure level of 787 hPa when the surface pressure is 1000 hPa (see Table 1) and has been chosen as an example of the temperature and humidity layer histograms.

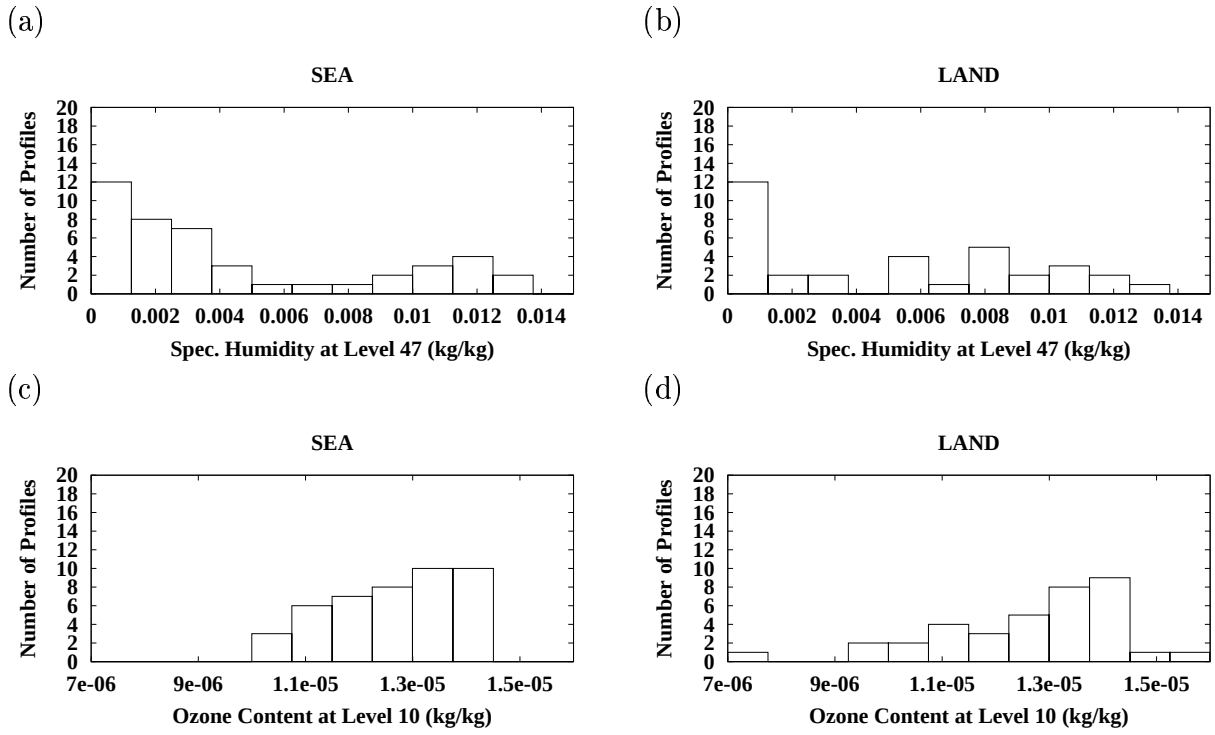

Figure 14: Same as previous. Layer 10 corresponds to a pressure level of 4 hPa when the surface pressure is 1000 hPa (see Table 1) and has been chosen as an example of the ozone layer histograms.

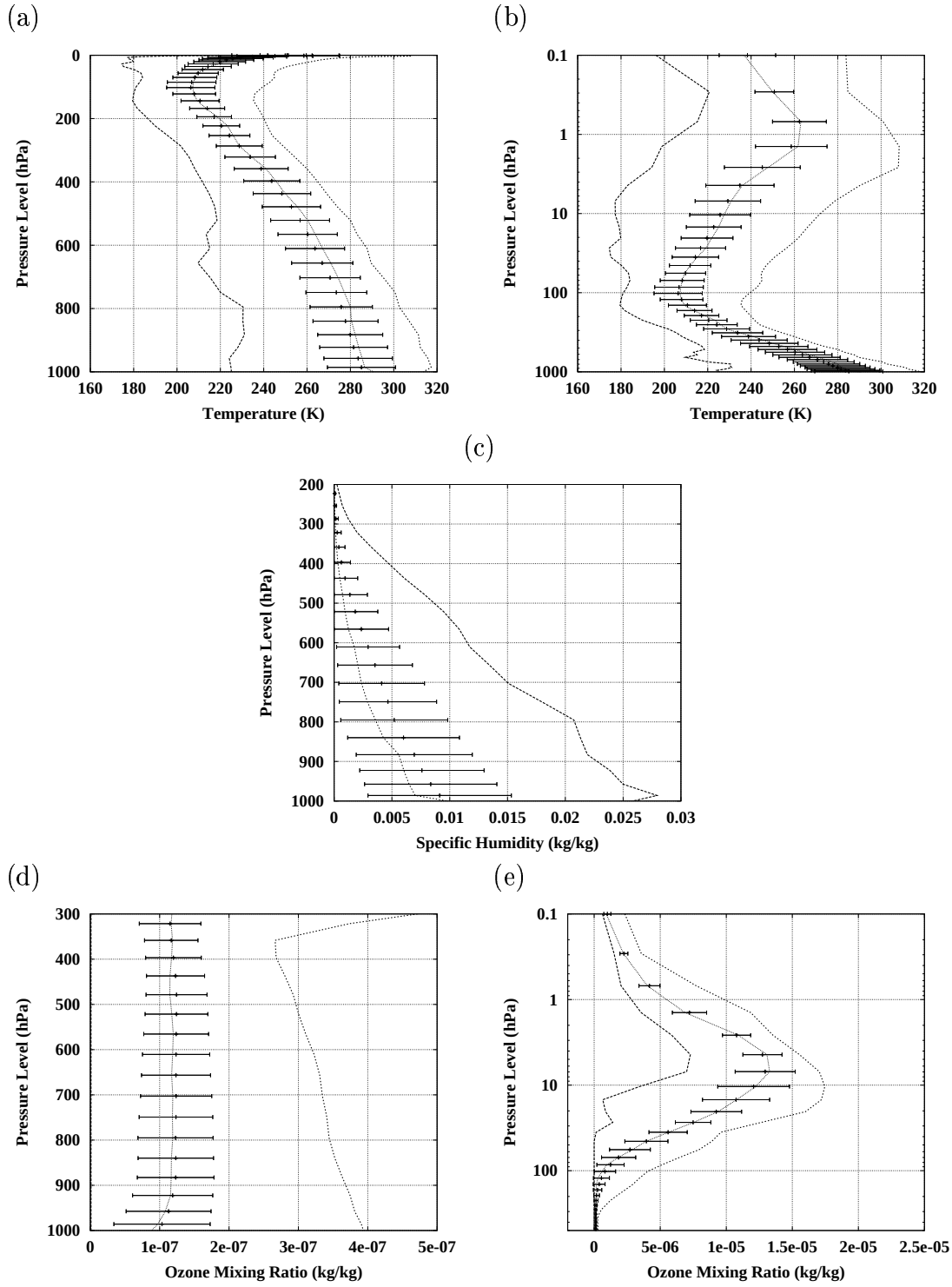

Figure 15: Statistics of the reduced 60-level sampled database (noted 60L-SDr). The outer curves show the minimum and maximum values. The horizontal bars have a length of twice the standard deviation and are centered at the mean. The inner curve is the median. The dataset is interpolated on a fixed pressure level grid: the 43-level RTTOV grid. Note that the 60L-SDr has values below 1013 hPa, that are not shown here.

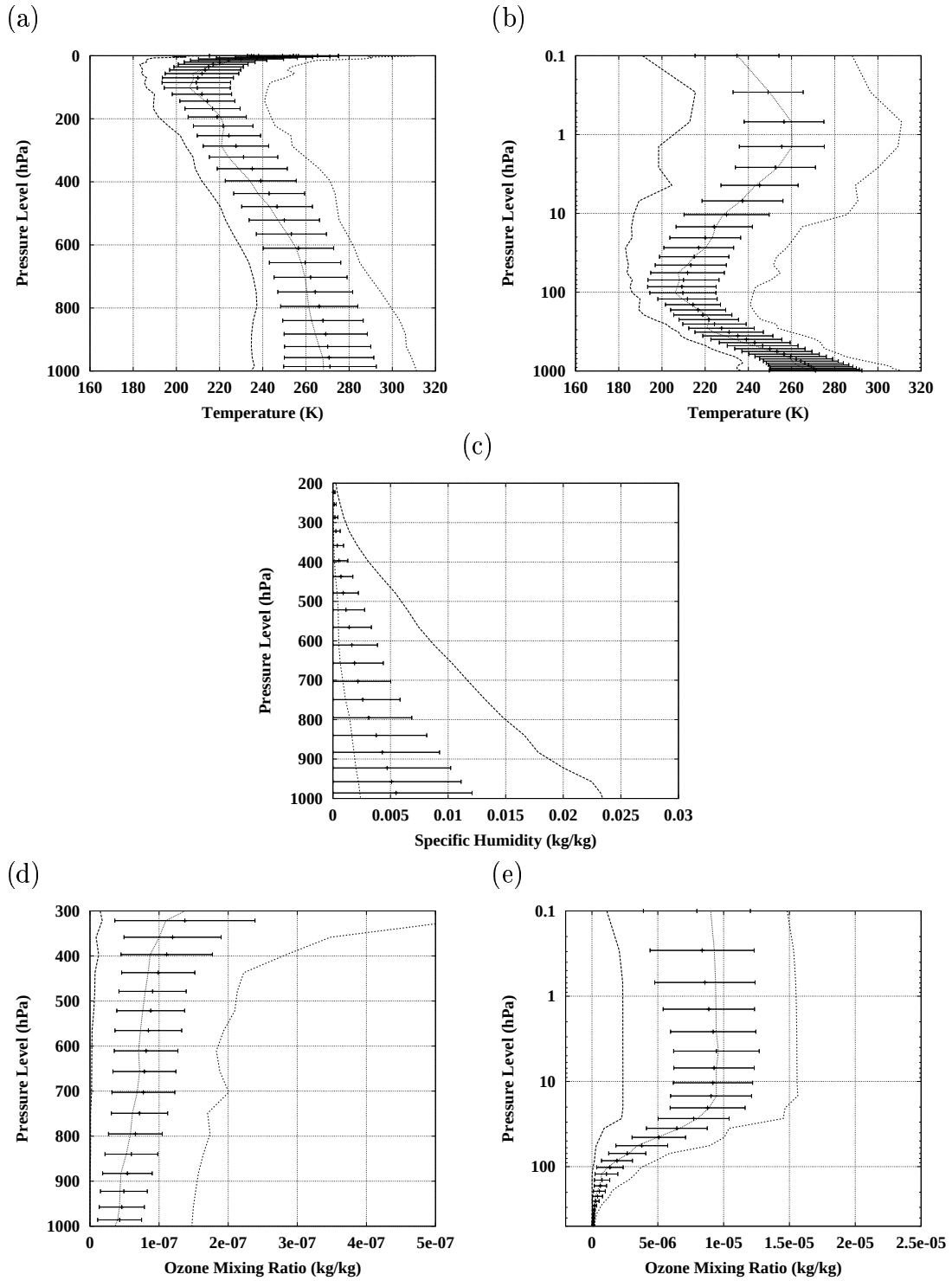

Figure 16: Same as previous, but for the RTTOV training sets.