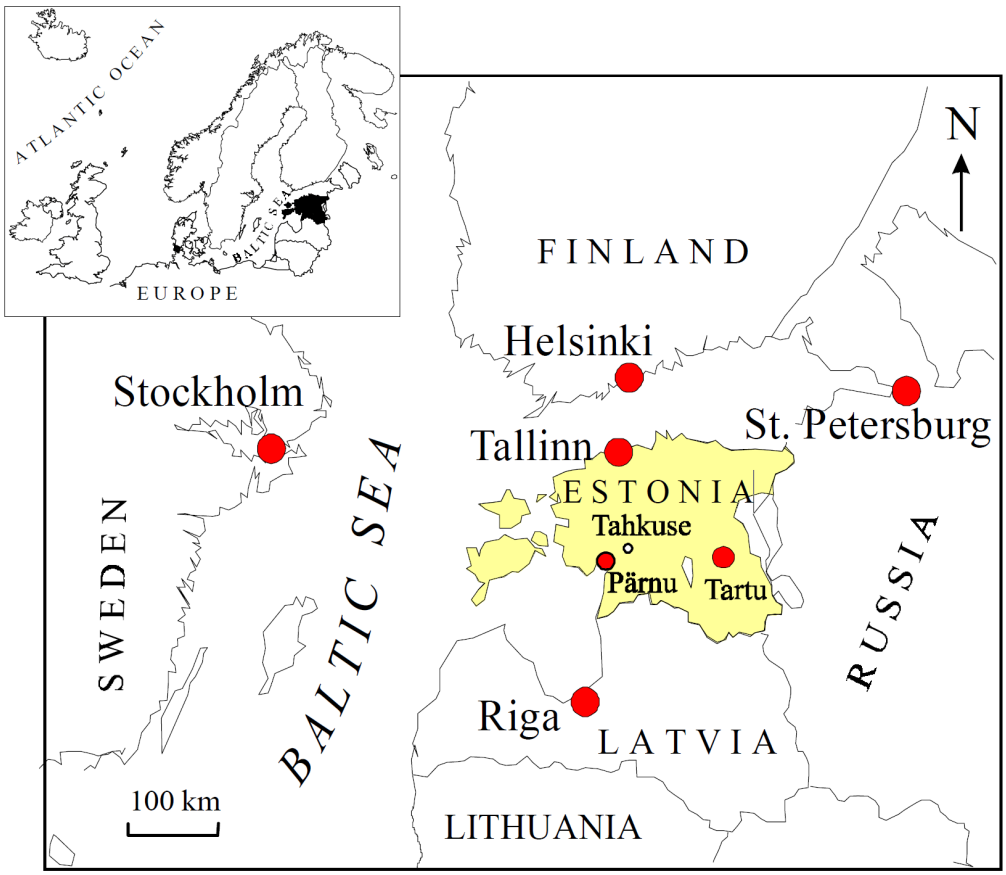
Dataset Tahkuse_2004_2014

U. Hõrrak, H. Iher, K. Komsaare, E. Tamm, H. Tammet
University of Tartu, Estonia

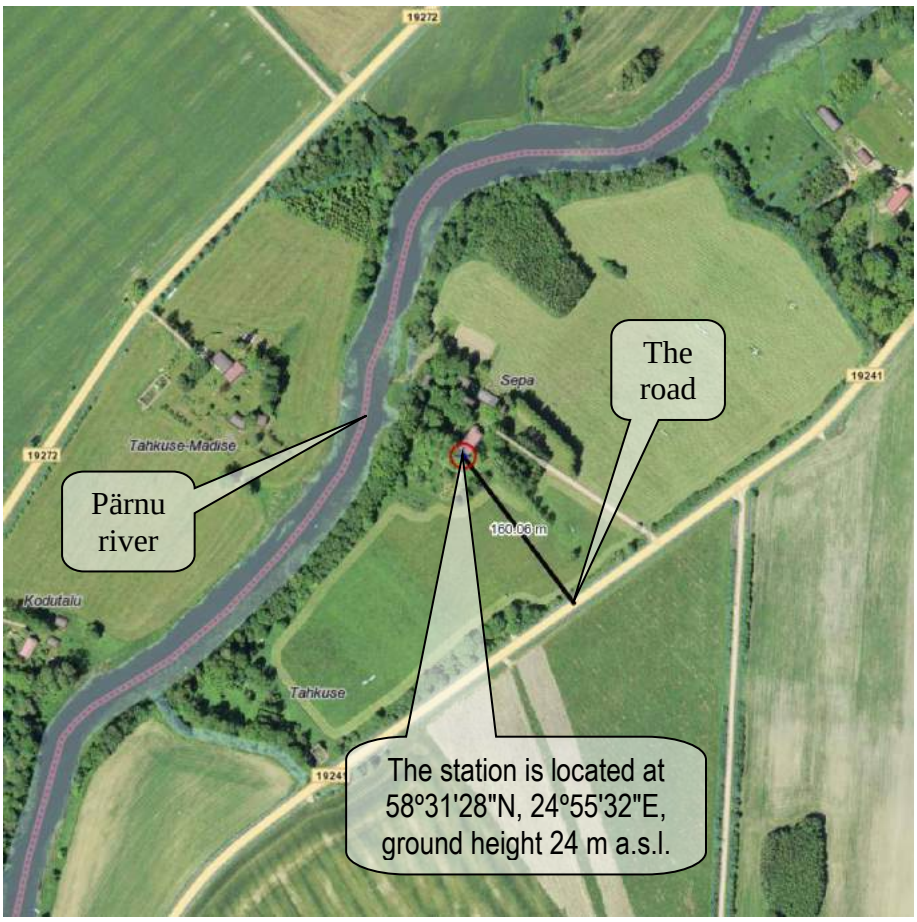
Contents

Introduction

The dataset makes available measurements, which were carried out after the control and data recording equipment of air ion spectrometers in the Tahkuse Air Monitoring Station was modernized in 2003 by Komsaare (2005). The dataset includes hourly averages of measurements for every hour from 01.01.2004 to 31.12.2014.



The station is located in the Sepa farmhouse in Tahkuse village about 27 km northeast of the city of Pärnu, Estonia. The farmhouse is surrounded with flat open landscape with some tree groups, grassland and agricultural land. Neighboring farmhouses are separated with distances of few hundreds of meter. Few cars per hour pass along the road about 160 m from the station. A detailed description of the landscape, station, and instruments is presented in the PhD thesis by Hõrrak (2001).



Location of the station

Air intake in the Sepa farmhouse

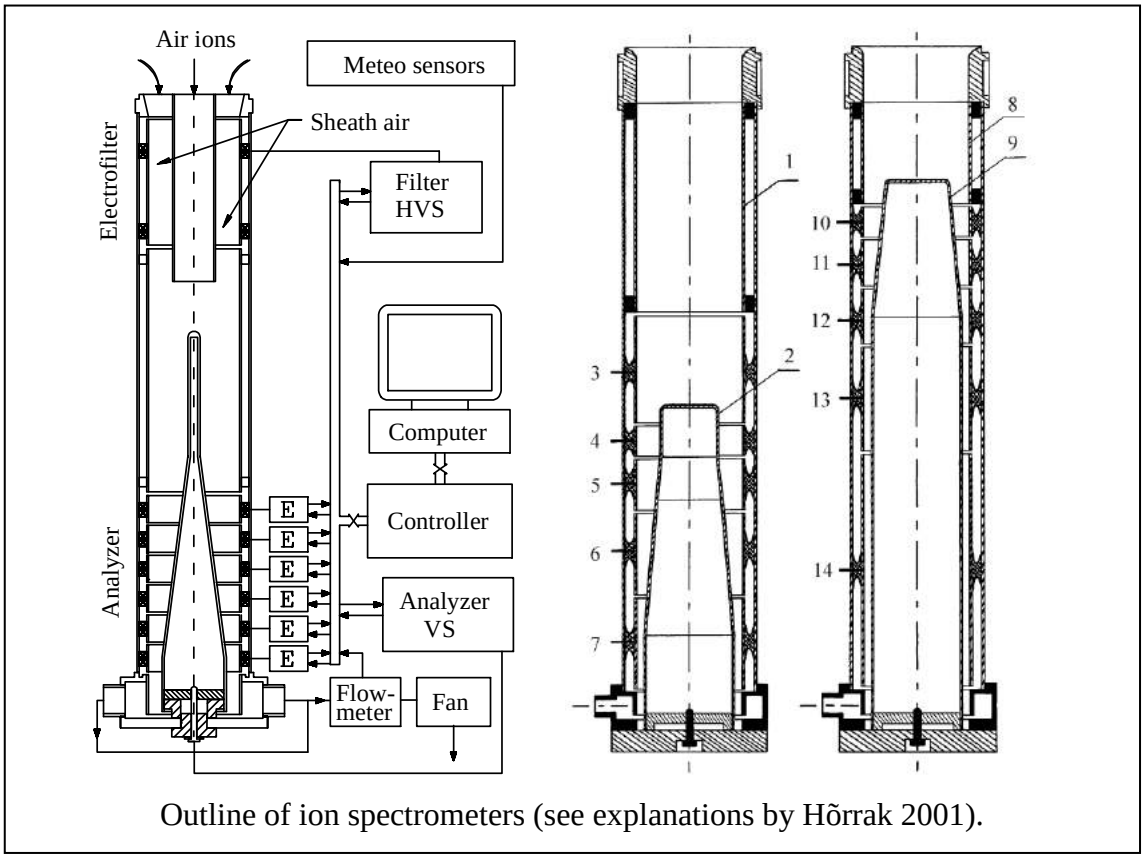
Service of instruments through a door in the container.

The first mobility spectrometer of natural air ions was installed in Tahkuse June 10, 1985 (Tammet et al., 1987). An important early success was discovery of intermediate ion bursts (Tammet et al., 1988; Hõrrak et al. 1998), which are known today as the new particle formation events or NPF events. An extended instrumentation for measurements of the air ion spectra in a wide mobility range was set up in July 1988. The control and data recording equipment was modernized in 2003 by Komsaare (2005). The air ion mobility spectrometers

are installed in a container in the attic of the building. The temperature in the container can be stabilized. The air intake is at a height of 5 m above the ground. The ion mobility spectrometers and arrangement of measurements are described in papers by Hõrrak et al. (1990, 2001), Tammet (1990), and Komsaare (2005). Additionally were recorded wind speed, wind direction, atmospheric pressure, relative humidity, and temperature of air in several locations. An analyzer of atmospheric NO_2, designed and manufactured in the University of Turku, Finland, was installed 1991. Regular measurement of aerosol particle size distribution by means of the electrical aerosol spectrometer EAS (Mirme, 1994; Tammet et al., 2002) started 2003. Monitoring of radon activity by means of the AlphaGUARD started 2007.

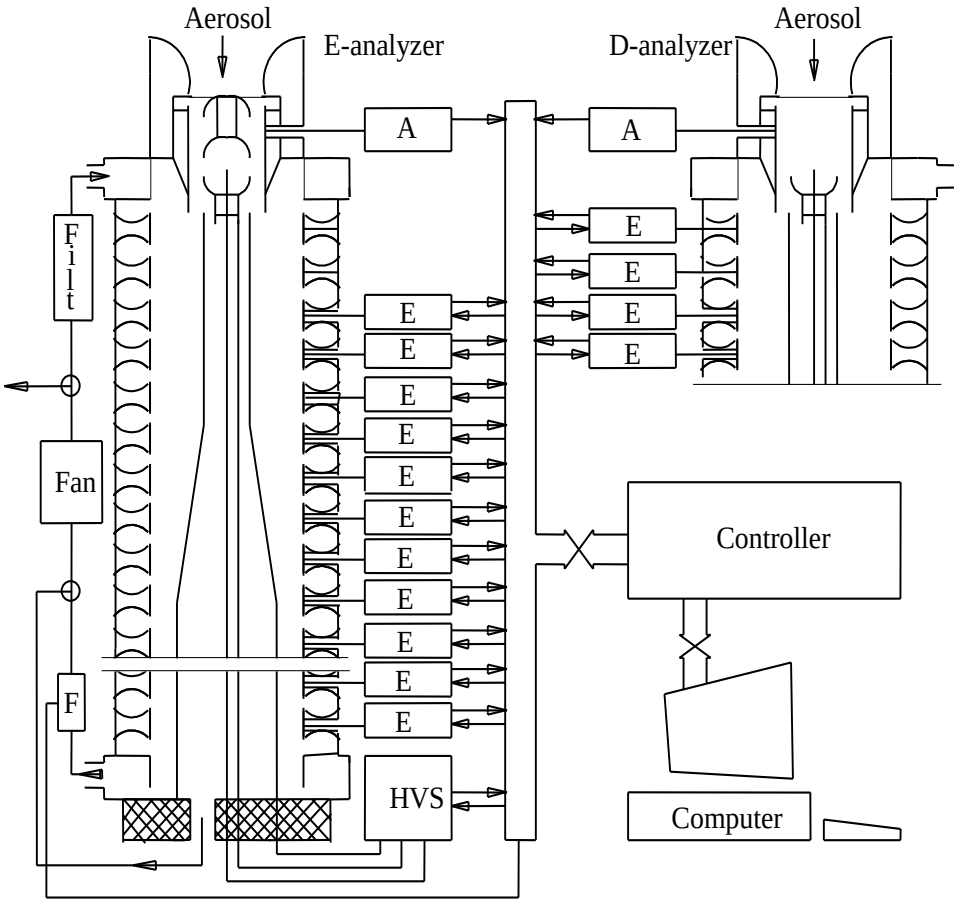
Today, the equipment of Tahkuse station is essentially complemented installing several sensors of air pollution gases, see http://www.envir.ee/sites/default/files/e-raamat_est.pdf and http://meteo.physic.ut.ee/tahkuse/#week. The present dataset extends until 2014 when the gas sensors were still not in use.

A large amount of data records was collected during long years. Most of the raw data is stored in the archive folder *Tahkuse_1991_2014_a1* in computers of Laboratory of Environmental Physics, University of Tartu. This folder contains 2,808 subfolders and 144,234 files with total capacity of 56,925,492,093 bytes. The structure of archive is complicated and it is not easy task to single out a necessary subset of data for a specific scientific analysis. An attempt to organize and publish the air ion measurements in an easy accessible format was formerly made for the period of 2003–2006. This data is included into the dataset *FinEstIon2003_2006* published in the *DataCite* web (Tammet, Hõrrak, 2015). Unfortunately, the Tahkuse data in *FinEstIon2003_2006* were not well preprocessed and contain some unexplained systematic errors.



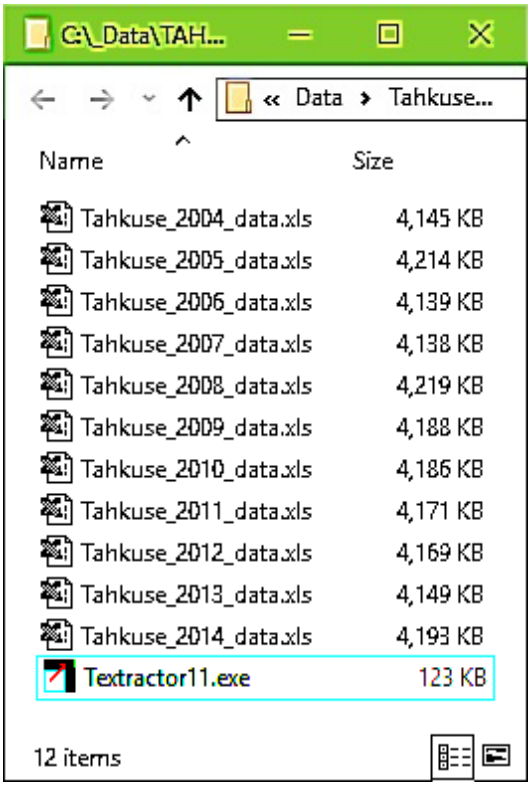
Outline of ion spectrometers (see explanations by Hõrrak 2001).

The present dataset covers 11 years. The detailed explanation of the data selection and organization is documented in the manuscripts by Tammet (2016a, 2016b), which are available only in Estonian. The algorithms of data processing are presented in a separate document *Tahkuse_data_2004_2014_supplement.pdf*.



Design of the EAS. A - corona current amplifiers,
E - electrometric amplifiers, VS - voltage supply, F - flow sensor.

Data structure



The data is saved in 11 tab-separated ANSI text files while every file covers just one year. The year is shown in the filename. Every file can be immediately opened and analyzed by MS Excel as well as by means of different computer programs, which are able to open and process the ANSI text. The extension of the filename "xls" facilitates usage of MS Excel. If the user changes the extension to "txt" then a double click on the file icon will open the file in the MS Notepad window.

A yearly file is composed as a table, which consist of 114 columns and 8761 or 8785 rows. The columns correspond to variables and rows to measurement hours. The first row is the header, which consists of short names of the variables. Following rows contain numbers, which present the one-hour average values of the variables. The rows are written without omissions for 24 hours per day even when the measurements were stopped and the values of variables are unknown. The cells with unknown (or missing in another terms) values are filled with the code of –99. If the user will prefer different code then the value –99 may be easily replaced with any specific code

using the "replace" operation in MS Excel. The new code of missing value can be numeric or not, e.g. expressions "NaA" or "NA", the symbol "?" as well as the empty cell "" are acceptable.

The measurements of ion and aerosol particle fractions contain high instrumental noise. As a result some measurements of the inherently positive quantity may appear negative and coincide with the missing value code of –99. In this case the measurement was deliberately shifted to –98 or–100, which excludes coincidences whereas brings about a unnoticeable low distortion when compared with the instrumental noise.

An excerpt of 12 columns and 1 + 8 rows from beginning of the data table is presented below:

year	month	day	hour	weekday	p:mb	T:C	RH%	v:m/s	vmax:m/s	v:deg	Tbox:C
2012	1	1	0	7	1006.2	-2.7	97.4	0.7	3.2	167	11.2
2012	1	1	1	7	1006.5	-2.8	97.9	1	3.3	259	10.8
2012	1	1	2	7	1006.9	-2.9	97.8	0.8	3.2	264	10.9
2012	1	1	3	7	1007.3	-2.7	97.7	0.8	2.9	267	11.1
2012	1	1	4	7	1007.7	-2.6	97.6	0.5	2.9	261	10.7
2012	1	1	5	7	1007.9	-2.8	97.5	1.1	3.7	245	10.7
2012	1	1	6	7	1008.1	-3.1	96.5	1	3.1	255	10.8
2012	1	1	7	7	1008.5	-3.2	96.1	0.8	3	271	11.1

Five first variables show the time of the measurement. The hours in a day are counted 0...23 and the weekdays 1..7, where 1 is Monday and 7 is Sunday. Air ion and aerosol particle concentrations are everywhere expressed in cm^{-3}. The ion mobility is expressed in $cm^2V^{-1}s^{-1}$. The air humidity is measured at a height of 2 m above the ground about 10 m away from the building. The air temperature is measured close to the air inlet. The wind sensor is installed above the flat open grassland on a mast in distance about 120 m from the building.

The dataset presents one row of measurement data for one hour. Actually, the size distribution of aerosol particles was recorded in Tahkuse every 5 minutes and the mobility distribution of air ions every 10 minutes. The process of compilation of the dataset was started with assembling of a table which contains six 10-minute data rows for every hour. This large dataset is stored in computers of Laboratory of Environmental physics, University of Tartu, and is accessible only in the laboratory internal network. The one-hour average values of variables were calculated according to the rule:

- at first the numbers of valid 10-minute measurements of a variable were counted for the first half of the hour $n1$ and for the second half of the hour $n2$,
- if $n1 = 0$ or $n2 = 0$ or $n1 + n2 < 3$ then the hourly average was decided missing and the code –99 was entered in the resulting table. Otherwise the hourly value was calculated as the arithmetic average of $n1 + n2$ valid ten-minute measurements.

Warning

The measurements of air ion mobility distribution in Tahkuse after 2003 may consist of unidentified and unexplained calibration errors in four fractions with nominal mobilities of 0.024, 0.28, 0.44, and 0.35 $cm^2V^{-1}s^{-1}$ (Tammet, 2016a, 2016b). The measurements of these fractions in the present dataset as well as in the dataset *FinEstIon2003_2006* (Tammet & Hõrrak, 2015) should not be immediately used in the substantial data analysis. The errors can be filtered out by disregarding the defective fractions when fitting the data with the two-power model of the distribution function, see page 12 in the present document. Thus the immediate measurements of intermediate and large air ions with $Z < 0.5$ $cm^2V^{-1}s^{-1}$should be replaced with the values of the approximation function, which parameters are tabulated as variables #100–113 in the dataset. The two-power model is not valid at $Z > 0.5$ $cm^2V^{-1}s^{-1}$, but the calibration of the instrument in the range of small ions is still reliable and corresponding measurements do not need the correction.

List of variables

The variables are briefly explained in the table. The space in the table is limited and extended explanations are presented in the section next to the table. NB: the values of *?*-marked variables may be incorrect and should be replaced with the two-power approximation.

#	Short name	Excel	Brief explanation (complements follow in the next section)
1	year	A	The years are 2004 ... 2014
2	month	B	The months are numbered 1 ... 12
3	day	C	The days in a month are numbered from 1 to 28, 29, 30, or 31
4	hour	D	The hours in a day are numbered 0 ... 23
5	weekday	E	The days in a week are numbered 1 (Monday) ... 7 (Sunday)
6	p:mb	F	Atmospheric pressure at a height of 30 m a.s.l.
7	T:C	G	Temperature of outdoor air near the inlet of spectrometers
8	RH%	H	Relative humidity of air at a height of 2 m in distance of 10 m
9	v:m/s	I	Average wind speed at a height of 12 m over the flat ground
10	vmax:m/s	J	Speed of wind bursts, average of six 10-minute maxima in a hour
11	v:deg	K	Wind direction at a height of 12 m over the flat ground 0 ... 360
12	Tbox:C	L	Temperature inside the container of air ion spectrometers
13	Tattic:C	M	Temperature of the attic near the instrument container
14	Tchimney:C	N	Temperature inside the chimney used to detect the burning of wood
15	rad:W/m2	O	Sun radiation intensity measured with a pyranometer
16	NO2:ug/m3	P	Concentration of NO_2 $\mu g/m^3$
17	flow:lpm	Q	Air flow in NO_2-meter. Value < 0.07 lpm indicates the malfunction of the instrument
18	Rn:Bq/m3	R	Concentration of radon activity measured with AlphaGUARD
19	Rn-s:Bq/m3	S	Estimate of instrumental noise in radon activity measurement
20	d@4.2	T	Aerosol particle size distribution $dN/d(\lg d)$ at d = 4.2 nm, cm^{-3}
21	d@7.5	U	Aerosol particle size distribution $dN/d(\lg d)$ at d = 7.5 nm, cm^{-3}
22	d@13	V	Aerosol particle size distribution $dN/d(\lg d)$ at d = 13 nm, cm^{-3}
23	d@24	W	Aerosol particle size distribution $dN/d(\lg d)$ at d = 24 nm, cm^{-3}
24	d@42	X	Aerosol particle size distribution $dN/d(\lg d)$ at d = 42 nm, cm^{-3}
25	d@75	Y	Aerosol particle size distribution $dN/d(\lg d)$ at d = 75 nm, cm^{-3}
26	d@133	Z	Aerosol particle size distribution $dN/d(\lg d)$ at d = 133 nm, cm^{-3}
27	d@237	AA	Aerosol particle size distribution $dN/d(\lg d)$ at d = 237nm, cm^{-3}
28	d@422	AB	Aerosol particle size distribution $dN/d(\lg d)$ at d = 422 nm, cm^{-3}
29	d@750	AC	Aerosol particle size distribution $dN/d(\lg d)$ at d = 750 nm, cm^{-3}
30	d@1334	AD	Aerosol particle size distribution $dN/d(\lg d)$ at d = 1.33 μm, cm^{-3}
31	d@2371	AE	Aerosol particle size distribution $dN/d(\lg d)$ at d = 2.37 μm, cm^{-3}
32	d@4217	AF	Aerosol particle size distribution $dN/d(\lg d)$ at d = 4.22 μm, cm^{-3}
33	d@7500	AG	Aerosol particle size distribution $dN/d(\lg d)$ at d = 7.5 μm, cm^{-3}
34	d:defect	AH	EAS noise index. The measurements with defect > 5000 were replaced with –99
35	Z+2.77	AI	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 2.77 $cm^2V^{-1}s^{-1}$, cm^{-3}
36	Z+2.22	AJ	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 2.22 $cm^2V^{-1}s^{-1}$, cm^{-3}
37	Z+1.77	AK	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 1.77 $cm^2V^{-1}s^{-1}$, cm^{-3}
38	Z+1.41	AL	Positive ion mobility distribution $dN/d(\lg Z)$ at Z =1.41 $cm^2V^{-1}s^{-1}$, cm^{-3}
39	Z+1.13	AM	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 1.13 $cm^2V^{-1}s^{-1}$, cm^{-3}
40	Z+0.88	AN	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.88 $cm^2V^{-1}s^{-1}$, cm^{-3}
41	Z+0.70	AO	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.70 $cm^2V^{-1}s^{-1}$, cm^{-3}
42	Z+0.56	AP	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.56 $cm^2V^{-1}s^{-1}$, cm^{-3}
43	Z+0.44	AQ	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.44 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
44	Z+0.35	AR	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.35 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
45	Z+0.28	AS	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.28 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
46	Z+0.213	AT	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.213 $cm^2V^{-1}s^{-1}$, cm^{-3}
47	Z+0.107	AU	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.107 $cm^2V^{-1}s^{-1}$, cm^{-3}
48	Z+0.051	AV	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.051 $cm^2V^{-1}s^{-1}$, cm^{-3}
49	Z+0.024	AW	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.024 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
50	Z+0.015	AX	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.015 $cm^2V^{-1}s^{-1}$, cm^{-3}
51	Z+0.0068	AY	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0068 $cm^2V^{-1}s^{-1}$, cm^{-3}
52	Z+0.0031	AZ	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0031 $cm^2V^{-1}s^{-1}$, cm^{-3}
53	Z+0.0014	BA	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0014 $cm^2V^{-1}s^{-1}$, cm^{-3}
54	Z+0.0007	BB	Positive ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0007 $cm^2V^{-1}s^{-1}$, cm^{-3}

#	Short name	Excel	Brief explanation (complements follow in the next section)
55	sum+0.3	BC	Apparent concentration in the integral condensor at Z_0 = 0.3 $cm^2V^{-1}s^{-1}$, cm^{-3}
56	Z–2.77	BD	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 2.77 $cm^2V^{-1}s^{-1}$, cm^{-3}
57	Z–2.22	BE	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 2.22 $cm^2V^{-1}s^{-1}$, cm^{-3}
58	Z–1.77	BF	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 1.77 $cm^2V^{-1}s^{-1}$, cm^{-3}
59	Z–1.41	BG	Negative ion mobility distribution $dN/d(\lg Z)$ at Z =1.41 $cm^2V^{-1}s^{-1}$, cm^{-3}
60	Z–1.13	BH	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 1.13 $cm^2V^{-1}s^{-1}$, cm^{-3}
61	Z–0.88	BI	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.88 $cm^2V^{-1}s^{-1}$, cm^{-3}
62	Z–0.70	BJ	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.70 $cm^2V^{-1}s^{-1}$, cm^{-3}
63	Z–0.56	BK	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.56 $cm^2V^{-1}s^{-1}$, cm^{-3}
64	Z–0.44	BL	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.44 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
65	Z–0.35	BM	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.35 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
66	Z–0.28	BN	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.28 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
67	Z–0.213	BO	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.213 $cm^2V^{-1}s^{-1}$, cm^{-3}
68	Z–0.107	BP	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.107 $cm^2V^{-1}s^{-1}$, cm^{-3}
69	Z–0.051	BQ	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.051 $cm^2V^{-1}s^{-1}$, cm^{-3}
70	Z–0.024	BR	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.024 $cm^2V^{-1}s^{-1}$, cm^{-3} *?*
71	Z–0.015	BS	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.015 $cm^2V^{-1}s^{-1}$, cm^{-3}
72	Z–0.0068	BT	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0068 $cm^2V^{-1}s^{-1}$, cm^{-3}
73	Z–0.0031	BU	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0031 $cm^2V^{-1}s^{-1}$, cm^{-3}
74	Z–0.0014	BV	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0014 $cm^2V^{-1}s^{-1}$, cm^{-3}
75	Z–0.0007	BW	Negative ion mobility distribution $dN/d(\lg Z)$ at Z = 0.0007 $cm^2V^{-1}s^{-1}$, cm^{-3}
76	sum–0.3	BX	Apparent concentration in the integral condensor at Z_0 = 0.3 $cm^2V^{-1}s^{-1}$, cm^{-3}
77	Z1noise	BY	Air ion spectrometer noise index for Z > 0.005 $cm^2V^{-1}s^{-1}$
78	Z2noise	BZ	Air ion spectrometer noise index for Z < 0.005 $cm^2V^{-1}s^{-1}$
79	n+	CA	Concentration of positive small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), cm^{-3}
80	Z+	CB	Average mobility of positive small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$)
81	fS/m+	CC	Conductivity of positive small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), fS/m
82	n–	CD	Concentration of negative small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), cm^{-3}
83	Z–	CE	Average mobility of negative small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$)
84	fS/m–	CF	Conductivity of negative small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), fS/m
85	n&	CG	Concentration of all small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), cm^{-3}
86	Z&	CH	Average mobility of all small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$)
87	fS/m&	CI	Conductivity of all small ions (Z > 0.5 $cm^2V^{-1}s^{-1}$), fS/m
88	q-eas	CJ	Quality index of EAS measurements
89	q-ion1	CK	Quality index of ion measurements at Z > 0.005 $cm^2V^{-1}s^{-1}$
90	q-ion2	CL	Quality index of ion measurements at Z < 0.005 $cm^2V^{-1}s^{-1}$
91	q-eas-st	CM	Stability of 10–300 nm aerosol particles during the preceding hour
92	q-ion-st	CN	Stability of 0.005–0.15 $cm^2V^{-1}s^{-1}$ ions during the preceding hour
93	q-ion-sym	CO	Polar symmetry of 0.005–0.15 $cm^2V^{-1}s^{-1}$ ions during the preceding hour
94	q-low	CP	Lowest value in the set of 6 above listed indices
95	a	CQ	Parameter of aerosol particle size distribution two-power-model
96	b	CR	Parameter of aerosol particle size distribution two-power-model
97	p	CS	Parameter of aerosol particle size distribution two-power-model
98	d0	CT	Parameter of aerosol particle size distribution two-power-model
99	dev%	CU	Deviation of aerosol measurements from the two-power-model
100	c+	CV	Parameter of positive ion mobility distribution two-power-model
101	g+	CW	Parameter of positive ion mobility distribution two-power-model
102	k+	CX	Parameter of positive ion mobility distribution two-power-model
103	Z0+	CY	Parameter of positive ion mobility distribution two-power-model
104	dev+%	CZ	Deviation of positive ion measurements from the two-power-model
105	c–	DA	Parameter of negative ion mobility distribution two-power-model
106	g–	DB	Parameter of negative ion mobility distribution two-power-model
107	k–	DC	Parameter of negative ion mobility distribution two-power-model
108	Z0–	DD	Parameter of negative ion mobility distribution two-power-model
109	dev–%	DE	Deviation of negative ion measurements from the two-power-model
110	c&	DF	Parameter of summary ion mobility distribution two-power-model
111	g&	DG	Parameter of summary ion mobility distribution two-power-model
112	k&	DH	Parameter of summary ion mobility distribution two-power-model
113	Z0&	DI	Parameter of summary ion mobility distribution two-power-model
114	dev&%	DJ	Deviation of summary ion measurements from the two-power-model

Explanation of variables

#1 – #13) The brief explanation in the table above is sufficient.

#14 Tchimney:C) Temperature sensor installed inside the chimney of the farmhouse. The smoke has a strong effect in the aerosol particle and air ion concentrations. The temperature of the chimney is the simplest indicator that allows to identify the periods when the particle and ion distributions are distorted by the local smoke emission.

#15 rad:W/m2) Tahkuse is not equipped for high quality actinometrical measurements. The intensity of radiation was measured by means of an old pyranometer without certificate and usage of the records is limited with study of correlations of air ion and aerosol parameters and sun radiation.

#16 NO2:ug/m3 & #17 flow:lpm) Concentration of air pollution gas NO_2 was measured according to the Salzman method by means of a computer controlled instrument designed and manufactured in the University of Turku, Finland. A critical process in this instrument is the ventilation. The results are considered reliable when the flow rate is at least 0.07 lpm. The value of the flow rate is recorded as the variable #17, which can be used for assessing the quality of NO_2 records.

#18 Rn:Bq/m3 & #19 Rn-s:Bq/m3) The radon concentration was measured at a height of 5 m from the ground by the AlphaGUARD monitor, which has reputation as an exact and stable instrument. However, the concentration of radon in Tahkuse is very low and comparable with the instrumental noise. The noise estimate delivered by the instrument itself is recorded as variable #19 with aim to provide possibility to assess the reliability of measurements in different situations. The radon monitor was installed in Tahkuse later than other instruments and the measurements of radon are not available for years of 2004–2006.

#20 d@4.2 – #33 d@7500) The aerosol particle size distribution was measured by means of Electric Aerosol Spectrometer designed in the University of Tartu and manufactured by company Airel Ltd (http://www.airel.ee). The instrument is described in papers by Mirme (1994) and by Tammet et al. (2002). It delivers values of the distribution function of particle concentration according the decimal logarithm of the diameter $dN/d(\lg d)$. The particle diameter is expressed in nanometers and the distribution function in cm^{-3}.

#34) d:defect shows an estimate of the instrumental noise in the EAS. The data can be discriminated according the value of this parameter:

Percentage of measurements	Critical level of *defect*
50%	93
90%	540
99%	3440
99.38	5000

0.62% of measurements with *defect* > 5000 were disqualified and the values of variables #20 – 34 in corresponding rows of the data table were replaced with the code of missing value –99.

#35 Z+2.77 – #54 Z+0.0007) Atmospheric ion mobility distribution is presented with the distribution function of elementary charge number concentration according the decimal logarithm of the mobility $dN/d(\lg Z)$. The concentration is expressed in cm^{-3}. The symbol + or – next to Z in the short name marks the polarity. The number next the polarity indicates the mobility expressed in $cm^2V^{-1}s^{-1}$. The mobility distribution was measured by means of mobility spectrometers designed and manufactured in the University of Tartu. Measuring principle and the control algorithm are described in papers Hõrrak et al (1990), Tammet (1990), Hõrrak (2001) and Komsaare (2005). Technical details of data preparation are explained in manuscripts by Tammet (2016a, 2016b) written in Estonian. The mobility spectrometer has 20 differential channels, which geometric centers and logarithmic widths are presented in the table:

Channel	Z_c	$\Delta_{lg}Z$
1	2.77	0.0964
2	2.22	0.0975
3	1.77	0.0978
4	1.41	0.0989
5	1.13	0.0976
6	0.88	0.1006
7	0.696	0.0975
8	0.556	0.0978
9*	0.443	0.0989
10*	0.353	0.0976

Channel	Z_c	$\Delta_{lg}Z$
11*	0.280	0.1051
12	0.213	0.2911
13	0.1067	0.3093
14	0.0507	0.3374
15*	0.0238	0.3208
16	0.01481	0.3542
17	0.00667	0.3382
18	0.00307	0.336
19	0.00141	0.3421
20	0.00065	0.3325

NB: the records of *-marked channels should be replaced with two-power approximation.

#55 sum+0.3) Apparent concentration of positive small ions as measured with an integral aspiration condenser at the limiting mobility of $Z > 0.3$ $cm^2V^{-1}s^{-1}$.

#56 Z–2.77 – #75 Z–0.0007) Negative ion mobility distribution. The variables 56–75 have the same meaning as variables 35–54 but express the distribution of negative ions.

#76 sum–0.3) Apparent concentration of negative small ions as measured with an integral aspiration condenser at the limiting mobility of $Z > 0.3$ $cm^2V^{-1}s^{-1}$.

#77 Z1noise) Noise index considering as positive as negative ion fractions with $Z > 0.005$ $cm^2V^{-1}s^{-1}$. A rough recommendation is to rate the measurements good when Z1noise < 10 and bad when Z1noise > 100.

#78 Z2noise) Noise index considering as positive as negative ion fractions with $Z < 0.005$ $cm^2V^{-1}s^{-1}$ (this is the last three fractions). A rough recommendation is to rate the measurements good when Z2noise < 20 and bad when Z2noise > 200.

#79 n+) Concentration of positive small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) calculated as the sum of concentrations of first 8 mobility fractions and presented in cm^{-3}.

#80 Z+) Average mobility of positive small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) calculated considering the first 8 mobility fractions and presented in $cm^2V^{-1}s^{-1}$.

#81 fS/m+) The fraction of air conductivity caused by positive small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) calculated considering the first 8 mobility fractions and presented in fS m^{-1}.

#82 n-) Concentration of negative small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) calculated as the sum of concentrations of first 8 mobility fractions and presented in cm^{-3}.

#83 Z-) Average mobility of negative small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) calculated considering the first 8 mobility fractions and presented in $cm^2V^{-1}s^{-1}$.

#84 fS/m-) The fraction of air conductivity caused by negative small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$), calculated considering the first 8 mobility fractions, and presented in fS m^{-1}.

#85 n&) Concentration of small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) of both polarities, calculated as the sum of n+ and n–, and presented in cm^{-3}.

#86 Z&) Average mobility of small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) of both polarities, calculated considering the first 8 mobility fractions and presented in $cm^2V^{-1}s^{-1}$.

#87 fS/m&) The fraction of air conductivity caused by small ions ($Z > 0.5$ $cm^2V^{-1}s^{-1}$) of both polarities, and presented in fS m^{-1}.

Q-indices) Following 6 variables a rank-type quality indices, which characterize the position of noise or irregularity of the selected variable in a downward sorted list of measurements available in the dataset. The quality of a missing measurement was equated to zero. Two alternative approaches are used. Three indices are based on the estimates of instrumental

noise available in the original set of measurements. Other three indices are based on the dispersion of values inside the one hour time interval. The dispersion of a distribution function during a specific hour is estimated taking into account a sample composed from 6 ten-minute measurements of 6 fractions, which consists of 36 numbers. If any of these numbers equals to –99 then the index is left unknown. Otherwise the dispersion is estimated as the mean-square deviation in the sample of the 36 numbers. After the dispersions for all hours are estimated, the measurements of a variable were sorted according to the dispersion through the full dataset including all 11 years. The value of the index is the ratio of the position of the specific measurement to the total number of measurements and scaled so that the value 1 corresponds to the most noisy instance and the value of 1000 to the most smooth and uniform hourly sample. The q-indices have uniform distribution in the interval 1...1000 and their arithmetic averages are 500.5. The application field of q-indices is discrimination of the time periods when the hypothesis of stationary state of atmospheric aerosol and ionization can be applied in theoretical models. An interpretation example: the value of 900 shows that 90% of measurements are worse than the considered specific measurement or in other words, the specific measurement belongs to the 10% of the best measurements.

#88 q-eas) Quality index of size distribution measurements evaluated according to variable #34 called the *defect*.

#89 q-ion1) Quality index of mobility distribution measurements evaluated according to variable #77, which represents instrumental noise for ions of mobility $Z > 0.005$ $cm^2V^{-1}s^{-1}$.

#90 q-ion2) Quality index of mobility distribution measurements evaluated according to variable #78, which represents instrumental noise for ions of mobility $Z < 0.005$ $cm^2V^{-1}s^{-1}$.

#91 q-eas-st) Quality index which shows the stability of particle size distribution during the running hour. The dataset is composed on the basis of measurements of six size fractions in the size interval of 7–100 nm. The instability is defined as the dispersion of the sample of 36 numbers.

#92 q-ion-st) An analogue of q-eas-st but applied to one hour sample of four fractions of positive and four fractions of negative ion mobility distribution in the mobility interval of 0.005–0.015 $cm^2V^{-1}s^{-1}$.

#93 q-ion-sym) Polar asymmetry is a characteristic of the difference between positive and negative small ion concentrations (variables #55&76). The measure of the asymmetry is calculated as the mean square of the differences between concentrations on ions of opposite polarity in the sample of six measurements made during one hour.

#94 q-low) This index is defined as the lowest of 6 previous indices. If q-low is high then any q-index of six cannot be low and the quality of data may be assessed good. Example: the requirement $q\text{-}low > 500$ means that every of the 6 specific q-indices must belong to the top 50%. In a set of uncorrelated variables it happens with probability of 1.6%. Due the correlation the actual probability of such event is 3.8%.

Next 20 variables are parameters of two-power models described in the following two sections.

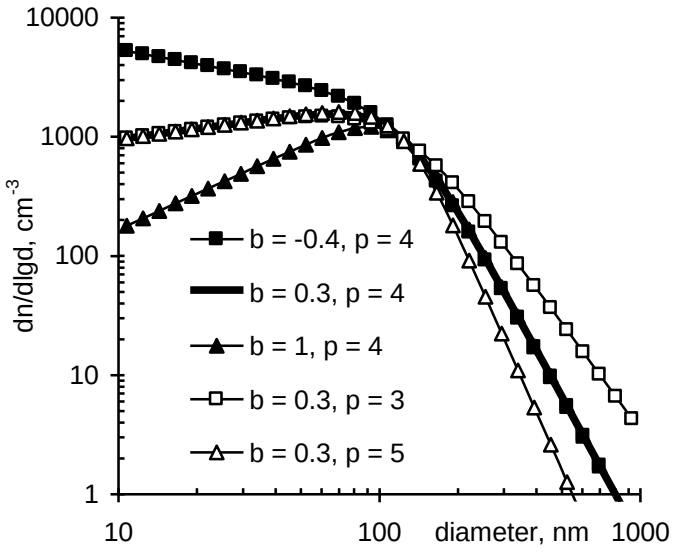
#95 a) Parameter of two-power model of aerosol size distribution.
#96 b) Parameter of two-power model of aerosol size distribution.
#97 p) Parameter of two-power model of aerosol size distribution.
#98 d0) Parameter of two-power model of aerosol size distribution.
#99 dev%) Parameter of two-power model of aerosol size distribution.
#100 c+) Parameter of two-power model of positive air ion mobility distribution.
#101 g+) Parameter of two-power model of positive air ion mobility distribution.
#102 k+) Parameter of two-power model of positive air ion mobility distribution.
#103 Z0+) Parameter of two-power model of positive air ion mobility distribution.

#104 dev+%) Parameter of two-power model of positive air ion mobility distribution.
#105 c-) Parameter of two-power model of negative air ion mobility distribution.
#106 g-) Parameter of two-power model of negative air ion mobility distribution.
#107 k-) Parameter of two-power model of negative air ion mobility distribution.
#108 Z0-) Parameter of two-power model of negative air ion mobility distribution.
#109 dev-%) Parameter of two-power model of negative air ion mobility distribution.
#110 c&) Parameter of two-power model of bipolar air ion mobility distribution.
#111 g&) Parameter of two-power model of bipolar air ion mobility distribution.
#112 k&) Parameter of two-power model of bipolar air ion mobility distribution.
#113 Z0&) Parameter of two-power model of bipolar air ion mobility distribution.
#114 dev&% Parameter of two-power model of bipolar air ion mobility distribution.

Two-power model of atmospheric aerosol particle size distribution

The number of aerosol particles in clean atmospheric air is low. Increase in the size resolution of the instrument is accompanied with decrease in the number of particles in size fractions and corresponding increase in the relative instrumental errors. Thus the smoothing of distribution curve according to least squares principle is often recommended. The best smoothing functions are numerical models of the size distribution. In the range of large particles with d > 500 nm the negative power law proposed by Junge is most popular. In the range of fine particles the Junge law is usually replaced with the modified gamma law proposed by Deirmendjian. The low-diameter asymptote of the modified gamma is the positive power law. The two models can be joined into the two-power model proposed by Tammet (1988, 1992) and analyzed by Tammet & Kulmala (2014). The model has four parameters *a, b, p, d*$_0$ and is presented with equation

$$\frac{dN}{d(\lg d)} = \frac{a}{(d/d_0)^{-b} + (d/d_0)^{p}}, \; b + p > 0.$$



lllustration of effect of parameters *b* and *p* on the two-power distribution on a typical case of *a* = 2000 cm^{-3} and d_0 = 120 nm.

The parameters *a, b, p* ja *d*0 were estimated according to the algorithm published together with the paper by Tammet & Kulmala (2014). The deviations of the model from measurements were calculated in the logarithmic scale

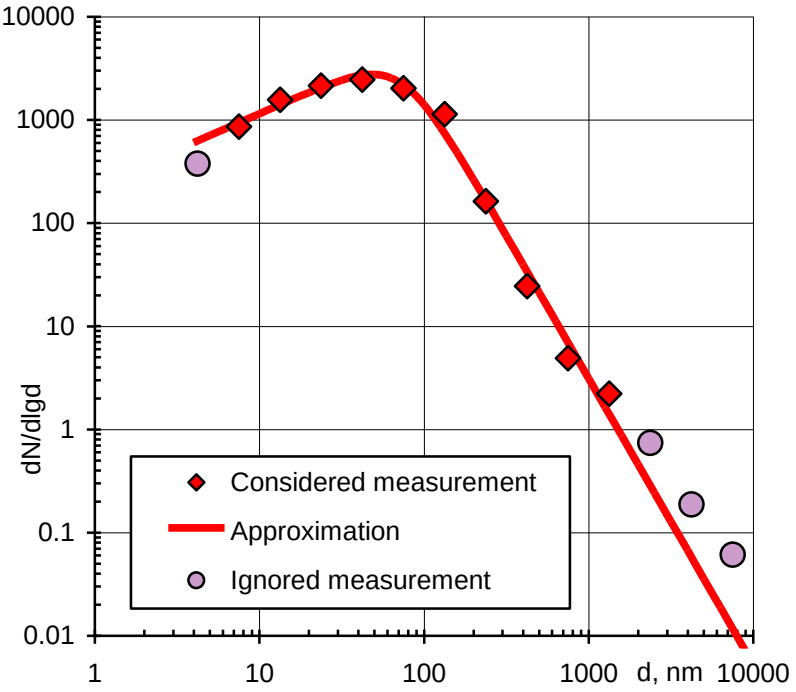
$$\Delta = \lg\left(\frac{dN}{d\lg d}\right)_{\text{measured}} - \lg\left(\frac{dN}{d\lg d}\right)_{\text{model}}$$

The algorithm determines the value of the parameter *a* so that the arithmetic average of deviations for all *m* fractions is zero. Other three parameters are estimated in this way that the root-square deviation

$$dev = \sqrt{\frac{\sum_{i=1}^{m} \Delta_i^2}{m}}$$

reaches the minimum. The size distribution is presented with 14 fractions in the size range of 4 – 7500 nm. The range of best approximation by the two-power model is a little narrower. Thus only 10 fractions with diameters 7 – 1400 nm were used when minimizing the criterion *dev*. The first fraction with diameter of 4 nm was disregarded because it is out of the application range of the model. The three coarse fractions are affected with the primary dust

and the difference between measurements and the model curve consists some information about the primary component of aerosol.



Example:
The approximated particle size distribution measured 09.05.2009, hour 2.

The measurements include instrumental noise and some recorded values of $dN/d\lg d$ may appear negative. Negative values cannot be approximated with the two-power law. Thus some records were disregarded and the table cells of the two-power law parameters were filled with the code of –99. The full record was disregarded when more than one of 10 fraction values appeared less than 0.01 cm^{-3}. If only one fraction was negative or too low then its value was replaced with 0.01 cm^{-3} and the approximation procedure was still performed.

Two-power model of atmospheric ion mobility distribution

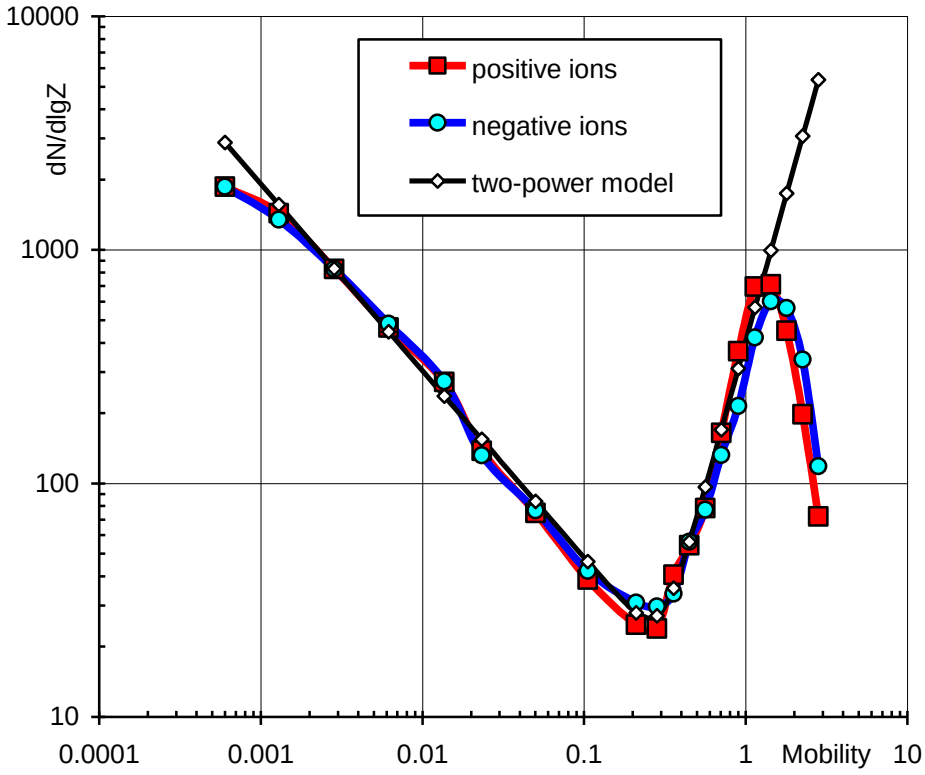
The calibration of Tahkuse air ion spectrometers during the years 2004 – 2014 contains some systematic errors in few channels (0.44, 0.35, 0.28 and 0.024 $cm^2V^{-1}s^{-1}$), discussed in the manuscript by Tammet (2016b). A simplest way to avoid these errors is to disregard the defective channels and use a model distribution as an interpolation compiled on basis of reliable channels. We recommend the two-power model for mobility, which differs from the two-power mobility of size distribution. The four-parameter model function

$$dN/d\lg Z \approx c\sqrt{(Z/Z_0)^{-g}+(Z/Z_0)^{k}}$$

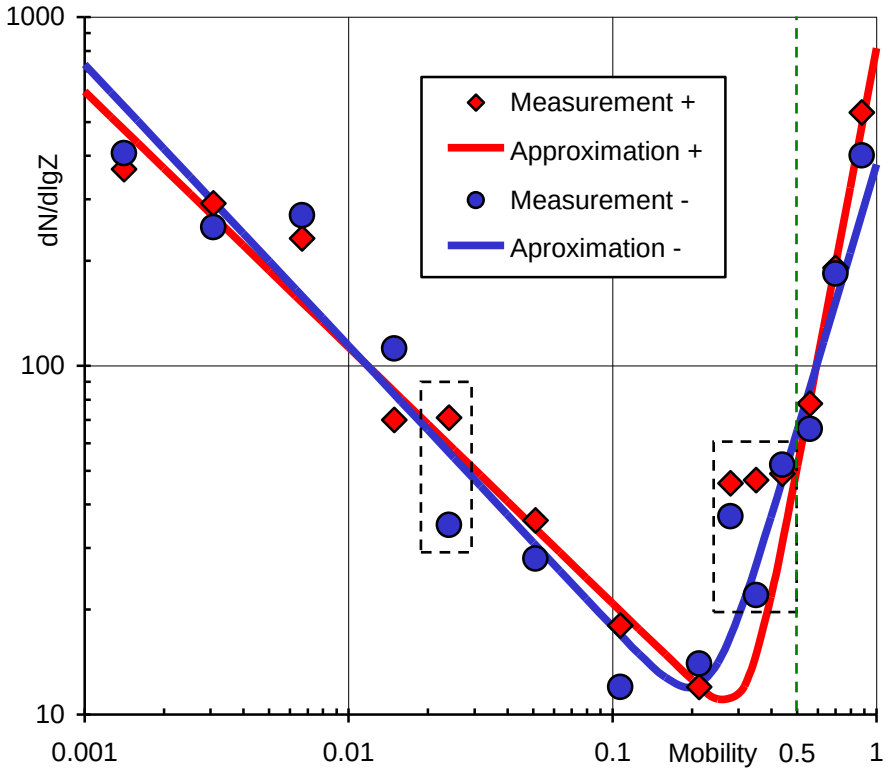
is a pretty good approximation of measurements in the range of 0.001 – 0.7 $cm^2V^{-1}s^{-1}$ as illustrated in Figure below. The figure is compiled for measurements carried out with reliably calibrated instrument in Tahkuse 1993–1994.

The parameters can be estimated minimizing the logarithmic square deviation between the measurements and the values of the model function for the 8 reliable channels. The incorrect fractions (0.44, 0.35, 0.28 and 0.024 $cm^2V^{-1}s^{-1}$), the fraction with the lowest mobility of 0.0007 $cm^2V^{-1}s^{-1}$, and fractions of small ions were not considered as arguments of the approximation. Thus only 9 fractions were really used. The nominal mobilities of these fractions are 0.00141, 0.00307, 0.00667, 0.0148, 0.0507, 0.107, 0.213, 0.556, and 0.696 $cm^2V^{-1}s^{-1}$. Some measurements of $dN/d\lg Z$ appear negative due to the instrumental noise. The bad records of mobility distribution were disregarded in the same way as in case of the size distribution. However, the critical level of 0.01 cm^{-3} was replaced here with 2 cm^{-3}.

Subsequent test showed that the root-mean-square of the difference between recorded measurements and approximation through the full dataset is about the same as the arithmetic average of the parameter *dev* delivered by the approximation procedure.



Atmospheric ion average mobility distribution at Tahkuse 1993–1994 and two-power approximation with parameters c = 20, g = 1.6, k = 5, Z_o = 0.3.



Example. The measurements at 09.05.2009, hour = 2, and the two-power approximation. The disregarded measurements are enclosed into the dashed rectangles in the figure.

References

Hõrrak, U. (2001) *Air ion mobility spectrum at a rural area*. Dissertationes Geophysicales Universitatis Tartuensis, vol. **15**, 157 pp., Tartu Univ. Press, Tartu, Estonia. https://dspace.ut.ee/handle/10062/50218.

Hõrrak, U., Miller, F., Mirme, A., Salm, J., Tammet, H. (1990) Air ion observatory at Tahkuse: Instrumentation. *Acta Comm. Univ. Tartu* **880**, 33–43. http://hdl.handle.net/10062/34419.

Hõrrak, U., Iher, H., Luts, A., Salm, J., Tammet, H. (1994) and Mobility spectrum of air ions at Tahkuse Observatory, *J. Geophys. Res.*, **99**, 10679–10700. http://onlinelibrary.wiley.com/doi/10.1029/93JD02291/abstract.

Hõrrak, U., Salm, J., Tammet, H. (1998) Bursts of intermediate ions in atmospheric air, J. *Geophys. Res. Atmospheres*,**103**, 13909-13915. http://onlinelibrary.wiley.com/doi/10.1029/97JD01570/pdf.

Hõrrak, U., Salm, J., Tammet, H. (2000). Statistical characterization of air ion mobility spectra at Tahkuse Observatory: Classification of air ions. *J. Geophys. Res. Atmospheres*, **105**, 9291–9302. http://onlinelibrary.wiley.com/doi/10.1029/1999JD901197/pdf.

Hõrrak, U., Salm, J., Tammet, H (2003) Diurnal variation in the concentration of air ions of different mobility classes at a rural area. *J. Geophys. Res. Atmospheres*, 108(D20), 4653, 11 pp, http://onlinelibrary.wiley.com/doi/10.1029/2002JD003240/pdf.

Komsaare, K. (2005) *Tahkuse Õhuseirejaama ioonispektromeetria kompleksi automatiseeritud andmehõivesüsteem* (Estonian) Magistritöö. Tartu Ülikool. 95 lk. http://ael.physic.ut.ee/tammet/library/Komsaare%202005%20Tahkuse%20ioonispektromeetria%20systeem.pdf.

Mirme, A. (1994) *Electric aerosol spectrometry*. Dissertationes Geophysicales Universitatis Tartuensis, vol. 6, 127 pp., Tartu Univ. Press, Tartu, Estonia. http://ael.physic.ut.ee/tammet/library/Mirme%20A%201994%20Thesis.pdf.

Tammet, H. (1988) Models of size spectrum of tropospheric aerosol. In *Atmospheric Aerosols and Nucleation. Lecture Notes in Physics*, Springer-Verlag, Vienna, **309**, pp. 75–78. http://www.springerlink.com/content/p7l948j8m356605k.

Tammet, H. (1990) Air Ion Observatory at Tahkuse: Software. *Acta Comm. Univ. Tartu* **880**, 44–51. http://hdl.handle.net/10062/34419.

Tammet, H. (1992) Comparison of model distributions of aerosol particle sizes. *Acta Comm. Univ. Tartu* **947**, 136–149. http://hdl.handle.net/10062/30272.

Tammet, H. (2006) Air ions. In *CRC Handbook of Chemistry and Physics, 74th edition*, CRC Press, Boca Raton, Ann Arbor, London, Tokyo, pp. Sect. 14, pp. 33–35. (The same article is published in subsequent editions since 1993). http://wenku.baidu.com/view/445e5a87ec3a87c24028c453.

Tammet, H. (2016a) Tahkuse koondandmete ehitus ja kasutamine (Estonian). Manuscript version 20161110, 15 pp.

Tammet, H. (2016b) Ülevaade Tahkuse vaatlusjaama 2002-2014 mõõtmistulemuste korrastamisest (Estonian). Manuscript version 20161107, 51 pp.

Tammet, H.F., Miller, F.G., Tamm, E.I., Bernotas, T.P., Mirme, A.A., Salm, J.J. (1987) Apparatura i metodika spektrometrii podvizhnosteï legkikh aéroionov (in Russian, transl.: Instrumentation and methods for mobility spectrometry of small air ions). *Acta Comm. Univ. Tartu* **755**, 18–28. http://hdl.handle.net/10062/34008.

Tammet, H., Salm, J., Iher, H. (1988) Observation of condensation on small air ions in the atmosphere. In *Atmospheric Aerosols and Nucleation. Lecture Notes in Physics*, Springer-Verlag, Vienna, **309**, pp. 239–240. http://www.springerlink.com/content/p436249m54004451.

Tammet, H., Mirme, A., Tamm, E. (2002) Electrical aerosol spectrometer of Tartu University. *Atmospheric Research*, **62**, 315–324. http://www.sciencedirect.com/science/article/pii/S0169809502000170

Tammet, H., Kulmala, M. (2014) Performance of four-parameter analytical models of atmospheric aerosol particle size distribution. *J. Aerosol Sci*., **77**, 145–157. http://dx.doi.org/10.1016/j.jaerosci.2014.08.001.

Tammet, H., Hõrrak, U. (2015) Dataset FinEstIon2003_06, http://dx.doi.org/10.15155/repo-4.

Basic statistics of full 11-year dataset

The last column shows the percentage of hours with missing data.

Variable	min	p01%	p10%	p50%	p90%	p99%	max	ave	std	?%
6) p:mb	952.8	982	997	1010.8	1024.1	1037	1052.1	1010.6	16.5	3.4
7) T:C	-31.1	-18.7	-6	6.2	18.8	26.3	31.6	6.3	9.8	1
8) RH%	-6.1	32.8	60.8	94	98.4	99.9	100.9	86.3	16.2	1.1
9) v:m/s	0	0	0.6	2.1	4.3	6.7	29.6	2.3	1.5	1.6
10) vmax:m/s	0	1.7	3.7	5.8	8.9	12.3	34.9	6.1	2.2	1.6
11) v:deg	0	0	58	184	281	328	360	175	83	1.5
12) Tbox:C	-9.9	2.5	8.3	13.8	21.8	26.9	29.9	14.5	5.3	7.1
13) Tattic:C	-31.8	-16.7	-3.7	8.9	20.3	25.9	30.7	8.5	9.6	7.1
14) Tchimney:C	-29.4	-11.1	-2.4	9.9	30	58	60	12.1	14	11.9
15) rad:W/m2	0	0	29	63	760	1285	1482	223	319	21.3
16) NO2:ug/m3	0.08	1.22	2.07	4.22	14.05	35.93	80.97	6.6	6.84	14.3
17) flow:lpm	0.05	0.05	0.07	0.11	0.2	0.25	0.72	0.12	0.05	13.2
18) Rn:Bq/m3	-0.7	0.5	1.8	5.3	17.5	38.9	91.9	8	7.8	37.4
19) Rn-s:Bq/m3	3	3	3.1	3.9	7.2	13.4	30.4	4.7	2.1	37.4
20) d@4.2	-252	0	1	224	732	4533	29326	413	1044	11.5
21) d@7.5	-802	0	51	412	1645	10222	42543	888	2120	11.5
22) d@13	-584	0	105	868	3660	16618	63179	1747	3277	11.5
23) d@24	-754	9	458	2562	7097	15610	71852	3379	3274	11.5
24) d@42	-469	105	1412	4755	10803	20165	88070	5655	4315	11.5
25) d@75	-985	56	1469	5124	11601	21783	89399	6090	4752	11.5
26) d@133	-528	16	621	2493	6312	12063	35034	3126	2635	11.5
27) d@237	-143	2	63	296	956	2078	4273	427	429	11.5
28) d@422	-76.3	-7.9	5.3	29.9	121.7	276.3	787.6	50	58.6	11.5
29) d@750	-1.6	0	1.9	7.4	24.6	60.2	272.5	11.2	12.5	11.5
30) d@1334	-0.35	0	0.96	3.29	9.04	19.16	75.89	4.37	3.99	11.5
31) d@2371	-0.29	0	0.212	0.797	2.005	4.512	20.015	1.012	0.926	11.5
32) d@4217	-0.16	0	0.056	0.223	0.585	1.685	10.873	0.304	0.362	11.5
33) d@7500	-0.061	0	0.016	0.074	0.201	0.604	3.924	0.104	0.132	11.5
34) d:defect	3	9	26	111	576	2157	4892	247	406	11.5
35) Z+2.77	-2896	-95	77	116	176	293	5093	120	90	8.9
36) Z+2.22	-2475	119	206	314	496	761	8017	337	140	8.9
37) Z+1.77	-2227	105	464	708	1097	1597	7884	750	278	8.9
38) Z+1.41	-3479	369	631	961	1454	2117	10489	1012	362	8.9
39) Z+1.13	-9999	252	512	883	1370	2162	85938	944	774	8.9
40) Z+0.88	-2183	103	235	456	787	1208	7077	490	233	8.9
41) Z+0.70	-3223	49	104	190	354	601	7412	214	127	8.9
42) Z+0.56	-2419	22	67	126	215	398	6524	138	94	8.9
43) Z+0.44	-6670	-40	56	102	177	377	8387	111	122	8.9
44) Z+0.35	-3104	-10	63	118	208	715	99999	170	1214	8.9
45) Z+0.28	-4288	1	64	114	216	571	12302	136	149	8.9
46) Z+0.213	-1710	0	16	30	83	308	31373	49	201	8.9
47) Z+0.107	-9999	-6	20	41	124	577	64329	122	1256	8.9
48) Z+0.051	-2580	-29	36	74	220	850	13114	117	211	8.9
49) Z+0.024	-5138	-112	64	133	396	1619	27587	216	438	8.9
50) Z+0.015	-8629	-82	83	193	402	1136	99999	242	1008	8.9
51) Z+0.0068	-9999	-279	153	372	746	2035	99999	442	902	8.9
52) Z+0.0031	-9999	-564	320	772	1646	6450	99999	1119	3376	8.9
53) Z+0.0014	-9999	-3251	449	1350	2983	10537	99999	1759	3744	8.9
54) Z+0.0007	-9999	-1408	106	1877	5181	36322	99999	3140	7475	8.9
55) sum+0.3	5	154	247	381	574	847	6023	401	152	9.4
56) Z-2.77	-3313	-108	43	103	186	966	7941	120	170	8.9
57) Z-2.22	-2235	170	284	434	702	1396	5536	478	229	8.9
58) Z-1.77	-2277	51	488	762	1212	1898	7051	815	331	8.9

Variable	min	p01%	p10%	p50%	p90%	p99%	max	ave	std	?%
59) Z-1.41	-2574	229	438	748	1225	1773	14959	798	335	8.9
60) Z-1.13	-9999	80	224	504	957	1518	75251	572	676	8.9
61) Z-0.88	-3513	-2	60	211	513	938	14033	258	213	8.9
62) Z-0.70	-2008	-30	12	82	234	533	5300	109	122	8.9
63) Z-0.56	-2033	-71	-20	40	130	340	6815	52	100	8.9
64) Z-0.44	-3672	-139	-12	37	113	345	6650	45	117	8.9
65) Z-0.35	-3807	-115	-16	40	133	667	99999	88	1151	8.9
66) Z-0.28	-3226	-75	-4	46	145	527	9260	67	147	8.9
67) Z-0.213	-2155	-2	13	24	90	400	29293	48	198	8.9
68) Z-0.107	-9999	-19	13	31	124	606	65881	114	1236	8.9
69) Z-0.051	-3969	-54	21	54	205	836	7023	98	188	8.9
70) Z-0.024	-6160	-251	16	81	333	1604	14702	159	399	8.9
71) Z-0.015	-8827	-87	77	186	380	1085	33983	223	309	8.9
72) Z-0.0068	-9999	-271	156	372	738	1990	57909	442	794	8.9
73) Z-0.0031	-9999	-705	252	704	1566	5917	99999	1053	3612	8.9
74) Z-0.0014	-9999	-3284	367	1299	2950	11514	99999	1729	3817	8.9
75) Z-0.0007	-9999	-2147	8	1781	4880	20074	99999	2545	4500	8.9
76) sum-0.3	5	132	212	336	525	769	7295	358	149	9.4
77) Z1noise	1	3	4	7	16	69	3169	11	31	8.9
78) Z2noise	0	0	1	5	44	286	2746	22	69	8.9
79) n+	10	169	250	372	561	833	8624	395	148	9.5
80) Z+	0.786	1.23	1.303	1.397	1.487	1.562	2.054	1.395	0.072	9.5
81) fS/m+	0.22	3.93	5.68	8.26	12.42	18.21	156.94	8.78	3.16	9.5
82) n-	11	128	187	301	495	767	7535	327	146	16.7
83) Z-	0.859	1.289	1.407	1.552	1.687	1.792	2.436	1.549	0.111	16.7
84) fS/m-	0.25	3.33	4.78	7.4	11.87	20.19	137.52	8.05	3.48	16.7
85) n&	27	289	430	665	1043	1548	16160	711	285	9.8
86) Z&	0.866	1.275	1.357	1.47	1.574	1.655	2.204	1.467	0.085	9.8
87) fS/m&	0.64	7.07	10.32	15.49	24.13	35.92	294.47	16.6	6.32	9.8
88) q-eas	1	11	101	501	900	990	1000	500	289	11.5
89) q-ion1	1	11	101	501	900	990	1000	501	289	8.9
90) q-ion2	1	11	101	501	900	990	1000	501	289	8.9
91) q-eas-st	1	11	101	501	900	990	1000	501	289	12.6
92) q-ion-st	1	11	101	500	900	990	1000	500	289	10.5
93) q-ion-sym	1	11	101	500	900	990	1000	500	289	10.5
94) q-low	1	3	25	147	432	704	982	193	165	20.2
95) a	0	862	4285	13687	30859	60580	191438	16312	12427	14.9
96) b	-2.605	-0.245	0.728	1.611	2.574	7.854	13.786	1.773	1.266	14.9
97) p	0.781	2.013	2.573	2.942	3.259	3.859	8.59	2.932	0.332	14.9
98) d0	10	23.9	43	67.3	95.9	134.4	9998.6	71.2	120.1	14.9
99) dev%	1.9	6.4	10.3	16.6	26.1	44.4	131	17.9	7.8	14.9
100) c+	0.56	5.58	11.34	21.45	63.25	250.79	3144.79	35.04	55.81	15
101) g+	-0.813	0.607	1.114	1.674	2.12	4.609	10	1.72	0.791	15
102) k+	0.1	0.1	2.023	3.908	6.349	11.436	21	4.145	2.302	15
103) Z0+	0.001	0.006	0.142	0.223	0.364	0.637	0.696	0.24	0.106	15
104) dev+%	0.4	1.4	2.6	5.9	15.3	45.6	112.9	8.2	8.2	15
105) c-	0.53	1.97	3.76	12.34	49.59	218.15	2174.93	24.6	47.34	22.8
106) g-	-0.419	0.623	1.189	1.831	2.317	3.777	10	1.848	0.761	22.8
107) k-	0.1	0.1	1.147	5.02	21	21	21	8.442	7.579	22.8
108) Z0-	0.001	0.007	0.119	0.338	0.61	0.696	0.696	0.35	0.196	22.8
109) dev-%	0.4	1.9	4	9.1	27.7	52.5	126.2	12.9	10.8	22.8
110) c&	0.4	11.06	20.17	36.44	120.99	455.97	3419.4	62.71	94.41	14
111) g&	-0.278	0.643	1.143	1.714	2.166	4.586	10	1.755	0.782	14
112) k&	0.1	0.1	1.275	3.503	7.163	21	21	4.103	3.022	14
113) Z0&	0.001	0.007	0.117	0.224	0.403	0.696	0.696	0.253	0.126	14
114) dev&%	0.2	1.4	2.7	5.9	14.7	48.2	119.5	8.1	8.4	14

Making excerpts from the dataset

The data can be processed by means of different universal programs and packages including the MS Excel. The universal software is not adapted to the specific dataset and may be inconvenient for reorganization of data covering several years and for making excerpts including limited subset of variables sorted in a modified order. Thus a specific service program "Textractor11.exe" is added to the dataset. The program offers a convenient way to extract data into a single file from any start date to any end date between 01.01.2004 and 31.12.2014 while picking a specific subset of variables in an arbitrary order.
The rules of application are:

- The program file "Textractor11.exe" must be located in the same folder as all 11 data files "Tahkuse_2004_data.xls" to "Tahkuse_2014_data.xls".
- If you have formulated the specific task in the file "Textractor11.ini" then the ini-file must be located in the same folder. Otherwise the folder must not contain any file with the name "Textractor11.ini".
- Double-click the icon of "Textractor11.exe".
- If the folder contains the task file "Textractor11.ini" then the program will create and fill the excerpt file as subscribed with the task file and the job is finished.
- If the folder didn't contain the task file "Textractor11.ini" then the program will create the standard sample file "Textractor11.ini", which must be edited according the specific task and saved. Next the user must double-click again the icon of "Textractor11.exe" and now the job will be automatically accomplished.

The standard task file "Textractor11.ini" contains the sample values of controls as well as the explanations of the controls. The text of the standard ini-file is:

```
A pattern of Textractor11.ini:

The extraction of a table is controlled with phrases in the 6 control lines.
Everything else in the control file has no effect on the extraction process.
NB: Edit and save the control file before running the Textractor11!

================================ CONTROL LINES ================================
First day to extract YYYYMMDD: 20090301
 Last day to extract YYYYMMDD: 20101031
   List of columns to extract: 1-4, 33-20, 34
Delimiter in the output table: (     )
  Missing entry in the output: (?)
      Name of the output file: Tahkuse test.xl
================================ EXPLANATIONS ================================

The variables to be extracted should be indicated with numbers of columns in
the data files. The numbers should be delimited with commas or hyphens. A pair
of hyphen-limited numbers denotes a range from the first to the second number
in the similar way as in the list of pages when printing a document in MS Word.
An extra option is possibility to define an inverted range like 34-20 in the
example above. Here the variables will be extracted in order 34, 33, 32, .. 20.

The delimiters and missing entry codes can include invisible symbols and must
be written between parentheses. E.g. the output delimiter above is the invisible
symbol of tabulator. The parentheses are here used only as a frame and do not
belong to the code. A delimiter should be presented with one symbol while
a missing entry code may consist of several symbols, one symbol or nothing.
In the last case a delimiter will immediately follow the preceding delimiter.

NB: the cells of a table are interpreted and compared as the text.
Thus the cells containing 999 and 999.0 are considered not equal.

NB: the files Textractor11.exe, Textractor11.ini and all 11 Tahkuse data files
(Tahkuse_2004.xls ... Tahkuse_2014.xls) must be located in the common folder.
```

The Pascal-code of the program is presented in the technical supplement in favor of possible future development.

Compact table of variables

#) Variable	min	p50%	max
1) year	2004	2009	2014
2) month	1	7	12
3) day	1	16	31
4) hour	0	12	23
5) weekday	1	4	7
6) p:mb	953	1013	1052
7) T:C	-31.1	6.2	31.6
8) RH%	-6.1	94	100.9
9) v:m/s	0	2.1	29.6
10) vmax:m/s	0	5.8	34.9
11) v:deg	0	184	360
12) Tbox:C	-9.9	13.8	29.9
13) Tattic:C	-31.8	8.9	30.7
14) Tchimney:C	-29.4	9.9	60
15) rad:W/m2	0	63	1482
16) NO2:ug/m3	0.08	4.22	80.97
17) flow:lpm	0.05	0.11	0.72
18) Rn:Bq/m3	-0.7	5.3	91.9
19) Rn-s:Bq/m3	3	3.9	30.4
20) d@4.2	-252	224	29326
21) d@7.5	-802	412	42543
22) d@13	-584	868	63179
23) d@24	-754	2562	71852
24) d@42	-469	4755	88070
25) d@75	-985	5124	89399
26) d@133	-528	2493	35034
27) d@237	-143	296	4273
28) d@422	-76.3	29.9	787.6
29) d@750	-1.6	7.4	272.5
30) d@1334	-0.35	3.29	75.89
31) d@2371	-0.29	0.797	20.015
32) d@4217	-0.16	0.223	10.873
33) d@7500	-0.061	0.074	3.924
34) d:defect	3	111	4892
35) Z+2.77	-2896	116	5093
36) Z+2.22	-2475	314	8017
37) Z+1.77	-2227	708	7884
38) Z+1.41	-3479	961	10489
39) Z+1.13	-9999	883	85938
40) Z+0.88	-2183	456	7077
41) Z+0.70	-3223	190	7412
42) Z+0.56	-2419	126	6524
43) Z+0.44	-6670	102	8387
44) Z+0.35	-3104	118	99999
45) Z+0.28	-4288	114	12302
46) Z+0.213	-1710	30	31373
47) Z+0.107	-9999	41	64329
48) Z+0.051	-2580	74	13114
49) Z+0.024	-5138	133	27587
50) Z+0.015	-8629	193	99999
51) Z+0.0068	-9999	372	99999
52) Z+0.0031	-9999	772	99999
53) Z+0.0014	-9999	1350	99999
54) Z+0.0007	-9999	1877	99999
55) sum+0.3	5	381	6023

#) Variable	min	50%	max
56) Z-2.77	-3313	103	7941
57) Z-2.22	-2235	434	5536
58) Z-1.77	-2277	762	7051
59) Z-1.41	-2574	748	14959
60) Z-1.13	-9999	504	75251
61) Z-0.88	-3513	211	14033
62) Z-0.70	-2008	82	5300
63) Z-0.56	-2033	40	6815
64) Z-0.44	-3672	37	6650
65) Z-0.35	-3807	40	99999
66) Z-0.28	-3226	46	9260
67) Z-0.213	-2155	24	29293
68) Z-0.107	-9999	31	65881
69) Z-0.051	-3969	54	7023
70) Z-0.024	-6160	81	14702
71) Z-0.015	-8827	186	33983
72) Z-0.0068	-9999	372	57909
73) Z-0.0031	-9999	704	99999
74) Z-0.0014	-9999	1299	99999
75) Z-0.0007	-9999	1781	99999
76) sum-0.3	5	336	7295
77) Z1noise	1	7	3169
78) Z2noise	0	5	2746
79) n+	10	372	8624
80) Z+	0.786	1.397	2.054
81) fS/m+	0.22	8.26	156.94
82) n-	11	301	7535
83) Z-	0.859	1.552	2.436
84) fS/m-	0.25	7.4	137.52
85) n&	27	665	16160
86) Z&	0.866	1.47	2.204
87) fS/m&	0.64	15.49	294.47
88) q-eas	1	501	1000
89) q-ion1	1	501	1000
90) q-ion2	1	501	1000
91) q-eas-st	1	501	1000
92) q-ion-st	1	500	1000
93) q-ion-sym	1	500	1000
94) q-low	1	147	982
95) a	0	13687	191438
96) b	-2.605	1.611	13.786
97) p	0.781	2.942	8.59
98) d0	10	67.3	9998.6
99) dev%	1.9	16.6	131
100) c+	0.56	21.45	3144.79
101) g+	-0.813	1.674	10
102) k+	0.1	3.908	21
103) Z0+	0.001	0.223	0.696
104) dev+%	0.4	5.9	112.9
105) c-	0.53	12.34	2174.93
106) g-	-0.419	1.831	10
107) k-	0.1	5.02	21
108) Z0-	0.001	0.338	0.696
109) dev-%	0.4	9.1	126.2
110) c&	0.4	36.44	3419.4
111) g&	-0.278	1.714	10
112) k&	0.1	3.503	21
113) Z0&	0.001	0.224	0.696
114) dev&%	0.2	5.9	119.5