Dataset Hyytiala08_10aerosol

Hannes Tammet, University of Tartu,
Markku Kulmala, University of Helsinki

Annotation

The dataset **Hyytiala08_10aerosol** contains results of routine measurements of atmospheric aerosols carried on in a well equipped boreal research station during 3 years. The particle size range from 3 nm to 15μm is split into 60 fractions and the records of distribution function are presented for 21682 hours of measurements. The dataset provides scientists with a tool for exploring the structure and dynamics of atmospheric aerosol size distribution. Additionally, it can serve as a basis for data analysis exercises for students in field of environmental sciences.

The dataset includes three files:

Data_Hyytiala08_10aerosol.xls – a table, which contains 60 columns of values of the particle size distribution function and 30 columns of complementary variables.

Description_Hyytiala08_10aerosol.pdf – detailed description of origin and structure the data. Additionally includes sample diagrams, which illustrate the data and may provoke new ideas for studies on atmospheric aerosol.

Package_Hyytiala08_10aerosol.zip – a compressed package, which contains both the data file and the description file. The package is to be downloaded to a personal computer, unzipped and used offline.

Origin of the data

The core of the dataset is a table of hourly mean size distributions of aerosol particles at Hyytiälä researh station, Finland. The table was compiled in process of studies on coagulation sink of fine nanoparticles by Tammet and Kulmala (2014a) and proposed as a reference for examination of aerosol size distribution models. An example of usage of the dataset is analysis and comparison of parametric models of atmospheric aerosol size distribution (Tammet and Kulmala, 2014b). The dataset includes results of routine measurements of atmospheric aerosols during three years (2008–2010) at the SMEAR II station (Station for Measuring Forest Ecosystem-Atmosphere Relations) located in Hyytiälä, Finland (61°31' N, 24°17' E, 181 m a.s.l., 220 km NW from Helsinki), and run by the University of Helsinki (Hari and Kulmala, 2005). The station provided data for the networks of European Supersites for Atmospheric Aerosol Research (EUSAAR, see http://www.eusaar.net/upload/SMEAR2.pdf) and EMEP/GAW joint super sites (Asmi et al., 2011). The air quality in Hyytiälä represents typical regional background conditions in southern boreal zone of Europe. The instruments are located in a cottage within a Scots pine stand. Conditions of measurements were the same as described by Virkkula et al. (2011). Distribution of particles according to the diameter was measured using the Differential Mobility Particle Sizers (DMPS) with Hauke-type DMAs and the Aerodynamic Particle Sizer (APS) TSI model 3321. The DMPS data of the distribution function $dN/d(\log(d_p))$ is given at 38 geometric diameters in the range of 3–980 nm and the APS data at 52 aerodynamic diameters in the range of 520 nm –19.8 μm. During data preparation, the two datasets were merged by converting the aerodynamic diameters to geometric diameters by the factor of 0.82, following the argumentation by Virkkula et al. (2011). Present dataset contains only a minor part of scientific data acquired at Hyytiälä. Information about the research on atmospheric aerosols in the University of Helsinki can be found in web, search Google including the keywords "aerosol" and "Hyytiälä".

Size distribution of atmospheric aerosol particles

Distribution of aerosol particles according to their size d_p is described with fraction concentrations n (d_{pa}, d_{pb}) or distribution function:

$$n(d_p) = \frac{dN}{d(\lg d_p)} \approx \frac{n(d_{pa}, d_{pb})}{\lg \frac{d_{pb}}{d_{pa}}}$$

The values of distribution function are presented in cm^{-3}.
Part of the water absorbed in particles is lost in the measuring process and the size d_p shown in the dataset is characteristic of slightly shrunken particles as they were detected in the instruments. In different applications the hygroscopic growth of particles is to be considered. Tammet and Kulmala (2014a) proposed an approximation

$$d_{humid} = (1 + ((1.5 - 0.3 \exp(-d_p / 100 \text{ nm})) - 1) / 2)\, d_p ,$$

which is based on research by Fors et al. (2011).

Mean distributions of number concentration and power-weighted concentrations for the period of 3 years are shown in Fig. 1 while Fig. 2 illustrates the relative variation of the number concentration in the full set of measured distributions.

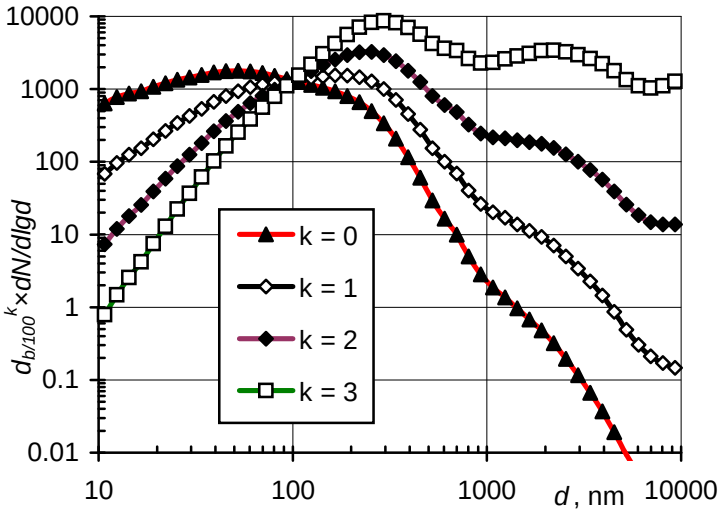


Fig. 1. Mean size distributions $d_{b/100}^{k} \times dN/dlgd$ in the reference data where $d_{b/100} = d$ / 100 nm. Markers are depicted according to 1/16-decade fractions and roughly estimated natural humid diameter d_{humid} corresponds to the shift of curves by about one marker.

The relative standard distribution characterizes both the natural variation of the distribution function and the instrumental noise. In the size range of 32...3200 nm the natural variation dominates and a noticeable instrumental variation in Fig. 2 may call in question only close to the joint of the measuring ranges of DMPS and APS around the diameter of 700–800 nm. The fraction concentrations of very small and very large particles are low and the Poisson noise of counting the particles involves increase in the instrumental noise. The number of detected particles in size fractions below 5 nm often tends to fluctuate between zero and one. The low-noise range of particle sizes is limited with diameters of 10...5000 nm. Analysis of the size distribution outside of this range requires special attention on the effect of random noise in the data.

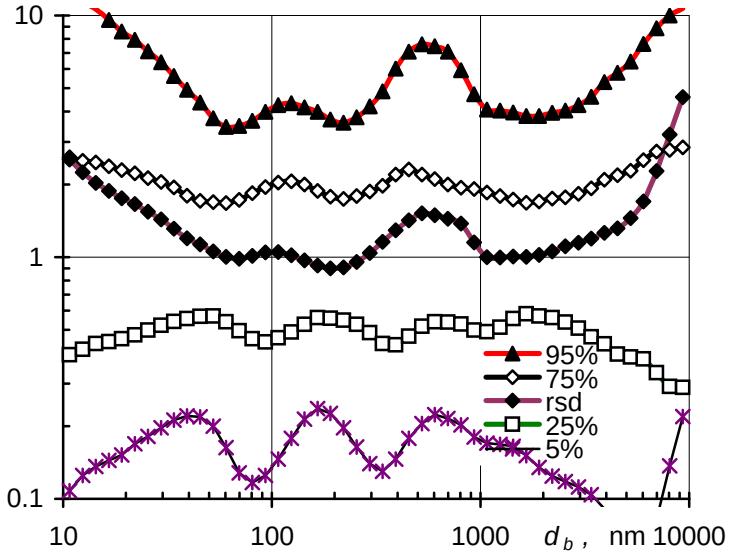


Fig. 2. Relative percentiles and relative standard deviation of hourly size distribution in the reference data. Relative percentile is the ratio of the specified percentile and median.

Preparation of the reference data

Data was originally recorded with the time interval of 10 min and presented with 1107 diurnal files for DMPS measurements and 1051 diurnal files for APS measurements. About 90% of available time is covered with both DMPS and APS measurements. The reference dataset was compiled in the following way: First, the hourly averages were calculated, including only the hours that had at least 3 datapoints. Next, the distributions of particles according to the measured diameter were converted into the logarithmically uniform grid with 16 size intervals in a diameter decade, where the geometric centers of the intervals from 2.94 nm to 14.3 μm are $d_{pi} = 10^{(i + 6.5) / 16}$, i = 1...60. The diameter grid was chosen to match the original grid of most of the instrument recordings. The simultaneous DMPS and APS data were merged so that the overlapping size range of 500–1000 nm was filled with the weighted averages

$$n\,(d_p) = ((1000\text{ nm} - d_p)\,n_{\text{DMPS}}\,(d_p) + (d_p - 500\text{ nm})\,n_{\text{APS}}\,(d_p))\,/\,500\text{ nm}.$$

Here $n_{\text{DMPS}}\,(d_p)$ and $n_{\text{APS}}\,(d_p)$ are the distribution functions from the DMPS and APS instruments. Finally, the data was analyzed with the aim of detecting possible defects. During this step about 7% of records were eliminated because recognizable distortions were detected, mostly lack of data in some size segment. The final dataset includes 21682 hours of data, which covers 82.5% of possible hours during the considered period.

Structure of the table Data_Hyytiala2008_10aerosol.xls

The size distribution is presented in the table with 60 values of the distribution function. The diameter grid is log-uniform and $\lg\,(d_{pi+1} / d_{pi}) = 1/16$. Thus a fraction concentrations can be estimated as $n\,(d_{pa}, d_{pb}) = (dn / d\lg d) / 16$. The fraction borders and geometric centers are:

```
    lower border := exp (((i + 6.0) / 16)* ln (10)) nm;
fraction center := exp (((i + 6.5) / 16)* ln (10)) nm;
   upper border := exp (((i + 7.0) / 16)* ln (10)) nm;
```

where the fraction numbers are i = 1...60. The size distribution is presented in 60 columns of the table with column numbers k = 11...70. Thus $i = k - 10$.

The table file contains plain ASCII text where the columns are tab-delimited. First line is the header and includes identifiers of the variables. The beginning of the header is:

```
# Year Month Day Hour DOY Week DOW Season. . . . .
```

Following 21682 lines contain 90 tab-delimited numbers in every line, which make of 90 columns of the table. Measurements are recorded according to local zone time UTC+02 independent of the season. The variable columns are explained in Table:

Columns of table *Data_hyytiala2008_10hours.xls*

No	Excel	Header	Variable
1	A	#	Number of data line 1...21682 (see comments)
2	B	Year	Year 2008...2010
3	C	Month	Month of year 1...12
4	D	Day	Day of month 1...31
5	E	Hour	Hour of day 0...23 (see comments)
6	F	DOY	Day of year 1.021...366.979
7	G	Week	Week of year 0...53 (see comments)
8	H	DOW	Day of week 1...7 (1 = Mon, 7 = Sun)
9	I	Season	Season of year 1...4 (see comments)
10	J	DayQrt	Quarter of day 1...4 (see comments)
11	K	d2.9	dn / d lg d (2.7 ... 3.2 nm), cm^{-3}
12	L	d3.4	dn / d lg d (3.2 ... 3.7 nm), cm^{-3}
13	M	d3.9	dn / d lg d (3.7 ... 4.2 nm), cm^{-3}
14	N	d4.5	dn / d lg d (4.2 ... 4.9 nm), cm^{-3}
15	O	d5.2	dn / d lg d (4.9 ... 5.6 nm), cm^{-3}
16	P	d6.0	dn / d lg d (5.6 ... 6.5 nm), cm^{-3}
17	Q	d7.0	dn / d lg d (6.5 ... 7.5 nm), cm^{-3}
18	R	d8.1	dn / d lg d (7.5 ... 8.7 nm), cm^{-3}
19	S	d9.3	dn / d lg d (8.7 ... 10.0 nm), cm^{-3}
20	T	d10.7	dn / d lg d (10.0 ... 11.5 nm), cm^{-3}
21	U	d12.4	dn / d lg d (11.5 ... 13.3 nm), cm^{-3}
22	V	d14.3	dn / d lg d (13.3 ... 15.4 nm), cm^{-3}
23	W	d16.5	dn / d lg d (15.4 ... 17.8 nm), cm^{-3}
24	X	d19.1	dn / d lg d (17.8 ... 20.5 nm), cm^{-3}
25	Y	d22.1	dn / d lg d (20.5 ... 23.7 nm), cm^{-3}
26	Z	d25.5	dn / d lg d (23.7 ... 27.4 nm), cm^{-3}
27	AA	d29.4	dn / d lg d (27.4 ... 31.6 nm), cm^{-3}
28	AB	d34.0	dn / d lg d (31.6 ... 36.5 nm), cm^{-3}
29	AC	d39.2	dn / d lg d (36.5 ... 42.2 nm), cm^{-3}
30	AD	d45.3	dn / d lg d (42.2 ... 48.7 nm), cm^{-3}
31	AE	d52.3	dn / d lg d (48.7 ... 56.2 nm), cm^{-3}
32	AF	d60.4	dn / d lg d (56.2 ... 64.9 nm), cm^{-3}
33	AG	d69.8	dn / d lg d (64.9 ... 75.0 nm), cm^{-3}
34	AH	d80.6	dn / d lg d (75.0 ... 86.6 nm), cm^{-3}
35	AI	d93.1	dn / d lg d (86.6 ... 100 nm), cm^{-3}
36	AJ	d107	dn / d lg d (100 ... 116 nm), cm^{-3}
37	AK	d124	dn / d lg d (116 ... 133 nm), cm^{-3}
38	AL	d143	dn / d lg d (133 ... 154 nm), cm^{-3}
39	AM	d166	dn / d lg d (154 ... 179 nm), cm^{-3}
40	AN	d191	dn / d lg d (178 ... 205 nm), cm^{-3}
41	AO	d221	dn / d lg d (205 ... 237 nm), cm^{-3}

No	Excel	Header	Variable
42	AP	d255	d*n* / d lg *d* (237 ... 274 nm), cm^{-3}
43	AQ	d294	d*n* / d lg *d* (274 ... 316 nm), cm^{-3}
44	AR	d340	d*n* / d lg *d* (316 ... 365 nm), cm^{-3}
45	AS	d392	d*n* / d lg *d* (365 ... 422 nm), cm^{-3}
46	AT	d453	d*n* / d lg *d* (422 ... 487 nm), cm^{-3}
47	AU	d523	d*n* / d lg *d* (487 ... 562 nm), cm^{-3}
48	AV	d604	d*n* / d lg *d* (562 ... 649 nm), cm^{-3}
49	AW	d698	d*n* / d lg *d* (649 ... 750 nm), cm^{-3}
50	AX	d806	d*n* / d lg *d* (750 ... 866 nm), cm^{-3}
51	AY	d931	d*n* / d lg *d* (866 ... 1000 nm), cm^{-3}
52	AZ	d1075	d*n* / d lg *d* (1000 ... 1155 nm), cm^{-3}
53	BA	d1241	d*n* / d lg *d* (1155 ... 1334 nm), cm^{-3}
54	BB	d1433	d*n* / d lg *d* (1334 ... 1540 nm), cm^{-3}
55	BC	d1655	d*n* / d lg *d* (1540 ... 1778 nm), cm^{-3}
56	BD	d1911	d*n* / d lg *d* (1778 ... 2054 nm), cm^{-3}
57	BE	d2207	d*n* / d lg *d* (2054 ... 2371 nm), cm^{-3}
58	BF	d2548	d*n* / d lg *d* (2371 ... 2738 nm), cm^{-3}
59	BG	d2943	d*n* / d lg *d* (2738 ... 3162 nm), cm^{-3}
60	BH	d3398	d*n* / d lg *d* (3162 ... 3652 nm), cm^{-3}
61	BI	d3924	d*n* / d lg *d* (3652 ... 4217 nm), cm^{-3}
62	BJ	d4532	d*n* / d lg *d* (4217 ... 4870 nm), cm^{-3}
63	BK	d5233	d*n* / d lg *d* (4870 ... 5623 nm), cm^{-3}
64	BL	d6043	d*n* / d lg *d* (5623 ... 6494 nm), cm^{-3}
65	BM	d6978	d*n* / d lg *d* (6494 ... 7499 nm), cm^{-3}
66	BN	d8058	d*n* / d lg *d* (7499 ... 8660 nm), cm^{-3}
67	BO	d9306	d*n* / d lg *d* (8660 ... 10000 nm), cm^{-3}
68	BP	d10746	d*n* / d lg *d* (10000 ... 11548 nm), cm^{-3}
69	BQ	d12409	d*n* / d lg *d* (11548 ... 13335 nm), cm^{-3}
70	BR	d14330	d*n* / d lg *d* (13335 ... 15399 nm), cm^{-3}
71	BS	N	Total number concentration, cm^{-3}
72	BT	a	Two-power law *a* (*b* in KL terms), cm¯3
73	BU	b	Two-power law *b (L* in KL terms*)*
74	BV	p	Two-power law *p (K* in KL terms*)*
75	BW	d0	Two-power law *d0 (dx* in KL terms*)*, nm
76	BX	dev	Deviation from the two-power law
77	BY	TZ2.0	Coagulation time for ions $Z = 2.0$ $cm^2V^{-1}s^{-1}$, min
78	BZ	TZ1.5	Coagulation time for ions $Z = 1.5$ $cm^2V^{-1}s^{-1}$, min
79	CA	TZ1.0	Coagulation time for ions $Z = 1.0$ $cm^2V^{-1}s^{-1}$, min
80	CB	Td1	Coagulation time for nanoparticles *d* = 1 nm, min
81	CC	Td2	Coagulation time for nanoparticles *d* = 2 nm, min
82	CD	Td3	Coagulation time for nanoparticles *d* = 3 nm, min
83	CE	Td4	Coagulation time for nanoparticles *d* = 4 nm, min
84	CF	Td5	Coagulation time for nanoparticles *d* = 5 nm, min
85	CG	Td6	Coagulation time for nanoparticles *d* = 6 nm, min
86	CH	Td7	Coagulation time for nanoparticles *d* = 7 nm, min
87	CI	Td8	Coagulation time for nanoparticles *d* = 8 nm, min
88	CJ	Td9	Coagulation time for nanoparticles *d* = 9 nm, min
89	CK	Td10	Coagulation time for nanoparticles *d* = 10 nm, min
90	CL	NP%	New particle index, % (see comments)

Comments

Measurements are recorded according to local zone time UTC+02 independent of the season.

***Column* 1**: MS Excel allows easily sort the data according to a specific column, which is often used for estimating the percentiles of the speficied parameter. Column 1 helps the user to restore the initial order of the rows resorting the table according to column 1. Additionally, the numbers of data rows may be used to give a quick reference to specific time.

***Column* 5**: The data are mean values for full hours. The hours in day are numbered 0...23. The numbers can be transformed to the alternative format 1...24 adding 1 to the value of the hour.

***Column* 7**: The week numbers are included for convenience of graphical representation of annual variations. If the first week of year contains less than 4 days, then it is traditionally denoted as week 53 of the preceding year. This is unfavourable when compiling diagrams. Thus the number of an incomplete short week (less than 4 days) in beginning of January is counted as week = 0 and an incomplete short week in end of December as week = 53.

***Column* 9**: Season are: 1 = (Dec+Jan+Feb), 2 = (Mar+Apr+May), 3 = (Jun+Jul+Aug), and 4 = (Sep+Oct+Nov).

***Column* 10**: Quarters of day are composed of hours numbered as 0...23 in following way: 1 = (22, 23, 0..3), 2= (4...9), 3= (10...15), 4= (16...21),

***Columns* 11...70**: The columns of distribution function do not contain negative or zero values. A zero found in original records is replaced with 1 (if d < 100 nm) or 0.0001 (if d > 100 nm). $dn / d\lg d$ is presented with 0...6 decimals depending on the value.

***Column* 71**: N = round ((sum of columns 11...70) / 16)

***Columns* 72...76**: The two-power approximation

$$n_{\lg}(d_p) = \frac{a}{\left(d_p/d_0\right)^{-b} + \left(d_p/d_0\right)^{p}}$$

is explained in a paper by Tammet and Kulmala (2014b). The values of parameters are estimated fitting the measurements in size range of 10–3200 nm, smaller and bigger particles were ignored.

***Columns* 77...89**: Coagulation loss of air ions and finest nanoparticles on the background aerosol is characterized with the coagulation sink S_{coag} or characterictic coagulation time $T_{coag} = 1 / S_{coag}$. The sink is measured in s^{-1} and defined as the relative intensity of loss in an imaginary situation when there is no supply of nanoparticles

$$S_{coag} = -\frac{1}{N}\frac{dN}{dt}.$$

Here N is the concentration of considered ions or nanoparticles. Values of coagulation sink for small ions and neutral nanoparticles were calculated according to Tammet and Kulmala (2014b) assuming air temperature 4ºC and pressure 1010 mb. The values of sink were converted to the values of characteristic coagulation time and expressed in minutes in the table. The coagulation sink in units s^{-1} can be calculated as

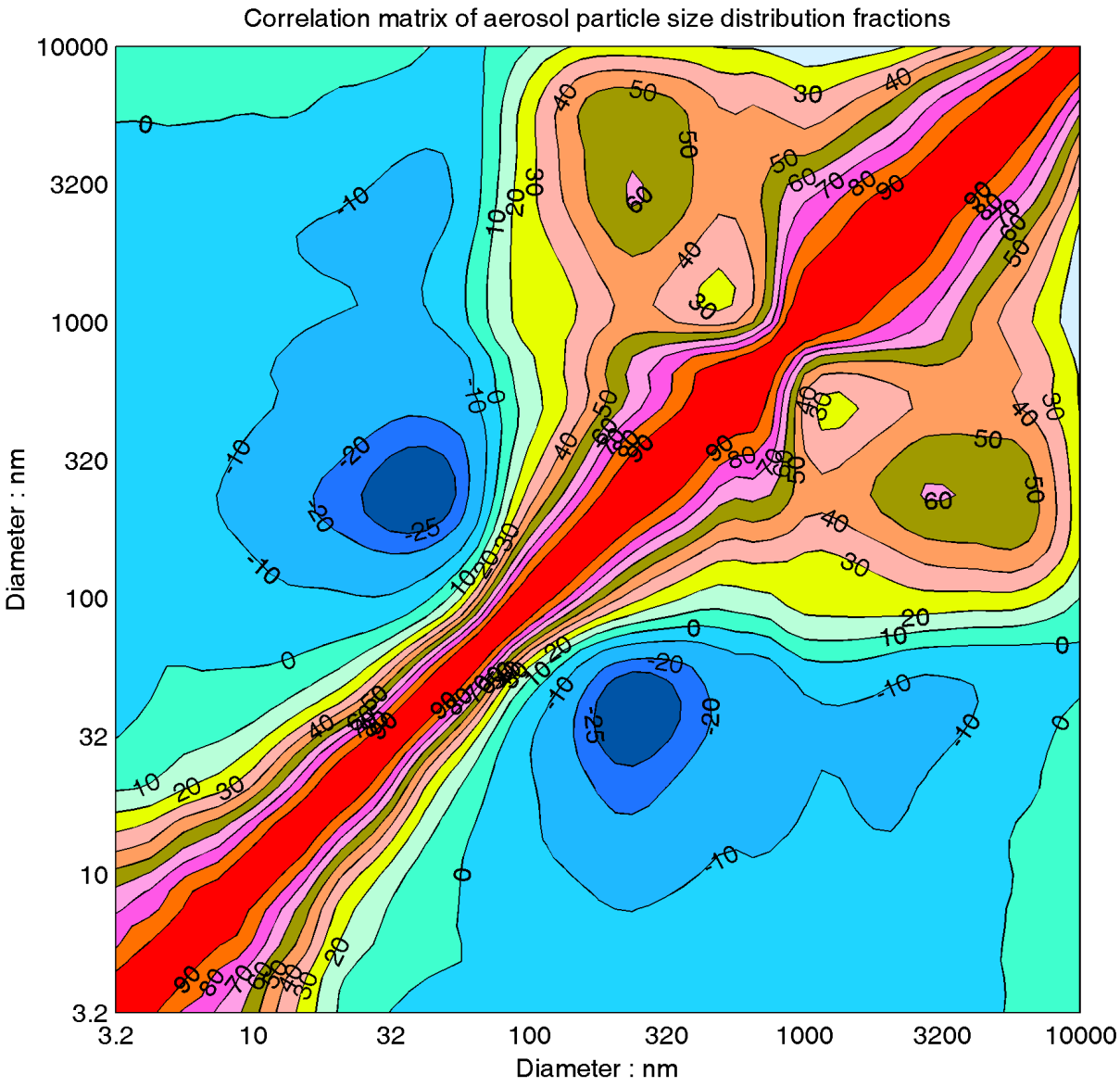
$$S_{coag} = \frac{1}{60 \times T_{coag}},$$

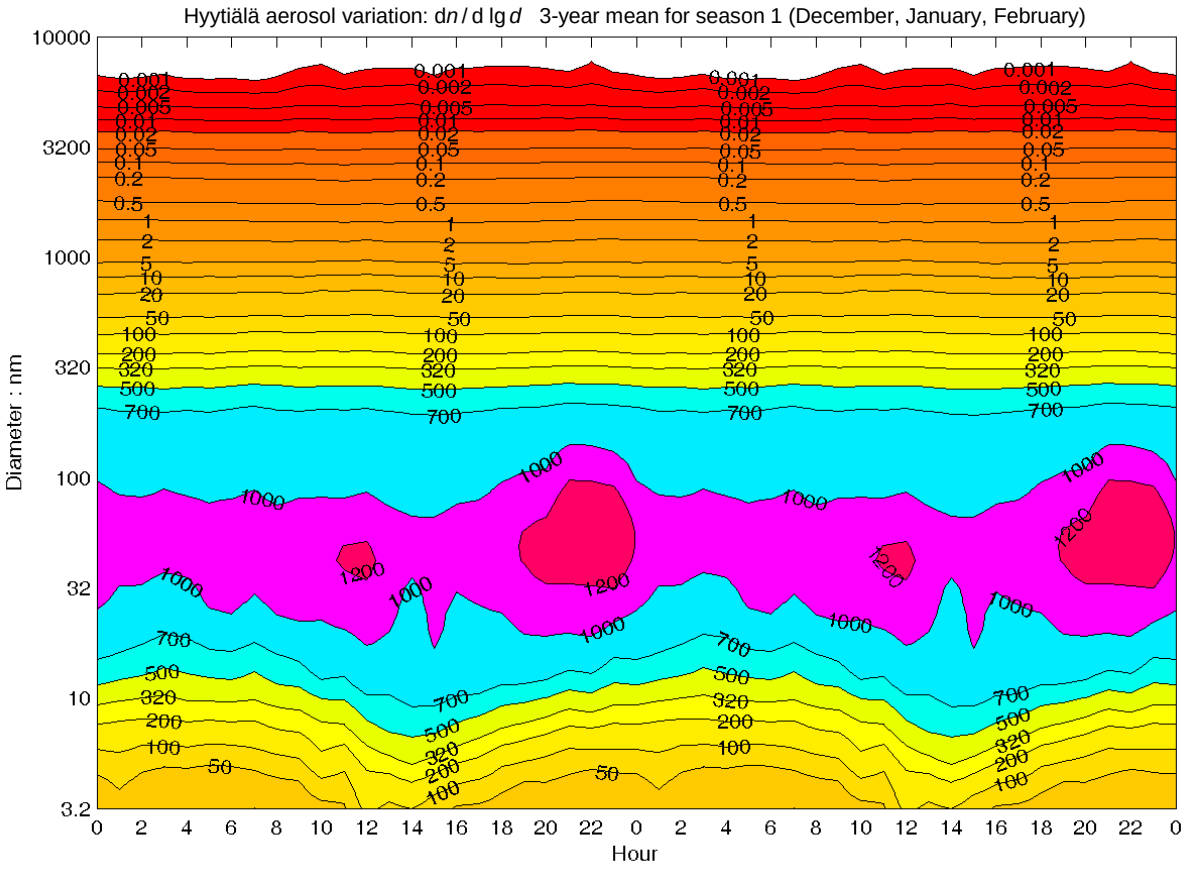
where T_{coag} is the number presented in the table.

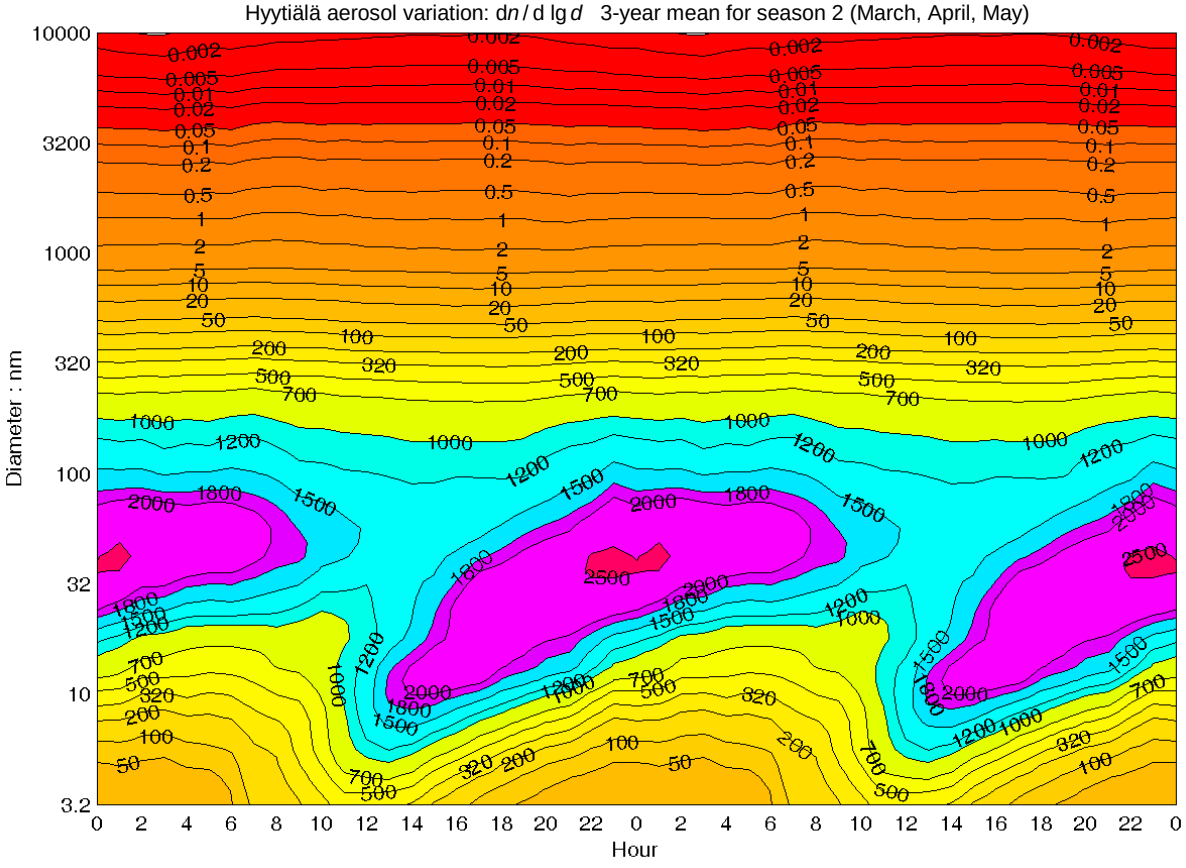
***Column* 90**: New particle index is calculated as per cent of 10...32 nm particles (columns 20...27) in the number concentration in the extended range of 10...3200 nm (columns 20...59):

$$\mathrm{NP\%} = 100 \times N_{10...32\ \mathrm{nm}} / N_{10...3200\ \mathrm{nm}}.$$

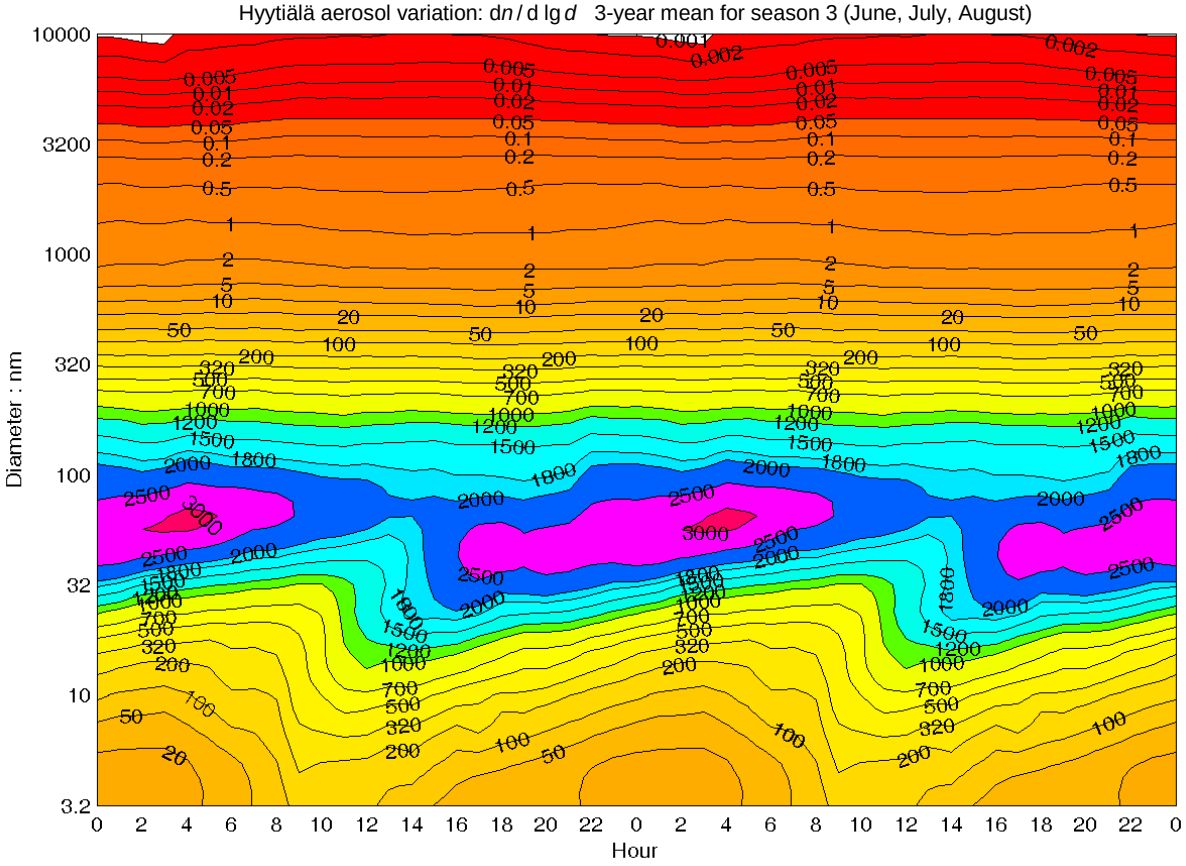
Sample diagrams



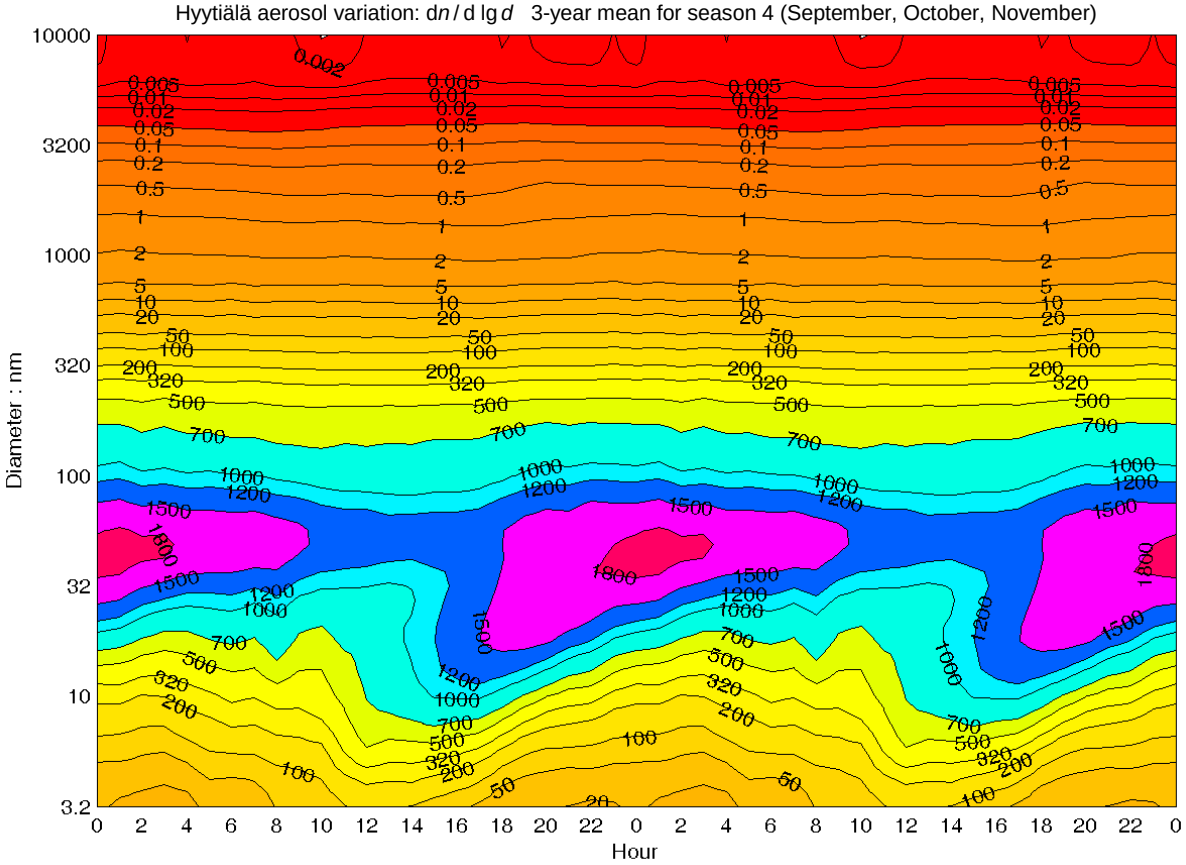

Hyytiälä aerosol variation: dn / d lg d 3-year mean for season 1 (December, January, February)
Diameter : nm
Hour

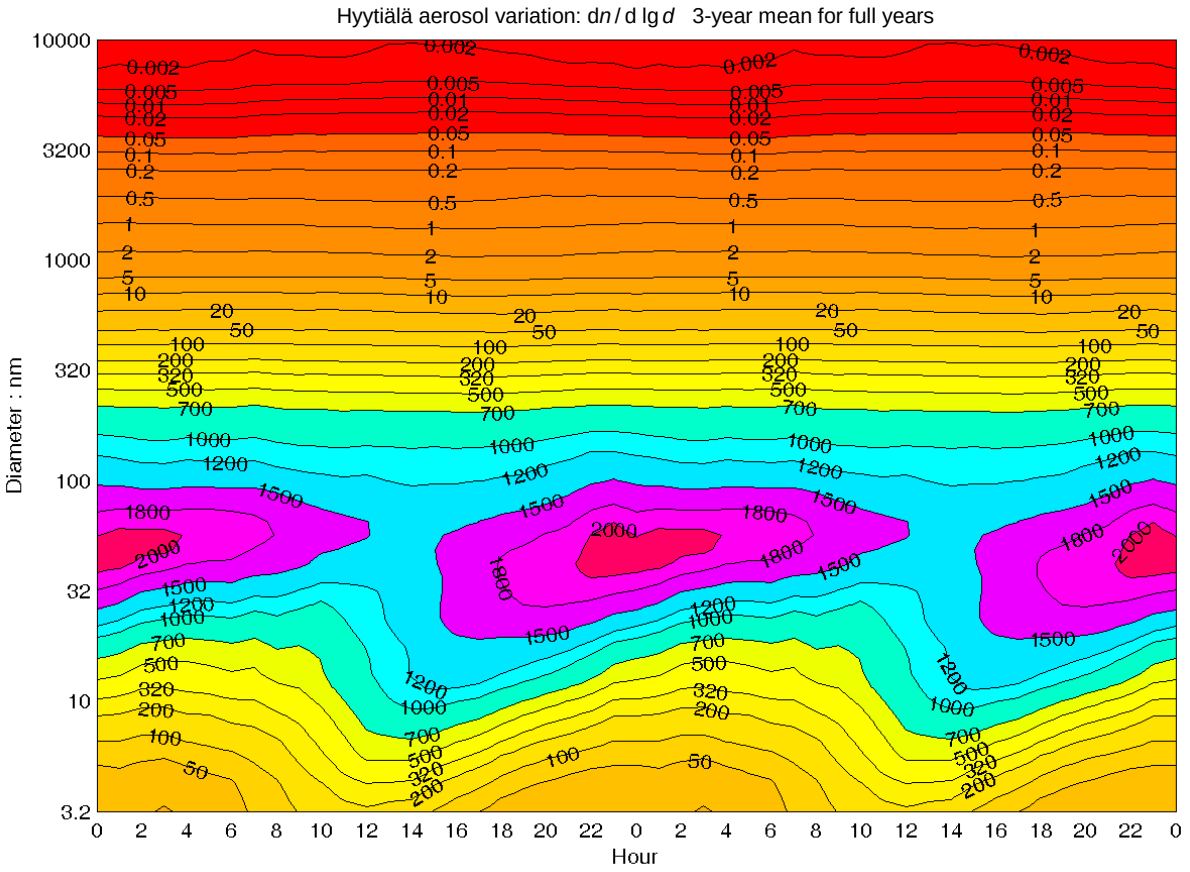

Hyytiälä aerosol variation: dn / d lg d 3-year mean for season 2 (March, April, May)
Diameter : nm
Hour

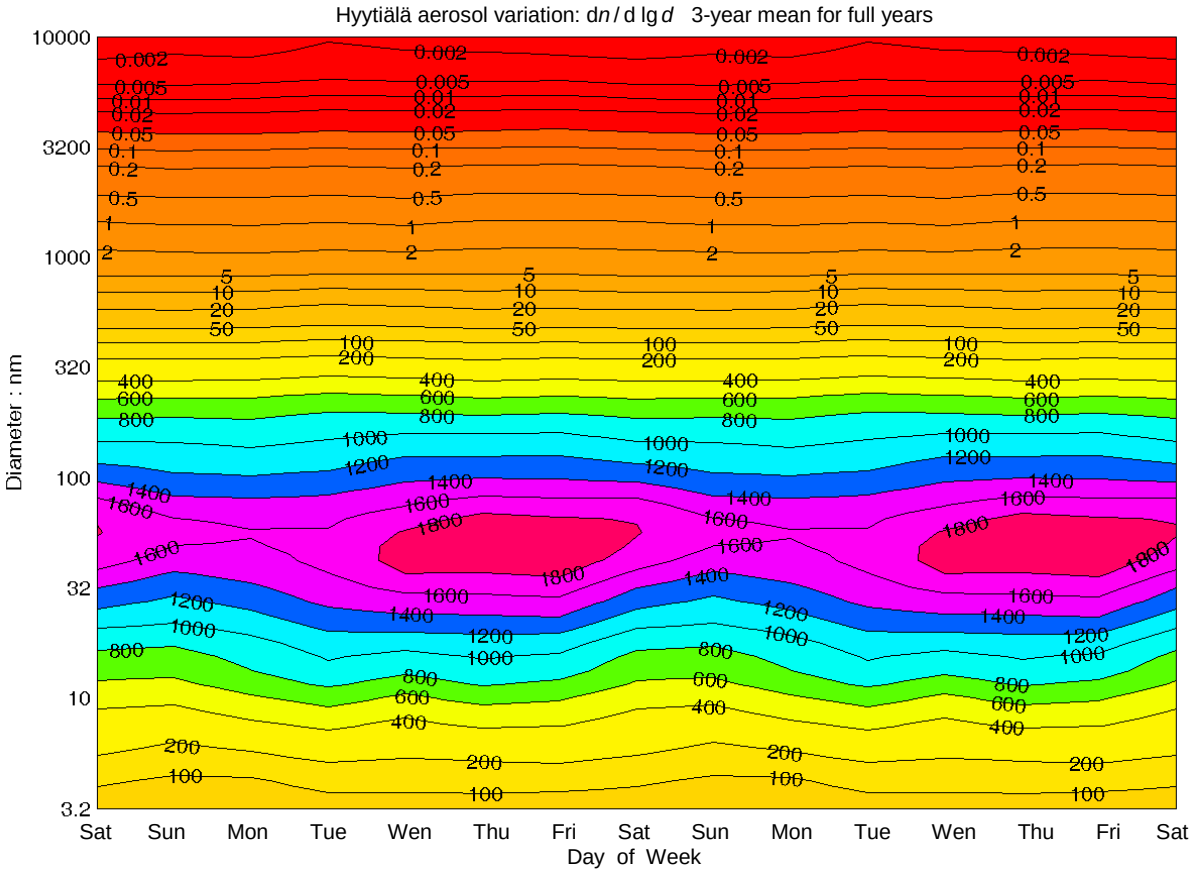
Hyytiälä aerosol variation: d*n* / d lg *d* 3-year mean for season 3 (June, July, August)

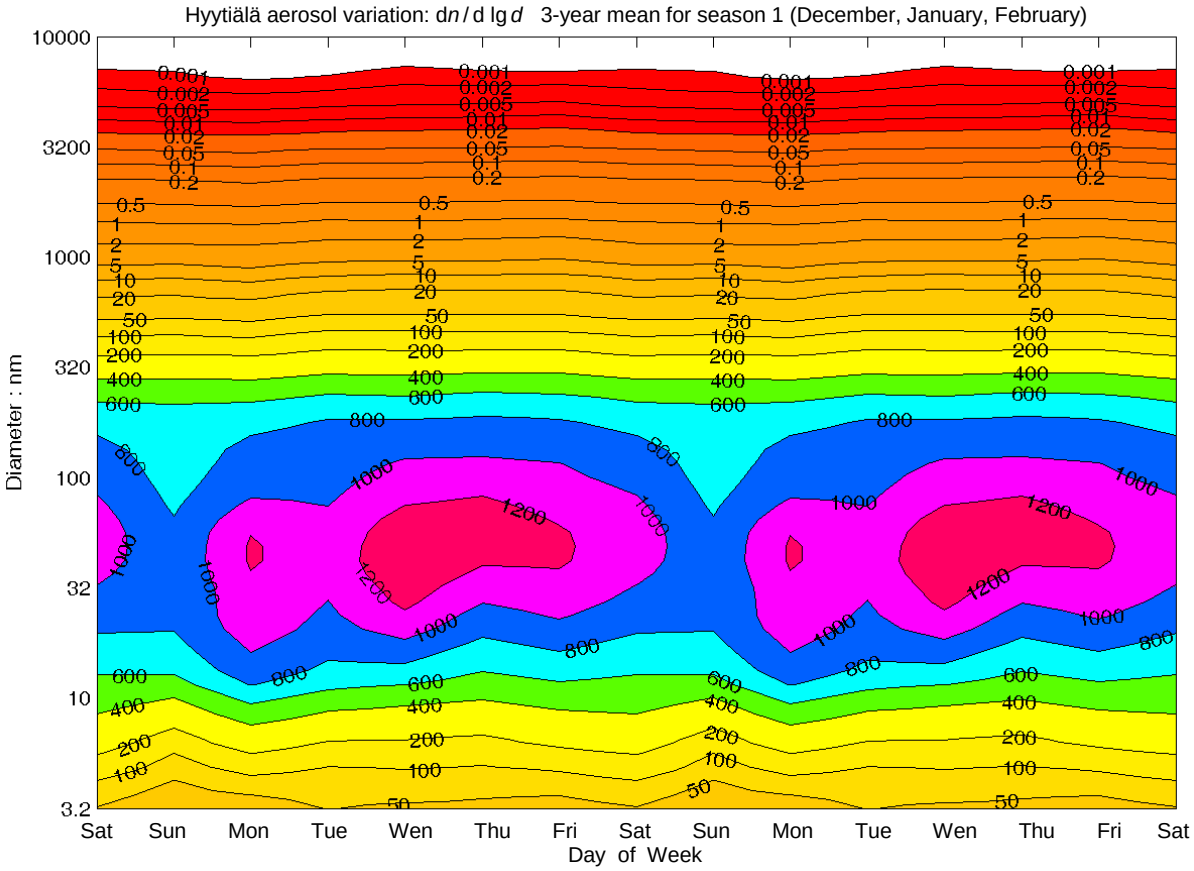


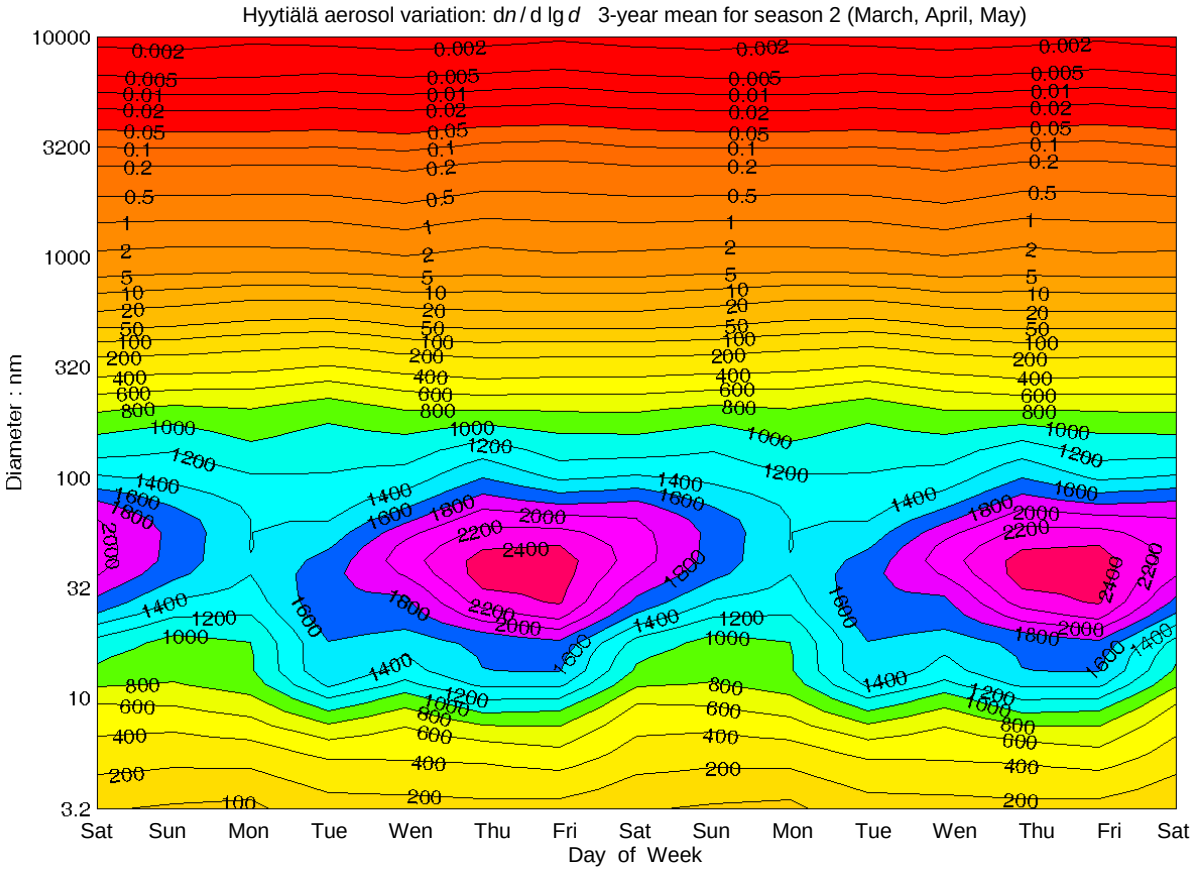
Hyytiälä aerosol variation: d*n* / d lg *d* 3-year mean for season 4 (September, October, November)

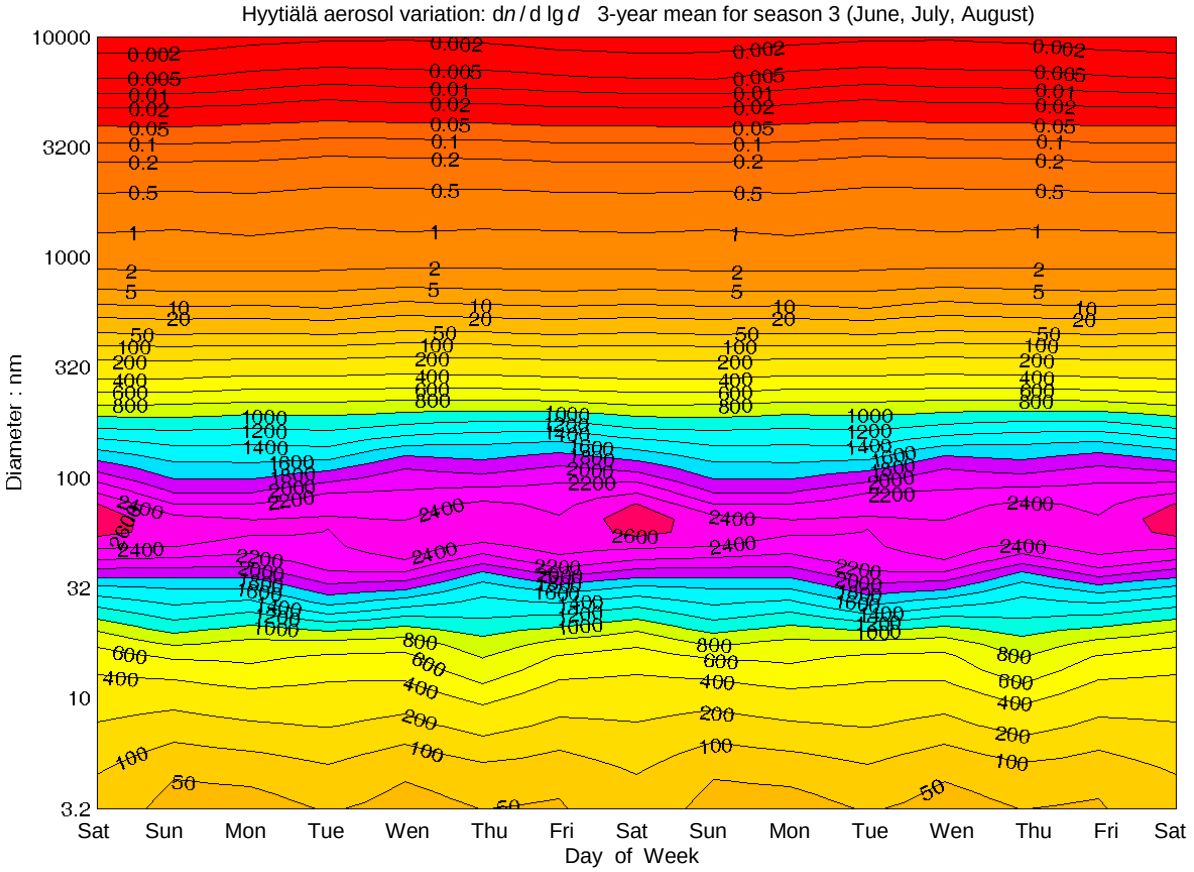


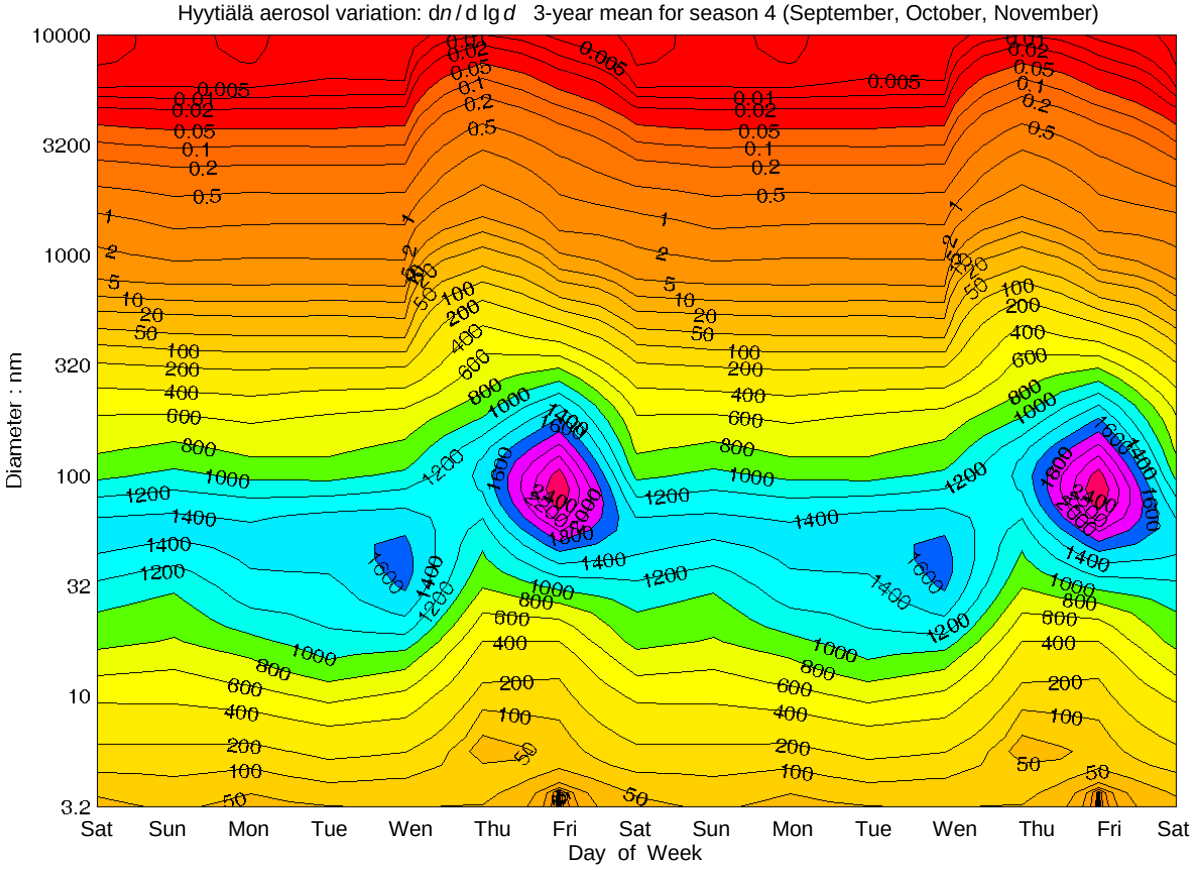

Hyytiälä aerosol variation: dn / d lg d 3-year mean for full years
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
0 2 4 6 8 10 12 14 16 18 20 22 0 2 4 6 8 10 12 14 16 18 20 22 0
Hour


Hyytiälä aerosol variation: dn / d lg d 3-year mean for full years
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
Sat Sun Mon Tue Wen Thu Fri Sat Sun Mon Tue Wen Thu Fri Sat
Day of Week


Hyytiälä aerosol variation: d*n* / d lg *d* 3-year mean for season 1 (December, January, February)
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
Sat Sun Mon Tue Wen Thu Fri Sat Sun Mon Tue Wen Thu Fri Sat
Day of Week


Hyytiälä aerosol variation: d*n* / d lg *d* 3-year mean for season 2 (March, April, May)
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
Sat Sun Mon Tue Wen Thu Fri Sat Sun Mon Tue Wen Thu Fri Sat
Day of Week


Hyytiälä aerosol variation: dn / d lg d 3-year mean for season 3 (June, July, August)
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
Sat Sun Mon Tue Wen Thu Fri Sat Sun Mon Tue Wen Thu Fri Sat
Day of Week


Hyytiälä aerosol variation: dn / d lg d 3-year mean for season 4 (September, October, November)
Diameter : nm
10000
3200
1000
320
100
32
10
3.2
Sat Sun Mon Tue Wen Thu Fri Sat Sun Mon Tue Wen Thu Fri Sat
Day of Week

References

Asmi A., Wiedensohler A., Laj P., Fjaeraa A.-M., Sellegri K., Birmili W., Weingartner E., Baltensperger U., Zdimal V., Zikova N., Putaud J.-P., Marinoni A., Tunved P., Hansson H.-C., Fiebig M., Kivekäs N., Lihavainen H., Asmi E., Ulevicius V., Aalto P.P., Swietlicki E., Kristensson A., Mihalopoulos N., Kalivitis N., Kalapov I., Kiss G., de Leeuw G., Henzing B., Harrison R.M., Beddows D., O'Dowd C., Jennings S.G., Flentje H., Weinhold K., Meinhardt F., Ries L., Kulmala M. (2011). Number size distributions and seasonality of submicron particles in Europe 2008–2009. *Atmos. Chem. Phys*., 11: 5505–5538, doi:10.5194/acp-11-5505-2011.

Fors E.O. , Swietlicki E., Svenningsson B., Kristensson A., Frank G.P., Sporre M. (2011). Hygroscopic properties of the ambient aerosol in southern Sweden – a two year study. *Atmos. Chem. Phys*., 11: 8343–8361.

Hari P., Kulmala M. (2005). Station for Measuring Ecosystem-Atmosphere Relations (SMEAR II). *Boreal Env. Res*., 10: 315–322.

Tammet, H., Komsaare, K., Hõrrak, U. (2014). Intermediate ions in the atmosphere. *Atmos. Res*., 135–136, 263–273. http://dx.doi.org/10.1016/j.atmosres.2012.09.009.

Tammet, H., Kulmala, M. (2014a). Empiric equations of coagulation sink of fine nanoparticles on background aerosol optimized for boreal zone. *Boreal Environ. Res*., 19, 115–126.

Tammet, H., Kulmala, M. (2014b). Performance of four-parameter analytical models of atmospheric aerosol particle size distribution. *J. Aerosol Sci*., 77, 145–157.

Virkkula A., Backman J., Aalto P.P., Hulkkonen M., Riuttanen L., Nieminen T., Dal Maso M., Sogacheva, L. De Leeuw G., Kulmala M. (2011). Seasonal cycle, size dependencies, and source analyses of aerosol optical properties at the SMEAR II measurement station in Hyytiälä, Finland. *Atmos. Chem. Phys*., 11: 4445–4468.